

# Measuring Tensions and Intentions

---

Mixing methods in the impact  
evaluation of development  
support to farmer organisations

Giel Ton



# **Measuring Tensions and Intentions**

**Mixing methods in the impact evaluation of  
development support to farmer organisations**

Giel Ton

## **Thesis committee**

### **Promotor**

Prof. Dr E.H. Bulte  
Professor of Development Economics Group  
Wageningen University

### **Other members**

Prof. Dr E. Stern, University of Bristol, United Kingdom  
Dr W.J.J. Bijman, Wageningen University  
Prof. Dr W.K. Olsen, University of Manchester, United Kingdom  
Prof. Dr H.A.J. Bras, Wageningen University

This research was conducted under the auspices of the Wageningen  
School of Social Sciences (WASS)

# **Measuring Tensions and Intentions**

**Mixing methods in the impact evaluation of  
development support to farmer organisations**

Giel Ton

## **Thesis**

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Thursday 1 December 2015

at 13:30 p.m. in the Aula.

Giel Ton

Measuring Tensions and Intentions:

Mixing methods in the impact evaluation of development support to farmer organisations

296 pages

PhD thesis, Wageningen University, Wageningen, NL (2015)

With references, with summaries in Dutch and English

ISBN 978-94-6257-573-8

# Table of contents

<b>Acknowledgements</b>	V
<b>1. Introduction</b>	1
<b>1.1 Motivation for the study</b>	2
Credible evidence	2
Intermediate outcomes	3
Empirical focus	5
<b>1.2 Research objectives</b>	7
<b>1.3 Theoretical background</b>	8
Tensions between approaches to impact evaluation	8
Tensions in economic farmer organisations	13
<b>1.4 Study design</b>	17
Research objective 1: Identify design principles for credible and lean impact evaluation	17
Research objective 2: Develop and validate a measure to assess organisational strength	19
Research objective 3: Present credible empirical evidence on the effectiveness of FONDOECAS	19
<b>1.5 Thesis structure</b>	20
Part 1 – Introduction	20
Part 2 – Taking stock	20
Part 3 – Research design	21
Part 4 – Empirical research	22
Part 5 – Conclusions	23
<b>2. Innovation Grants to Smallholder Farmers: Revisiting the key assumptions in the impact pathways</b>	25
<b>2.1 Introduction</b>	26
<b>2.2 Methods</b>	28
<b>2.3 Results</b>	30
Impact pathways type A: Voucher grant systems	31
Impact pathways type B: Business development matching grants	36
Impact pathways type C: Farmer-driven innovation support funds	41
<b>2.4 Discussion and conclusion</b>	48
<b>3. The Triviality of Measuring Ultimate Outcomes: Acknowledging the span of direct influence</b>	51
<b>3.1 Introduction</b>	52
<b>3.2 Setting: Impact evaluation of certification</b>	53
Public accountability requirements	54
Harmonised indicators and rigorous measurement	54
Refined theories of change	55
The evaluation challenge in certification	57
<b>3.3 Case study: Evaluation of training for certification in Ivory Coast</b>	58
Multiple actors, treatments and contexts	58
Sample sizes for rigorous measurements	60
<b>3.4 Discussion</b>	61
<b>3.5 Conclusions</b>	63

<b>4. Development Impacts of Value Chain Interventions: How to collect credible evidence and draw valid conclusions in impact evaluations?</b>	65
4.1 Introduction	66
4.2 Impact evaluations: Credible evidence and valid conclusions	67
Asking evaluative questions	67
Drawing valid conclusions	68
4.3 Methodology challenges	71
Measuring outcome patterns	71
Attribution in open systems	74
Social embeddedness and generalisation	77
4.4 Application	77
FONDOECAS: A small-grants fund for collective marketing	78
Intervention theory	78
Key evaluation questions	80
Core methodology	80
Measurement challenge	81
Attribution challenge	82
Generalisation challenge	82
4.5 Conclusions: Towards a theory-based mixed methods design	83
<b>5. The Mixing of Methods: A three-step process for improving rigour in impact evaluations</b>	85
5.1 Introduction	86
5.2 Approach	87
5.3 A step-wise process to improve designs	88
Step 1: Refine the evaluation questions based on the intervention logic	88
Step 2: Anticipate the validity threats to the expected type of conclusions	90
Step 3: Maximise the scope for comparative research	92
5.4 Application	93
Example 1: Impact evaluation of micro-irrigation technology supply	94
Example 2: Impact evaluation of small grants for collective processing	100
5.5 Concluding remarks	104
<b>6. Tension Containment Capacity: Towards a measure of organisational social capital in economic farmer organisations</b>	107
6.1 Introduction	108
6.2 Farmer organisations in collective marketing	109
6.3 Comparing organisational strength	110
6.4 Tension Containment Capacity (TCC)	113
Ten agency dilemmas	113
Field test in Bolivia	115
Calibrating the measure	117
Correlation with economic performance	121
6.5 Reliability in longitudinal analysis	122
Reality check	124
Towards a more robust measure	129
6.6 Discussion and conclusions	131

<b>7. Using Qualitative Comparative Analysis to Explore Outcome Patterns of Grant Support to Farmer Organisations in Bolivia</b>	133
7.1 Introduction	134
7.2 Qualitative Comparative Analysis	135
7.3 FONDOECAS	138
7.4 Conditions used in QCA	139
Fuzzy-set calibration	141
Crisp-set outcomes	145
Description of the data set	148
7.5 Truth table analysis	150
Outcome 1: Increased market access of members	151
Outcome 2: Improve organisational capacities	155
Outcome 3: Increase capacity to pay organisational expenses	159
7.6 Why are grants to well-endowed organisations unsuccessful?	161
7.7 Triangulating the results of QCA with logistic regression	162
7.8 Discussion and conclusions	165
<b>8. Innovation Grants to Farmer Groups for Collective Marketing: A contribution analysis of the FONDOECAS small-grant fund in Bolivia</b>	167
8.1 Introduction	168
8.2 Context	169
Bolivia's rural structure	169
Economic farmer organisations	170
FONDOECAS	172
8.3 Study design	175
Objectives	175
Contribution Analysis	175
Intervention logic and key assumptions	177
Sample selection	179
8.4 Verifying the assumption of relevance	180
Household survey	180
Do economic farmer organisations contribute to local economic development?	181
Conclusions on relevance	184
8.5 Verifying the assumption of effectiveness	184
Case-based comparative analysis	184
Did the grants improve organisational capacities?	186
Did the grants improve access to markets?	190
Did the grants increase the capacity to pay organisational expenses?	194
Did the grants improve access to financial services?	195
Conclusions on effectiveness	197
8.6 Verifying the assumption of efficiency	199
Analysis of monitoring data	199
Did feasibility scores predict implementation progress?	199
Conclusions on efficiency	202
8.7 Contribution story	202
8.8 Conclusion	203



<b>9. Synthesis and Discussion</b>	205
<b>9.1 Introduction</b>	206
<b>9.2 Main findings</b>	206
<b>9.3 Design principles</b>	209
Anticipate generalisation	209
Map the intervention logic	209
Use theory-based evaluation	210
Explore the existing literature	211
Define the span of direct influence	212
Organise a threats to validity check	212
Combine different logics to detect causal patterns	212
<b>9.4 Realist analysis of organisational social capital</b>	213
<b>9.5 Credibility of evidence</b>	218
<b>References</b>	221
<b>Annex 1. Case study summaries</b>	242
<b>Annex 2. Changes in the sample</b>	281
<b>Annex 3. Organisational radiography instrument</b>	282
<b>Annex 4. Time-series data sheet</b>	283
<b>Annex 5. Questionnaire included in the AR-LAT household survey</b>	284
<b>Summary</b>	286
<b>Samenvatting</b>	290
<b>About the author</b>	295

## Acknowledgements

This book reflects the results of my research during the last five years, when I had the opportunity to spend part of my time exploring evaluation approaches, learning new research methods and contributing to the search for effective ways to empower smallholder farmers in markets. I enjoyed it!

The person whom I would like to thank most is my partner Gerda Zijm, who at crucial moments pointed to the merits of the research and the logic of its argument. She combined analytical sharpness and practical knowledge on economic farmer organisations with loyal household management and enduring love. Without doubt, this exceptional combination has been my main support throughout this research endeavour.

I also acknowledge the support of my daughters, Tymke and Gonneke. Their classes on philosophy at secondary school stimulated discussions about methods in science. Later, my knowledge on statistics was refreshed by the books on multivariate analysis that Gonneke brought home from Leiden. Both also helped me with the transcription and coding of FONDOECAS monitoring data.

I would like to thank Erwin Bulte, my promotor, for his scrutiny of logical reasoning and his attention to possible sources of bias in the data analysis. I appreciate his openness and flexibility, and his positive critical stance toward unconventional efforts to detect patterns in data. Sietze Vellema, long-time friend, former colleague and co-author of two papers in this thesis, was an important source of inspiration. Many of my ideas became clearer during our conversations on the train to the Hague, on the buses in Uganda and during our lunch-time walks around the Leeuwenborch. I hope that Erwin and Sietze will not identify themselves too much with the fictional caricatures ‘randomista’ and ‘realista’. The usual disclaimer applies: “Any resemblance to actual people, living or dead, is purely coincidental.”

At LEI, I would like to thank Nico de Groot, who created the possibility for my part-time PhD-trajectory, and Marieke de Ruijter de Wildt, Gerdien Meijerink and Ruerd Ruben, who tried to shield me somewhat from the hectic reality of applied research. Other (former) LEI colleagues experimented with design principles and research tools developed in my research, especially Bart Doorneweert, Lucas Judge, Sabine Hiller, Verina Ingram, Fédés van Rijn, and Yuca Waarts. And, of course, thanks to Karin de Grip and Marie Luise Rau, who helped in the search and screening of impact studies on innovation grants, and co-authored Chapter 2. LEI colleague Lan Ge helped me when I felt unsure about my statistics and regressions and co-authored Chapter 5.

The empirical research was made possible, thanks to my previous work as a policy analyst in the national platform of economic farmer organisations CIOEC-Bolivia, between 2000 and 2004. I want to thank Maria Julia Jiménez, at that time the executive director of CIOEC, for the trust that she placed in me, her inspiring management style and political sensitivity. This gave me the opportunity to get acquainted with many economic farmer organisations and discuss with them the implications of policies for their organisational development and business strategies.

ICCO welcomed me, when I suggested starting an impact evaluation on FONDOECAS. My strongest supporter was Conny Toornstra, ICCO's regional director, with whom I share a personal history of supporting farmer groups in Nicaragua and Bolivia. Other ICCO staff were also very supportive, such as Johannes Solf and Ángel Ramos, who helped to start up the research.

Within FONDOECAS, I would like to give special regards to Freddy Ticona, Oscar Chambi, Richard Arguedas and Zulema Morales, who helped me to organise the research and provided monitoring information for my analysis. At CIOEC, I also acknowledge the support of Rubén Monasterios, Javier Valda and Sergio Gutierrez, the executive directors during the research period, and Primo Nina, Carlos Leon and Justino Loayza, the successive elected chairpersons of the CIOEC board. In AOPEB, I thank Carmen Sotomayor, executive director of AOPEB during the research period.

Several independent consultants helped me with the data collection: Evaristo Yana and Lithzy Flores did most of the first round of qualitative interviews with the 40 farmer organisations in my panel, in 2011. The second round of interviews, in 2013, was done exclusively by Lithzy Flores, who made a tremendous effort travelling to all those remote places. Thanks for your flexibility, perseverance, tact and analytical capacity. Also, I thank Rafael Aramayo for his efforts to coordinate the household survey and data processing.

I owe special recognition to my English editors, Françoise Kaminker and Mundie Salm (Chapter 5), who not only improved the wording but also helped to check my arguments, and to Mariëtte Boomgaard, who designed the nice lay-out of this book.

This thesis is to a large extent the product of nature and nurture. I had the extreme good luck to be born in the Netherlands, and in a horticultural family in Middelburg. During my childhood, the mood at lunchtime was influenced by market prices, when my father called the auction in order to assess the income effect of his latest tomato consignment. The auction was a cooperative, in which my father was board member, just like my grandfather and my great-grandfather, who was one of its founders. I learned that collective action is not easy and needs organisational intelligence. Technological innovation, collective action, and impact on household income may well have become the main subjects in this thesis because I internalised its relevance during childhood. Therefore, last but not least, I deeply acknowledge the influence of both my parents. My father, who passed away 15 years ago, is present between the lines.



# 1 | Introduction

Giel Ton

Agricultural Economics Research Institute, LEI Wageningen

*“... it is possible to discern a yawning gap between those who seek to control or overcome complexity in order to establish certitude and those inclined to settle for a social science of multiple truths, normative standpoints and politicised inquiry. I have sought to argue, from the middle, against absolute truths and against relative truths and for the idea that only partial truths emerge from evaluative inquiry. I take this to be the orthodox position, indeed the humdrum expectation, in scientific inquiry.”*

Ray Pawson (2013), *The Science of Evaluation: A realist manifesto*, p. 191-192

## 1.1 Motivation for the study

### Credible evidence

This thesis is about the methodological challenges to evaluating agricultural development interventions that intend to empower smallholder farmers in markets. Empowering smallholders is a lofty goal, and is essential for agricultural development (World Bank, 2007). There are divergent opinions about the long-term prospects of smallholders as the prime actors to sustain the future food economy (Collier and Dercon, 2014). However, for several decades to come, small-scale family farming will undoubtedly remain the major source of employment and food security in rural areas (IFAD-UNEP, 2013), and as such will remain in need of support for improvement and innovation. In many places, governments, development organisations, farmer organisations, firms and farmers themselves try to influence the institutional arrangements in agriculture, the ‘rules of the game’ (North, 1990) with widely different approaches and objectives. A diversity of institutional arrangements – regulations, policies and support instruments – are designed or refined with the intention to support smallholders’ access to markets. Often they fail, sometimes they succeed.

The UN Millennium Development Goals (MDGs) formulated during the United Nations Millennium Summit in 2000, stimulated the emergence of large-scale approaches for tackling poverty with proven interventions and ‘quick wins’ (Sachs and McArthur, 2005). The Millennium Development Goals were coupled with an influx of new, well-endowed development funds, such as the Bill and Melinda Gates Foundation (BMGF), launched in 2000, having the explicit objective of upscaling promising approaches for poverty reduction.

This new impetus for development support to agriculture also created a demand for ‘evidence-based’ policies and technologies that were proven to work, and could be replicated or scaled up. This evidence, however, was lacking for many of the interventions that had been implemented in the past, in spite of the many evaluations that had been carried out. According to the OECD guidelines (OECD, 2001), evaluations need to cover the relevance, effectiveness, efficiency, impact and sustainability of support. Generally, these evaluations take place in the last years of a project or programme, using the qualitative and quantitative information available to come to conclusions. However, the data available to evaluators is often limited, and, therefore, their conclusions about effectiveness and impact lack ‘credible evidence’. The Evidence-based policy movement arose from this lack of systematic knowledge and analysis that could be used by policy makers on the one hand, and the lack of ways to judge the trustworthiness of findings with which policy makers were bombarded by academics, pressure groups and lobbyists on the other hand (Stern et al., 2012: 9).

The international donor community recognised this ‘evaluation gap’, and concluded that governments or donors did not demand or produce enough high-quality impact evaluations (Savedoff et al., 2005). To fill this gap, in 2006, the international network of evaluators NONIE (Network of Networks for Impact Evaluation) was established to foster more and better impact evaluations. In 2009, NONIE published a guidance note (Leeuw and Vaessen, 2009), in which the authors recommended research designs to test intervention theories rather than projects or programmes. Impact evaluations should also address the issue of attribution of outcomes

with a design that assesses the ‘counterfactual’, defined as a hypothetical statement of what would have happened, had the program not been implemented (USAID, 2009). In 2008, the international donor community established the International Initiative on Impact Evaluation (3ie) to produce credible evidence by funding and backstopping impact evaluations with a design to reduce biases in the estimates of effectiveness. In the Netherlands, the Policy and Operations Evaluations Department (IOB) started, in 2010, to require all public-funded development organisations to implement a rigorous impact evaluation of their support, with a research design that included methods to assess the counterfactual situation.

The importance attached to quantitative impact estimates from independent external researchers generated strong reactions, especially by evaluators who stressed the importance of monitoring and evaluation processes as a tool for participatory learning and programme steering (Guijt, 2008; Patton, 2002). For these evaluators, data-collection on differential effects and dynamics among beneficiaries was considered more important than the comparison of outcomes between beneficiaries and non-beneficiaries. The discussion often centred on the strengths and weaknesses of the randomised controlled trial (RCT) in development evaluation, a method that started to be promoted as the design that could provide the most credible, unbiased estimates of effectiveness (Duflo et al., 2007; Khandker et al., 2009; Lensink, 2014). For example, in 2007, the European Evaluation Society issued a declaration (EES, 2007; Donaldson et al., 2008) in which they questioned the dominance of RCTs and pointed to several issues for which this design is considered inappropriate (complex situations where outcomes are the result of multiple factors acting simultaneously) or counter-productive (undermining the willingness of development partners to undertake impact evaluation). In 2012, DFID published a study by Stern et al. (2012) explicitly geared to broadening the range of research designs and methods for impact evaluation.

This thesis originated in the context of these debates on impact evaluation of development assistance. Since 2006, I have been working in the applied research institute LEI Wageningen UR, and in this work, I was involved in the design of several impact evaluations that had to comply with these stricter requirements for methodological rigour. This created the need to take position in this debate, and to balance the demand for ‘credible counterfactual research’ with the demand for research ‘geared towards learning’. I found this tension stimulating. With the research project of this thesis, I could organise my work in a way that gave me the opportunity to better understand the arguments in this debate, and it gave me the possibility to explore the potential to combine impact evaluation for accountability with impact evaluation for learning, and to experiment with methods that might bridge the gap between qualitative and quantitative research traditions.

## **Intermediate outcomes**

Impact evaluators need to collect information on outcomes. Often, the effects of the support are located in various outcome areas. Commissioners of impact evaluations are typically more interested in the longer-term outcomes, whereas the implementers of the support are more interested in a mapping of the short and mid-term outcomes, especially when this information is useful for them to steer and adapt the intervention to increase its effectiveness.

Impact evaluations must, therefore, try to ‘capture’ a logical sequence of outcomes. Authors differ in wording to describe this sequence, and use for example the term result chain (Mayne, 2001) (DCED, 2010), programme theory (Rogers, 2008), logic model (Rogers, 2008), theory of change (Connell and Kubish, 1998; Mayne, 2011) or intervention theory (Pawson, 2013). There are differences in these definitions, especially the inclusion of contextual factors and unintended outcomes in the graphic representation, but - in practice – they are often used as equivalents (Funnell and Rogers, 2011). In this thesis, I will mostly use the term intervention logic. An intervention logic generally presents several nested sequences of outcomes, each related to clusters of activities for specific stakeholder groups (Mayne, 2001), called ‘impact pathways’ (DCED, 2010) or ‘sub-theories’ (Pawson, 2013).

Theory-based impact evaluation verifies the key assumptions in this intervention logic (Chen, 1994; White, 2009), usually this includes the assumption that intervention is effective in ‘causing’ certain outcomes. Commissioners of impact evaluations are often primarily interested in impact on poverty alleviation, biodiversity or local economic development. However, even when development support has as its main rationale to generate development impact, e.g., to improve individual livelihoods or company performance, attribution of effects at the level of these ultimate outcomes is often impossible. Ultimate outcomes and development impact are often a result of multiple development supporters working together and in a context of multiple external influences, where the effects of one of these interventions cannot be separated, and where each support intervention is, at most, a contributory cause of the change. Attribution of effects to a support intervention is only possible within a span of direct influence (Mayne, 2011) and, of course, only when outcomes can be properly measured or observed. Each impact pathway of an intervention will have a different boundary of this span of direct influence. And, of course, some research methods are better able to capture some of these ‘borderline’ outcomes than others.



Source: Author’s own elaboration

Figure 1.1 Overall intervention logic of the four impact evaluations mentioned in this thesis

As shown in Figure 1.1, my thesis covers research on various agricultural development interventions to support smallholders' markets access. It covers innovation grants, certification schemes, micro-irrigation technology supply, and investment subsidies to economic farmer groups. I observed in all these intervention logics that the impact on smallholder market access was mediated or moderated by farmer organisations. Often, the short-term outcome of the development support was the strengthening of farmer organisations that could act as market channels for smallholder production (Ton et al., 2007), to represent smallholders in their interaction with donors and policy makers (Ton and Bijman, 2006; Ton et al., 2014b), or to broker agricultural innovation (Ton et al., 2015; Ton, 2007b).

All these impact evaluations started with a commissioner who asked to measure impact on livelihoods, principally with respect to household income and poverty status. Only a few commissioners asked to measure organisational strengthening, in spite of this being key intended outcomes of the interventions. The obvious reason for this emphasis on livelihood impacts is the accountability requirements of these development organisations to their back donors, governments or international donors who had committed themselves to using the Millennium Development Goals as their overarching reporting framework (DCED, 2010). But it was also due to the lack of common indicators to report impact and benchmark effectiveness. The MDGs provided common indicators for such ultimate outcomes as household income, but there were no indicators for intermediate outcomes such as organisational strength. This limited the attention of implementers to the monitor organisational strengthening processes, and constrained their possibility to reflect on the effectiveness of their activities with farmer groups in the field. Moreover, in academic studies the moderating influence of this organisational social capital is rarely included in econometric models used to estimate farmer-level effects of support interventions. I consider economic farmer organisations to be a key instrument for empowering smallholders in markets, and, therefore, I felt the need for instruments that would make it possible to measure and compare the organisational strength of farmer organisations.

## Empirical focus

Collective action is not easy. Keeping a group together requires efforts and money. Especially in collective marketing, the organisations need to find 'fair' rules and regulations that result in cost-effective processing and intermediation of member products. The competition is 'out there', offering sales alternatives to the members. The struggle of farmers to improve their position through collective action is fascinating. Groups need to find logistic processes, rules and incentives that generate benefits for members and keep the organisation competitive in the market. Strong organisations need to be resilient to external turbulence and have an internal organisational structure ensuring their continued functioning, despite disintegrative tendencies (Ostrom, 1990). These inherent tensions between group and individual in collective action are called 'agency dilemmas' (Shapiro, 2005) and need to be contained in order to prevent the organisation from falling apart.

The main empirical research in this thesis focused on a peculiar type of farmer organisations, specifically, membership organisations that engage in bulking, processing and collective marketing. The most prominent legal status of such organisations is the cooperative. However,



especially in developing countries a whole range of different organisational formats are commonly used, often initiated and supported by governments or donors that see them as instrumental for the empowerment of small scale farmers in markets. In this thesis, I will call these groups ‘economic farmer organisations’. The main empirical research in this thesis took place in Bolivia. Between 1999 and 2004, I was employed by CIOEC-Bolivia, the national coordinating platform of economic peasant organisations (*Coordinadora de Integración de Organizaciones Económicas Campesinas, Indígenas y Originarias de Bolivia*), as policy analyst, paid by the Dutch development organisation ICCO. This gave me the opportunity to interact with many farmer organisations engaged in collective marketing activities, looking for their common policy interests and shared organisational problems (CIOEC-Bolivia, 2000, 2004). Bolivia has changed very much since 2004. The social contradictions and revolts that took place when I worked with CIOEC contributed to transform the country, in 2008, into the Plurinational State of Bolivia. The Morales Government, which took office in 2006, reintroduced government interference with price setting in key agricultural markets, preferential credit facilities for smallholder farmers and deepened the preferential access of smallholders to the niche market of government procurement, especially in nutritional programmes (Cordoba, 2014). These policies modified the institutional environment for economic farmer organisations, and, together with a period of steady economic growth after 2004, has improved their access to markets and support services (Prudencio and Elías, 2014).

These new government policies also created contradictions. Due to ideological preference and political strategy, the sector of economic farmer organisations became side-lined in the allocation of public investments and subsidies. The government preferred to support the strong and politically connected network of territorially-based, all-inclusive village groups, organised around the control of natural resources (land, water, minerals) and the access to social investments. The government gave meagre support to the sector of functionally organised economic farmer organisations, who represented a self-selected sub-group of households in each village, those who produce a certain crop and, through voluntary membership, committed themselves to collective action in markets.

After 2004, based in The Netherlands, I could follow the social changes in Bolivia through my involvement in several applied research assignments, in which I helped several development organisations to design an impact evaluation, and for which I had the opportunity to travel to Bolivia. Bolivia was, thus, a logical choice as the location for my PhD-research. My previous work experience in Bolivia made it more easy to contact farmer organisations, local researchers and other key informants for field research. Furthermore, because I anticipated using comparative case studies to address the role of organisational strength of farmer organisations, I could benefit from my experiences and substantive knowledge on the geographical, social and political context in which these change processes would unfold, to complement and understand the information collected by the implementers of the intervention and local researchers.

Most important for my focus on the FONDOECAS grant fund as object of the impact evaluation was the fact that, in 2005, I had assisted ICCO and CIOEC to resolve a deadlock in the formulation of a support project, which subsequently became the FONDOECAS grant fund (*Fondo de Fortalecimiento Económico para las Organizaciones Económicas Campesinas*). Research on this grant system, which catered specifically to the sector of economic farmer organisations,

seemed relevant and timely, because, as an instrument to empower smallholder farmers in markets, it had the potential to be replicated or scaled up. Evidence on the FONDOECAS grant fund could help to refine the emerging policy instruments of the Bolivian government. And, because it concerned a relatively uniform intervention (fixed amount of money to invest in business plans of farmer organisations), it seemed a promising intervention for the use of a quasi-experimental research design.

Last but not least, I expected that an impact study on FONDOECAS would provide evidence that grant support to farmer organisations could empower smallholder farmers in markets, a conviction that I share with many others (World Bank, 2007; Bosc et al., 2001; Markelova et al., 2009; Bebbington, 1996; Bernard et al., 2010; Shiferaw et al., 2011): grants strengthen economic activities of farmer groups, and stronger economic farmer organisations are key institutions to improve market access and local economic development. My study could help to increase this type of support in the future.

## 1.2 Research objectives

The research had three interlinked objectives:

1. Identify design principles for credible and lean impact evaluation, appropriate for interventions that aim to improve smallholders' access to markets.

I expected that contribution analysis (Mayne, 2001) would be a useful framework to analyse the effectiveness of interventions. I also expected that quasi-experimental methods would facilitate counterfactual thinking, and that regression-analytic and configurational comparative approaches provide complementary insights on the effectiveness of interventions.

2. Develop and validate a measure to assess organisational strength of collective marketing groups, appropriate for cross-sectoral comparative and longitudinal analysis.

I assumed that common indicators for measuring effects of development interventions on organisational strength of economic farmer groups were lacking, and that semi-structured interviews to check the dynamics surrounding agency dilemmas in collective marketing could be used to derive a comparative measure of organisational strength of economic farmer organisations.

3. Present credible empirical evidence on the effectiveness of the FONDOECAS subsidy fund, which provides grants to economic farmer organisations, in attaining its intended outcomes.

I expected that FONDOECAS grants were invested in business plans of economic farmer organisations, and that the grant-supported business plans created access to markets for their members, resulted in organisational strengthening, and created group income to pay the costs of collective action.

## 1.3 Theoretical background

In this thesis, I situate my research in a context of ‘tensions’. These tensions manifest themselves in two ways. First, I propose to bridge two different approaches towards impact evaluations - two different schools of thought. One school is organised around experimental and quasi-experimental methods to impact evaluation, with Shadish, Cook and Campbell’s handbook as a key reference (Shadish et al., 2002). The other school is characterised by causal explanation, with Pawson and Tilley’s book on realist evaluation as landmark publication (Pawson and Tilley, 2006). Often, these two schools in evaluation are methodologically contrasted by their focus on quantitative research and hypothesis testing versus qualitative research and causal explanation. Apart from the inevitable debates on the appropriateness of each school’s favoured methods for answering evaluation questions (Guijt et al., 2011), the tension and debates between these two schools hold the potential to create synergy in mixed-method designs (Brady and Collier, 2004; Brady et al., 2006; White, 2011; Pawson, 2013; Shadish et al., 2002).

Second, I focus my attention on the inherent tensions of collective action by farmer organisations in markets. Organisations face inherent pressures of disintegration, similar to entropy in natural systems. Strong organisations need to be resilient so as to have an internal organisational structure that supports their social reproduction (Giddens, 1979), their ongoing functioning as a collective action, in the face of disintegrative tendencies (Ostrom, 1990). These inherent tensions between the group and individuals in collective action are called ‘agency dilemmas’ (Shapiro, 2005). In economic farmer groups that are active in collective marketing of member products, these agency dilemmas are related to economic transactions. Trust, institutions and social networks are important determinants of this organisational capacity for collective action (Ostrom and Ahn, 2009) but may result in multiple viable internal institutional arrangements and organisational forms.

### Tensions between approaches to impact evaluation

For clarity, I reduce the debate in impact evaluation to only two archetypical positions, the ‘randomistas’ and ‘realistas’. I will present them as caricatures to augment their differences in order to explore for synergy.

In the discussion on approaches to impact evaluation, scholars looking for unbiased, average effects through experimental or quasi-experimental methods are often labelled as ‘randomistas’ (2009; Patton, 2008); this is in reference to the randomised control trial that is regarded as the highest-quality research design to compute average effects by most systematic review bodies (e.g., Cochrane Collaboration, Campbell Collaboration). No sensible researcher will identify completely with this archetypical image of ‘randomista’ and restrict his or her analysis to ‘RCT-for-average-effects’ only. Many of the scholars, who favour randomisation in quasi-experimental research designs, also reject the exclusive focus on average effects in impact evaluations. These scholars point to the importance of analyses of heterogeneity in contexts and implementation modalities of social interventions, next to analyses of average effects (Heckman and Smith, 1995; Ravallion, 2009; Deaton, 2010).

I will call the other school the ‘realistas’. The word ‘realistas’ refer obviously to ‘realist evaluation’, with Pawson and Tilley’s book ‘Realistic Evaluation’ as the landmark publication (Pawson and Tilley, 1997). Realist evaluation is the pragmatic operationalisation of the theory of science developed by Bashkar (1978) and Sayer (2000), known as ‘Critical Realism’, to the evaluation of social policies and programmes (Pawson, 2013). While Critical Realism is considered as postmodern (Blaikie, 2000), Pawson and Tilley are rather modernist in their approach to scientific inquiry. Pawson and Tilley (1997) proposed as main objective of impact evaluation the distillation of different Context-Mechanism-Outcome Configurations, as ‘middle-range theories’: What works for whom, under what conditions, and why? In their view, a development intervention aims to change the incentive structure that predicts, to a certain extent, how people behave. Interventions include activities that aim to trigger changes in this behaviour. In realist evaluation, the causality through which the incentive structure predicts behaviour is covered by the term ‘causal mechanisms’, and is ‘located in the head of people’ (Pawson and Tilley, 2006). “Incentivisation is thus a generic mechanism in programmes, and ‘behavioural change’ the generic intended outcome” (Pawson, 2013: :84).

Both archetypical schools of impact evaluation have commonalities. In contrast to social constructivists, who consider reality to be a subjective, culture-specific interpretation of individual observations, both ‘randomistas’ and ‘realistas’ consider that there is a ‘reality’ that exists independently from the observers (Shadish et al., 2002; Pawson and Tilley, 1997). Scientific methods can partially observe/discover this reality in order to identify regularities, patterns and causalities that help to steer or anticipate social processes. Scientific research has a role in identifying, testing and modifying causal models that describe, explain or predict these regularities in the social reality, imperfect as these models will inevitably be (Rohlfinger, 2012; Pawson, 2013).

In spite of their shared goal of evaluation science for evidence-based policies, the ‘randomistas’ and ‘realistas’ differ in the kind of evidence that they prioritise in impact evaluation, and the criteria used to judge the rigour or credibility of conclusions (see Figure 1.2). The ‘randomistas’ focus on the measurement of net-effects of an intervention and analyse the differences between beneficiaries and non-beneficiaries of the support to verify if the support proves a relevant causal factor (Are the effects significantly different from zero?), and how much it did help to produce these development outcomes (What is the average effect?). ‘Realistas’, in contrast, will highlight the differential effects that an intervention may have in different contexts, and focus especially on the exploration and explanation of causal configurations that define effectiveness (Why does it work, and under what conditions?). They are less interested in average effects, but more in the characteristics of the groups that proved to show (most) effects versus those that show no (or fewer) effects under the same conditions, or have the same effects while having different conditions (De Meur et al., 2006).

This relates to differences in the preferred way that causal relations are being analysed in both archetypical approaches to impact evaluation, different conceptualisations of causality in complex social systems, and differences in the preferred way to express this complexity in causal models and hypotheses that can be verified or tested. ‘Randomistas’ use regression analytical methods to detect or verify causality, whereas ‘realistas’ prefer configurational comparative methods (Byrne and Ragin, 2009; Thiem et al., 2015).

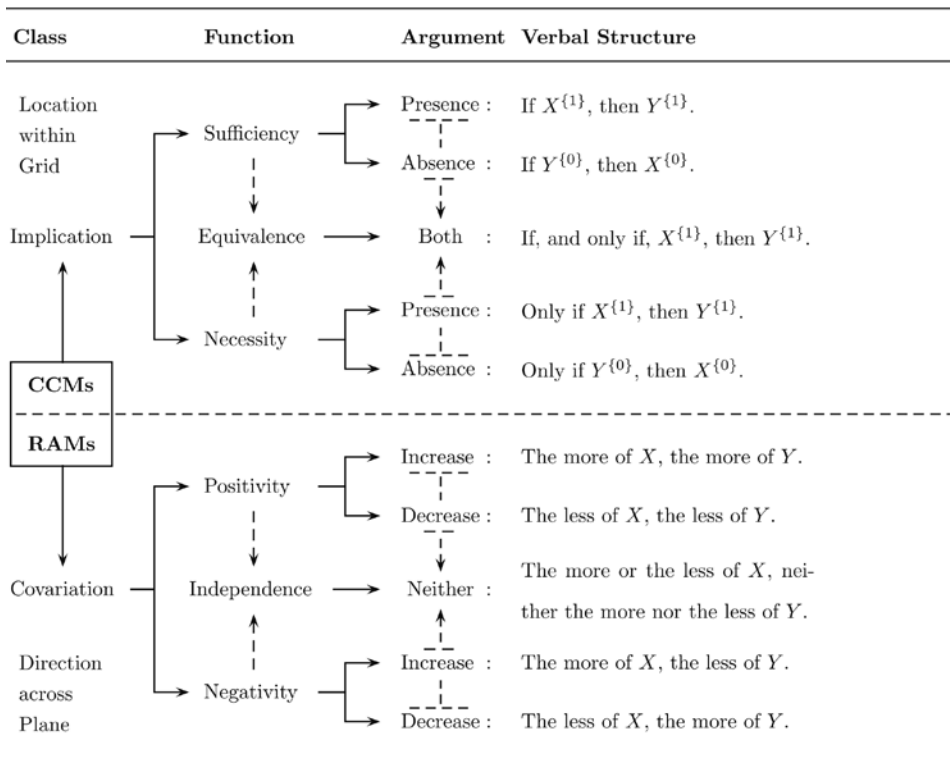
There are widely divergent strategies for collecting information on effectiveness but, generally, impact evaluations include the analysis of data-sets. ‘Randomistas’ will typically take a random sample of supported (‘treated’) and unsupported units (‘comparison group’) to test a theory about causal relations, whereas realists tend to use comparative case-studies. Both schools will use qualitative case studies to verify whether the theorised causal process indeed took place. Brady, Collier and Seawright (2006), who promoted a pluralistic vision on methodology, argue that for any generalised causal inference there is a need to combine Causal Process Observations (CPOs) with Data-Set Observations (DSOs).

Mill identified three basic conditions that define causality (Mill, 2009 (1843)): the cause needs to be active before the effect is produced; the cause must be related to the effect produced; and alternative explanations for the effect must be discarded. Both schools will adhere to these principles for determining if a factor is a cause. However, the preferred way to explore or test this in data-sets differs between these archetypical approaches (Befani, 2012). ‘Randomistas’ will prefer to explore for statistically significant differences between beneficiaries and non-beneficiaries of support in data-sets, while realists will focus on the causes or packages of causes (configurations). Thiem et al. (2015) show that these ways to infer causality are fundamentally different. They made an overview of these differences in causality, comparing the causal models and associated types of causal inference. They contrasted configurational comparative methods (CCMs), associated with the ‘realistas’, and regressional analytic methods (RAMs), related with the ‘randomistas’, and explain the function of a causal factor in the respective causal models, the argument used for considering a factor as being a ‘causal’ factor, and the verbal structure of the causal inference. They also suggested a coherent syntactic structure associated with each causal claim, but with a view to simplification, I left this out in Figure 1.3.

These two ways of conceptualising causes – by implication or by covariation – are applied in different, complementary ways for detecting patterns in data. The analysis of causes by covariation is done by conventional statistical packages that use linear algebra, like SPSS (IBM Corp., 2012). The causal analysis by implication is done by computer applications that use Boolean algebra, like Qualitative Comparative Analysis -QCA (Ragin and Davey, 2009; Reichert and Rubinson, 2014; Duşa and Thiem, 2014). In this thesis I will apply both data-analysing methods to explore for causality in data-sets (chapter 7 and 8). I present this tool in more detail in the next section on study design.

Impact evaluations generally want to draw generalised inferences from impact data. Patterns in data suggest causes, causes suggest predictors of effects. Shadish et al. (2002) stress that any generalised causal inference will have threats to validity. But, they emphasize that even though there will always remain some threats to the validity, rigorous research designs incorporate design elements that reduce at least the most pressing validity threats. The issue of methodological rigour is a central concern in systematic reviews. Systematic reviews have as their objective to capitalise on the accumulated knowledge about the effectiveness of interventions in high-quality research (Higgins and Green, 2011), and need to have a transparent and reproducible methodology to do so. To judge the merit for inclusion in a review, the relevance and rigour of a study is rated according to predefined quality criteria. There are different types of systematic reviews (Gough et al., 2012), associated with the archetypical approaches to

impact evaluation described above. The ‘randomistas’ use meta-analysis as their preferred systematic research method, whereas the ‘realistas’ favour more explorative systematic review approaches, such as thematic reviews or mapping studies (Gough et al., 2012; Snilstveit, 2012). The ‘randomistas’ and ‘realistas’ have their favoured Systematic Review Bodies to check the quality of the protocol for and final reports of systematic reviews, with the Campbell and Cochrane Collaboration more focussed on meta-analysis (Higgins and Green, 2011) and the RAMESES guidelines (Wong et al., 2013) promoted to judge the rigour of ‘realist synthesis’ (Pawson, 2013).

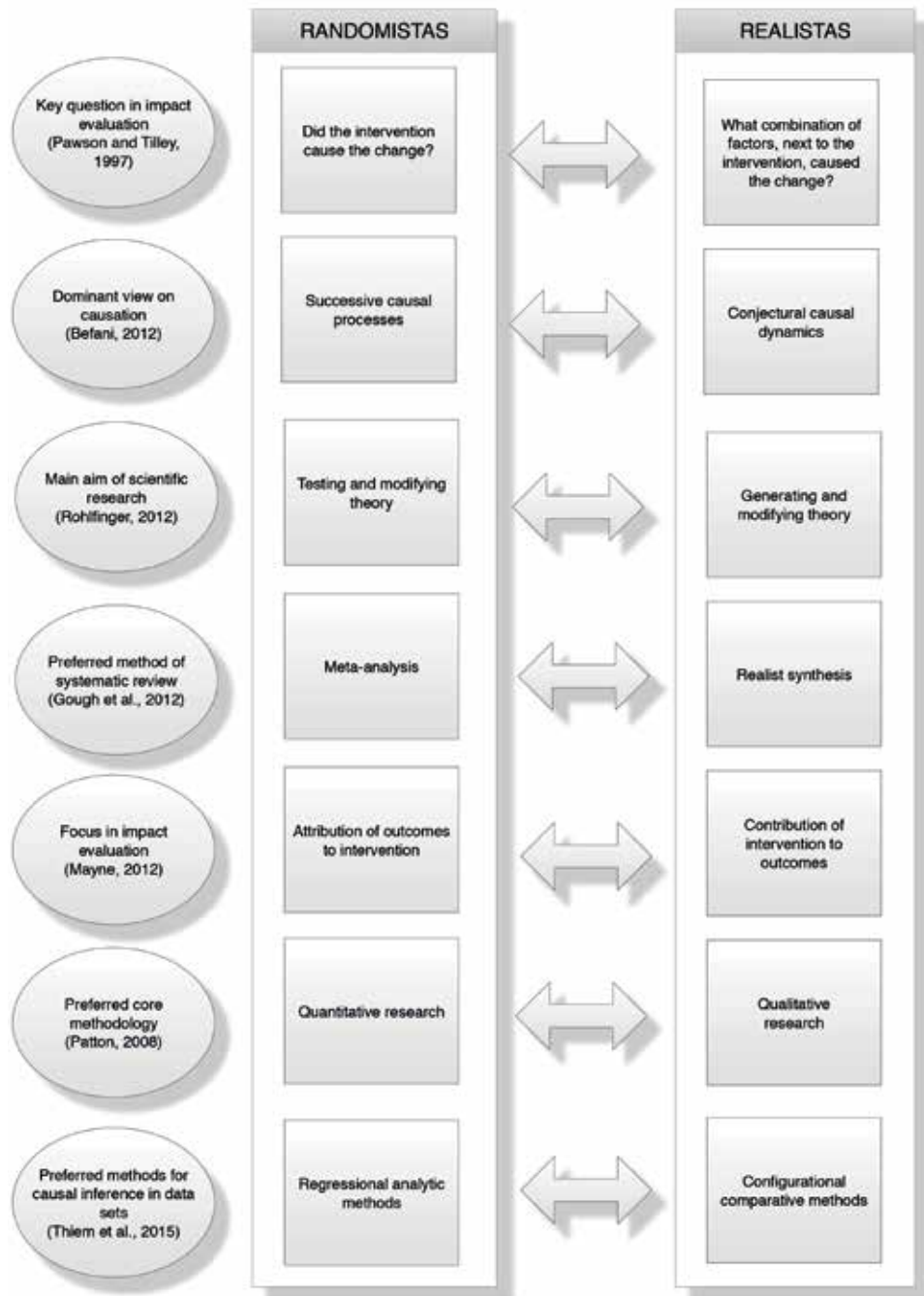


Note:  $X^{\{0\}}$  refers to the absence and  $X^{\{1\}}$  to the presence of condition  $X$ ;  $Y^{\{0\}}$  to the absence and  $Y^{\{1\}}$  the presence of outcome  $Y$ .

Source: Modified from Thiem, Baumgartner and Bol, 2015.

Figure 1.3 Two approaches to derive causal inferences from data set observations

For proper systematic review and meta-analysis, access to studies that found no effects or negative results is also needed. These studies tend to be relatively scarce due to publication bias: researchers, commissioners and journals prefer reports about interventions that were proven to work. Apart from the scarcity of studies with a research design having a low risk of bias in their effect estimates, systematic reviewers face the challenge that most studies published on development interventions document (partly) successful interventions, and very few studies relate



Source: Author's own elaboration

Figure 1.2 Stylised representation of two approaches to impact evaluation

to development interventions that failed. The ‘randomistas’ may have addressed this problem, at least partly, through the pre-analysis plans as requirement for funding or publication in high-ranked journals, at least for the analysis of the prime outcome of less-complex interventions (Olken, 2015). The ‘realistas’ are less inclined to test one theory with one design, for which pre-analysis plans are most suited. However, this explorative objective of research carries a high risk for data fishing and model fitting, which may result in spurious causes and effects, and add to the publication bias of studies that ‘find’ a causal relation between an intervention and an outcome, while these do not exist in reality. Pawson would emphasises the need for an organised, constant critical scrutiny of methods and research results (Pawson, 2013).

It is obvious that, based on the same quality of data and analysis, the more careful and cautious conclusions will have fewer threats to validity. However, ‘bolder’ conclusions are generally more appreciated by the commissioners of impact evaluations, and more easily communicated to a wider audience. This creates an additional tension for the evaluator, who needs to navigate between cautious academic conclusions, with a deliberately constrained generalisation domain to limit threats to validity, and bolder inference with a deliberately stretched generalisation domain and, thus, open to more validity threats.

## Tensions in economic farmer organisations

The second area of tensions is the central focus of the empirical research in this thesis: the inherent tensions and disintegrative tendencies in organisations. Economic organisations need to compete in the market with alternative institutional arrangements that channel production from the producer to the consumer. Transaction costs between producers and consumers in markets are the *raison d’être* for firms (Coase, 1937). If there were no transaction costs for farmers to sell their products, economic farmer organisations would not be needed. Firms and organisations will take up economic activities when they can do so more efficiently than alternative market channels (Williamson, 2000); they need to avoid the tendency to organise their services cost-effectively. Transaction costs are, thus, an important element to explain their existence but also their disappearance due to competition. Product attributes, technologies and logistics are important determinants of these transaction costs, and determine to a large extent the modes of inter-firm coordination in value chains (Gereffi et al., 2005).

However, firms and organisations are not uniform, rational, technical entities; they are made up of persons that interact. The quality of internal social relationships influences the performance of an organisation and the cost-effectiveness of its operations. Ashby et al., distilling principles of success of groups active in sustainable production and trade, call this the ‘internal social capital’ of a group (Ashby et al., 2009). I prefer the term ‘organisational social capital’ used by Leana and van Buren (1999) for this property of social interactions between persons in an economic organisation. Organisational social capital is defined as *‘a resource that reflects the character of social relations within the organization. It is realised through members’ levels of collective goal orientation and shared trust, which create value by facilitating successful collective action’* (Leana and Van Buren, 1999). They propose organisational social capital also as better construct than personal characteristics of the leader (‘leadership’) or procedures of internal organisation (‘bureaucracy’), often used in the literature to characterise organisational strength.



In this study, I intended to develop a measure of organisational social capital, appropriate for economic farmer organisations that are active in various economic sectors, with widely different technologies and logistic functions. Organisational social capital refers to the capabilities needed to resolve both the challenge of competition and the tensions between the group and supplying members.

Groups that work collectively to attain benefits for their members face a major challenge in the differing commitment of members. Passive members tend to benefit from the efforts of active members, while these active members bear the lion's share of the costs of collective action. The efforts of active members can be diverse: they may be an investment in cash or kind, but may also consist of time spent on organisational issues. This key problem of collective action was put on the research agenda by Olson (1965), who showed that smaller groups are better able to contain these tensions, through social pressure, than larger groups.

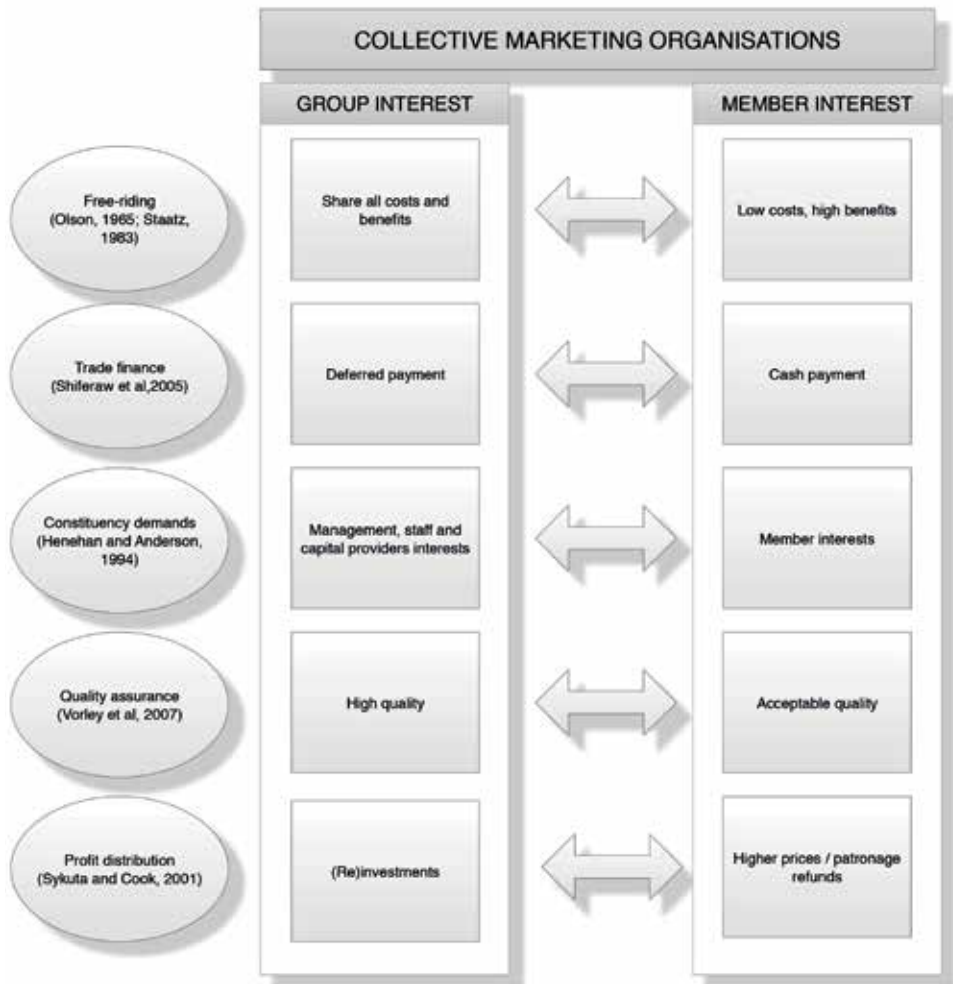
In developing countries, the main alternative to collective marketing is located in spot markets and their associated agent-trader networks, generally working in the informal economy (Fafchamps, 2004; Ton et al., 2010). Because formal institutions often do not provide financial services in rural areas, traditional traders respond to the immediate cash needs of farmers (Peppelenbos, 2008). The issue of working capital for cash payments to members needs to be addressed by any group willing to engage in collective marketing (Shiferaw et al., 2006; Ruben and Heras, 2012). A group needs to balance the collective costs of trade capital with the cash needs of the members.

Especially in larger and more developed economic farmer organisations, additional tensions may arise between the commercial negotiators or logistic operators, represented by the elected board members or professional staff, and the individual members that supply products (Henehan and Anderson, 1994), see Figure 1.4. Members must accept the prices and quality requirements that have been agreed upon with the buyer, as well as the deduction of a margin to pay for the collective marketing services. In their governance system, economic farmer organisations need to align the interests of different constituencies and groups of members with heterogeneous interests (Hendrikse and Bijman, 2002; Henehan and Anderson, 1994).

Other scholars point to agency dilemmas surrounding quality and quantity. The capacity to comply opportunely with the quality requirements of buyers proved key to successful linking of smallholders into modern markets, which increasingly 're-govern' the traditional market relations (Vorley et al., 2007; Reardon and Berdegúe, 2002; Bienabé et al., 2004; Bijman et al., 2011). However, farmers supplying commodities have an individual interest in also disposing of a portion of their sub-grade products (Ton, 2008).

Many scholars that study cooperatives stress the peculiar tensions related to the allocation of profits. Income from service provisioning to members is used to bear the cost associated with collective action and the remainder tends to be distributed to members, partly by increasing the price for the supplied products, and partly by profit redistribution. This twofold way of using economic rent makes economic farmer organisations different from conventional firms. Profit maximization, as the guiding strategy for private firms, is mixed with the objectives of maximizing turn-over and improving input price levels to members. Several other tensions may

exist. For example, some researchers also point to the trade-off between short term interests of members and targeted investments to seize business opportunities (Sykuta and Cook, 2001).



Source: Author's own elaboration

Figure 1.4 Common tensions in collective marketing organisations discussed in the literature

Organisational social capital is needed to contain these disintegrative tendencies inherent to collective marketing. Many organisations have this social capital. Bachman writes, similarly to Ostrom (1990), '*.. the simple fact that stable organisations exist, and even more so, that many of them are quite successful with their activities, can be taken as a strong indication that this co-ordination problem can be tackled and in fact effectively solved every day all around the globe*' (Bachmann, 2003: :58). Both Ostrom and Bachman point to the importance of trust, and

the importance of trust-enhancing mechanisms such as internal regulations and organizational procedures.

Not all these agency dilemmas will manifest themselves as a problem in the real-world activities of economic farmer organisations, and they also depend on the stage of the organisations' development (Henehan and Anderson, 2001). But organisations need to be resilient to the dynamics that may convert these agency dilemmas into a real threat to long-term survival. Internal rules and regulations need to balance trust and sanctioning, and be cost-effective. This implies a learning process similar to adaptive learning, where destabilization is both an inevitable part of the cycle of change, as well as an extraordinary learning opportunity (Karkkainen, 2005). Small, controlled perturbations in the system, such as grant-supported business plans, may spur innovation and learning to craft more resilient internal organisational systems.

**Table 1.1** Research design process in four impact evaluations

	Micro-irrigation technology supply	Training for certification (various projects)	Grants for smallholder innovation	Grants for processing and collective marketing
	2006-2010	2010-2014	2010-2013	2010-2014
Drawing the intervention logic	Kick-off meeting with commissioners and programme implementers, December 2007	Workshops with commissioners; focus group discussions with implementing agencies	Workshop with an international Advisory Board reflecting on impact pathways developed by the research team, April 2012	FONDOECAS Evaluation meeting, December 2011
Resulting key evaluation question	Does micro-irrigation technology supply increase farmer income?	Does (training for) certification result in better agricultural practices and improved income?	Are innovation grants to smallholders effective in facilitating agricultural innovation?	What is the relevance, effectiveness and efficiency of the grant fund, for whom and under what conditions?
Core methodology	Cohort design with before-after measurements	(Matched) difference in difference design with focus group discussions	Systematic literature review with realist synthesis	Comparative case studies on random sample
Mix of methods	Survey, focus group discussions, story harvesting	Survey, focus group discussions, monitoring data	Literature review	Case studies, time series data, monitoring data, survey
More information on the design process in:	Chapter 5, and Ton et al. (2012)	Chapter 3, and e.g., Waarts et al. (2013a; 2013b)	Chapter 2, and Ton et al. (2011a)	Chapter 4, 5 and 8

## 1.4 Study design

### Research objective 1: Identify design principles for credible and lean impact evaluation

For the first research objective, I used a theory-based impact evaluation design (White, 2009), which starts with a process to distil the ‘theory’ behind a support intervention and summarise it in an intervention logic. Generally, this step requires intensive consultation with the implementers and commissioners. Commissioners are key to defining the ultimate outcomes, whereas the implementers are key to defining the sequence of immediate and intermediate outcomes that are expected to result from their activities with specific stakeholder groups (Mayne, 2001). The implementers of these interventions were generally professional staff of development NGOs or farmer organisations. Their ways of participation were more varied than that of the commissioners. They often reacted to a draft intervention logic through e-mail contacts, focus groups, or multi-stakeholder meetings (see Table 1.1).

Intervention logics show many causal relations. To focus the impact evaluation, only the key causal assumptions in the intervention logic were considered for in-depth data collection; the remaining assumptions were less contested or less fundamental, and had less priority in the impact evaluation design. The overall framework used to organise the evidence on effectiveness is Contribution Analysis (Mayne, 2001; 2011; 2012). Mayne (2001) describes Contribution Analysis as a logical sequence of six steps to obtain a convincing ‘contribution story’. These steps describe an iterative process of building and refining the intervention logic. It identifies the key assumptions of impact that need to be verified and bolstered, gathers evidence to verify these; and reflects on results. The six key steps in Contribution Analysis are (Mayne, 2001; Mayne, 2012):

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

To seek out additional evidence to verify the selected key assumptions in the intervention logic (theory of change), I followed Brady and Collier (Brady and Collier, 2004; Brady et al., 2006), who state that causal inferences need both causal process observations and data-set observations. In practice, this meant that the designs included in-depth qualitative case studies and survey data. Doing so, I deliberately combined methods that are typically used by both ‘randomistas’ and ‘realistas’, the two archetypical evaluation approaches described in the preceding section. The preferred method by the ‘randomistas’ is the randomised control trial, considered to be the only truly ‘experimental’ design (Shadish et al., 2002; Khandker et al., 2009). Second best are quasi-experimental designs in which randomisation is used to select the treatment and comparison groups that are compared in time, such as difference-in-difference designs. The selection bias between both groups needs to be controlled for by instrumental variables, propensity score matching or other econometric techniques that use

the least-squares method (e.g., ANCOVA), or estimate a maximum likelihood function (e.g., Logistic regression). The more the treatment and comparison groups are similar, the better a credible average effect can be computed as the difference between both groups. In contrast, ‘realistas’ favour purposive sampling instances such as cases in a comparative case study design, because they look for the greatest diversity of contextual conditions that can explain effectiveness. The more diversity is covered in the data-set, the larger the generalisation domain of inferences about causal factors that need to accompany the intervention in order for it to become effective (context-mechanism-outcome configurations). Qualitative Comparative Analysis (QCA) was developed to analyse data sets for explorative research to identify causal ‘packages’. In short, the method creates a table with all possible combinations of factors included in the analysis, called ‘truth table’. The rows of the table show the number of observations that have this configuration of factors, and the outcome in question, in common. A Boolean minimisation algorithm searches for the most parsimonious formula to describe all the cases, combining rows of the truth table. This formula will have several terms, all potential causal configurations, and their respective consistency and coverage. I will describe QCA in more detail in Chapter 7.

## **Research objective 2: Develop and validate a measure to assess organisational strength**

For this research objective, I developed a tool that makes a ‘radiography’ of the capacities of organisations to address the inherent tensions between the members and the group in collective marketing. The design and field-testing of this tool is described in detail in Chapter 6. Through in-depth interviews with board members and staff, local researchers made detailed descriptions of the rules and regulations developed by each group to resolve agency dilemmas in a number of areas. I summarised the information in order to make a quantitative measure of organisational social capital. Therefore, each interview report ended with a one-page summary sheet, called ‘organisational radiography’, with two assessment questions, each with three answer options (see Annex 3) to be filled in by the researcher after the interview. The first question captured the ‘presence’ of each of the agency dilemmas in the practice of the organisation, and the second the ‘effectiveness of the organisational solution’. I used this information to derive a quantitative measure (TCC), see Table 1.2

This measure is field-tested by correlating it with economic performance indicators, and an independent ranking of FONDOECAS staff. The difference in TCC-score ( $\Delta TCC$ ) between two measurement moments in the same organisation (2011 and 2013) are compared with the qualitative information on the real dynamics and change processes documented in the in-depth interview reports.

**Table 1.2** Composition of the measure of Tension Containment Capacity (TCC)

	Question 1 (Q <sub>1</sub> )			Question 2 (Q <sub>2</sub> )			
Agency dilemmas	The tension comes up in the activities of our organisation	The tension hardly comes up	The tension never comes up	We managed to resolve the tension with agreements and organisational arrangements	We are looking for a way to resolve it	We do not need to resolve it	
T <sub>1</sub>	3 points	2 points	0 points	3 points	2 points	1 point	Score TC <sub>1</sub> =Q <sub>1</sub> *Q <sub>2</sub>
T <sub>2</sub>							TC <sub>2</sub>
...							...
T <sub>n</sub>							TC <sub>n</sub>
Tension containment capacity (TCC) =							$\sum_{i=1}^n TC_i$

### Research objective 3: Present credible empirical evidence on the effectiveness of FONDOECAS

I proposed a comparative case study methodology on a random sample of beneficiaries and non-beneficiaries. Randomness in sampling, rather than purposively selecting the sample, was expected to improve the generalisation domain of our conclusion, because I wanted to generalise the results to the entire population of eligible economic farmer organisations, in order to predict effectiveness and improve the targeting of grants in FONDOECAS. The sample of case studies initially selected consisted of 30 supported organisations and 20 unsupported ones.

Time series data on sales, membership and patrimony was collected, using the data available in the business plan proposals completed by the local researchers. The TCC-tool, described above, was applied in 2011 and 2013, in order to have baseline and end-line data on outcomes. Each case study was qualitatively analysed with process tracing methods (Beach and Pedersen, 2013) in order to explain the outcomes of the grant support in the organisations. This analysis includes counterfactual reasoning to discard alternative explanations (Yin, 2013; Vellema et al., 2013). After the reconciliation of two independent evaluations of each case, in 2014, I determined if the grant could be considered as a plausible contributing factor for these outcomes.

I analysed this information as a data-set with Qualitative Comparative Analysis software, fsQCA 2,5 (Ragin and Davey, 2009) and Kirq 2.1.12 (Reichert and Rubinson, 2014), and with logistic regression in SPSS 21.0 (IBM Corp., 2012). To prepare the data for QCA, I used the data-reduction technique Principal Component Analysis and converted the information into fuzzy-set conditions.

In Part 3 (Chapter 4, 5 and 6), I further detail the methodological design of the empirical research. In Part 4 (Chapter 7 and 8), I will reflect on the difficulties that I had in operationalising this design and the adaptations made. In the final part (Chapter 9), I will synthesise these

experiences, in order to distil some design principles as lessons learnt on the design of impact studies and discuss some limitations.

## 1.5 Thesis structure

The chapters in this book are all related to the discussion on methods in impact evaluation, especially on the objectives of impact evaluation (Part 2) and the challenges in impact research design (Part 3). The results of the empirical research on the FONDOECAS grant fund in Bolivia (Part 4) are followed by a conclusion (Part 5) which refers back to the issues and objectives covered in this introduction (Part 1). Figure 1.5 gives a graphical overview of this thesis structure.

### Part 1 – Introduction

#### *Chapter 1*

In this part of the thesis, I identify the context, debates and approaches that stimulated me during the research: the ‘tensions around the measurement of intentions’, in reference to the debates on impact evaluation, and the ‘intention to measure the tensions in organisations’, in reference to my quest to measure the organisations’ capacities to contain agency dilemmas in collective action.

### Part 2 – Taking stock

#### *Chapter 2*

Ton, G. , L. Klerkx, K. de Grip and M.L. Rau (2015). “Innovation grants to smallholder farmers: revisiting the key assumptions in the impact pathways.” *Food Policy* 51(1): 9-23.

This chapter reflects findings of my first systematic reviews, centred on the effectiveness of innovation grants to smallholder farmers. The peer-review process on the protocol for this systematic review was a good learning experience that deepened my insights into the differences between explorative realist evaluation and confirmative meta-analysis. The intervention ‘innovation grants’ was a container concept that included widely divergent interventions. Innovation grants are common as part of innovation policies but they are seldom evaluated as a separate component. The number of studies, therefore, was limited and the indicators used to evaluate their effectiveness differed greatly. Though we wanted to include a meta-analysis on similar interventions and similar outcome indicators, this was impossible. The paper documents the results of what became an explorative realist synthesis, using the available evidence to reflect on the causal assumption in the intervention logics and impact pathways.

*Chapter 3*

Ton, G., S. Vellema and L. Ge (2014). “The Triviality of Measuring Ultimate Outcomes: Acknowledging the Span of Direct Influence.” *IDS Bulletin* 45(6): 37-48.

This chapter underlines the importance in differentiating between outcome levels. The differentiation between immediate, intermediate and ultimate outcomes is useful to draw the ‘span of direct influence’ of an intervention, to reflect on the appropriateness of research designs with a focus on attribution of net-effects through quasi-experimental methods versus the verification of the claim to being a contributory cause in a complex process of change. The paper showed that even for an apparently simple intervention, that of farmers being supported by training for certification, the treatment was complex. We used baseline data to make statistical power calculations, which suggests that within real-world budgetary constraints, the precise measurement of net-effects on farmer income may be an unattainable goal. Instead of a focus on ultimate outcomes in farm income or yields, common indicators on intermediate outcomes, farmer knowledge and practices, could increase the possibilities of comparison among different support strategies and inform implementers on effectiveness. Contribution analysis is proposed as an overall approach to verify assumption in the intervention logic, combining precise net-effect measurement on intermediate outcomes with less precise, lean monitoring of indicators to verify the contributory role on outcomes that are outside the span of direct influence, such as poverty alleviation.

**Part 3 – Research design***Chapter 4*

Ton, G., S. Vellema and M de Ruyter de Wildt (2011). “Development impacts of value chain interventions: how to collect credible evidence and draw valid conclusions in impact evaluations?” *Journal on Chain and Network Studies* 11(1):69-84.

I wrote this chapter when I started with my PhD research, and worked together with my co-authors. We found common ground in our interest in the realist approach to study ‘mechanisms in context’ and theory-based evaluation. We discovered Chen’s difference in descriptive and normative programme prospective theory (Chen, 1994), the usefulness of realist case studies and causal process tracing (Perri 6, 2006), and Shadish, Cook and Campbell’s plea to add methods to reduce validity threats (Shadish et al., 2002). The chapter illustrates the arguments with design elements in the empirical research on the FONDOECAS grant fund.

*Chapter 5*

Ton, G. (2012). “The mixing of methods: A three-step process for improving rigour in impact evaluations.” *Evaluation* 18(1):5-25.

This discovery, and increased mastering, of approaches and methodologies to improve rigour in impact evaluation had implications for my work. In the usual very tight constraints on budgets, research logistics and time frames in applied research projects, the check on validity threats to the anticipated type of conclusion proved a useful tool to improve and focus impact evaluation assignments. We used it to harness research proposals on applied research work.



One of the additional benefits of this validity check was that it opened up the creativity of researchers having different research paradigms to accept different research methods as part of the impact evaluation design. The chapter presents two experiences of impact evaluation design in which this process has been used, for an intervention that consisted in micro-irrigation technology supply to smallholder farmers, and for the FONDOECAS impact evaluation.

#### *Chapter 6*

Ton, G. (2015). “Tension Containment Capacity: Towards a measure of organisational social capital in economic farmer organisations.” This thesis.

This chapter documents the field-test of a novel instrument– Tension Containment Capacity (TCC) – as a proxy for organisational social capital. I aimed to convert the tool into a useful cross-sectoral measure to assess organisational strength, which would permit benchmarking the effectiveness of various support modalities in a ‘lean’ way. The results of the tests point to five core agency dilemmas to be included in the TCC-measure for Bolivia: quality assurance, payment systems, side-selling, task delegation and political representation. The chapter explores the validity threats to the construct in longitudinal analysis, especially measurement error and researcher bias, when comparing independent measurements.

### **Part 4 – Empirical research**

#### *Chapter 7*

Ton, G. (2015). “Using Qualitative Comparative Analysis To Explore Outcome Patterns of Grant Support to Farmer Organisations in Bolivia.” This thesis.

The FONDOECAS grant system allocated grants to approximately 150 organisations between 2007 and 2014. I studied the outcomes of the grant in a sample of 26 of these, who had received a grant before 2011, exploring for conditions that could explain/predict success or failure of the grant to achieve intended outcomes. I explored this for three outcomes: improved market access for members, increased organisational strength, and capacity to pay organisational expenses. One of the conditions used as potential explanatory factor was these groups’ organisational social capital in 2011, with the TCC-score as their proxy-indicator, described above. Other candidate causal conditions were derived from the information available to the grant allocation committee when deciding on the business proposal, and related to group sales, group membership, patrimony and whether an organisation sourced their raw material from members or from non-members. The core method used in this explorative analysis, Qualitative Comparative Analysis (QCA), developed by Ragin (1987, 2008) in political science. The analysis is an illustration of the differences and complementarities of regression-analytic and configurational-comparative methods for exploring predictors of success or failure.

### *Chapter 8*

Ton, G. (2015). “Innovation Grants to Farmer Groups for Collective Marketing: A contribution analysis of the FONDOECAS small-grant fund in Bolivia.” This thesis.

This chapter reflects on the effectiveness of the grant system, using Contribution Analysis as evaluation framework, and a mix of methods for data collection and analysis. It uses a large household survey to verify the relevance of the support to economic farmer organisations. It uses time-series data combined with qualitative causal process tracing on 32 case-studies to verify effectiveness of the grants. And, it uses the monitoring data of FONDOECAS to reflect on the efficiency of the FONDOECAS grant allocation system. The results show that the assumption of relevance is largely supported by the data, even though it appears that market access is not the only, nor the main function of these organisations. The evidence on effectiveness is mixed, but especially disappointing when considering the expected increase in market access of members. Also, the efficiency of FONDOECAS’ grant allocation process seems quite low, although we could not compare and benchmark its efficiency with other grant funds.

## **Part 5 – Conclusions**

### *Chapter 9*

I use the results of the empirical research to reflect on the research objectives, and refine the methodological issues described in the introduction. Based on the research experiences, I identify some ‘design principles’ that I consider fruitful for impact evaluation of development interventions in value chains. I situate the approach to assess organisational social capital in the social science literature, and I reflect on the credibility of FONDOECAS research findings and discuss some limitations of the study.

PART 1: INTRODUCTION	PART 2: TAKING STOCK	PART 3: RESEARCH DESIGN	PART 4: EMPIRICAL RESEARCH	PART 5: CONCLUSIONS
<p>CHAPTER 1</p> <p>Introduction: tensions and intentions</p>	<p>CHAPTER 2</p> <p>Innovation Grants to Smallholder Farmers: Revisiting the key assumptions in the impact pathways</p> <p>Ton, Klerck, de Grip and Rau (2015): Food Policy 51(1): 9-23</p> <p>CHAPTER 3</p> <p>The Triviality of Measuring Ultimate Outcomes: Acknowledging the span of direct influence</p> <p>Ton, Vellema and Ge (2014) IDS Bulletin 45(6): 37-48.</p>	<p>CHAPTER 4</p> <p>Development Impacts of Value Chain Interventions: How to collect credible evidence and draw valid conclusions in impact evaluations?</p> <p>Ton, Vellema and deRuiterdeWit (2011) Journal on Chain and Network Studies 11(1):99-84</p> <p>CHAPTER 5</p> <p>The mixing of methods: A three-step process for improving rigour in impact evaluations.</p> <p>Ton (2012) Evaluation 18(1):5-25</p> <p>CHAPTER 6</p> <p>Tension Containment Capacity: Towards a measure of organisational social capital in economic farmer organisations</p> <p>Ton This thesis</p>	<p>CHAPTER 7</p> <p>Using Qualitative Comparative Analysis to Explore Outcome Patterns of Grant Support to Farmer Organisations in Bolivia</p> <p>Ton This thesis</p> <p>CHAPTER 8</p> <p>Innovation Grants to Farmer Groups for Collective Marketing: A contribution analysis of the FONDOECAS small-grant fund in Bolivia</p> <p>Ton This thesis</p>	<p>CHAPTER 8</p> <p>Synthesis and discussion</p>

Source: Author's own elaboration

Figure 1.5 Thesis structure



## 2 | Innovation Grants to Smallholder Farmers: Revisiting the key assumptions in the impact pathways<sup>1</sup>

Giel Ton<sup>a</sup>, Laurens Klerkx<sup>b</sup>, Karin de Grip<sup>a</sup> and Marie-Luise Rau<sup>a</sup>

<sup>a</sup>. Agricultural Economics Research Institute, LEI Wageningen UR

<sup>b</sup>. Knowledge, Technology and Innovation Group, Wageningen University, The Netherlands

### Abstract

*Grant funds specifically targeted to smallholder farmers to facilitate innovation are a promising agricultural policy instrument. They stimulate smallholders to experiment with improved practices, and to engage with research, extension and business development services providers. However, evidence on impact and effectiveness of these grants is scarce. Partly, because attribution of changes in practices and performance to the grant alone is challenging, and the grant is often invested in innovation processes that benefitted from other support in the past. We discuss three modalities: vouchers, business development matching grants and farmer-driven innovation support funds. Our review points to an important and transversal outcome area of innovation grant systems: the creation of human and social capital to sustain creative thinking and innovative practices. Harmonising measurement on these outcomes could enhance the usefulness and comparability of impact studies and facilitate benchmarking of different policy options for smallholder innovation.*

---

<sup>1</sup> Published as: Ton G, Klerkx L, de Grip K, et al. (2015) Innovation grants to smallholder farmers: Revisiting the key assumptions in the impact pathways. *Food Policy* 51: 9-23.

## 2.1 Introduction

There is widespread consensus that users need to be endowed with decision-making authority to influence the research processes and other service provision (extension, business services, inputs) that support innovation (Douthwaite, 2002; Klerkx and Leeuwis, 2008; Neef and Neubert, 2011; Poulton et al., 2010). This makes it essential to have, besides financial support for the formal research, extension and business development organisations, research and extension approaches to support experimentation and innovation for and by smallholder agricultural producers (Hall et al., 2007; Wongtschowski et al., 2010). Innovation grant funds are receiving increasing recognition as a promising avenue for agricultural innovation (World Bank, 2012). Nevertheless, funds that are specifically targeted to smallholder farmers are quite rare.

Grants for agricultural innovation are used to stimulate private sector and farmer engagement in activities related to technology generation, technology dissemination and overall innovation processes. The increased use of innovation grants in the last decade is a result of two tendencies that shape policies on agricultural extension and advisory services. Firstly, many countries have shifted to a more demand-led agricultural research system, in which users of research have a voice in determining research and innovation priorities or even decision-making authority (Klerkx and Leeuwis, 2009; Neef and Neubert, 2011). Simultaneously, also extension and business development support systems (including input supply) are moving towards demand-driven systems (Kilelu et al., 2011; Kilelu et al., 2013; Minh et al., 2014). Secondly, there is growing awareness that agricultural development is not only driven by technology produced by agricultural research but also encompasses organisational and institutional change (Hounkonnou et al., 2012; Klerkx and Nettle, 2013). Agricultural innovation is, therefore, not only about adopting new technologies; it also requires a balance among new technical practices and alternative ways of organising, for example, markets, labour, land tenure and distribution of benefits (Adjei-Nsiah et al., 2008; Dormon et al., 2004; Pamuk et al., 2014). Agricultural innovation is a co-evolutionary process, i.e. combined technological, social, economic and institutional change (Kilelu et al., 2013), which may be both driven by top-down interventions, and bottom-up farmer's grassroots activities (Smith et al., 2014). The process of obtaining and using the grants stimulates smallholders to be more pro-active and critical towards research and extension providers instead of being passive recipients of top-down technological recipes (Heemskerk and Wennink, 2005; Rivera, 2000). A key premise is that the separation of the funding of the research, extension or business service provision from the provisioning of the research would make service provision more demand-driven. Also, because, in a market setting, several providers may compete for the contract, this would enhance the performance of the provider and its orientation towards the wishes of the smallholder clients (Klerkx and Leeuwis, 2008; Klerkx et al., 2006).

While there has been considerable policy attention to the importance of co-funding the innovation processes by smallholders, and governments and donors are experimenting with different grant modalities, there is little and dispersed information on the impact and effectiveness of these grants in facilitating agricultural innovation. In 2011 we reviewed the studies that analysed the impacts of innovation grants to smallholders in developing countries (Ton et al., 2013b). The systematic review combined an electronic search in academic data-bases with follow-up searches of grey literature.

To facilitate comparative analysis, we divided the innovation grant systems into three types, each with a different funding modality and objectives.

*A = Voucher systems:* These provide grants directly to the end-users to enable them to procure goods (e.g. fertilizers) or services (e.g., research and extension) either in the form of vouchers that represent a certain monetary value or through reimbursement of investments after proof of the transaction has been provided. Users can try out a service without investment risk, create access for people who previously did not have sufficient purchasing power, and facilitate a relationship of accountability between the service provider and the client (Bebbington and Sotomayor, 1998b; Kidd et al., 2000; Klerkx et al., 2006). For proper functioning of voucher systems, potential users of services must learn to identify and articulate their needs, negotiate with service providers and judge and control service quality, and service providers must have the right skills and knowledge to provide the required services. In order to ensure the longevity of demand-side financed extension systems, generally a financial contribution from the end-user is required.

*B = Business development matching grants:* Often these grant fund ask for business proposals for which co-funding is needed. These grants are seldom directed to individual smallholders but to organised groups, like cooperatives, associations or village organisations that coordinate input provisioning, marketing or added-value production (Donovan et al., 2008; Poulton et al., 2010; Ton et al., 2014a; Yang et al., 2014). Often international donors, like IFAD or World Bank, contribute development funds that are used to establish dedicated governmental and non-governmental business plan competitions. These tend to collaborate with intermediary organisations to help local groups of smallholders to generate a feasible business proposal eligible for funding. These grant systems do not focus on a predefined menu of technological options, and therefore are more flexible and functional for smallholder specialisation in markets.

*C = Farmer-driven agricultural innovation funds:* As Neef and Neubert (2011) argue, one important dimension of participatory research and innovation is the extent to which farmers have an institutionalised influence on the whole process of research agenda setting (i.e. query generation, prioritization and fund decision making – see Klerkx and Leeuwis, 2008) and research execution. Farmer-driven funds are those grant funds where smallholders take part in the governance of the grant fund. This gives farmers the possibility to determine what type of research is needed and to represent the interests of smallholders in their relationship with research providers. Often, these funds are multi-stakeholder partnerships in which there is a facilitating role by a research institute or development NGO (Gandarillas et al., 2007; Klerkx and Leeuwis, 2009). This facilitation may conflict with the capacity of the farmer organisations in effectively influencing the governance of the innovation fund, creating blurred line between farmer-driven or facilitator-driven agricultural innovation funds (Cordoba, 2014). Therefore, we only included studies on those grant funds where, according to the studies, smallholder organisations had a visible and active role in the grant governance system.

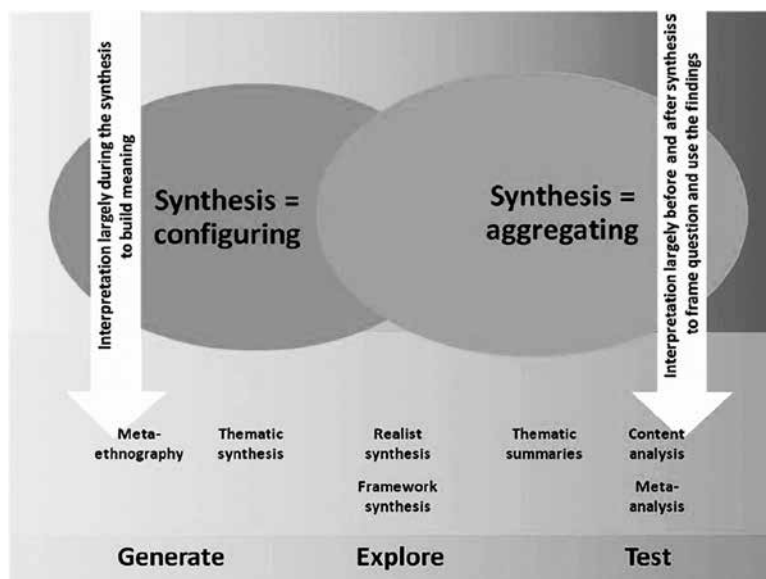
Each type has its specific way(s) of facilitating innovation. To review the evidence on effectiveness in facilitating agricultural innovation, as recommended by Snilstveit (2012) and White (2011), we developed a core impact logic for each of these types. These impact logics relate

to the causal steps that are expected to translate the grant for innovation into outcomes for smallholders (Fig. 2.2, Fig. 2.3 and Fig. 2.4).

## 2.2 Methods

The paper discusses the findings of a systematic review of impact studies done in 2011–2012, published as Ton et al. (2013b), that synthesised the available literature in order to explore under what conditions innovation funds to smallholder farmers tend to be effective in facilitating agricultural innovation.

There are widely divergent methods of systematic review (Gough et al., 2012; Pawson, 2002; Thomas and Harden, 2008). Thomas et al. (2012) differentiate between those reviews that aggregate the evidence in studies on similar treatments to make more generic inferences (meta-analysis), explorative systematic reviews that make a typology of the evidence provided to reflect on causal pathways (realist synthesis, framework synthesis) and interpretive systematic reviews that makes sense of the literature without an a priori defined framework or typology. Fig. 1 presents this methodological continuum of approaches to systematic review



Source: Snilstveit et al. (2012). Based on Thomas et al. (2012)

Figure 2.1 Methodological continuum of synthesis approaches and methods

Our explorative systematic review organised the studies according to core impact logics, which reflect different rationales behind the support, and associated with different grant implementation modalities. Within this framework, we reviewed the evidence on effectiveness and the information on the processes and conditions that influenced the effectiveness of the

funds. We wanted to avoid a situation, common to several systematic reviews on international development interventions (Hagen-Zanker, 2012), where a sole focus on studies with a (quasi)-experimental design reduces the richness of information in such a way that it proves fairly uninformative for practitioners that want to learn about the reasons why some grant systems are more effective than others. The low number of rigorous studies that remains after a very strict screening of the impact evaluation research design may negatively affect the possibility to provide answers to this question (Woolcock, 2013). Therefore, we included, in addition to the quantitative effectiveness studies, also more process-oriented qualitative studies in our systematic review (see Table 2.1).

**Table 2.1** Mapping of retrieved studies

Type of study	A: Voucher systems	B: Business development grants	C: Farmer-driven innovation support funds	General studies	Subtotal
Impact studies	8	3	9	0	20
Outcome monitoring reports	0	0	6	0	6
Descriptive studies	5	5	17	9	36
Subtotal	13	8	32	9	62

The search strategy, as defined in the systematic research protocol (Ton et al., 2011a), resulted in 4322 studies retrieved from the electronic data bases. As can be expected, when using a wide number of broadly defined search terms, like ‘innovation’ and ‘experimentation’, most of these studies had no relation at all to the object of our study. The screening of the information in the title and abstract, using the EPPI Reviewer 4 software application (Thomas et al., 2010), helped to reduce the number of relevant studies that were retrieved from the electronic data-bases to 227. Additionally, 41 studies were retrieved through searching of web-sites of development organisation, international research institutes and practitioner networks. The full-text screening of these 268 studies resulted in 53 studies included in the review. Most studies that were excluded in this full-text screening had no information on the grant system, had no grants to smallholder farmers or were not directed to smallholder farmer innovation. Based on the names of the authors and the grant system, we retrieved 11 more studies, making a total of 62 studies.

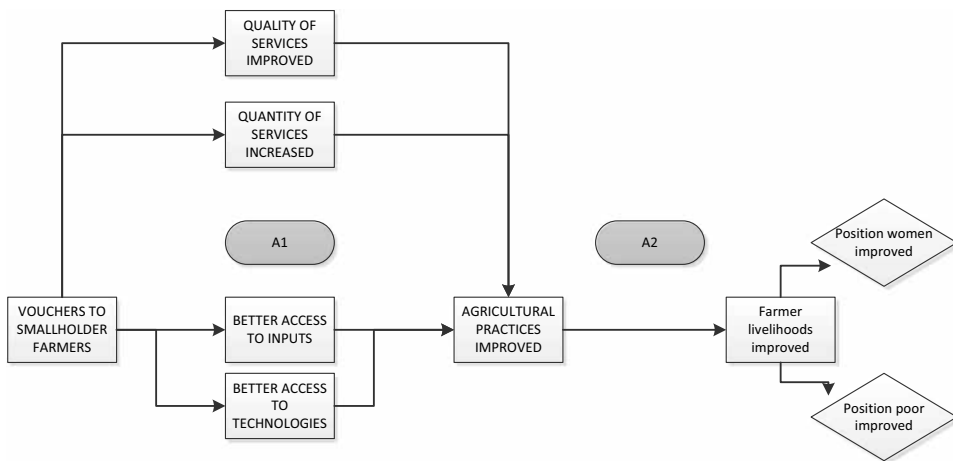
We mapped the relevant studies in three categories: (1) impact studies – studies with a structured process of data collection on outcomes of the grant system on agricultural innovation; (2) outcome monitoring reports – studies that present monitoring data without conclusions about impacts or effectiveness of the grant system; (3) descriptive studies – studies that discuss the merits and effectiveness of grant systems but without a systematic way of presenting evidence on outcomes of the grant system on agricultural innovation (see Table 2.1). Several of these studies were working papers that have been published in academic journals later on. In this article we refer to the published version when possible. For each impact study, we scrutinised the way in which evidence on outcomes was collected and how claims of attribution were made. We listed the outcome areas and proxy-indicators used in the impact evaluation.



## 2.3 Results

The twenty impact studies included in the synthesis related to a relatively small number of eleven empirical innovation funds. However, as we show in this paper, they had widely divergent indicators of outcomes, which prevented the use of systematic comparisons, like quantitative meta-analysis. Our synthesis of the evidence became essentially explorative and qualitative in nature. We discuss the evidence on the effectiveness of grant systems according to these three modalities, and illustrate them presenting some empirical instances in text boxes (Box 2.1–2.5).

In the Table 2.2, Table 2.3 and Table 2.4, we present the evidence, outcome areas covered in the impact study and the proxy-indicators used to assess changes in these outcome areas. For each of these proxy-indicators we indicate the direction of change and the rigour of the underlying study design. Authors often draw conclusions from a long list of outcome indicators in one and the same study, and these are not always assessed with the same research design. Therefore, a study can be strong on the measurement of impact on one proxy-indicator, while being weak on another. We considered a method as being strong in rigour when the design had a process to assess the counterfactual with a design that addressed the issue of selection bias – the special characteristics of being a beneficiary of the grant. We considered a study to be moderate in rigour when a comparison group is used but without a procedure to eliminate the most obvious sources of selection bias. A study is considered weak when only the change in the beneficiary group are reported on, or comparisons are made with population averages.



Source: Authors' own elaboration

Figure 2.2 Impact pathway A: voucher grant systems

## Impact pathways type A: Voucher grant systems

This type of innovation grant provides vouchers to distribute subsidies on inputs, technologies and/or services to trigger innovation in agriculture. For example, voucher programmes are used to subsidise the distribution of quality seeds and fertilisers, to promote micro-irrigation, to hand out tools and seeds after conflicts or natural disasters, to distribute heifers in dairy expansion programmes, etc. While in the absolute sense the degree of innovation might seem low, at the local level it does imply major changes in the socio-institutional and technical agricultural system around smallholder farming, and thus facilitates innovation at local level. The objective of input voucher programmes is to impact directly by improving on-farm production, productivity and income/food security. The vouchers are a way to target the subsidies to the recipient groups. A subtype of voucher scheme targets the development of an enabling institutional environment for farmers to produce. Fostering demand from smallholders, the vouchers are used to encourage a sector of service providers to develop knowledge and routines to target farmers, such as private extension services or business development services. This triggers the development of institutions and institutional arrangements that facilitate the innovation by farmers. Vouchers provide a means of ‘incubating’ a service sector for farmers and an incentive for experimenting with these services by farmers. Generally, they are intended to develop a sector that becomes economically sustainable when the voucher system ends.

Vouchers are used to increase the uptake of inputs or support use of services by a target group. The key causal steps in the impact logic behind voucher grant systems are related to the way that impact on agricultural practices is realised, and the way that these practices translate to improvements in farmer livelihoods.

- Causal assumption A1: *The quantity and quality of inputs and services provided to smallholder farmers are enhanced as a result of the voucher system and can be sustained in the future.*
- Causal assumption A2: *Farmers’ livelihoods, and in particular those of the poor and women, start to change as a result of the improved agricultural practices enabled by these inputs and services.*

The included impact studies (see Table 2.2) are all independent evaluations and have random selection of respondents as part of their research design. Outcome indicators that proxy for agricultural innovation by smallholders are input use, especially fertiliser and improved seeds. Other proxy-indicators for farmer well-being covered in the study by Holden and Lunduka (2012) are crime levels, health and education, food security – though with evidence collected and analysed using weaker research methods than for the production and income indicators. As an indicator to monitor impact in the institutional organisation in the value chain, all impact studies mention the impact of the input voucher on agrodealers.

Table 2.2 Grant impact studies on voucher schemes

Grant system	Evidence base	Outcome areas	Proxy-indicators	Change	Methods used	Rigour	Independent evaluation		
Malawi input starter pack	Holden and Lunduka (2010b)	Input use	Application of fertiliser	Positive	Matched comparison	Strong	Yes		
			Use of improved seeds	Positive	Matched comparison	Strong			
Sierra Leone – war rehabilitation	Holden and Luduka (2010a)	Biodiversity effects	Intensified maize production	Positive	Matched comparison	Strong			
			Application of manure	Positive	Matched comparison	Strong			
		Input market	Indigenous tree vegetation on farms	Negative	Matched comparison	Strong			
			Intercropping system in maize	Positive	Matched comparison	Strong			
		Farmer livelihoods	Size of secondary market for inputs and input vouchers	Negative	T-test	Strong	Yes		
			Self-perception on food security	Positive	T-test	Strong			
		Natural resource base	Agricultural production	Positive	Regression	Strong			
			Perceived food security	Positive	Simple tabulation	Moderate			
			Health/school	Positive	Simple tabulation	Moderate			
			Crime level	Negative	Simple tabulation	Moderate			
Ricker-Gilbert and Jayne (2009) (Dorward et al., 2008)	Farmer livelihoods	Assets accumulation	Neutral	Regression	Strong				
		Area planted	Neutral	Simple tabulation	Moderate				
		Use of organic manure	Neutral	Simple tabulation	Moderate				
		Use of fertiliser	Positive	Regression	Strong	Yes			
Kenya inputs access programme	KENFAP (2010)	Farmer income	Maize production per acre	Positive	Panel survey	Moderate	No		
			Gross margin	Positive	Panel survey	Weak			
		Innovation system	Effective co-financing extension services by farmers	Neutral	Interviews	Strong	Yes		
			Household income	Positive	Random survey	Moderate			
		Chilean extension vouchers	Bebbington and Sotomayor (1998a) referring to unavailable study by MIDEPLAN (1994)	Input market	Household food security	Positive	Regression and simulation	Strong	
					Number of agrodealers	Positive	Simulation	Strong	
				Input use	Changes in sales volume per agrodealer	Positive	Time-series	Strong	
					Use of fertiliser	Positive	Regression	Strong	
				Farmer organisation	Knowledge on composition village council	Positive	Interviews	Moderate	Not clear
					Perception of function	Positive	Interviews	Moderate	
Seed diversity	Quality of deliberation	Positive	Observations	Strong					
	Number of rice varieties used per farmer	Positive	Interviews	Moderate					
Kenya inputs access programme	KENFAP (2010)	Farmer income	Range of seed varieties requested	Positive	Interviews	Strong			
			Maize production per acre	Positive	Panel survey	Moderate	No		
Chilean extension vouchers	Bebbington and Sotomayor (1998a) referring to unavailable study by MIDEPLAN (1994)	Innovation system	Gross margin	Positive	Panel survey	Weak			
			Effective co-financing extension services by farmers	Neutral	Interviews	Strong	Yes		
Chilean extension vouchers	Bebbington and Sotomayor (1998a) referring to unavailable study by MIDEPLAN (1994)	Farmer livelihoods	Household income	Positive	Random survey	Moderate			
			Household income	Positive	Random survey	Moderate			

**BOX 2.1 MALAWI AGRICULTURAL INPUT SUPPLY PROGRAMME**

The Malawian government started implementing the Agricultural Input Supply Programme (AISP) in the 2005/2006 season with the stated objectives of improving smallholder productivity and food and cash crop production, and of reducing vulnerability to food insecurity and hunger. Other objectives were promotion of food self-sufficiency, development of the private sector input markets, and wider growth and development. Different suppliers offered different pack sizes of OPV (open pollinated variety) and hybrid seed and fertilisers (2 kg of hybrid seed or 2 kg or 3 kg of OPV seed, depending on supplier costs). The seed system introduced an element of farmer choice, with competition between suppliers. In the 2006/2007 growing season, the programme allocated two million seed and three million fertiliser coupons to districts and areas within districts for distribution to targeted households.

All impact studies on the Malawi voucher scheme (see Box 2.1) – Dorward et al., 2008, Holden and Lunduka, 2010a, Holden and Lunduka, 2010b, Holden and Lunduka, 2012 and Ricker-Gilbert and Jayne, 2009 - present evidence for positive impacts on yields and household income when farmers used new seeds and fertiliser. This observation has been affirmed in more recent review articles by Ricker-Gilbert et al. (2013). The impact of the vouchers on farmers' asset accumulation is less strong, implying that the cash generated from production (or from the sale of the subsidised inputs on the market) was probably spent on household expenses and food security. The studies on voucher systems show ample evidence that the vouchers indeed lead to the uptake of practices that enhance innovation in the smallholder farming system. The impact studies support the assumption in the impact pathway that vouchers increase inputs or services to farmers. The vouchers facilitate the growth of an agro-input 'market': the vouchers are a way to establish input supply chains in rural areas that need a threshold market demand; that is, they provide an effective demand for inputs for private investors (agrodealers) to come in with their investments.

Whereas the link between fertiliser use and yields seems quite positive, the picture of impacts on poverty (income) is less clear. Though it seems plausible that the voucher scheme has a positive impact on income, the analytical techniques to capture this impact (or refute the claim of impact) are prone with methodological challenges, especially those related with the definition of the treatment (e.g. Do the authors measure the access to or use of the voucher or to the use of the voucher or to the application of the inputs?) and the instrumental variables used to control for selection bias (Ricker-Gilbert et al., 2013).

The main mechanism that can prevent the increased yield from translating into better well-being is the impact of a better harvest on farm prices. Market conditions are thus especially important as a moderating factor. In some regions of Kenya, market prices are mentioned as being negatively affected by the increased supply of the crop to remote local markets (KENFAP, 2010) in such a way that these locations did not benefit. Even in these places, however, the farmers that apply the package will be better off than those who do not, because the additional yields would provide these farmers with more income and food than farmers who had not applied the package. Most studies, therefore, point to the necessary complementing of a voucher scheme with effective market-stabilising local institutions and infrastructure, such as

storage facilities and roads and regional trading networks. Without this market infrastructure, a rapid increase in production of one specific crop in an isolated locality can lead to a very low price and provide negative incentives for future farm investments.

All impact studies mention the possibility that cash transfers instead of vouchers could have similar effects on food security and household well-being, though they would have had less effect on increased agricultural production. If farmer livelihoods and well-being were the sole objective of the implementing agency, cash transfers would be an alternative option that would increase farmer decision making on the grant use.

For a voucher scheme for seeds in Sierra Leone, (Richards, 2007) points to the risk that the distribution through vouchers of a 'one size fits all' seed variety may tend to reduce farmer experimentation instead of facilitating it. The technology is introduced in a context where farmer innovation practices are already in place, e.g. the simultaneous cropping of a diversity of seeds in small plots, a practice that might be lost due to the cheap flow of the variety provided through the voucher system. The use of the voucher in a context of choice may remediate this. Linking up with seed fairs seems an effective way to provide a choice of seeds; this is a very promising activity that embeds the vouchers in a broader context of local farmer innovation (Remington et al., 2002), and generates possibilities for enabling both external certified seeds and locally improved varieties, in addition to possibilities for using the same venues to provide access to other technologies, such as ox-traction and storage facilities, that could trigger innovation by smallholders.

Though vouchers stimulate the settlement of agrodealers in rural areas, there are also victims of these dynamics (Govere et al., 2009; Holden and Lunduka, 2010b). Competition between established and new agrodealers can force some previously existing agrodealers to shut down, especially when they are bypassed by the voucher system for political reasons. In Zambia, Xu et al. (2009) found that input vouchers stimulated agro-dealers in remote areas but crowded out the existing commercial fertilizer distribution in areas where agro-input dealers were already in place. For Malawi, Holden and Lunduka (2010b; 2010a) suggest that the emergence of a secondary market of inputs is a threat to existing input-provisioning channels but do not provide evidence for this in their paper. More convincing is the argument that this secondary market undermines the targeting mechanism. The studies show that without effective targeting mechanisms to ensure they benefit the current non-users of inputs, the distribution of vouchers tends to be directed to the farmers who already use the inputs and technologies, substituting part of their cash expenses with government subsidy support, without facilitating agricultural innovation per se (Mason and Ricker-Gilbert, 2012). Voucher schemes that want to avoid subsidising farmers that already use these technologies need to target only the group of smallholders that is currently not using the inputs and can be expected to start doing so as a result of the vouchers (Dorward and Chirwa, 2011; Ricker-Gilbert et al., 2013).

There is a risk that vouchers are allocated in ways that strengthen existing power relations of exclusive clans (Richards, 2007) or influence party politics (Banful Afua, 2011; Ricker-Gilbert and Jayne, 2009). Other studies in the East African region (Pan and Christiaensen, 2012) confirm this bias in the allocation of the voucher due political and social networks that tend to give some farmers better access to the vouchers than farmers without this social capital.

However, vouchers can also be used to change power relations. Richards (2007) gives an example where he points to the importance of transparency and ‘ritual’ in the distribution of seeds and inputs as a way to build more robust local institutions that might take up other roles and functions than ‘just’ channelling input subsidies. Dorward et al. (2008) and Denning et al. (2009) provide evidence that there might be some limits in the mechanisms to target the beneficiaries of the government subsidies, although this does not imply that there is a negative impact due to the voucher schemes on the related local institutions.

### ***Conclusions on the assumptions in the impact logic for ‘Voucher schemes’***

Vouchers are used to increase the uptake of inputs or support services by a target group. They represent a certain monetary value with which purchases can be made or because farmers get a reimbursement of investments after proof of the transaction has been provided. As farmers become direct purchasers of inputs, a market develops in which there is a better match between demand and supply. The key causal steps in the impact logic behind voucher grant systems are related to the way that impact on agricultural practices is realised, and the way that these practices translate to improvements in farmer livelihoods.

Causal assumption A1: *The quantity and quality of inputs and services provided to smallholder farmers are enhanced as a result of the voucher system and can be sustained in the future.*

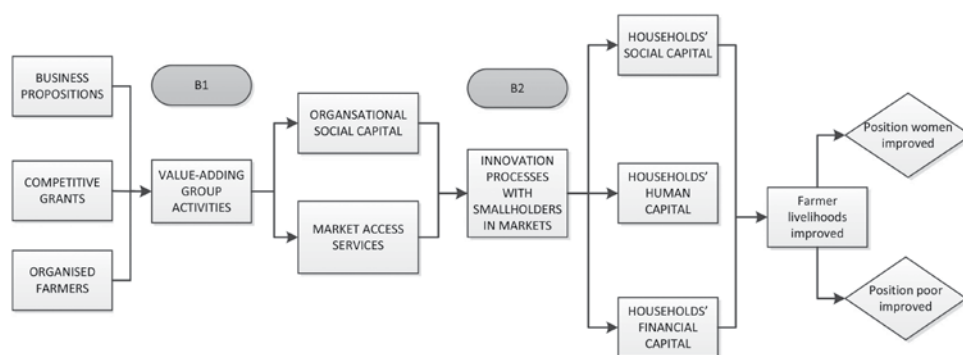
The studies on voucher systems show ample evidence that the vouchers indeed lead to the uptake of practices that enhance innovation in the smallholder farming system. Effective targeting mechanisms to reach non-users are key.

Conclusion: *strong support in studies.*

Causal assumption A2: *Farmers’ livelihoods, and in particular those of the poor and women, start to change as a result of the improved agricultural practices enabled by these inputs and services.*

The studies show positive impact on key elements of farmer livelihoods, except when prices fall in response to an increase in production in a context of limited markets outside the production area. The content of ‘one size fits all’ technology package supplied through a voucher system could constrain agricultural innovation, while offering a menu of options to choose from would enhance innovation.

Conclusion: *moderate support in studies.*



Source: Authors' own elaboration

Figure 2.3 Impact pathway B: business development matching grants

### Impact pathways type B: Business development matching grants

Increasingly, the execution of the agricultural support activities is delegated to implementing agencies through competitive grants systems (Garforth et al., 2003; Klerkx and Jansen, 2010). Through 'matching grants', these funds provide co-financing for agricultural business development (see Box 2.2). These activities can vary from research and extension support for companies and farmer organisations as well as for training workshops, pre-professional internships, and even direct subsidies, to necessary infrastructural investments. Generally, these grant funds are managed by decentralised, specialised governmental entities.

Many value-chain development projects have a grant component to help farmers overcome threshold investments hurdles to entering other (urban, regional, international) markets. Business plan competitions are a common term for this type of grant system. The short-term outcomes of these grants are not necessarily located in the farmer households but related to the economic and organisational performance of the group/business. Mid-term direct outcomes for farmers' livelihoods are reflected in better prices and increased sales through the marketing arrangement.

Business development matching grants are intended to enable farmer organisations to seize business opportunities that facilitate innovation processes in rural areas. The key assumptions relate to the impact on the capabilities of the group and the impact that these have on farmer livelihoods.

- Causal assumption B1: *Competitive matching grants trigger value-adding business activities by (groups of) farmers as a way to facilitate innovation processes for smallholder farmers in markets.*
- Causal assumption B2: *Farmers' livelihoods improve as a result of social activities and economic returns derived from the new value-adding business activities.*

Van der Meer and Noordam (2004) reviewed the World Bank portfolio of projects to address market failures, in which competitive grants for business development, ‘productive-type projects’ are a minor though growing part. They concluded that very few studies look at outcomes that have an economic character; most outcomes reported for this type of project are of a qualitative nature. Likewise, the Donor Committee for Enterprise Development (DCED) calls for more attention and more methodological rigour in monitoring and evaluation the outcomes of these type of funds Kessler (2013). However, it is difficult to capture the effects of these grants, especially as the number of beneficiaries of the business opportunities in the short term tends to be limited and comparison with non-supported business lack meaning due to essential heterogeneity. And, even more important, the effects of the business plans need time to mature. Scale can be reached only after some time, when other support and a range of other market factors will have complemented the grant support. Counterfactual designs with control groups at the level of the household are therefore ineffective to capture the impact of business grants, as these outcomes are beyond the span of direct influence (Ton et al., 2014c).

**Table 2.3** Grant impact studies on business development grants

Grant system	Evidence base	Outcome areas	Proxy-indicators	Change	Methods used	Rigour	Independent evaluation
Corredor Puno – Cusco	Sotomayor et al. (2008)	Farmer livelihood	Farmer sales volume	Positive	Self-assessment	Moderate	No
			Food security	Positive	Household survey	Strong	
			Household income	Positive	Household survey	Strong	
		Farmer organisation	Use of business planning tools	Positive	Self-assessment	Moderate	
Inspección de Calidad Agrícola (INCAGRO), Peru	IEG-World Bank (2009), referring to Escobal unpublished data (2003, 2005)	Farmer livelihood	Net income per hectare	Neutral	Matched comparison	Moderate	Yes
			Producer income	Positive	Household survey	Strong	
			Technology adoption	Positive	Household survey	Moderate	
Centros de Gestión, INDAP, Chile	Fundación Chile (2009)	Farmer organisation	Profits	Positive	Business survey	Weak	No
			Use of business planning tools	Positive	Business survey	Weak	



**BOX 2.2 MATCHING GRANTS TO FARMER GROUPS IN LATIN AMERICA**

Most impact studies focus on the experiences in Latin America, where in the last decades quite similar business grant systems were introduced in several countries. Through ‘matching grants’, these funds provide co-financing for agricultural business development. These activities can vary from research and extension support for companies and farmer organisations as well as for training workshops, pre-professional internships, and even direct subsidies to necessary infrastructural investments. Generally, these grant funds are managed by decentralised, specialised governmental entities. Producer organisations can apply for grants to fund applied technology innovation proposals. They are eligible for funding if they co-finance a percentage of the total funding requested (often 30%), either from their own resources or with the support of third parties, and, of course, if their business proposal meets minimum quality requirements defined by the grant system. When a proposal is deemed eligible for a matching grant, a public call for proposals is sometimes issued to invite private service providers to further elaborate the business case in co-ordination with the farmer organisation that originally submitted the business proposal.

The need to report within the project period means that most evaluation reports cover a short time span. They focus on disbursement of funds and the outreach (the number of smallholders involved) but only very superficially on the outcomes in terms of farmer livelihoods or changes in the innovation systems. Another reason is the ‘embeddedness’ of the grant systems in wider systems of support to agriculture, which induces the implementation agencies to evaluate not the effectiveness of the grant modality as a separate instrument, but the impacts on rural development of the total support package.

Most studies covered by the review (Table 2.3) highlight the outreach of their business development grant systems and the diversity of business proposals that have been funded. The evidence that the grant effectively triggers the start-up of value-adding activities by the target group is convincing. However, the evidence on the impact of the activities after start-up on farmer livelihoods is far less convincing, not the least because almost all studies reflected on the performance of the grant fund during its operation and did not follow up the businesses supported by it. Little is known about the performance of the business venture after the period of support. Berdegué (2001) conducts one of the few studies with a longer time-frame that examines this type of external support to farmer group business activities in a more-than-anecdotal manner. He concludes that the grants to associative business are more effective when they relate to activities in higher-end markets. He states that, in a market economy open to international competition, organisations involved in non-traditional products and in markets with high transaction costs will have more economic impact on their members’ farms and households. His description of the support package to small-scale producers in Chile points to the importance of a range of supporting services, in addition to the financial grant.

Support for the assumption that the business grants trigger changes at the household level is even weaker. Impact depends a lot on the performance of the farmer group that handles the grant. The performance of the group is influenced by many more factors than just the grant and, as a result, so are the quantity and quality of their services to their members, the smallholder farmers. The positive influence of farmer groups on their members is an axiom gener-

ally considered to be self-explanatory: if this were not the case, the members would withdraw their support. Clarification of membership and the development of internal regulations to sanction deviant behaviour are considered to be essential elements for farmer organisations' business plans to be successful (Berdegué, 2001; Lyon, 2000; Ostrom and Ahn, 2009; Ton, 2008; Ton, 2007b). Trust in and commitment of smallholders to their organisation are strong mediating factors for grants to farmer groups to be effective.

Bebbington and Sotomayor (1998a) and Toro and Espinosa (2003) indicate a weak point in the Chilean and similar systems, where the limited market of service providers leads to a situation where farmers are already 'married' to a service provider when submitting a proposal to the grant system, and the co-financing requirements – an essential element for determining the seriousness of the proposal – exist only on paper. They are in fact co-financed by the service provider, not by the farmer group. Hartwich et al. (2007) and Ton (2007b) report on similar processes in Bolivia. Hartwich et al. (Hartwich et al., 2007) highlight the unintended effects of strict eligibility criteria for service providers being used during the bidding process. It tends to generate operational antagonism between the (locally scarce) service providers.

Because of the inherent dynamic nature of the business environment, time lags between the initial business proposal and the implementation of the plan tend to create a need for modifications of already approved proposals. However, as the evaluation criteria were applied to a written document and the verdict has to be 'fair' to proposals that were discarded, the room for such adaptability is generally constrained. This inflexibility creates room for 'white elephants', of unused or over-dimensioned infrastructure. Toro and Espinosa (2003) advise instituting an independent 'flexibility committee' to make decisions on this.

The World Bank report on the design of agricultural innovation funds (World Bank, 2010) stresses the need to embed the matching grants of business development in a wider context of support, with specific attention to value-chain development platforms and the use of brokers in supporting the applicants to generate better business plans and comply with other fund requirements. While promoting this type of fund because of its flexibility in adapting to demands in diverse and changing contexts, the report advises concentrating the investments in sectors or clusters to generate multiplicity of experiences and a more developed market of service providers and market outlets. This may feed sector dynamics with spill overs and synergies beyond the direct applicants. The World Bank (2010) also stresses the need for field appraisals of the applicant's situation before approving the concept note for further development. The information provided by the applicant on paper may differ quite dramatically from the reality on the ground.

Perrett (2004), reflecting on IFAD experiences with community development funds, is concerned about the mushrooming of this type of grant fund in the absence of a good initial understanding of whether a sufficiently enabling political, institutional and social environment exists for its use. He notes that these funds have generally performed better on short-term infrastructural and tangible achievements than on capacity building for longer-term impact, and are better at disbursing funds than channelling benefits to the targeted poor. And he points to another unintended effect, where the provision of a large number of grants may potentially undermine the credit culture and repayment rates for related programmes.

### **Conclusions on the assumptions in the impact pathway for ‘Business development grants’**

Causal assumption B1: *Competitive matching grants trigger value-adding business activities by (groups of) farmers as a way to facilitate innovation processes with smallholder farmers in markets.*

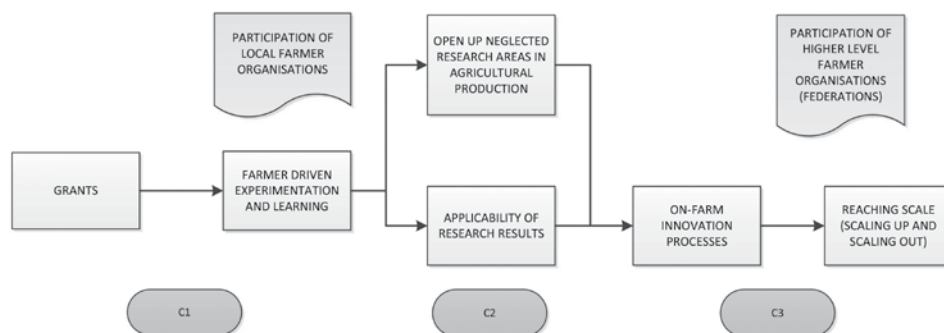
The studies on business development matching grants show that the grants indeed translate into investments in technology or support services for business proposals from farmer groups. Initial organisational social capital within the groups is a necessary precondition to develop these proposals and to handle the grants. Grants tend to be only one in a wider constellation of factors that make business proposals successful. Therefore, outcomes of the grant system on organisational social capital that provide the context for further development of these business initiatives are important. The necessary transparent and sustained procedures needed for business support grants place high demands on the governance system. Participation of farmer organisations in the governing body is valued positively by most authors.

Conclusion: *strong support in studies.*

Causal assumption B2: *Farmers’ livelihoods improve as a result of social activities and economic returns derived from the new value-adding business activities.*

The three studies that analysed the impact of the business proposals supported by these grants documented positive impacts on producers, though their methodologies suffer from the absence in their research design of comparison groups or other methods of counterfactual reasoning. The change in income through the grant-supported business proposals is not necessarily attributable to the grant, and definitely not to the grant alone.

Conclusion: *weak support in studies.*



Source: Authors’ own elaboration

Figure 2.4 Impact pathway C: farmer-driven agricultural innovation grants

## Impact pathways type C: Farmer-driven innovation support funds

This type of grant system covers research support to farmers for experimentation enabled by the provision of a grant. The logic behind this type of grant system is based on the assumption that farmer experimentation is key to developing, testing and/or adapting innovations that respond to the constraints experienced by the farmers.

Farmer-driven agricultural innovation grants are directed at learning about, and experimenting on, key constraints in the farmers' agricultural system. The difference from other agricultural extension and innovation approaches is the assumption that the participation of farmers, through their organisations, in the steering and governance of the grant system makes them effective to reach smallholder farmers who tend to be bypassed in traditional government-led or private-sector-led interventions. The issues that farmer-driven grant systems address are assumed to be different from, or complementary to, the issues that would normally be addressed in research and extension.

- Causal assumption C1: *Grants to facilitate farmer-driven experimentation and learning open up neglected research areas in agricultural production and enhance the applicability of research results.*
- Causal assumption C2: *Participation of local farmer organisations in decision making about research grant funds is effective in (re-)directing the research to critical constraints in on-farm agricultural innovation, and particularly to the needs of the poor and women.*
- Causal assumption C3: *Participation of higher-level farmer organisations in decision making about research grant funds is effective in scaling-up and scaling-out on-farm agricultural innovation processes.*

The result of the studies on these types of funds are summarised in Table 2.4. Most of the impact studies retrieved under this grant modality relate to the NAADS programme (see Box 2.3). The objectives are to enable the 'economically active poor' farmers of Uganda to increase their agricultural productivity and incomes in a sustainable manner. Under the NAADS approach, farmer groups contract private sector service providers (including NGOs) who are awarded short-term contracts to promote specific agricultural activities (called 'enterprises') and provide advisory services. The two of the retrieved impact studies on NAADS, by IFPRI, the International Food Policy Research Institute, used a rigorous quasi-experimental research design (Benin et al., 2007; Benin et al., 2008). The 2008 report was published in a peer-reviewed article (Benin et al., 2011). The 2007 study used household surveys without matching the characteristics of the NAADS beneficiaries and the control group. The study was informative about the mixed results of NAADS. The follow-up report in 2008 (Benin et al., 2008), later published as Benin et al. (2011) applied robustness checks on the difference-in-difference regression using four different econometric methods for estimating average treatment effects. They corrected the differences in outcomes between participants and non-participants through a matching procedure. The four different estimation procedures, each with different matching algorithms, result in tables with mixed evidence on impact: some changes in outcome indicators are not significant with some estimation procedures while they are with others.

Table 2.4 Grant impact studies on farmer-driven innovation support funds

Grant system	Evidence base	Outcome areas	Proxy-indicators	Valuation	Methods used	Rigour	Independent evaluation			
NAADS	Benin et al. (2007)	Farmer livelihoods	Awareness of improved practices	Negative	Household survey	Strong	Yes			
			Use of improved practices	Negative	Household survey	Strong				
			Participation in markets	Positive	Household survey	Strong				
			Farmer income	Positive	Household survey	Strong				
			Food security	Positive	Household survey	Strong				
			Soil management	Negative	Household survey	Moderate				
			Farmer empowerment	Neutral	Group survey	Moderate				
			Access to services and institutions	Positive	Group survey	Moderate				
			Quality of advisory services	Positive	Group survey	Moderate				
			Participation in community activities	Neutral	Group survey	Moderate				
			Farmer empowerment	Neutral	Group survey	Moderate				
			Quality of advisory services	Positive	Household survey	Strong	Yes			
			Improved agricultural practices	Positive	Household survey	Strong				
			Soil conservation	Neutral	Regression (IV, 2SWR)	Strong				
FFS/NAADS-Uganda	Friis-Hansen (2008)	Innovation system	Crop productivity	Positive	Regression (IV, 2SWR)	Strong				
			Participation in markets	Neutral	Regression (IV, 2SWR)	Strong				
			Livestock productivity	Neutral	Regression (IV, 2SWR)	Strong				
			Farmer income	Negative	Regression (IV, 2SWR)	Strong				
			Number of private extension providers	Neutral	Regression (IV, 2SWR)	Strong				
			Analytical and organisational skills	Positive	Interviews	Moderate	Yes			
			Trust among group members	Positive	Interviews	Moderate				
			Change in poverty status	Positive	Interviews	Moderate				
			Improved agricultural practices	Positive	Household survey and lifecycle interviews and well-being ranking	Strong				
			Empowerment of farmers	Positive	Household survey	Strong				
			Engagement with research	Positive	Interviews	Weak	Not clear			
			Income from innovations	Neutral	Interviews	Moderate				
			Experimenting with new agricultural practice	Negative	Interviews	Weak				
			CIAL	Kaaria et al. (2006)	Farmer livelihoods	Adoption of new agricultural practice	Positive	Household survey	Moderate	No
Participation in community organisations	Positive	Household survey				Moderate				
Farmer-to-farmer extension	Positive	Household survey				Weak				
Adoption of new seed varieties	Neutral	Household survey				Strong	No			
Adoption of other new agricultural practice	Positive	Household survey				Moderate				
Target crop yields (beans)	Neutral	Household survey				Moderate				
Non-target crop yields	Positive	Household survey				Moderate				
Analytical and organisational skills	Positive	Household survey				Moderate				
Crop diversification	Positive	Household survey				Moderate				
Yields	Positive	Household survey				Moderate				
Prolinnova	Shroff et al. (2012)	Farmer livelihoods				Engagement with research	Positive	Interviews	Weak	Not clear
						Income from innovations	Neutral	Interviews	Moderate	
						Experimenting with new agricultural practice	Negative	Interviews	Weak	
						Adoption of new agricultural practice	Positive	Household survey	Moderate	
			Participation in community organisations	Positive	Household survey	Moderate				
			Farmer-to-farmer extension	Positive	Household survey	Weak				
			Adoption of new seed varieties	Neutral	Household survey	Strong	No			
			Adoption of other new agricultural practice	Positive	Household survey	Moderate				
			Target crop yields (beans)	Neutral	Household survey	Moderate				
			Non-target crop yields	Positive	Household survey	Moderate				
			Analytical and organisational skills	Positive	Household survey	Moderate				
			Crop diversification	Positive	Household survey	Moderate				
			Yields	Positive	Household survey	Moderate				

**BOX 2.3 NATIONAL AGRICULTURAL ADVISORY SERVICES IN UGANDA**

The NAADS programme in Uganda is a public-private extension delivery approach with the goal of providing a decentralised, farmer-owned and private sector extension system. When a farmer decides to participate, he or she has to do so through membership of a farmer group. Then, together with other NAADS-participating groups in the sub-county, they request specific technologies and advisory services associated with their preferred enterprises and also obtain grants to support acquisition and development of those technologies. The grants are mostly used to finance the establishment of an experimental plot, the proceeds of which become a revolving fund for members.

The study by Friis-Hansen (2008) focuses on one of the districts where NAADS was considered to be most successful. In his study he explored the heterogeneity of the impact for groups defined on a poverty ranking based on farmers' own perception of well-being. The complementary use of life-cycle interviews supports his inferences of positive impacts. The survey design included a group of respondents who were not member of a NAADS group.

Friis-Hansen refers especially to the processes and mechanism that created different responses of farmers to NAADS, which were induced by a former experience with FFS, Farmer Field Schools. He shows the role of an earlier FFS in improving the groups' organisational and productive capacities. The FFS-groups had already created a bank account and related administrative procedures that proved an advantage for obtaining access to support through the new NAADS system. Likewise, several of the FFS in Kenya, documented in the descriptive study by Gustafson(2002), have continued beyond the initial year as such, self-financed by commercial activities that FFS members implement with the knowledge or technologies they have acquired. Gustafson indicates that the size of a typical FFS has been designed to provide a critical mass that enables the group to continue when a support project withdraws. The sustainability of the innovation process is in the heads of the farmers, not in the support project that is by definition temporal. Gustafson(2002) mentions the establishing of personal links to the Kenya Agricultural Research Institute (KARI) staff as an important success, both for the farmer groups that obtain access to wider support than just the FFS, and for the KARI staff that can use them as social infrastructure for research and outreach with and beyond FFS. These FFS members are relatively affluent, but do not form a self-contained group, being in frequent contact with the poorer farmers (Braun et al., 2000; Friis-Hansen and Egelyng, 2006; Nathaniels, 2005). More than a way to open up neglected research areas, the FFS seem a way to articulate demand for already developed technologies (Gustafson, 2002), to test and 'peer-review' the innovations that are already available, and to promote 'first see then believe' outreach to the wider farming community. Interestingly, the FFS-led farmers seem particularly effective as facilitators of innovation when they share knowledge and experiences with farmers outside their own villages (Braun et al., 2000), when they are treated as knowledgeable experimenters only, free from other cultural stigmas that may influence the interaction with their neighbours.

The demand-led character of NAADS and the process of prioritising 'enterprises' (crops or livestock sectors to be developed as commercial farming activities by the community) indeed

created room for manoeuvre for farmers to get the advisory services adapted to issues that they see as important. Opondo et al. (2006) point to the fact that the constrained number of enterprises from which farmers could choose led to the exclusion of certain social groups that had limited ability to work on these enterprises, especially due to limited access to land and labour for commercial crops. However, over time, enterprises selected under NAADS have tended to include activities with lower cost of adoption. Overall, NAADS seems to have made a difference in smallholder farming practices. A study by Ekwamu and Brown (2005) report that only 22% of the households in the NAADS districts have the same top two crops as 10 years ago, suggesting farm households are willing to change production in favour of crop and livestock activities that yield higher returns. Benin et al. (2008; 2011) stress that the quality of advisory services is nevertheless not the only important factor influencing this technology adoption. Credit, access to inputs, adequate access to farming land and mechanisms to cope with unfavourable weather patterns and the incidence of pests are all mentioned as factors for which other government interventions are needed to complement and reinforce the extension support.

The technology areas ('enterprises') that were supported were selected through a participatory dialogue between three actors: sub-county farmer fora (representing all farmer groups), private service provider companies and district NAADS staff. All three actors influenced the technology enterprise selection and development. Over time, the control of the process by farmers' institutions (farmer fora) gradually increased. Especially after 2007, the emphasis became more on the transfer of technology. The key assumption that farmer-driven innovation grants would lead to a shift in research focus became, therefore, less relevant for NAADS, as the main characteristic became not the generation of improved knowledge and appropriate technologies, but the creation of awareness in farmers of existing technologies and knowledge and linking farmers with service providers that could train the groups on these issues, and/or provide the inputs to experiment with them. This feature also explains the problems encountered by NAADS where knowledge is less uniform and codified and where more interaction between the knowledge of the extension worker (private service provider) and the knowledge of farmers is needed, such as the experimentation and learning related to marketing, an issue which has consistently featured as a low priority in the implementation of the programme (Benin et al., 2007; Benin et al., 2008; Bukenya, 2010; Opondo et al., 2006).

The original NAADS guidelines called for formation of new agriculturally oriented farmer groups disregarding existing groups or assuming that there were none. Nevertheless, Opondo et al. (2006) and Friis-Hansen (2008) point to the fact that the districts where NAADS groups emerged often built on pre-existing groups and networks. As NAADS groups were formed by election in the village, people and groups that had previous experience in organisations and networks tended to be elected. As a result, the NAADS group members tend also to be more affluent than the average farmers in the area (Friis-Hansen, 2008). The initial high expectations (Bukenya, 2010; Opondo et al., 2006) motivated farmers to become active in the groups to obtain access to the (expected/promised) credit and technologies. This 'pull factor' was reinforced by the initial practice of paying farmers for their attendance at NAADS sensitisation sessions (Opondo et al., 2006). As the programme progressed, these groups tended to reorganise themselves in response to the reality of limited access to tangible inputs, with farmers that stopped their participation, leaving a core group, primarily motivated by agricultural experimentation.

NAADS is based on farmer groups managed through farmer representatives at sub-county and district levels known as 'farmer fora'. Opondo et al. (2006) point to the fact that these farmer fora did function, though with responsibilities for which capacity and 'voice' were initially quite low. The role assumed by the farmer fora was especially to monitor the performance of the service providers. It indeed reflected an empowerment of farmers in the advisory system but created also a certain antagonism between the farmer organisations at the higher level and the service providers and their client groups in the villages (Opondo et al., 2006). Friis-Hansen (2008) is more positive about the farmer fora and the empowerment that resulted from the NAADS governance structure, especially in the first phase of NAADS, till 2007.

#### **BOX 2.4 PROMOTING LOCAL INNOVATION**

PROLINNOVA is an NGO-initiated multi-stakeholder network to stimulate local innovation in ecologically oriented agriculture and natural resource management. The network builds on and scales up farmer-led approaches to development, which starts with finding out how farmers do informal experiments to develop and test ideas for better use of natural resources. The small grants (typically a few hundred dollars per group of farmers) specifically targeted poor and vulnerable households and focused more (but not exclusively) on local ideas and technologies. Grants can go up to US\$2000 or more when stakeholders other than farmers are also involved in the joint experimentation.

Promoting Local Innovation – PROLINNOVA – (see Box 2.4) was mentioned in several comparative studies (Friis-Hansen and Egelyng, 2006; World Bank, 2012) as a promising example of farmer-driven innovation support funds. Our initial search did not produce any studies that described outcomes of the LISFs in a systematic way. The document most closely resembling an external evaluation (Shroff et al., 2012) has a strong focus on assessing changes as a result of the PROLINNOVA-supported grant system but is not based on structured data collection. The arguments are supported mainly by results of interviews with field staff in two countries, which focused more on the process than the impacts. It is not clear if it is an external and independent evaluation, as Rockefeller Foundation was the main donor of the intervention and sponsor of the study. The local support funds promoted by PROLINNOVA cover, compared to the other innovation grant systems studied in this review, a very broad pallet of innovations on crops, technologies and organisation. The comparative literature (Friis-Hansen and Egelyng, 2006; Triomphe et al., 2012; van Veldhuizen et al., 2005) suggests that assumption that these grant systems open-up of research to critical constraints of smallholders seems indeed effective and promising, though there is still very little systematic evidence on the discrete innovation processes funded with the grants, nor on the novelty of the experiments of the farmers for the formal research community. The grant amounts involved are also very small, which may make it challenging for implementing NGOs to allocate sufficient resources to structured monitoring and reporting.

The development of alternative farmer-governed funding mechanisms for local agricultural research for development is the stated objective of PROLINNOVA (Wongtschowski et al., 2010).



The experiences from the PROLINNOVA programme are expected to lead to grant management formats that are easily manageable and will not need expensive local support by NGOs. If this indeed proves possible, the scaling-out through existing networks of farmer organisations or farmer federations looks promising in the future. PROLINNOVA objectives and future plans are assuming the above pathway, but studies do not yet provide the evidence to support or challenge the assumption that this farmer participation in the governance structure proves indeed more effective. PROLINNOVA facilitates an interface between farmers and support organisations in rural innovation. The pilots differ greatly in the way they relate to the wider innovation system. The links with the national research community seem less close than in other innovation grant funds covered in this review (CIALs, FFS, NAADS). The diversity of topics and the relatively unstructured and interactive process of experimentation will make it more difficult to establish closer links with current formal agricultural research, which has organisational and institutional limitations to dealing with these dynamic changes in research questions and research process.

This articulation of farmers with researchers is a more prominent feature covered by the studies about the Local agricultural research committees (further referred to as CIAL, with its Spanish acronym – see Box 2.5). The descriptive studies on the experiences in CIALs (Ashby et al., 2000; Braun et al., 2000; Humphries et al., 2005; Humphries et al., 2000) support the causal assumption that this approach generates a different research agenda and a different relationship between the researchers and the farmers. A high degree of pre-existing social capital is considered an asset that makes the innovation fund more effective. The regular meetings, inherent to the CIAL approach, build on social capital and, in doing so, help to enhance it. The organisational and leadership skills required to conduct the weekly/monthly meetings are strengthened and can help its members to become involved in a range of other social and economic activities. These skills are evidence of organisational maturity and capacity for collective action, which is helping to build social capital more broadly in the communities (Humphries et al., 2000; Ashby et al., 2001).

#### **BOX 2.5 LOCAL AGRICULTURAL RESEARCH COMMITTEES**

The CIAL-approach was developed at CIAT (International Center for Tropical Agriculture) in Colombia in the 1990s, with the goal of increasing the efficiency of the agricultural research and technology development system by integrating farmers better into the process. The CIAL conducts research on priority issues identified through a diagnostic process, in which all are invited to participate. The community monitors the performance of the CIAL and is free to add, remove or replace committee members at any time. Each CIAL is supported by an agronomist or extension agent who trains the committee members in research design (controls, replicates, systematic evaluation of results) and who visits their trials regularly to provide technical support. Support for the agronomist comes from the institution supporting the CIAL, usually an NGO, the national research or extension service, or some other institution involved in technology development and transfer.

Comparing the two impact studies on CIALs that we retrieved proves interesting. The studies reflect the efforts of the authors to increase the validity of the evaluative findings; the first (2006) study was improved in the subsequent (2009) paper with additional data. The main difference in the analyses is the use, in the later study, of a comparison group of villages that are not supported by a CIAL to allow counterfactual reasoning about impact. Sandoval et al. (2009), using data from a comparison group, come to somewhat different conclusions to those in the earlier study (Kaaria et al., 2006). The difference in conclusions between the studies provides food for thought, especially with respect to the lack of significant impact on crop yields, and the small difference in the adoption of new agricultural practices between the treatment and control groups. This small difference is explained by the authors as a consequence of the fact that organisations and institutions other than the CIALs were working on agricultural development in the nearby ‘control’ areas. This illustrates the difficulty of applying a counterfactual design with an ‘untreated’ control group in rural development, as the counterfactual might not be the absence of a treatment but the presence of another type of treatment.

Different from most other studies covered in the review, the studies on CIAL pay particular attention to impacts on social capital, especially on the organisation of farmers. In both Colombia and Honduras, where the CIAL approach was implemented on a relatively large scale, second-order farmer organisations were created on the basis of the local CIAL groups (Ashby et al., 2000; Humphries et al., 2000). The two documented higher-level organisations, in Colombia and Honduras, have the maintenance of the CIAL network as their prime focus. They did not exist before CIAL groups were formed. As such, these higher-level organisations cannot be considered as a moderating factor for faster scaling-up and scaling-out. Instead, it is the result of the scaling itself, realised through other mechanisms, principally through the networking with local development NGOs and local research institutes.

### ***Conclusions on key assumptions in the impact logic for ‘Farmer-driven innovation support funds’***

Causal assumption C1: *Grants to facilitate farmer-driven experimentation and learning open up neglected research areas in agricultural production and enhance the applicability of research results.*

The studies on farmer-driven innovation support funds in this review all made reference to the difference that doing this type of participatory research made compared with traditional research in the area and to the benefits of an interactive relationship between the farmers and the technical supporters or researchers. No study had a design that permitted counterfactual reasoning about which other research areas would or would not have been opened up without the grant. Impact studies provide weak support but the causal assumption is considered to be valid by most authors.

Conclusion: *moderate support in studies.*

Causal assumption C2: *Participation of local farmer organisations in decision making about research grant funds is effective in (re-)directing the research to critical constraints in on-farm agricultural innovation, and particularly to the needs of the poor and women.*

The review only examined the studies where farmers participated in the governance structure. The studies show that this participation indeed defines the research activities in ways that make them more in line with their priorities.

Conclusion: *strong support in studies.*

Causal assumption C3: *Participation of higher-level farmer organisations in decision making about research grant funds is effective in scaling-up and scaling-out on-farm agricultural innovation processes.*

The studies all mentioned the progressive involvement of higher-level farmer organisations in the scaling-up and scaling-out of the innovation grant activities. The organisations mentioned in the studies, however, are more a result of the scaling process itself, not the drivers of it. Supporting institutions (NGOs, governments) are more important in this respect.

Conclusion: *weak support in studies.*

## 2.4 Discussion and conclusion

Given the scarcity of studies that related to the same treatment/innovation grant system, the same context and the same outcome areas, we decided for an explorative synthesis of the available evidence (Ton et al., 2011a). Interpretation of the findings in the studies may therefore be subjective (Wong et al., 2010). We distilled some issues from the studies that helped to improve our understanding of the diversity in innovation grant systems, and that tend to be relevant for the study of impacts of innovation grants to smallholders.

First, we found no study that challenged the relevance or effectiveness of innovation grant systems for smallholder farmers, as compared to conventional research and extension approaches. Though the evidence base is rather thin, the assumptions in the rationale, on which the decision to implement innovation grant systems is based, remain largely unchallenged. All studies present evidence of the positive changes as a result of these investments in agricultural innovation. Some of the impact studies show mixed impacts on natural resources, especially due to land clearing of tree species or increased cultivation without soil conservation. The negative outcomes reported in these studies are, however, always accompanied by a positive outcome in another area, such as an increase in yields or income. This general positive attitude by the authors of the studies contrasts with the policy reality, where still only a minor share of the funding on agricultural research and development is invested in this type of grant funds for smallholder innovation. This may point to a publication bias, but may also indicate that there exists a development potential. However, as a result of the wide diversity in contexts and implementation modalities of such funds, it is very difficult to assess their cost-effectiveness compared to other innovation policy instruments.

This leads to our second point. The evidence from the impact studies shows that input vouchers as such indeed cause better yields and, in doing so, trigger innovation in agriculture, but the studies do not provide the means to evaluate if these effects are commensurate to the

investments made. There are some critical remarks in the studies, e.g. in the studies on input vouchers (Holden and Lunduka, 2010a; Mason and Ricker-Gilbert, 2012; Ricker-Gilbert et al., 2013) that question the political priority of funding innovation grant systems compared to infrastructural investments or cash transfers. This triggered a debate on the use of vouchers as a means to spur innovation in East African countries, especially in relation to the share of the government budget used to fund it, compared with infrastructural investments or market enabling policies. Unfortunately, none of the studies that we reviewed had a research design that could generate comparative information about the cost-effectiveness of these alternative policies on smallholder innovation.

Thirdly, we see that grant systems add to a pre-existing capacity for innovation. Grant systems that target lead farmers or farmer experimenters often build on the capacities created by earlier projects or programmes. Their main outcomes might also be realised in follow-up activities of farmers. Experiences with innovative practices will feed into a process of enhanced learning. Friis-Hansen (2008) points to the fact that in Uganda the FFS provided the social capital that explains positive outcomes of the NAADS systems that was implemented later on. Gustafson (2002) suggests to use the innovative behaviour and innovation capabilities of farmer groups a prime indicator of success. When considered as such, the innovation grant systems may contribute beyond the specific project and period.

As a fourth observation, we note that most of the quasi-experimental impact studies focus on field-level impacts only, and use household survey data to support their inferences. When the impact on households is more indirect, for example when facilitating innovation by farmer groups, household survey may not be able to capture the outcomes of the grant. The outcomes are group-based and need time to generate changes at household level. This partly explains the scarcity of impact studies on business development grants and innovation support funds. Often, these grant modalities explicitly target on-going innovation processes that had been started or shaped in cooperation with other support entities, next to the grant fund. Difference-in-difference designs, household surveys with treatment and comparison groups, may be appropriate for the assessment of short-term impact in common outcomes that directly result specific technology packages or other similar uniform treatments. However, they are not appropriate for measuring outcomes that need more time to mature, and that result from more complex and diverse innovation processes (Ton et al., 2014c). For the latter, complex multi-stakeholder processes, the major gains in the quality and usefulness of evaluations, will lie in the accuracy and comparability of the monitoring of short-term outcomes in the group of direct beneficiaries. The studies point to one important and transversal and relatively short-term outcome of innovation grant systems that may be put more central in impact evaluation, at least in addition to the assessment of field-level impacts: the creation of human and social capital to sustain experimentation, creative thinking and innovative practices. Currently, the operationalisation of these indicators for human and social capital differs a lot between the studies.

This leads to our fifth and final point for discussion. The outcome areas and proxy-indicators used in the studies vary widely, even when researchers study the same type of intervention. Table 2.5 gives an overview of the different outcome areas and proxy-indicators used to assess the effectiveness of the innovation grant system. In line with the findings of Ricker-Gilbert et al. (2013), we argue that the empirical challenges for the evaluation of impacts of innovation

**Table 2.5** Proxy-indicators used in the impact studies to capture impact

Agricultural practices	Environmental impacts	Knowledge on agricultural practices	Farmer organisation	Structure of the innovation system	Farmer income	Farmer livelihoods
Application of fertiliser	Application of manure	Awareness of improved practices	Participation in community organisations	Effective co-financing of extension services by farmers	Yields	Household food security
Use of improved seeds	Indigenous tree vegetation on farms	Farmers' analytical and organisational skills	Knowledge on composition village council	Access to services and institutions	Gross-margin	Farmer empowerment
Intensified maize production	Intercropping system in maize	Farmers' engagement with research	Perception of functions	Quality of advisory services	Household income	Livestock productivity
Technology adoption	Area planted (agricultural frontier)		Quality of deliberation	Number of private service providers	Farmer sales volume	Change in poverty status
Use of improved practices	Use of organic manure		Use of business planning tools	Size of secondary market of inputs	Net-income per hectare	Health/schooling
Soil management			Profits made by the farmer groups	Number of agro-dealers	Farmer income	Crime levels
Adoption of new seed varieties			Participation in community activities	Sales volume per agro-dealer	Participation in markets	Assets accumulation
Adoption of new practices			Trust among group members		Agricultural revenue per adult equivalent	Crop diversification
Number of rice varieties			Farmer-to-farmer extension			
Range of seed qualities requested						

grants are huge and that comparative research and common methods are needed to provide benchmark information about the effectiveness of these type of interventions in a way that facilitates informed decision making by policy makers and development cooperation. If common proxy-indicators to measure changes in this capacity for innovation could be developed, they would enable the comparison between alternative policies and projects. Potential transversal indicators to measure these outcomes are knowledge on good agricultural practices, implementation of these agricultural practices, capacities of farmers to learn and adapt, and capabilities of farmer groups to sustain collective action. Policy-makers and grant system designers may need to specify these areas as a major indicator of success and, doing so, create an incentive for projects to monitor human and social capital regularly.

## Acknowledgements

*The systematic literature review was funded by Australian Agency for International Development (AusAID ABN 62 921 558 838). Members of the advisory board to the review were Esbern Friis-Hansen (DIIS), Bernard Triomphe (CIRAD), Ann Waters-Bayer (ETC AgriCulture), and Mariana Wongtschowski (KIT). Complementary funding was provided by the Netherlands' Ministry of Economic Affairs (KB-11-004). We acknowledge the support by Hans van Hoeven (LEI Wageningen UR) in the search of electronic libraries, and by the late Marieke Douma for preparing most of the retrieved studies for synthesis.*



# 3 | The Triviality of Measuring Ultimate Outcomes: Acknowledging the span of direct influence<sup>2</sup>

Giel Ton<sup>a</sup>, Sietze Vellema<sup>b</sup> and Lan Ge<sup>a</sup>

<sup>a</sup>. Agricultural Economics Research Institute, LEI Wageningen UR

<sup>b</sup>. Wageningen University, Chair group Knowledge, Technology and Innovation

## Abstract

*Sustainability standards and certification schemes have been promoted as a market-driven instrument for realising development impacts and receive public funding. As a result, companies, NGOs and supporting donors and governments want to know if these ambitions have been fulfilled. Their tendency is to commission household surveys to assess net effects of certification in areas such as poverty, productivity and food security. This paper argues that, rather than trying to measure precise net effects on farmer income, the focus should be on detailed measurement of more immediate outcomes in terms of knowledge and implementation of good agricultural practices. Contribution analysis is proposed as an overall approach to verify the theory of change, combining survey-based net-effect measurement of these immediate and intermediate outcomes with less precise, lean monitoring of indicators to verify the contributory role of these outcomes that are outside the span of direct influence, like household income and poverty alleviation.*

---

2 Published as: Ton G, Vellema S and Ge L. (2014) The Triviality of Measuring Ultimate Outcomes: Acknowledging the Span of Direct Influence. *IDS Bulletin* 45: 37-48.

### 3.1 Introduction

The private sector and market-led strategies have become increasingly central to development policy and practice. Moreover, non-governmental organisations (NGOs) are teaming up with companies or private-public partnerships. This shift from public to private-led development strategies is based on changing expectations of the role of trade versus aid for poverty alleviation. In many donor countries, this policy is increasingly based on the assumption that the private sector is more effective in reaching development goals than development aid through governments or NGOs. Accordingly, donor agencies have begun to re-allocate public resources to companies and private-public partnerships. From a public perspective, the obvious question for impact evaluation is how to demonstrate this assumed effectiveness.

In general, donor agencies prefer precise measurements of net effects in relation to the Millennium Development Goals, with income generation and poverty reduction as main objectives (DCEP, 2010; DGIS, 2011). This often translates into survey-based research designs, including baseline studies, randomised sampling, and comparison groups. This paper challenges the exclusive emphasis on precise measurement of income effects in quasi-experimental evaluation designs. Net effects, especially those related to business performance and income, are influenced by a wide range of intervening factors that are impossible to control for under real-world conditions. This makes it difficult to attribute effects to the actual interventions and provides little information on the effectiveness of developmental activities. Based on our experiences with impact studies of certification-induced training programmes for farmers (Ton et al., 2011b; Waarts et al., 2012; Ton, 2012b; Waarts et al., 2013a; 2013b), we argue that there are good reasons to limit this dominant focus on measuring net-effects in ultimate outcomes, and propose to shift attention to the domain of immediate and intermediate outcomes.

This paper uses the example of certification to discuss the methodological challenges for impact evaluation of market-led development interventions. Sustainability standards and the related certification schemes, implemented in tropical commodity chains such as cocoa and coffee, aim to enhance environmental sustainability, social justice and economic viability. Multinational firms and global NGOs partner in defining and implementing these sustainability standards (Vellema and van Wijk, 2014). Government and donor agencies are motivated to support such endeavours because they believe that implementation of these standards is instrumental to achieving development goals. Standards systems aims to enhance their public accountability, but also to shift attention to more their impact on more intermediate outcomes. We describe recent advances in these efforts by certification schemes and illustrate the challenges in impact evaluation of these type of interventions.

The challenge addressed by this paper is to find ways to get credible data on outcomes that are still attributable to the support interventions that are related with certification, and to do so in a way that allows comparison between different possible support modalities that may lead to the same type of outcomes. We propose to measure and compare the effectiveness of activities foremost on the increase in knowledge (immediate outcomes) and improved business practices (intermediate outcomes). Further, we aim to verify the contribution of these intermediate outcomes to business performance (ultimate outcomes) and development impact. This entails a combined use of the realist notion of verifying and refining programme

theories (Pawson and Tilley, 1997; Rogers, 2009; Vellema et al., 2013; Ton, 2012b) and a mix of methods to collect evidence that bolsters the ‘contribution story’ (Mayne, 2012; Mayne, 2001). Data collection in an impact evaluation along the lines of contribution analysis uses multiple strands of evidence to verify, support or challenge the key assumptions in the intervention logic. *“The research builds a compelling case with evidence from which it is reasonable to conclude with confidence that the intervention has made a contribution and why”* (Mayne, 2012). Contribution analysis combines the precise measurement of the outcomes and the analysis of the causal processes set in motion within the span of direct influence of an intervention, with the monitoring of outcomes and influencing factors outside the span of direct influence. In addition to survey-based research, a mix of methods is used to enable cross-case comparative analysis, as well as for finding out from stakeholders and expert panels how relevant the intervention is compared to alternative strategies (benchmarking).

The paper first contextualises the challenge addressed in this paper. It reviews some important initiatives to improve reporting on the impacts of standard setting and certification in the cocoa and chocolate industries. Second, we reflect on our experiences with the design and implementation of survey-based impact evaluation in cocoa production in Ivory Coast. Finally, we discuss the implications of our findings for future impact evaluations of private-sector support programmes. We propose to limit rigorous measurements of net-effects to outcomes and processes ‘within the realm of the programme’ and to use a mix of methods to collect information to verify the assumption that these business practices (intermediate outcomes) are contributing factors that together generates a change in business performance and development impact.

### 3.2 Setting: Impact evaluation of certification

Development impacts are generally framed in terms of the triple P (RSCE, 2009): People-Planet-Profit. Accordingly, texts accompanying certification schemes, such as UTZ Certified, Rainforest Alliance, Fairtrade, or Organic, suggest contributions to environmental sustainability (reflected in benign farming practices and conservation of forests, natural resources and biodiversity), social justice (reflected particularly in labour rights, improved working conditions and inclusion of marginalised groups), and economic fairness (reflected mainly in business opportunities for smallholder farmers, improved rural incomes and living conditions, and vitality of a sector). In addition to these developmental goals, standard systems have more internal objectives related to the logistics and verification of quality and quantity of transactions in the value chain, reliable and cost-efficient sourcing models, and traceability in the chain of custody.

The objective of fostering sustainability in the supply chain through certification is aligned with concerns for corporate social responsibility on the part of leading companies involved in global trade and processing of tropical commodities. Likewise, governments and public donor agencies support certification because they themselves are committed to sustainability as a public goal, and consider market-led intervention strategies as an effective vehicle to achieve this. As a consequence, NGOs, businesses and governments in the field of certification and sustainability in cocoa need to report on their achievements, both to account for public funding and to convince consumers of the benefits of paying an additional price for the products with a certificate.



## Public accountability requirements

In 2010, a group of cocoa-processing companies, retailers, chocolate manufacturers, non-governmental organisations, and Dutch ministries signed a letter of intent to support the revitalization of cocoa production in West-Africa to enhance the consumption of sustainably produced and certified chocolate in the Netherlands (Chocolate Working Group, 2010). The Netherlands is the world's largest importer of cocoa and is home to large processing facilities as well as the offices of several Voluntary Standards Bodies that govern the certification process, which explains the private and public interest in such a partnership in this country. The endeavour is linked to the Dutch Ministry of Economic Affairs' policy concerning International Cocoa Agreements and to public-private partnerships working on the Roundtable for a Sustainable Cocoa Economy, the World Cocoa Foundation and the Dutch Sustainable Trade Initiative (IDH). Stakeholders in the partnership have agreed to source only certified cocoa in 2025, as a joint commitment to enhance sustainability.

The assumption underlying this partnership is that an increase in market share of certified chocolate would lead to an increase in sustainability of the cocoa supply. The letter of intent places a strong emphasis on measuring the market share of certified chocolate products in the Dutch market as a proxy-indicator of impact. The following year, the Netherlands Ministry of Foreign Affairs gradually increased evaluation requirements for public funding of private-sector support programmes and required to report on impact on poverty and food security. In the 'Protocol on Evaluability and Attainment of Results' (DGIS, 2011), it demanded a monitoring and evaluation plan that included baseline, progress and end-of-project measurements and the use of control groups for robust net-effect measurements. The protocol suggested measuring and reporting on the impact of private-sector development support on nutritional status and household income of the beneficiary population. In addition to these mandatory impact areas, the protocol suggested, among several other things, to measure productivity of land use, input efficiency, access to training and finance, and quality of the business environment. This tendency to increase the requirements on private-sector recipients of development aid, in order to better elucidate the effectiveness of their interventions, is not specific to the Netherlands, but is a generalized trend among all OECD donors (DCED, 2010).

## Harmonised indicators and rigorous measurement

The International Social and Environmental Accreditation and Labelling Alliance (ISEAL Alliance) set out to improve the quality of impact evaluation of certification and to respond to accountability requirements with credible evaluation research.. ISEAL aims to introduce minimum quality requirements for monitoring and evaluation by standards systems and certification schemes (ISEAL Alliance, 2014) and to advance towards harmonising outcome indicators between sustainability systems (ISEAL Alliance, 2013).

ISEAL requires standards systems and certification schemes to ensure the quality of performance-monitoring data and of outcome and impact evaluations to guarantee transparency of the sustainability claims communicated to consumers (ISEAL Alliance, 2014). The scheme owner must ensure that at least some of these are independent impact evaluations. Harmonization of indicators used to track outcomes and impact would permit benchmarking and comparison between standards systems.

Where donor communities emphasized reporting on sustainable economic development and poverty reduction (DCED, 2010), discussions within ISEAL shifted attention to the measurement of more tangible outcome areas within the sphere of control of these voluntary standards organisations. This means evaluating such aspects as the adoption of conservation practices, yields, sales practices, satisfaction with crop profitability, perceptions about changes in natural resources, etc. Several of these common indicators are still in the process of being fine-tuned, e.g., comparative measurements to obtain knowledge on and adoption of good agricultural practices in specific crops (El Hage, 2012; Russillo and Pintér, 2009; Rigby et al., 2001), assessment of the capabilities of farmer groups to engage in marketing and value-chain coordination (Ton et al., 2014a; Donovan and Stoian, 2012), and the use of a common multi-dimensional poverty index.

As early as 2007, the demand for better impact evaluation of voluntary sustainability standards had led to an international, multi-stakeholder initiative to improve the quality these measurements: the Committee on Sustainability Assessment (COSA, <http://thecosa.org/>). COSA's main efforts have been to develop, pilot and implement metrics and indicators for measuring sustainability outcomes over time. COSA emphasises the need for times-series data and comparison groups and the use of econometric methods to limit selection bias. They propose to gather information on the different aspects of sustainability by using lists of questions, which are converted into dummy variables, and through Principal Component Analysis converted into factor scores representing the relative position of the respondents in relation to various aspects of sustainability (COSA, 2013: Appendix II). However, even these sophisticated quasi-experimental designs, or designs that deviate from random assignments of treatments, struggle with increasing numbers of observable and unobservable factors that influence the ultimate outcomes in farm performance that are only indirectly influenced by activities in the field.

### Refined theories of change

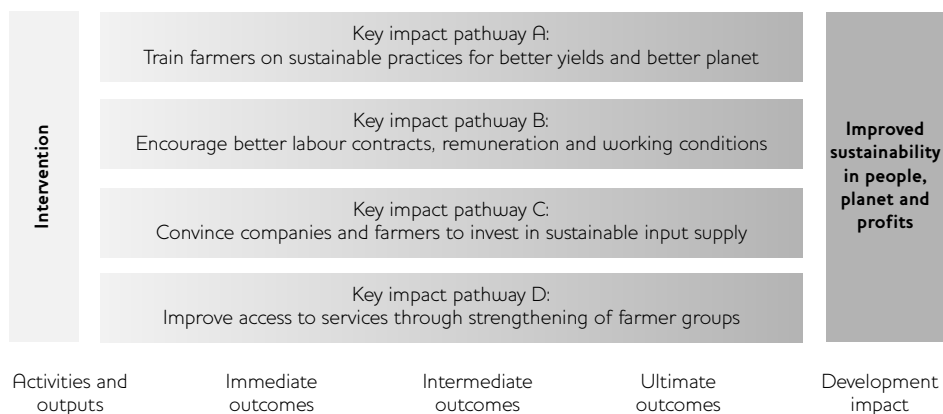
Methodological challenges related to the above initiatives on impact evaluation, encouraged a discussion between practitioners and researchers about feasible ways to register and attribute impact. Next to being a way to be publicly accountable for their role in reaching Millennium Development Goals, impact evaluation needed to be instrumental for gathering information that could help to improve the intervention strategy itself (Nelson and Martin, 2012). ISEAL initiated and supported a series of consultations to develop and modify an Impact Code (ISEAL Alliance, 2014), specifying how voluntary standards organisations should show the outcomes and impact of certification and standards in a credible way, while respecting the logistic and budgetary constraints of the implementing partners. The discussions in ISEAL stimulated a growing interest in learning how to construct and refine theories of change, and to select those performance indicators that would help to verify the key assumptions in their intervention logic (Ton, 2012b; Rogers, 2008; White, 2009).

In 2011, we helped several Voluntary standards organisations and their implementing partners to define their theory of change. For this, we used Mayne's framework (2001), which differentiates the main activities and outputs per stakeholder group, the immediate outcomes in knowledge of these stakeholders, the intermediate outcomes in improved practices of the

stakeholders and the ultimate outcomes in performance indicators related to these modified (business) practices. Based on a detailed ‘cloud’ of outcome areas, derived from their programme documents and mission statements, we developed a stylized representation (Figure 3.1), in which we identified several different impact pathways.

The stylized representation resulted from our exercise with UTZ Certified, and was later further refined and modified by UTZ in several of their communications, e.g., the 2014 Impact Report (UTZ Certified, 2014: 12-13). The impact logic of the support of UTZ Certified assumes several pathways that are expected lead to poverty reduction. For example, compliance with the prescribed agricultural practices and the provision of extension services is expected to increase the efficiency of cocoa production and consequently result in higher and more stable household incomes. Moreover, support to farmer organisations in managing an Internal Control System is expected to enhance their capacity to negotiate prices and/or to obtain better access to credit and agri-inputs. Reliable access to output markets and predictable incomes makes it possible for cocoa farmers to invest in their farms and offer better remuneration and working conditions to farm labourers.

This identification of various pathways proved useful to the key evaluation questions in commissioned research on impact, and helped to find appropriate outcome areas needing to be monitored. Each pathway embodies a specific sequence of causal steps and configurations of influencing factors, and each will have a specific result chain to graphically represent this causal logic. These result chains embody the assumptions about causal relationships between the main activities per stakeholder group, the main outputs of these activities, and the outcomes and development impacts.



Source: Authors' own elaboration, based on discussions with UTZ Certified. A more detailed version of their theory of change is published in UTZ Certified (2014:11-14)

**Figure 3.1** Stylised representation of the theory of change developed with UTZ Certified composed of distinct impact pathways

## The evaluation challenge in certification

Even when certification bodies communicate impressive differences in yields and income between certified farmers and comparison groups (UTZ Certified, 2014), the attribution claims reported in the more rigorous studies are more modest, as selection biases and the influence of confounding factors cannot be entirely ruled out. Whereas it would be ideal, from the perspective of donors and standard-setting bodies, that impact studies measure exact net effects on poverty and environment, in reality this becomes difficult or impossible, as these outcomes are in fact influenced only marginally by the certification-related activities in the field. The efficacy of the intervention becomes increasingly dependent on activities of other actors or factors.

Only truly experimental designs, such as randomized control trials (RCTs), are sufficiently robust to handle the influence of many observable and unobservable confounding factors, if they are based on random assignment of fairly uniform treatments. However, random assignment of the support is, from the perspective of the implementers of this support and the traders that need to sell the certified products logistically highly undesirable. Certification efforts need a crucial mass of farmers in a geographically constrained area and an internal control system that builds on the locally available organisational social capital. The context will vary and activities with farmers (treatments) tend to be fairly heterogeneous, as they are adapted to cope with these contextual differences.

Designs of evaluations, therefore, will need use quasi-experimental, observational studies, in which a group of beneficiaries is compared with a group of farmers that did not receive support. Several survey-based research designs are available to cope with differences in context and characteristics between these two groups (Khandker et al., 2009; Shadish et al., 2002). However, the econometric methods to find 'comparable' treatment and comparison households in these quasi-experimental designs depend on the limited information on key characteristics that is available. Even so, they are contingent on normative decisions about what to include or exclude as a variable in the matching model. Net effects on ultimate outcomes measured with quasi-experimental designs always have, therefore, a high level of inaccuracy and are subject to the positive or negative biases of enumerators, data analysts and inferring researchers.

Nevertheless, in the field of certification, reporting on net-income effects tends to be the prime focus of impact evaluations. Information on the mean and variations in income tend to be the anchors for statistical power calculations and the determination of minimum sample sizes. To calculate household income, fairly detailed quantitative information about crop revenue and input costs are needed. In the context of diversified farm systems with multiple crops, most common in smallholder agriculture, the disaggregation of labour time and input costs for the target crop is notoriously difficult. This results in ambiguity in the constructs that are used as proxies for income effects (Ton et al., 2012), e.g., net income including self-consumption of production, cash income, agricultural cash income, and income from target crops. The estimates given by the farmer of market prices, input costs and labour time spent is prone to recall bias, which results in unreliable income estimates when used in calculations.

Recent systematic reviews of the effectiveness of certification point to inconclusive results and, therefore, limited usefulness of net-income estimates in studies on the impact of certification (Alvarez and von Hagen, 2011; Alvarez and von Hagen, 2012; Crosse et al., 2012; SCSSC, 2012;

Blackman and Rivera, 2010; Blackmore et al., 2012). Except when price premiums are an important component of the intervention, as with Fairtrade and organic certification, the positive or negative changes in farmers' income are only very remotely related to the support and services provided to comply with the certification requirements. A wide range of factors determines income (yield, input use, costs, etc.). In most certification-related interventions, a change in yield or farm income will be influenced by a set of agricultural practices that are being promoted, such as the use of improved varieties, different handling protocols for plants and products, or new or enhanced soil conservation measures. However, even more important are the factors over which the intervention has no control at all, such as site-specific weather patterns that define yields, changes in market prices in response to site-specific trade dynamics, competition between buyers, or changes in crop patterns or off-farm income due to employment generating activities and seasonal out-migration. Multi-year agronomic research could provide more convincing evidence to support the assumed impact of improved agricultural practices on yields and income than the estimates derived from information collected through household surveys.

Our aim has been to link the above methodological considerations with on-going discussions on impact assessments within ISEAL and the practitioners' domain of standard-setting and certification (e.g. Vellema and Ton, 2012; Vellema, 2010). Our premise is that impact evaluation can better focus on the measurement of the change in knowledge on and implementation of agricultural practices, and 'reason through' the likely effect of these practices on yields and household income. We have illustrated the evaluation challenge by describing one of our experiences in the design and implementation of an evaluation study on certification of cocoa farmers in Ivory Coast.

### **3.3 Case study: Evaluation of training for certification in Ivory Coast**

As researchers of Wageningen UR, we were contracted by a number of organisations to co-design survey-based impact evaluations of cocoa certification initiatives in Ghana, Ivory Coast and Indonesia. Cocoa certification schemes typically require a codified set of good practices related to cocoa production and farm management, and include third-party auditing to confirm that the requirements are met or will be met within a specified time frame. In designing the methodology for these evaluation assignments, we attempted to identify the outcomes and processes in field-level certification initiatives that could be attributable to certification initiatives. In most cases, the main intervention was related to training of farmers in good agricultural practices - a mandatory requirement by certification schemes such as UTZ Certified - and the organisation of internal control systems.

#### **Multiple actors, treatments and contexts**

UTZ Certified is a certification scheme whose aim is to support sustainable farming worldwide. Its mission is to create a world in which sustainable farming is the norm— a world in which farmers implement good agricultural practices (GAPs) to manage their farms profitably with respect for people and planet, industry invests in and rewards sustainable production, and consumers can enjoy and trust the products they buy. In 2007, UTZ Certified launched

its cocoa programme with founding members Cargill, Ecom, Heinz, Mars, Nestle and Ahold and the not-for-profit organisations Solidaridad, Oxfam Novib and WWF, with the first pilots in Ivory Coast starting in 2008. In 2012, UTZ Certified and one of its implementing partners, the Dutch NGO Solidaridad, commissioned LEI-Wageningen UR to design and implement an impact evaluation of their cocoa programme in Ivory Coast.

Implementation of the cocoa programme involved a heterogeneous group of actors, each with different objectives, roles and responsibilities. NGOs, traders, private partnerships, governments and international public organisations were partners in implementing the certification programme. All cocoa farmers in Ivory Coast covered by UTZ Certified were organized as producer groups, which were generally cooperatives of varying sizes. Most producer groups were linked to particular traders, who assisted them in attaining certification. These traders could target more than one certification scheme (e.g., Rainforest Alliance Certification, Fairtrade). As a result, half of the 86 UTZ-Certified producer groups had multiple certifications with overlapping requirements. Approximately 21% of the farmers participating in the UTZ programme were also Rainforest Alliance certified, and 2% were both UTZ and Fairtrade certified. This resulted in a high level of diversity in the actual ‘treatment’ that the beneficiaries received (Table 3.1). The modalities of training differed from centralized sessions with professional agronomists to more intensive and participatory methods of knowledge exchange, such as field demonstrations in so-called Farmer Field Schools. Training was sometimes combined with additional support such as the supply of agro-inputs or credit.

This complexity of the interventions, which was detected at baseline, complicated the design of the impact study. The initial approach, which was to assess impact through a (matched) comparison of the ‘UTZ-programme group’ (the ‘treatment group’) and the ‘non-UTZ group’ (the ‘control group’) needed to be modified due to the diversity of treatments in the ‘certified group’.

Furthermore, both prior to and during the UTZ Certification programme, other activities had taken place relating to sustainable cocoa production, which had addressed the same type of practices that were promoted through the programme. Obviously, this history of support from multiple sources in different agro-ecological zones made it even more challenging to find treatment and comparison groups that would allow us to attribute changes in outcomes to the UTZ programme activities.

**Table 3.1** Example of the variation in treatments in support programmes targeting cocoa farmers in Ivory Coast

Type of treatment	Number	Explanation
Trader specific CSR programmes	8	Trading companies have their own CSR programme related to certification
Intervention activities per trader	2 to 5	Traders may provide training on business skills, organizing demo plots, providing gifts, inputs, nursery and seedling supply, etc.
Phases in certification schemes	6	Farmers can be in starting phase of the programme or have been in the certification programme for up to 5 years.
Training on different topics	>10	Topics including a variety of elements covering production methods, use of Personal Protection Equipment (PPE), waste management, etc.

Own elaboration, based on field data reported by (Ingram et al., 2014). Impact of UTZ Certification of Cocoa in Ivory Coast. The Hague: LEI Wageningen UR.

The variations in treatments were further compounded by differences in agro-ecological conditions in which the smallholder farmers operated. The statistical analysis of baseline indicators showed a very high level of variability (standard deviations of more than 75% of the mean) in indicators such as productivity and yield or net income. We used a regression analysis with more than 20 explanatory variables to detect the differences in outcomes that could be related to differences in characteristics between the group of farmers that were included in certification-related activities (treatments), and those that were not. The heterogeneity in treatments and farmer characteristics, meant that few treatments had a significant correlation, because of the dominant impact of being situated in specific agro-ecological zones having different production potentials. This pointed to the need to, at least, match treatment and comparison groups according to their agro-ecological zone.

### Sample sizes for rigorous measurements

Variability in treatments and contexts has substantial implications for the sample size required to detect a significant difference in income by comparing treatment and comparison groups. Table 3.2 presents an estimate of the minimum sample size that would be needed using a simple T-test to detect the expected differences in means between a trained and an untrained group. Although we used regressions to derive impact estimates, these minimum standard sizes were indicative of the minimum size of the household survey needed to detect significant differences between the different groups and reject the null hypothesis of no impact.

The measurement of net effects in knowledge increase and implementation of farming practices appear to be constrained by the logistics and budget of an impact evaluation (see Table 3.2). The minimum sample sizes needed to detect the (plausible) effects on yield and income between two groups were prohibitively high. Why should we collect very precise data on income and yields, with long interviews and burdensome data cleaning, when this would anyhow not give strong evidence of net effects? We decided that, for net-effect estimates, we could better restrict ourselves to the precise measurement of immediate and intermediate outcome indicators. With a reliable instrument to assess knowledge levels and farmer practices, we could verify the key assumption that training induced by certification is instrumental in changing farmer practices. The positive impact of these practices for yield and income would have to be verified with other methods, not household surveys.

Anticipating this, we had piloted an instrument during the baseline study to assess the level of knowledge on good agricultural practices in cocoa and the implementation of these practices by the farmers as immediate and intermediate outcome indicators. The farmers' knowledge level was estimated as a 'knowledge score' derived from their answers to a range of multiple-choice questions on good agricultural practices (GAPs) as required by UTZ Certified. Similarly, farmers' implementation of GAPs was similarly measured as an 'implementation score', based on their answers to questions about the practices that they had implemented on their fields. The valuation of the 'correctness' of the various answer categories for each of these questions was determined in consultation with agronomic specialists from UTZ Certified and local research institutes.

We reflected on the practices that were being promoted in consultations with groups of farmers and cocoa experts. Farmers and researchers may disagree on which practices are considered to be ‘good agricultural practices’, as they might use different criteria to judge some of these GAPs. Regular consultations with a multi-disciplinary expert panel to validate the local appropriate set(s) of good agricultural practices in cocoa is a cost-effective way to capitalize on the available experiences and evidence and ‘reason through’ the contribution of these modified practices to yields. To feed this discussion with farmers and experts, data was collected through household surveys on the reasons for implementing the practices or not. Special attention was paid to those practices that multiple training programmes have tried to convince farmers to introduce, but where implementation was low.

**Table 3.2** Minimum sample size calculations based on estimated effect sizes and baseline standard deviations

Outcome indicators		Trained group in same agro-ecological zone at baseline				Plausible net-effect to be captured in the research (a)	Minimal sample size needed for the hypothetical difference between the two groups to be statistically significant ( $p < 0.05$ ) using a two sample T-test*
Category	Indicator	Actual sample size (N)	Mean	Standard deviation	Variability (b)	%	Total N
Immediate outcomes	Knowledge score	436	0.246	0.110	0.45	30%	74
Intermediate outcomes	Implementation score	436	0.241	0.054	0.22	20%	42
Ultimate outcomes	Yield (kg/ha)	406	531	416	0.78	10%	1914
	Net income (USD/ha)	326	712	666	0.93	10%	2718

\* Sample Sizes were estimated using the statistical package Stata 13 [StataCorp. (2013). *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP] with the command: `power twomeans 1 1+a, power(0.8) sd (b)`, where *a* is the plausible net-effect, *b* is the variability of the indicator (calculated as the standard deviation divided by the mean).

Source: Own elaboration, based on Ingram et al. (2014). Impact of UTZ Certification of Cocoa in Ivory Coast. The Hague: LEI Wageningen UR.

### 3.4 Discussion

The evaluation challenge for certification, described in this paper, applies to a wider range of development interventions. Most impact studies of market-led development strategies tend to focus on outcomes related to the performance of business practices, such as rural incomes or well-being, which are difficult to attribute to the actual processes set in motion by the private sector support. Similar to the support to farmers in certification, these support interventions involve multiple actors and have many intervening factors that influence their performance. This makes it impossible to attribute changes in outcomes to one specific type of activities (treatment), or, even worse, to one specific supporting agency.

Because outcomes can be quite diverse, they may be difficult to simply compare between treatment and comparison groups. For example, the enhanced social network that results from certification-related activities may provide access to additional sources of credit, or,



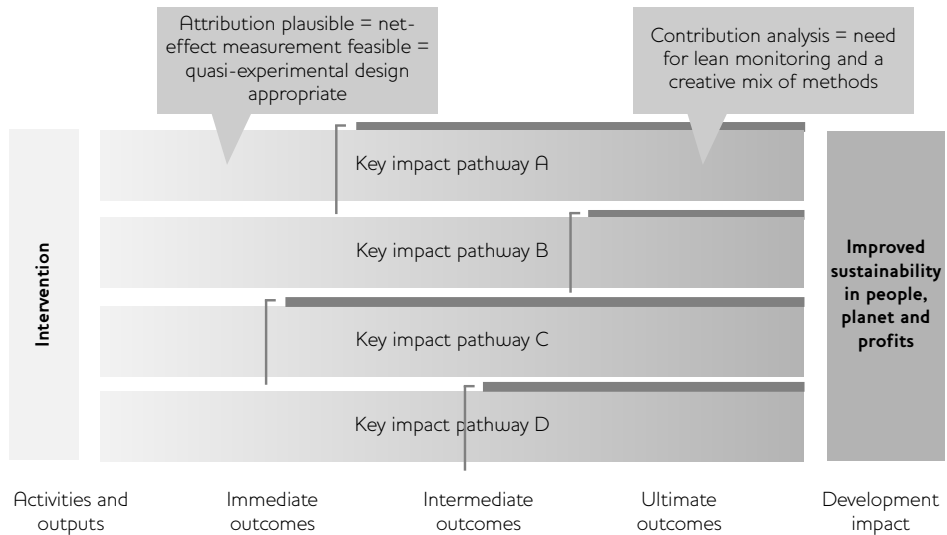
when the social network of a farmer is extended, this may stimulate the migration of their children to gain more promising livelihoods.

We showed that training in good agricultural practices is, at most, one of the contributory factors (Mayne, 2001; Stern et al., 2012) to outcomes, as are household income or yields. Monitoring changes in such ultimate outcomes may be possible, but deriving net effects and claiming attribution of changes in these outcomes to a single part of this complex of factors is not. Instead of attributing net effects using farmer or business surveys, other methods to verify the role of an intervention are needed. A research approach is needed that examines whether the type, amount and timing of support is right, instead of focusing on the causal effect of only one treatment — just as it takes matches, fuel and oxygen to start a fire (Mackie, 1965).

One of the possible ways to handle the evaluation challenges in private sector support, which emerged out of our collaborative research experience with certification organisations, is the importance of better explaining the theories of change. In this work, we identified impact pathways and discussed where to draw boundaries around the direct span of influence of the intervention in each of these pathways to impact. This exercise helped unravel plausible causal processes triggered by interventions and to identify multiple intervening factors that influenced outcomes; these factors, or combinations of factors, are often more essential to effecting changes and at the same time largely unpredictable.

It is evident that a creative mix of methods is needed to collect the multiple strands of evidence to ascertain the contribution story (Mayne, 2001; Mayne, 2012). We expect that large scale surveys of household income and expenditures will be less useful for reflecting on how certification improves sustainability and alleviates poverty than more qualitative approaches. It is also important to analyse how interventions contribute to wider development processes. Expert panels that reflect on sector dynamics and identify the strengths and weaknesses of various interventions can assess the achievement of intended ultimate outcomes; this can help to estimate the added value of the support.

The implication of the above is that, when moving towards ultimate outcomes, it is important to recognize where interdependencies of factors and actors become too dominant to make net-effect measurement feasible. This suggests a shift from 'impossible' quantitative-attribution-oriented research on ultimate outcomes to a precise identification of proxy-indicators for key immediate and intermediate outcome areas that are still (plausibly) attributable to the intervention. As illustrated in Figure 3.2, each pathway in the intervention logic will have a different boundary of the span of direct influence — a boundary for which net-effect measurements become impossible. This boundary will be a result of causal-theoretical logic, and a function of budgetary and methodological constraints (Bamberger et al., 2004). Quasi-experimental research designs may be appropriate when assessing net effects within this bounded span of influence. They will need to focus on those immediate and intermediate outcome indicators within the span of direct influence, where a change (+ or -) is still indicative for performance of the intervention. Only on those outcomes can these designs provide the 'credible counterfactual' (Ruben et al., 2009; Alvarez and von Hagen, 2011) and generate meaningful and informative net-effect estimates.



Source: Authors' own elaboration

**Figure 3.2** Different impact evaluation designs within and outside the span of direct influence of an intervention

Beyond this boundary of the span of direct influence, monitoring information on ultimate outcomes (poverty, income, yields) may still be informative, but not for establishing net effects attributable to the intervention. Therefore, rough indicative values are sufficient, for example to compare the participant group with others, exact measurements to calculate absolute values in net-effect calculations are not needed. Instead of collecting detailed data on these outcomes in household surveys, there is a need for lean proxy-indicators that help to map the relative poverty position of a household, such as the Progress-out-of-Poverty indicator (PPI). While the PPI is not appropriate nor intended for net-effect calculations (Chen and Schreiner, 2009), this simple questionnaire takes very little time for both the respondent and enumerator. Being a common indicator, it may provide useful information to compare the targeting of the support in relation to alternative interventions that have similar goals.

### 3.5 Conclusions

The paper builds the case to reconcile precise measurements of immediate and intermediate outcomes in business practices that are considered to be within the span of direct influence of an intervention, combined with other methods to verify the casual assumptions of contribution of these practices to development impacts, following the logic of 'contribution analysis'. We propose to refine the intervention logic in distinct impact pathways, to identify key assumptions and key outcome areas, and draw the boundary of the span of direct influence of the intervention. We propose contribution analysis as an overarching approach that combines

precise survey-based net-effect measurement of immediate and intermediate outcomes, with less precise, lean monitoring of indicators and the use of a creative mix of methods to verify the contributory role of these outcomes in household income and poverty alleviation. Using similar indicators in key outcome areas would create enhanced opportunities for systematic cross-case analysis and for benchmarking and learning.



## 4 | Development Impacts of Value Chain Interventions: How to collect credible evidence and draw valid conclusions in impact evaluations?<sup>3</sup>

Giel Ton<sup>a</sup>, Sietze Vellema<sup>b</sup> and Marieke de Ruyter de Wildt<sup>a</sup>

<sup>a</sup>. Agricultural Economics Research Institute, LEI Wageningen UR

<sup>b</sup>. Wageningen University, Chair group Knowledge, Technology and Innovation

### Abstract

*In development policy and practice, support to or interventions in value chains are considered to be instrumental for achieving outcomes such as poverty alleviation. This paper reviews methodological discussions on how to show the effects and workings of value chain support in a context of donors demanding rigorous impact evaluations. The paper starts a discussion with evaluation methods strongly anchored in ex-post statistical analysis of effect measurements, and argues in favour of a theory-based evaluation protocol, equipped to handle threats to valid conclusions. Value chains are open, multi-layered systems and development outcomes are multi-dimensional and contingent on contextual particularities. Moreover, development interventions in value chains are often time, place and commodity specific, unlikely to repeat in a similar way, which complicates generalization and constrains evaluative conclusions. The example of a small-grant fund promoting collective marketing by smallholder organisations illustrates these methodology challenges and shows the value of using a mix of methods for addressing the problems of outcome measuring, impact attribution, and generalisations from highly diverse contexts.*

---

<sup>3</sup> Published as: Ton G, Vellema S and de Ruyter de Wildt M. (2011) Development impacts of value chain interventions: how to collect credible evidence and draw valid conclusions in impact evaluations? *Journal on Chain and Network Studies* 11: 69-84.

## 4.1 Introduction

Value chain development has emerged as an important area of donor interventions for poverty reduction in developing countries. The World Development Report 2008 (World Bank, 2007) put it as a centrepiece to agricultural policy in developing countries. Value chain support focuses on capacities and capabilities of value chain actors and the enabling policies and institutions that facilitate change processes that benefit the poor. Donor interventions link actor specific effects, such as increased incomes for farmers, to overall performance of a specific value chain. Chain performance can be enhanced by policies and projects that, for example, increase the scale of operations, improve service provision to producers, develop capacities to comply with (buyer-driven) quality requirements or address the process of value creation and value distribution. This also makes value chain development or support container concepts that have strong parallels with generic policy approaches such as ‘private sector development’ (Donor Committee for Enterprise Development), ‘making markets work for the poor’ (DFID), ‘growing inclusive markets’ (UNDP), and ‘opportunities for the majority’ (IADB).

The discussion in this paper is motivated by the growing public pressures on development cooperation and aid money to show its worth, convincing evidence is asked for the effect on poverty alleviation (OECD 2008; SDC 2009). These calls for credible evidence have led to more stringent accountability requirements for agencies to defend the logic and demonstrate the impact of these interventions (Tanburn, 2008). However, generating convincing evidence on the link between development ‘outcomes’ and donor supported intervention ‘inputs’ is not easy.

Value chains are complex, multi-layered and open socio-technical systems that are influenced by a myriad of intervening actors, and are continuously shaped and reshaped to adapt to changing conditions; a whole range of unintended consequences of value chain support are difficult to grasp; and, there is the critique that value chain support picks ‘winners’ by focusing on a relatively small group of entrepreneurial poor and hence has a limited impact on average poverty levels (Humphrey and Navas-Aleman 2009).

Development outcomes are also dependent on the ways in which value chains touch down in specific contexts wherein governments, business associations, entrepreneurs, producers organisations and/or labour unions negotiate rules and institutions that shape development outcomes (Helmsing and Vellema, 2011). Attribution of impacts of interventions in this dynamic ‘cloud’ of complex and intertwined sets of institutional arrangements is difficult, though necessary to answer legitimate questions on relevance, effectiveness and replicability of value chain support (Roche and Roche, 1999; OECD, 2008).

One of the promising initiatives to generate credible and comparable information on donor funded value chain interventions originates from the Donor Committee for Enterprise Development (DCED, 2010). The initiative proposes a set of minimum standards for reporting on private sector development outcomes. When fully implemented, this could result in a body of evidence on plausible regularities resulting from value chain support, which lends itself for comparative analysis and benchmarking (Maredia, 2009). The argument developed in this paper is to warn against a one-sided emphasis on measuring outcomes, and against a ten-

gency to constrain the design of impact evaluations to one-method econometrics. We stress that a methodological design has to face the most common threats to validity to the evaluative conclusion (Bamberger et al., 2006; Creevey and Woller, 2006; Shadish et al., 2002).

The paper consists of three sections. First, we discuss the basic ingredients of impact evaluations, namely the questions asked and conclusion drawn, and illustrate the different threats to validity when concluding on evaluative questions. Second, we examine in more detail the methodological challenges for measuring outcomes patterns, attributing outcomes in open systems, and generalizing conclusion from diverse pilot experiences. Third, we illustrate the points raised with a real impact evaluation assignment of an intervention supporting collective marketing activities by farmers organisations, active in a wide range of different value chains and social contexts. In the conclusions, we stress the importance of linking ex-post impact evaluation processes with ex-ante constructions of plausible impact theories and credible outcome measurement methods.

## 4.2 Impact evaluations: Credible evidence and valid conclusions

There are many different reasons for doing an evaluation. Chelimsky and Shadish (1997) distinguish three types of evaluation: evaluations that primarily look for accountability, for knowledge, or for development. Accountability evaluations look at the value of public expenditures, focusing on issues of costs, efficiency and effectiveness; knowledge evaluations aim for insights into problems, policies, programs and processes, analysing old interventions in order to develop new ones; and, development evaluations seek to strengthen institutions and agencies in a particular evaluative area. The first two types are largely summative in nature, while the third type is largely formative. We focus in this paper especially on the first two types of evaluations, which are interested in measuring and attributing outcomes of interventions. This section examines two key aspects of such impact evaluations, what type of questions to start with, and how to end-up with answers that can face critical scrutiny.

### Asking evaluative questions

Summative impact evaluations address varying combinations of three basic questions for which information and evidence has to be collected:

- Does it work? What positive and negative changes did the intervention generate in the performance of the value chain?
- How does it work? What components of the support generated intended or unintended effects, for whom, and under what conditions?
- Will it work elsewhere? What components might work for whom under what conditions?

The first question is a quest for evidence of effects and especially relevant when public or private investments have alternatives and need an indication of the extent to which their support contributed to stated objectives. The second one is especially directed to the processes that make this happen: the generative mechanisms (Gerring, 2007). Thirdly, an impact evaluation is often commissioned to assess the possibilities to replicate an intervention in other contexts, or upscale it from 'pilot' to 'mainstream'. This third question is most directly related to the policy recommendations of an evaluation; often the best read part of any evaluation report and the most vulnerable to critique.

Each evaluation assignment will build upon a combination of these three basic questions. The 'weight' of each type of questions is decided at the start of the assignment. The first and third questions get generally more attention than the second, the 'How?' question, and this has major impacts on the research methodologies preferred. We share the observation by Pawson and Tilley (1997) that the first question is far too dominant in evaluation research whilst the second question is more productive in providing guidance to stakeholders and in generating useful policy recommendations. Also Ravallion (2009), chief evaluator at the World Bank, points to the dominance of evaluation assignments that limit themselves to show that policies work or not, without generating additional information on how they work and could work in other settings. He opposes, specifically, the dominance of econometric impact assessment methods that only compare average values of indicators between treated and control groups. According to Ravallion (2009), echoing similar comments by Heckman and Smith (1995: 95) the audience of most impact assessments, policy makers also want to know whether the intervention worked in the way it was intended to work, whether it worked the same for different groups, whether it still worked after the program was scaled up, and whether it can work differently.

The above presents one of the challenges central to this paper, how to arrive at an appropriate combination of evaluation questions with a proper research design to measure outcome patterns, gain insight in inner workings of the interventions, and an appreciation of the social embeddedness of impacts. The three questions require different kinds of information, or, at least, with different 'depth and detail'. Whereas the first question may treat the intervention as a one-package black box, the second question explicitly opens the black box to know what generative mechanisms are fired, or not. Information relevant for the second question documents processes, sequences of instructive events, time paths and outcome patterns differentiated for distinct groups or areas. The answer on the third question builds on the data and conclusions of the first two questions and needs information on the circumstantial conditions enabling or constraining these processes.

### **Drawing valid conclusions**

Above we argued that any evaluative assignment has to arrive at an appropriate combination of three questions. The answers to these questions need to be robust and be able to address the most obvious threats to validity. Shadish et al. (2002) argue that no generalised causal inference has absolute validity, there will always be some specific conditions that limit the generalisation domain of the conclusion. They stress the need to design procedures that (partially) control some of the limitations of the research methods used that may weaken the

validity claims of causal inferences. However, in many evaluations the validity issue is only applied to the tools and techniques used, not to any systematic assessment of validity threats on the conclusions. In this paper, we acknowledge that most impact evaluations have to make pragmatic choices related to budget, time, data (or sometimes political) constraints, and attention to how to still draw valid conclusions on the basis of available evidence seems timely (Bamberger and Rugh, 2008). Few evaluations in international development systematically address issues of validity; the field of value chain support is no exception to this (Zandniapur et al., 2004; Humphrey and Navas-Aleman, 2009).

We consider it useful to adopt the four threats to valid conclusions proposed by Shadish et al. (2002):

- **Statistical conclusion validity:** the way inferences about correlations are made in data-set observations. It emphasises the need to comply with proven methods to estimate association or correlation between variables.
- **Internal validity:** the way causality is attributed in the evaluation. This refers to the logic behind the observed correlations and explains why and how interventions contribute to the observed change.
- **Construct validity:** the way that generalisations are made from the categories used in the evaluation to broader units of representation. It stresses the importance of precise definitions and concepts.
- **External validity:** the way that the findings are generalizable to other persons, times and contexts. This requires being precise about conditions and requirements that define the generalization domain.

*Statistical conclusion validity* involves statistical analysis of data-sets, usually comparing groups of respondents and calculating averages or other measures of comparison in the sample population. We then use several tests to conclude on the probability or ‘significance’ of a correlation between their characteristics and the outcomes. Just producing an output table that indicates ‘significant’ relations is insufficient. All statistical tests have assumptions and pre-conditions related with the data, like the ‘normal distribution of the data’ or the ‘homogeneity of variance of the different groups’. Taking statistical conclusion validity seriously, we need to be explicit about such assumptions, and include analytical methods to check these assumptions.

*Internal validity* is intimately related to the argumentations to support a causal inference. It is important to be clear how the evaluative research makes the link between an intervention (cause) and specific outcomes in the value chain. There are three basic conditions that define causality (Mill, 2009 (1843)): the cause needs to be active before the effect is produced; the cause must be related to the effect produced; and alternative explanations of the effect must be discarded. In value chain development, it is unlikely that there is just one cause of the change. More likely it is ‘configurational’: several jointly influences are necessary for produc-



ing the observed outcomes (Blatter and Blume, 2008). Each individual factor in a configuration is a so-called *inus condition*: in itself *insufficient* to explain the outcomes of a support intervention, but a *non-redundant* part of a wider constellation of factors that is *unnecessary* but *sufficient* to produce the outcome (Mackie, 1965). For attributing outcomes to an intervention, hence, we will have to make plausible that the value chain support was indeed a necessary ingredient of the configuration of factors that produced the observed change in outcomes. Non-observables, characteristics or factors that are not (and often cannot be) registered in the data-set, may provide for alternative explanations of the observed effects. Strong evaluative conclusions, therefore, need to collect information on a wide range of factors that might be important. To support an evaluative conclusion on the effectiveness of a value chain support intervention, the non-redundancy of the intervention in this constellation of causal factors will have to be made plausible.

*Construct validity* is about whether we indeed collect information about categories and concepts used in the research design. The evaluators need to be explicit about the way they generalise the concepts and constructs that they use in the evaluation. For example, if we conclude something about the effectiveness of a certain intervention in the chain, e.g. 'investments in cooling tanks makes linkages of dairy producers to markets more remunerative, we immediately face several threats to construct validity'. Is 'dairy producers' a good construct, or do we need to make distinctions in small and bigger dairy farmers, diversified farms or specialized farms? Does the inference hold for all types of investment support that facilitate cooling tanks in this specific case, or do we need to make distinctions in grants and credit schemes, or farmer-driven and government-driven schemes? Is it valid for all markets, or only for the urban fresh milk markets and not for cheese and yoghurt markets? To face threats to construct validity, we need to be precise about the concepts and constructs used and design our research methods accordingly.

Even more challenging are the threats to *external validity*. When we come to the conclusion that in a specific context the intervention was a key factor with positive results, this will not necessarily hold in all other settings. Hence, we need to clarify why, and to what extent, the outcomes came about in this specific situation and whether the findings can be generalized and remain valid for other contexts and conditions. All 'best practices' and lessons learnt on value support can be questioned by indicating a 'peculiarity' related to the context. For example, the outcomes of what seems to be a technical intervention in a value chain, such as setting up a warehouse, can be dependent on the nature of collective action among farmers or on the articulation of farmer groups with business association and local government departments in a context of decentralization.

In this section, we discussed that impact evaluations require capacity to combine three different questions and to address four threats to drawing valid conclusions. Below, we will examine such a combinatory framework is linked to collecting evidence informing how we define, describe and defend our 'generalisation domain' (Chen, 1994).

### 4.3 Methodology challenges

We will now apply these validity checks to three methodology challenges that are core in evaluation research on value chain dynamics and that are intimately related with the evaluation questions: Does it work? How does it work?, and, Will it work elsewhere? Our first concern is the problem of measuring outcome patterns. Performance indicators vary between relative simple indicators to complex constructs that are difficult to operationalize. Second, we focus on the issue of attribution. In complex and multi-layered social systems like value chains, not one intervention functions in isolation: many stakeholders, prices and market trends influence value chains that are socially embedded in diverse cultural settings. More so, interventions have various components, implemented with different time frames, in varying combinations that interact with each other. We complement this with a third challenge: how to generate generalizable conclusions on interventions that are socially embedded.

#### Measuring outcome patterns

The first evaluation question, does it work, seeks to measure the change caused by the intervention. The DCED (2009) proposes some basic steps for this: define the impact model; define indicators of change (and projections); measure these indicators; and capture the wider change in the value chain. In value chains, support is often directed at actors and institutions in the environment of (poor) producers, like financial and non-financial business support services, rather than at producers themselves. All interventions will have an explicit or implicit 'theory of change' or impact model on how the support is expected to translate into desired outcomes on these service providers and institutions, and a theory on the way that these changes translate are expected to generate better incentives for producers in the value chain. This impact model helps to focus the evaluation on key assumptions in the logic; the evaluation is not necessarily comprehensive but, indeed, may better concentrate on 'critical' subsets of conditions, components of interventions, specific instruments, and types of outcome patterns that are expected to be present/working (Maredia, 2009).

In designing the concepts and indicators in impact assessments, construct validity is a key challenge. Performance of a value chain relates to different layers and dimensions of social interaction in the chain network. Similar to the challenges to assess other abstract attributes of social systems, like 'organisational strength', the immaterial aspect of chain performance makes it difficult to capture and measure. More so, concepts and indicators to assess performance are often influenced by the disciplinary background, ontological theories and personal interests of the evaluator (Vellema, 2011). For example, when looking for outcomes of support to multi-stakeholder chain platforms, an economist trained in transaction economics will look for 'trust' and 'coordination' between chain actors, while someone specialised in the analysis of group dynamics will focus on 'inclusion/exclusion' and 'synergy'. A political economist will see 'changing power relations' and a scholar in strategic marketing will look at 'innovativeness' and 'competitiveness'. All will see some of the outcomes of the intervention, but not the whole picture. It is, therefore, important to carefully select indicators and select an evaluation team and research methodology that is able to identify and operationalize the relevant performance indicators (Snodgrass, 2006), where triangulation of different sources of

information and metrics helps to better frame changes in value chain performance (Denzin, 1970), and multiple ways of approaching may help to identify causal factors, *in* conditions, that explain their realisation (Blaikie, 1991).

Even apparently straightforward indicators need to be well defined, according to a causal model that is comprehensive enough to include the most important outcomes, but lean enough to facilitate attribution. One of the three ‘universal’ indicators proposed by DCED (2010) is “additional net income (additional sales minus additional costs) accrued to targeted enterprises as a result of the programme per year”. This sounds straightforward, however, even here, the scope for varying interpretations is considerable. E.g. net additional income as a result of a dairy development intervention, can be restricted to net income growth from fresh milk sales. However, it can also be understood as the net income change of the whole agricultural system of the household, as increasing dairy production and increased animal feed production may impact horticultural production and family income. Positive spill-over effects may exist too, since farmers may have learned about milk quality issues, and, as a result of increased communication with other chain actors, may have improved their entrepreneurial skills and technology beyond dairy only. However, this more comprehensive way of calculating net income introduces a wider range of confounding factors, that complicate the attribution of the impact to the specific intervention: e.g. horticultural prices fluctuate a lot between weeks and seasons and are prone to natural conditions, and will influence incomes without any causal relation with the dairy support intervention being evaluated.

Measuring needs accuracy. Commonly, changes in value chain performance are assessed by subtracting or comparing indicator scores: at least a ‘before-after’ situation and, if possible, a ‘with-without’ estimate. Measuring *differences* in indicator scores with some accuracy is more important than measuring the absolute value of the indicator. Relatively small measurement errors in both indicators may translate in large errors in the calculated difference between them. Tracing these measurement errors in indicator averages between non-equivalent groups is difficult and often limited to outlier checks only. E.g. difference in income calculations from the same respondents in consecutive surveys can be influenced by recall bias or other critical estimation errors (Ton et al., 2012). Too often researchers analyse data that has not been checked sufficiently for data collection errors (Vaessen, 2010). Careful outlier elimination often improves the analysis lot. However, routine outlier eliminations can have adverse results; while the reasons for these outliers can indeed be related with data collection or data transcription, and elimination from the data-set is an improvement, but, on the other hand, outliers may well reflect real situations and be very useful for understanding the change processes and enabling conditions related to the interventions, e.g. for applying analysis of ‘contrasting cases’ (Lawson, 2009), and eliminating them reduces the information in and usefulness of the data-set..

To indicate the impact that is attributable to interventions, a comparison is needed with a fictive situation that the intervention had not been active (Khandker et al., 2009; Rihoux and Ragin, 2009). Ideally, this would be a group with similar characteristics that did not experience the working of interventions. The comparison between these two groups helps to assess if the outcomes can be attributed to any ‘exogenous’ or ‘unknown’ causal factor, not related to the intervention’s causal mechanisms. Experimental methods, with random assignment

to treatment and control groups (Duflo et al., 2006) are especially designed to facilitate this measurement of outcomes between treated and non-treated groups. However, this design is often impossible (Shadish et al., 2002; Bamberger et al., 2006) and deliberate exclusion of some groups of stakeholders in the value chain from the benefits of a support intervention (like coordination platforms, value chain financing, certification programs, investment subsidies) is often socially and politically unfeasible. Also, in many cases, there are important spill-over effects from pilot-intervention areas to other areas and chain actors that make the definition of who is a participant and who is not a gliding scale, and the distinction in 'treated' and 'control' groups unworkable (Ravallion, 2009). As random assignment of the intervention to a defined population is rarely possible and, therefore, other, quasi-experimental methods are, hence, more frequently used. However, research designs that deviate from random assignment face the risk of being affected by a selection bias, introducing differences between the treatment and the control group that are unrelated to the intervention, but important in producing the outcomes (e.g. attitude, resource base, etc.). This is a major threat to the validity of the statistical conclusion. A proper evaluation design will have to consider, limit and control for such a bias in data-set observations.

Generally, a survey ends up in a set of qualitatively distinct variables used as proxies for 'improved livelihood strategies of smallholder households'. Statistical analysis, with a set of distinct dependant outcome variables, generates additional threats to validity of the correlations found. Current software makes consecutive iterations of statistical analysis with changing combinations of variables so easy, that 'significant' correlation between variables may result from 'fishing the data': repeating statistical tests that analyse the significance of differences between groups by selectively re-grouping respondents, variables etc. Even if the intervention has no effect at all, in complex data sets, one or more significant correlations are likely to appear after a sufficient number of iterations (Shadish et al., 2002). Concluding on causal relations from such correlations may wrongly attribute these outcomes to the intervention. On the one hand, as conclusions tend to concentrate primarily on significant effects, this results in a bias in impact evaluations towards 'significant' though irrelevant conclusions. On the other hand, non-significant effects can be a result of low statistical power (low sample size) or measurement errors that could have been corrected when more deeply analysed. Large mean effects need attention even when they do not prove statistically significant. A recommended solution against 'fishing' is to specify, ex-ante, the hypothesis or theoretical model that is tested and to increase the threshold (significance level) of the correlation detected through iterative analysis. However, fishing is difficult to detect as often no ex-ante causal hypothesis exists or, more common, the hypothesis is adjusted during analysis and reporting the data. Interestingly, this temptation is even stronger for academics involved in evaluative research, as the chance of research results to be published in scientific journals is far higher with an argument that is supported with 'significant' statistical evidence, especially when sample sizes are low (Begg, 1994) This 'publication bias' creates incentives for ex-post modelling of hypothesis and generates a problem for meta-research, as there is an overestimation of changes as a causal result of interventions in the literature.

Only data-set observations from surveys with a sufficient sample size (*statistical power*) will make it possible to detect differences between subgroups in the survey population. Commonly, a minimum subgroup size of 30 is used as a rule-of-thumb (Creevey and Ndiaye, 2008). The sample size will have to consider attrition, some respondents will fall out of the sample due to

moving, passing away or changing activities. When one wants to compare between different subgroup locations ( $g$ ) disaggregated on different typology criteria ( $c$ ), this minimum total sample size will, thus, be, roughly,  $N = 30 * g * c$ . For explorative statistical analysis, and considering attrition, sample sizes are ideally larger than the minimal required size. In the 'real world', however, sample sizes are often restricted by resource constraints (financial, not enough people, too difficult to get to, etc.).

The DCED recommends to capture wider changes than just 'predicted' change by the logical model or intervention theory. The most obvious threat to validity of an evaluative conclusion is that it leaves important factors out of the equation, be it as confounding causal factors or as outcome indicators, and, so, weakening the internal validity of the findings. Unintended changes are unlikely to be captured by pre-established indicators in causal impact models. Additionally, more open and qualitative Causal-Process Observations (Brady et al., 2006) are needed to check for these unintended outcomes. The emphasis on documenting wider impacts is important; too often, evaluations restrict assessment designs to find proof for impact logic only (European Commission, 2008)

### Attribution in open systems

Data-set observations need causal theories to differentiate between co-linearity (it happens together) and causality. Analyses of the logic behind the observed changes are necessary to interpret these correlations and to identify causal relations. Significant correlations do not indicate causality, but at least indicate that there is, most probably, a relation between the intervention and the outcomes. This holds especially for relative simple or moderately complicated systems (Snowden and Boone, 2007; Rogers, 2009). However, this is far less realistic for interventions with a wide constellation of causes in systems that behave with increasing levels of complexity (Lawson, 2003; Pawson, 2002; Hospes, 2008; Snowden and Boone, 2007). If value chain support takes place with a high degree of contingency in system behaviour, experimental and quasi-experimental methods that rely on data-set regressions alone will have problems in the internal validity of their conclusions (Heckman, 2005).

The difficulty to grasp complexity of change process in econometric models holds also for the popular evaluation research designs based on comparing groups through 'matching' procedures, like Propensity Score Matching (PSM). In PSM, impact is assessed by measuring the outcome difference in pairs of respondents that 'match' on most of their characteristics, except their adoption of the innovations promoted by the intervention. The characteristics on which matching takes place are, ideally, derived from a model that comprises the whole 'constellation of factors' that are expected to lead to the measured outcomes (e.g. adoption of technology that leads to higher income levels). The matching is done through calculation of a 'propensity score' for all respondents on a construct of different variables that 'models' the context of the respondent. The respondents with a comparable score on the model's dimensions will form 'matched pairs' and are supposed to share the likelihood to have the same outcomes, except the ones that result from the adoption of the innovation promoted by the support intervention. The difference in outcomes between the 'matching pairs' of adopters and the non-adopters are considered to be attributable to the intervention. As will be clear from the above, these matching models are heavily theory-laden, and they suppose that

the matching is done on all relevant variables that will make the pairs react similar to the interventions incentives. This model to ‘capture context’ is ideally elaborated before the PSM survey data is gathered (because on all characteristic there need to be information from the survey), but, in practice, it is often only constructed after the survey, during data-analysis and the matching model is constrained to only those variables that are readily available in the data-set for both beneficiaries and non-beneficiaries. But even when the data collection is comprehensive, in complex systems, the model used to match respondents will always be incomplete and will suffer from ‘essential heterogeneity’ (Heckman, 2005): it may miss a latent, unobserved external that is key in the constellation of causal factors that determine the reactions of stakeholders to the interventions. Even the more sophisticated econometric methods that explicitly try to correct for the variance due to unobservable factors that influence a respondent’s behaviour will end up testing closed models of reality. Therefore, critics may always challenge the validity claims of causal inferences derived from econometric analysis of survey data, indicating that the model is too simplistic and that the context is far more complex to be captured in mathematical models (Lawson, 2003). A (partial) defence against these critics is to limit PSM to only those social processes that are relatively simple. Social systems with increased levels of complexity or chaos limit the possibility to generate credible and valid conclusions with quasi-experimental research designs (Maredia, 2009).

All value chain interventions are, ultimately, intended to change attitudes and behaviour in persons. The workings of the intervention are often implicitly assumed in the impact logic. Chen (1994) makes a useful distinction in two sets of intervention theories: *causal theories* and *normative theories* of program impact. Causal theories are descriptive of change processes in social systems, while normative theories are more prescriptive and action-oriented and represent the impact model behind an intervention. Obviously, the latter benefits from the first and normative theories improve when more causal theory is generated.

Realist evaluation (Pawson and Tilley, 1997) provides a useful framework for analysing specific mechanisms in an intervention that may be ‘fired’ in a specific context and that trigger behavioural change. It emphasises the need to build ex-ante hypotheses related to the (project) mechanisms that (are assumed to) motivate or influence stakeholders ‘to act differently’ and generate changes in outcomes. Realists propose to test key assumptions in these hypotheses with the concepts “Context-Mechanism-Outcome Configurations”. The realist concept of ‘mechanisms’ opens the black-box between intervention/treatment and outcome/impact. The concept ‘configuration’ indicates that mechanisms will only produce certain outcomes in certain contexts, making key discriminations that automatically limit the generalization domain of the causal inference. In contrast with the mainstream econometric approaches, realist evaluators concentrate on the ‘treatment’ and the different configuration of incentives for the ‘treated’, without bothering too much about a control group. They emphasize that mechanisms work under specific conditions, part of a wider constellation of causal factors.

In evaluations of value chain support interventions, the realist framework can be used to describe the workings of interventions in its context (Table 4.1). The detailed description and analysis of a pilot intervention may then provide the framework for a new intervention theory in new policies, and the research results may feed into learning process on good principles or practices in value chain development.

**Table 4.1** A 'realist' case study format for value chain research

	Domain of application	
	Causal Theories	Normative theories
Aim	Understanding pilot interventions	Designing policy and program
Context	Situation of the value chain stakeholders in the pilot experience	Situation of the value chain stakeholders in another setting where the support intervention will take place
Mechanism	Incentives that condition the behaviour of stakeholders in specific institutional arrangements that have emerged in and around the value chain	Intervention that changes the incentive structure for stakeholders and generates an improved institutional arrangements in and around the value chain
Outcome	Actual performance of these institutional arrangements in the value chain.	Intended outcomes of the intervention on institutional arrangements.
CMO-Configurations	Comparative case descriptions of causal connections between interventions and the performance of specific institutional arrangements.	Defined recommendation domain for replicable policies and interventions that enable effective and sustainable institutional arrangements in the value chain

Critics of realist evaluation point to the tendency to generate a whole range of hypothesis from qualitative case-studies that cannot be tested with the empirical evidence. "Realists may be strong in identifying rival explanations for the observed impacts and outcomes, but quite poor in convincingly testing and eliminating the erroneous ones" (Farrington, 2003). Farrington argues that with limited time and resources for evaluations, it is difficult to deal with multiplicity of contexts, mechanisms and outcome patterns. He argues that qualitative research methods alone lack the necessary procedures to answer the most obvious threats on internal validity. He strongly favours the use of statistical analysis of data set observations on ex-ante hypothesis on the effectiveness of mechanisms.

We agree with Farrington that, when designing impact evaluation research, it is recommended to generate comparative data-set observations with a broad enough set of variables to test plausible explanatory theories that can be used to support the validity claim of inferences from the more 'thicker' realist case-studies. However, ex-post hypothesizing remains important; a combination of 'deduction' and 'induction', or, in the realist tradition between 'interpretation' and 'abduction' (Blaikie, 2000) are key to make sense of the surprises and dynamic changes that characterize most value chains. We might also disagree with him on the type of statistical analysis that can be applied on this data-set to derive to meaningful conclusions on causality, where variable-based regression seems to be the golden standard (Lawson, 2003).

Recently, several new methods have been developed for statistical analysis of configurations of mechanisms, instead of discrete variables, that are active in each case-study. Instead of econometric regressions that look for significant correlations between case on one variable, they propose methods of classification that respects the integrity of each case as a system that can have different outcomes according to different sets of variables/factors. These 'Case-based Methods' (Byrne and Ragin, 2009) include diverse tools like cluster analysis, contrasting case methods and Qualitative Comparative Analysis (Ragin, 2000; Rihoux and Ragin, 2009) and are promising as part of a mixed-methods evaluation design capable of configurational explanations.

## Social embeddedness and generalisation

Threats to external validity arise when the conclusions of an evaluation are not bound to the population and or context from which observations are made, e.g. a sample has been taken, and are applied to contexts and conditions that are totally different in space and time. Unfortunately, this generalisation of conclusions to dissimilar contexts and conditions is often explicitly asked for when the evaluation of agricultural value chain support pilots are concerned. External validity of these generalised inferences might be threatened, as evidence of impact in one commodity chain will not necessarily be relevant for another commodity and evidence in one cultural setting or time period will not be generalizable to another. Policy makers are often especially interested in ‘best practices’, as it provides them with a menu of options. Evaluative conclusions to be used by policy makers, therefore, have to maximize the generalisation domain while maintaining validity and credibility. ‘Good Practices’, ‘Best Fit Solutions’ or ‘Principles’ are all concepts used to indicate mechanisms or interventions that proved to have worked in a certain setting, and that might work in others. Instead of strong causal inferences about ‘Best Practices’, only possible in relatively simple social change processes, these concepts apply to more complicated and complex ones.

In this section, we discussed in more detail methods for collecting credible evidence on outcome patterns and for detecting candidate mechanisms. To maximize the generalisation domain of the conclusions on impact while respecting its limits, Shadish et al (2002) present principles to check external validity of findings. Two of which are especially useful for comparative research on value support. First, to assess the apparent similarities between the context of the intervention that has been studied and the characteristics of the context targeted for replication (Surface Similarity). Second, to explicate those contextual differences that are irrelevant because they do not change a generalisation (Ruling Out Irrelevancies). We consider the challenge to generalize conclusion to other conditions as uncharted terrain where less prove tools are available. To facilitate this, the format of realist case-studies, explained above, seems helpful to generate comparable information on contexts and mechanisms that work in specific value chain support interventions.

## 4.4 Application

From the above, we can distil the main steps to get a rigorous methodology design for impact evaluation assignments in agricultural value chains. For evaluations that intend to conclude on the replicability or scalability of agricultural value chain interventions, we propose a design that is based on a combination of 1) impact models that reflect the intervention theory, 2) a realist focus on the mechanisms that are assumed to be fired by the intervention and the conditions under which these work, and 3) triangulation of data-set observations and casual process observations with, case-based, comparative statistics. The resulting mix of data-collection methods provides useful information for accountability purposes, for monitoring on-going interventions and for learning on good principles and good-fit practices that have potential to be effective when replicated in future interventions.



We summarize the methodology challenges and the evaluative questions as described above in the following Table 4.2, in which ‘++’ indicates the core challenges for a research design to provide answers on each of the questions.

**Table 4.2** Evaluation questions and their main methodology challenges

	Measuring outcome patterns	Attribution in open systems	Generalization and social embeddedness
Does it work?	++	++	+
How does it work?	+	++	+
Will it work elsewhere?	+	++	++

Thus, depending on the evaluative questions that are central in an evaluation assignment, the methodology design faces main challenges. The question of attribution is central to all rigorous evaluation assignments, with the measuring challenge and the generalisation issue as complementary challenge somewhat related to the kind of question that is central. This difference in emphasis helps to select the core methodology, and decisions to incorporate complementary methods by revising the validity threats to the expected type of conclusions to be derived from this core methodology.

### **FONDOECAS: A small-grants fund for collective marketing**

We illustrate this with an evaluation assignment on a dynamic change processes in fifty different value chains contexts in Bolivia. It couples a pretty ‘standard’ donor intervention with high diversity in contexts and high diversity in generative mechanism.

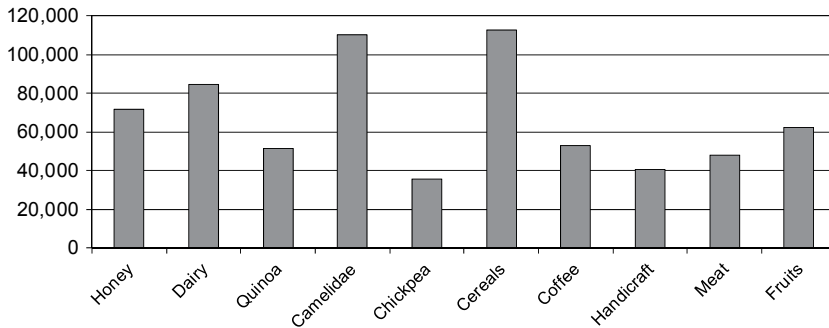
FONDOECAS is a semi-autonomous entity managing a small-grants programme that caters to economic smallholder associations. The small grant fund finances investments in processing equipment and related capacity building to organized groups that already for some years engage in collective marketing activities. The small-grants amount around US\$10,000 each. Between 2007 and 2010, FONDOECA reached around ninety organisations, with an estimated total membership of 18,000 families.

We start with a reflection on the logic model and the evaluation questions that became central to the assignment, and show that this translates in high demands to the methodology design, especially to find proper outcome indicators that enable cross-site comparisons, in a way that attribution with the support can be made plausible, and where the diversity in context-mechanisms-outcome configurations can be ‘captured’ in typologies that can serve as learning material on good practices.

### **Intervention theory**

The small-grants fund finances investments in processing equipment and related capacity building to organized groups that already engage in collective marketing activities. The fact

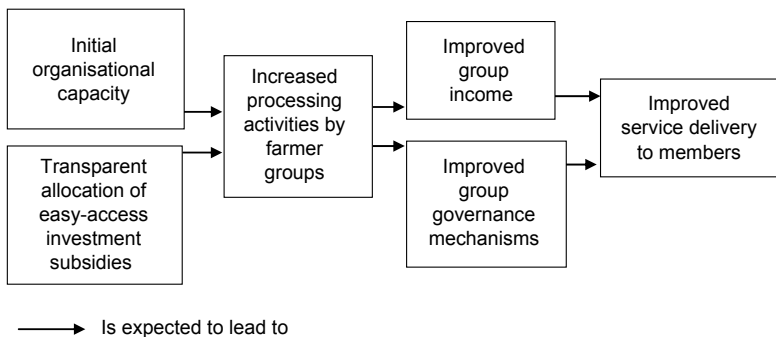
that the beneficiaries build on differing trajectories of organizational change, work in a wide range of economic sectors, and all propose to seize different niche-market business opportunities in highly dynamic and contested markets, poses huge challenges to the tools and methods proposed to identify and attribute impact. Figure 4.1 illustrates this diversity ‘in extremis’, typical for many value chain support interventions in the financial sphere.



Source: Authors' own elaboration

Figure 4.1 Distribution of FONDOECAS small-grants among different sectors (2007-2010, in US\$)

The two main requisites of formal existence as legal person and having a minimum of two years of collective marketing experiences, together with a lean and transparent fund management based on an impartial and knowledgeable analysis of the viability of the business plan submitted, are assumed to be sufficient guarantee that the small grant can really make a difference in the organizations. By generating or expanding the activities of processing of agricultural products both income and organizational capabilities of the organization are expected to be enhanced by experiential learning. And by doing so the service delivery to their members can be improved, translating in increased wellbeing. The impact logic of this intervention can, thus, be summarized in the following result chain, Figure 4.2:



Source: Authors' own elaboration

Figure 4.2 Impact logic of the FONDOECAS small-grants fund

## Key evaluation questions

Based on evidence of impact, the fund expects to attract new and major donors to FONDOECAS and/or to induce the government or donors to replicate or upscale this lean and transparent small grant allocation model for smallholder farmers' economic organisations. The core evaluation question arrived at, is "For what kind of organisations in what kind of contexts does a small-grant result in positive outcomes in the economic and organisational sphere?". Therefore, conclusions need to respond especially to the third question 'Will it work elsewhere?'. Attribution and generalisation are the key methodology challenges that need priority attention when deciding on the core methodology, next to the inevitable measurement challenge.

## Core methodology

To cover the major threats to validity related with the attribution issue, especially the issue of internal validity of the evaluative conclusion on impact, a quasi-experimental difference-in-difference design was chosen as the core methodology. Difference-in-difference implies a comparison between a baseline and a future situation for a group of fund beneficiaries (the treated) and a comparison group. The comparison group can be used for counterfactual reasoning: "What would have happened to the organisations without the small-grant?". As a second step, we explored the major threats to validity of the expected evaluative conclusion that would derive from our core methodology and measurement methods, and added methods as part of a mixed method design. Table 4.3 summarizes this exercise. The design embraces the notion of diversity in contexts and development trajectories, and collect information on a range of context variables, especially the mechanisms that are active in the baseline situation.

**Table 4.3** Threats to validities to the core methodological design

Type of validity threat	Main threat	Additional mixed methods	Result/observation
Statistical conclusion	selection bias between treatment and comparison group	Case-based statistics to maintain case integrity in group comparisons	Instead of measuring and comparing average impacts, we identify types of responses related with types of contexts and types of constellation of factors. These typologies are refined/validated in focus group discussions with key stakeholders.
Internal	attribution in complex systems	Process tracing based on case-descriptions of significant experiences in resolving agency dilemma's in collective action	Evidence on the ways that organizations change their organizational capabilities by collective marketing activities is collected, to test the 'intervention theory' that assumes that the agro-processing experience translates in learning and generates refined internal regulations and incentive structures.
Construct	measurement of organizational capabilities	Repetition of measurement with differing panel composition in the same organization	For each agency dilemma a description of the related incentive structure is made. A panel survey is applied to discover the agency dilemmas that are 'problematic' in the daily operation of the organisation.
External	Diversity in extremis	Structured case-studies, with due attention to incentive structures (internal organisational mechanisms) that limit opportunistic behaviour	By focusing on behavioural incentives for internal control instead of functional diversity in economic activities, common challenges of organizations are explored and solutions presented with a defined generalization domain (CMO-configurations)

## Measurement challenge

As an indicator of outcomes in the organisational capacities of the farmer organisations, we decided to focus on the organisational capacities to contain the disintegrative mechanisms in collective action, so-called agency dilemmas (Cornforth, 2004; Ostrom and Bloomington, 2009). We did so, as, in spite of all diversity, these agency dilemmas prove to be ‘inherent’ in most collective marketing experience, especially when they mature and scale-up in time (Bijman et al., 2011; Ton, 2008; Borgen, 2004; Cook, 1994; Olson, 1965). Therefore, we collect of time-series data on economic and organisational variables, and a before-after assessment of the capacities of the organisations through a panel scoring exercise. This resulted in a research tool to map the pertinence of common agency dilemmas, or disintegrative tendencies in collective marketing, summarized in Table 4.4, and descriptions of the way that the organisation has learned to cope with them. This learning is often codified and institutionalized in ‘better’ internal regulations and dispute settlement procedures. As farmer organizations are expected to adjust and improve their internal organization as a result of the new business development, current the baseline situation is compared with the situation after the new business plan has been put in practice.

**Table 4.4** Disintegrative tendencies in collective marketing (Ton, 2010b)

‘Regulating Member Supply’	Tensions can emerge when individual members increase their supply to the marketing organization, and, doing so, negatively affect the possibilities of other members to supply.
‘Quality Assurance Systems’	When a deal is made, the quality that the organisations has promised will have to be controlled for: individual members may tend to deposit lower quality and the organisations needs a system to maintain minimum quality requirements.
‘Coping with Working Capital Constraints’	Many smallholder farmers, tend to face cash constraints; and ask for fast payment, while the organisations need time to finish transactions with the ultimate buyer.
‘Anticipating Side-Selling’	The organisation might provide a credit service or advance payment system to enable production. However, there is a serious risk that farmers “side-sell” their product to competing traders or processors, to which they have no repayment obligation.
‘Ways to Dispose of Profits’	When the organisation makes profit, the organisation will tend to invest or increase capital reserves, while the member will have a tendency to prefer more short-term benefits, e.g. better prices.
‘Differentiating Services to Members and Non-Members’	Most economic organisations need contributions from members to realize their business opportunities. However, members face a number of disincentives to do so when benefits which flow from investment, accrue to investors and non-investors alike.
‘Decision Making on Activities that Benefit Only a Sub-group’	When the type of investment is not likely to benefit all members, investment decisions that seem economically optimal from the perspective of the management are not necessarily desirable from the standpoint of (sub-groups of) members.
‘Task Delegation and Supervision of Professional Staff’	Member-based organisations elect persons to supervise and support the management. However, the limited technical knowledge of board members and the lack of transparency of information disclosed by the management often limits the effectiveness of this governing structure.
‘Disclosure of Market Information’	Investments in market intelligence become an asset for the bearers of it, usually the sales persons. The group has to decide on partial or full disclosure of market information, motivating group investment in market intelligence and preventing defection of personnel.
‘Liability in Contracts and Loans’	There is an inherent tension between members that want to limit their liability for group actions and the need of the group as a whole to generate as much collateral as possible. Organisations use to specify procedures for decision making when the board is contracting on behalf of the group.
‘Managing Political Aspirations’	Economic smallholders’ organisations tend to take up a broader representative role next to their economic service provisioning to members. Members delegate their political voice to the organisation while the political representatives of the organisation may never fully discuss all political decisions with them.

## Attribution challenge

Each small-grant (around US\$10,000) is part of a configuration of factors leading to impact. A whole range of actors and factors are active in each organization, all contributing to the outcomes of a specific business strategy of the farmers' organisation and often complementing other co-investments, generally NGO's or local governments. The size of the grant and related business plans are sometimes of a size and scope that only fits a particular village or subgroup within the organization, while the major economic activities of the organisation remain relatively untouched (e.g. tourism as a complement to coffee exports; product development for indigenous handicraft; cheese making equipment for rejected deliveries to the dairy processor, etc.).

As Figure 4.2 shows, the funded processing activities are very much related to the characteristics of the commodity/product and its place in the members agricultural system (e.g. dairy as full time activity or honey as a complementary activity; coffee roasting or tourism development). This 'diversity in extremis' poses strong validity threats to claims of attribution and impact; different configurations of factors (size, networks, degree of professionalization, level of trust etc.) may lead to positive impacts in certain contexts/sectors/activities but may fail in others.

Therefore, the evaluation proposed a comparative case study design that could generate structured information to be analysed with 'configurational comparative methods' (Byrne and Ragin, 2009; Rihoux and Ragin, 2009), a wide term that covers a range of statistical procedures that look for causality/attribution between configurations of factors and outcome indicators in case-observations, instead of, or complementary to, the mainstream testing of differences on discrete variables through regression-based econometrics.

## Generalisation challenge

To maximize the generalisation domain, we selected a random sample of 30 organisations from the ones that had received a grant from FONDOECAS. Additionally, we selected randomly a comparison group from a list of organisations that could have applied but did not (yet), having similar characteristics, especially having a legal personality.

To enable the distillation of 'good practices' in collective marketing organisations, with a defined generalisation domain, we designed a simple comparative case study format to take stock of key organisational learning experiences on ways to resolve these agency dilemmas, presented in Table 4.5. These 'thick descriptions' complement the time-series data and panel scoring, and help to underpin the basic assumption in the intervention theory that the supported economic practices induce learning on the management of collective action problems.

**Table 4.5** Concise format for comparative case-study interviews

Context	Description of the problem that occurred Time and place when it occurred Involved stakeholders Contributing factors
Mechanisms	Pallet of options The main options discussed at that time envisaged to solve the problem Internal decision making process External influence on decision making process  Incentive structure introduced Internal arrangement chosen: what did they decide to do and how was it expected to be effective? What has been the way to formalize and communicate it to the members?
Outcomes	How has it worked out in practice? What effect did it have: organisational, economically, socially?
Recommendation domain	Do they recommend the solution to other organisations? If so, what would most probably be adjusted to the new context If not, what alternative solution do they suggest

## 4.5 Conclusions: Towards a theory-based mixed methods design

The increased attention of donors to standardised and rigorous impact assessments that can demonstrate impact of value chain support, builds momentum for the development of lean and effective tools and approaches. The existing lack of evidence does not necessarily reflect a low priority on measuring impact, but rather points to the lack of appropriate and credible instruments to do so, and to the complexity of social processes in value chain development processes (Vellema, 2011).

To define relevant and sharp questions that can shed light on replicability or scalability of value chain interventions, the need for ‘theory’ is paramount. (Donaldson et al., 2008; Rogers, 2009; Pawson, 2003). An impact evaluation needs to build on a logic model that indicates how the intervention is expected to influence the incentives for people’s behaviour. In the statistical analysis of data-set observation, this theory feeds the variables and matching process used, while in realist evaluation of causal-process observations, the theories are related to the workings of incentives provided and mechanisms triggered by the intervention. In impact evaluation, we need to make causal inferences about “what has worked for whom under what conditions”, and, concerning replicability, “what might work for whom under what conditions”.

A way to focus on the mechanisms that ‘trigger’ behaviour during or after an intervention is to use the realist concept of “Context-Mechanism-Outcome Configurations” in comparative case studies, analysing cases along the four aspects. To be useful prospectively, as a normative theory, these pilot case studies need to be written in a way that the contextual requirements for the intervention/mechanisms that triggers performance enhancing behavioural changes by chain actors are sufficiently explicit, and with a credible measurement of key outcome indicators. The case-studies will have to describe and unravel context-dependant processes and practices that generated the impact. In real world value chain dynamics, these case-studies

can suggest ‘good practices’ or promising ‘principles’ that may work in a comparable context with a similar configuration of conditions. They can be used as ‘food for thought’ in a learning process with stakeholders from other contexts. Information to conclude on comparability of the two configurations (the match between the case-study reality and the reality in the new intervention context) will always be incomplete, but the realist question ‘What works for whom under what conditions, is helpful to generate that information. Besides methods that make theories explicit, properly designed data collection tools and qualitative research techniques are needed that quantify or describe the outcomes and impacts of value chain interventions and that can be used to test the key assumption inherent in the impact models. Multiple methods are needed to support claims that something does work, and provide information that is useful to explore the real causal processes and compare them with the normative impact models.

The concept of threats to validity, developed by Shadish et al (2002) proved useful to check the design of the mix of methods. An appropriate mix of methods will have to respond to the phrasing of the evaluative question and the ‘kind of conclusion’ that the stakeholders inducing the evaluation are expecting. Starting from the core methods of collecting information (e.g. a survey; case studies; focus groups), the review of the threats to validity in its four dimensions (statistical conclusion validity, internal validity, construct validity and external validity) will indicate the need for additional research methods to underpin conclusions, and will lead to a set of mixed methods to collect evidence/information/observations.

Impact evaluation demands serious efforts from organisations to invest in critical reasoning while designing interventions, presenting an initial ‘intervention theory’ or ‘impact logic’, that can be tested and improved through monitoring and evaluation activities. Using a realist method to describe and analyse intervention pilots as comparative case-studies facilitates the exchange of experiences between development agencies with evidence-based research. Its restricted and defined generalisation domain may prevent uncritical embracement of good practices. For example, specific types of contract farming, branding, fair trade labelling prove to be viable and effective in a wide range of situations but are not the panacea, the standard solution, for creating market access; they all involve specific sets of institutional arrangements that create specific incentives to the stakeholders involved.

## Acknowledgements

*This paper is a result of the research programme Value Chains for Pro-poor Development (VC4PD) implemented under the Partnership Programme ‘Globalisation and Sustainable Rural Development’ between the Netherlands’ Ministry of Foreign Affairs and Wageningen UR and of the Knowledge Base Funds of the Netherlands Ministry of Economic Affairs, Agriculture and Innovation (project KB-03-008-004 in the programme Chains and Logistics). The authors would like to thank the anonymous referees of Journal of Chain and Network Science for their comments to improve this paper.*



## 5 | The Mixing of Methods: A three-step process for improving rigour in impact evaluations<sup>4</sup>

Giel Ton

Agricultural Economics Research Institute, LEI Wageningen UR

### Abstract

*This paper describes a systematic process that we have found helpful in improving impact evaluation assignments, within restricted budgets and timelines. It involves a rethink of the key questions of the evaluation and a way of designing a mix of research methods to generate evidence that supports more valid conclusions. The approach is illustrated through two examples: one on measuring income impacts in an irrigated horticulture programme in Nepal, Zambia and Ethiopia; and another on the assessment of changes in organisational capacities for collective marketing by smallholders in Bolivia. The simple, straightforward and structured three-step process developed has helped us to reduce the tendency to one-method designs. Enhanced critical reflection within the team allowed for greater sensitivity to validity threats and the creativity to find ways for handling them.*

---

4 Published as: Ton G. (2012) The Mixing of Methods: a three-step process for improving rigour in impact evaluations. *Evaluation* 18: 5-25.



## 5.1 Introduction

This paper reflects an evolving experience within LEI (Agricultural Economics Research Institute, part of Wageningen University and Research Centre) working out a step-wise process to add rigour to applied research assignments. Our work is specialised in change processes in agricultural value chains. We have a portfolio of research assignments from clients seeking impact evaluations with a strong methodological design, often including an assessment of the counterfactual situation: ‘what would have happened without our intervention’. Funding for impact evaluation has increased because, partly as a result of stronger public pressure on aid money to prove its worth, donor countries increasingly require convincing evidence of impact on poverty alleviation (OECD, 2001, 2008; DCED, 2010). Being able to assess ‘net effects’ of support interventions has increasingly become an important quality attribute of methods underlying impact evaluation (Leeuw and Vaessen, 2009, Khandker et al., 2009; Maredia, 2009; McKenzie, 2010). Generating convincing evidence on the link between development ‘outcomes’ and donor-supported intervention ‘inputs’ is not easy (Ton et al., 2011). Most change processes that are the objective of value chain development projects are at the very least ‘complicated’ (Rogers, 2008), influenced by a myriad of other factors, projects and policies; these all constrain any claim of attribution of impact to specific interventions. Furthermore, interventions in agricultural value chain development projects tend to be shaped and reshaped continuously to adapt to changing conditions, thus limiting the possibility to assess their effectiveness over time. For instance, support to ‘Round Tables’ for the sustainable production and sourcing of commodities such as cocoa, oil palm or shrimps, involves many different initiatives to respond to dynamics within and between the stakeholders involved (i.e., companies, unions, NGOs and farmers). On the other hand, support to an apparently more simple intervention, such as contract farming in African vegetable production is not so straightforward either; it must relate to the blurry concept of ‘contract farming’ that tends to be affected during the project period by recurrent changes in contract regulations, associated credit support, prices, etc. (Vellema, 2002, Ton and van der Mheen, 2009). It is very difficult to imagine any value chain support coming close to interventions such as the extension of medicines or seeds, where the ‘treatment’ is clearly specified.

Working in an applied research institute on contracted research assignments for a variety of public and private clients, our group has needed to find creative ways to cope with the inherent complexity and diversity of value chain development processes. Our clients hire us explicitly to reduce the complexity for them and to generate insights that help them draw lessons from past experiences and/or to develop clear scenarios for future developments. Luckily, our group includes a wide range of disciplinary backgrounds (economists, anthropologists, sociologists and agronomists) that generates a well-provisioned research methods toolbox. Nevertheless, we are often tempted to in practice use one-method research designs, partly because they are often easier to communicate to clients during the acquisition phase. Also, budgetary and time-constraints related to evaluation assignments generally do not allow for comprehensive evaluation designs (Bamberger et al., 2006). While we all believe in the strengths of combining different methods in one research design, we experience problems in implementing them in our assignments. Pragmatic and creative thinking are therefore necessary. The peer-to-peer process that we describe in this paper was developed in order to reduce the tendency toward one-method research designs, and to strengthen the overall methodolog-

ical rigour within the practical constraints. Through a structured process of internal reflection, we stimulate creative thinking to ‘stretch’ research designs, to include additional design elements and to come up with an appropriate mix of quantitative and/or qualitative methods.

This paper outlines how the design of evaluation frameworks, within constrained conditions, benefits from a critical review process that improves rigour. It describes a systematic process involving a rethink of the key questions of the evaluation assignment; and a way to design a mix of research methods that generates evidence to support more valid conclusions. The rest of this paper is structured as follows: Section 2 describes the theoretical rationale behind mixed-methods research designs; followed by the presentation of a three-step process in Section 3, that we have found useful for improving our evaluation practice. Section 4 illustrates how this process has been applied in two very different impact evaluation assignments: the first sought data on income impact in a micro-irrigation programme in Nepal, Zambia and Ethiopia to explore conditions for up-scaling and replication; and the second, to monitor impacts of a small grants fund in Bolivia on the organisational strength of farmers’ organisations. The paper concludes with a short section summarising findings.

## 5.2 Approach

Organisations build their monitoring and evaluation processes to meet a variety of objectives, such as good management practices, learning and reporting to donors. We acknowledge the crucial importance of utilisation-focused, participatory monitoring and evaluation processes to support reflexive learning (Guijt and Woodhill, 2002; Patton, 2008; Kusters, 2011). Nevertheless, this section focuses on the challenges of designing a strong research methodology for collecting evidence of impacts. Although this forms a modest part of a comprehensive monitoring and evaluation system, it is usually the part where we, as contracted applied researchers in LEI, start interacting with the client.

As explained above, rigour and relevance are important aspects of our impact-related research designs. The definition of ‘rigour’ is contested in different strands of the evaluation literature and involves different communities of practice that have difficulty relating to each other. Some are especially concerned about the methodology to resolve the issue of the ‘counterfactual’ (Khandker et al., 2009; White, 2010). Others stress the need to look for insight into contextual configurations rather than general conclusions on and overall averages of impact (Pawson and Tilley, 1997; Stame, 2004). Yet others stress the need to produce relevant outputs for stakeholders and policy makers to strengthen their capacity to learn and adapt to new, emerging situations (Morell, 2010; Guijt et al., 2011). Happily, many of the participants in this discussion admit that there can be synergy between different approaches to impact evaluation; this is especially the case when the complexity and dynamism of the change processes are duly recognised (Rogers, 2009) and specific key evaluation questions are formulated.

The debate in the evaluation community on preferred methods (such as those documented in Donaldson *et al.* (2008)) echoes the long-standing tradition in the social sciences to contrast quantitative and qualitative research paradigms, and related tools and methods to underpin claims of causality and attribution. Our researchers are of course also influenced by this

debate, and we therefore needed a process to break down the barriers between those using mostly quantitative or mostly qualitative research methods. In particular, reflecting on ‘what works for whom under what conditions’ instead of measuring and attributing impact of an intervention as a black box with the sole question of ‘what works’, tends to favour the process of mixing methods and becoming more creative with statistics (Ravallion, 2009; Byrne and Ragin, 2009). The evaluation researchers in the realist tradition (Bhaskar, 1978; Sayer, 2000,) focus on identifying configurations of causal mechanisms that are ‘fired’ under particular conditions to produce certain types of outcome patterns (Pawson and Tilley, 1997). This emphasis on contextual configurations gives room for synergy between otherwise quite distant research traditions. Paying specific attention to the diversity of outcomes due to different contextual conditions implies that a mechanism or intervention can never ‘always’ and ‘universally’ work; and any ‘average’ measure of impact needs to be defended by a clear specification of the context in which these apply.

Brady et al. (2006) provided another useful way to depolarise research positions as they reframe the false dichotomy between ‘quantitative’ and ‘qualitative’ research traditions; they distinguish instead between Causal-Process Observations (CPOs) and Data-Set Observations (DSOs), both of which can be collected with quantitative or qualitative methods. They argue that to make high-validity causal inferences, both types of observations are needed in a process of ‘nested inference’ or ‘triangulation’ (Brady and Collier, 2004). There is no necessary sequence nor hierarchy in either type of observations when used to claim causality and attribution. The key to this process is to involve different research methods that study the same mechanisms, processes and outcome patterns. ‘Triangulation’ thus focuses on the same research question from different perspectives, thereby anticipating alternative explanations and improving the validity of the evaluation conclusion. The arguments behind method triangulation and the combination of CPOs and DSOs for causal inference make a clear plea for mixed-method design.

Shadish et al. (2002) also call for a combination of different research ‘design elements’ in challenging the tendency toward simple one-method designs. The evidence base to support evaluation conclusions must be critically appraised, and other analytical procedures, methods and perspectives need to be brought together in the research design to at least anticipate the most salient threats to validity. In the next section, we present a step-wise process that includes a threats-to-validity ‘check’ to the methodologies chosen to answer the key questions.

## 5.3 A step-wise process to improve designs

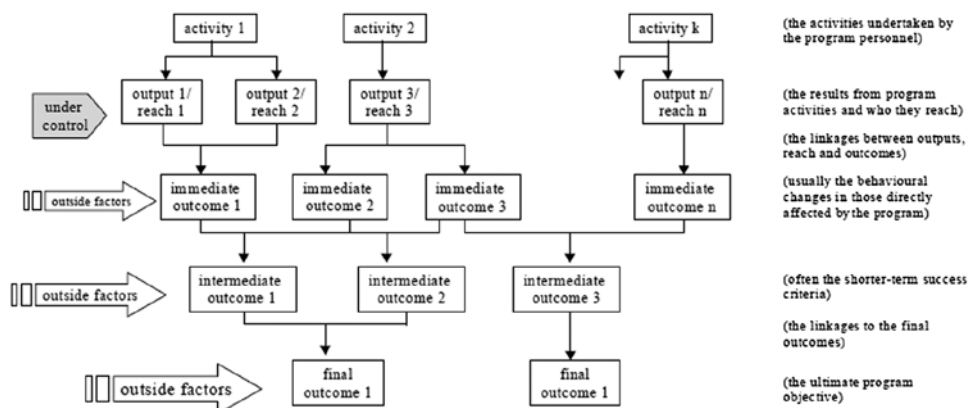
### Step 1: Refine the evaluation questions based on the intervention logic

Most of our group’s evaluation assignments start with quite comprehensive terms of reference, requesting an evaluation of the relevance and effectiveness of an intervention; the emphasis is on ‘does it work?’ types of questions that measure outcomes –often Millennium Development Goal-oriented. These then allow for the inclusion of each researcher’s favourite core methodology to measure/register impacts - be it qualitative or quantitative. This core methodology tends to be based on one of the methods described in impact evaluation handbooks (Campbell

et al., 1963; Shadish et al., 2002; Khandker et al., 2009). The result is a rather method-driven approach to research that tends to generate evaluation statements that are rather general in nature, without much attention to the dynamics involved in reaching these outcomes (Pawson and Tilley, 1997; Stame, 2004; Heckman and Smith, 1995). Later interventions in value chain development are for example often considered to be ‘immune’ to ‘yes-it-works / no-it-doesn’t-work’-types of conclusions from impact evaluations in past projects or interventions; this is so because they are often considered to be ‘better designed’ or having ‘better enabling conditions’, exactly the issues where the ‘did it work’ impact evaluation question does not generate much information.

This has motivated our group to come up with more relevant, specific questions together with our clients. Our first step in this process is therefore to refine the scope of the evaluation to only those aspects that really matter to the persons involved in the value chain development interventions (including those who decide on the particular assignment). For this, we use a process of logic modelling together with the stakeholders involved in the assignment to improve the quality and usefulness of the evaluation questions. In Theory-Based-Evaluation (TBE), this logic can be broken down into interchangeable concepts, as with a ‘logic chart’ (Mayne, 2001), ‘result chains’ (DCED, 2010), ‘programme theories’ (Chen, 1994; Rogers, 2009) or ‘theories of change’ (Weiss, 1997). This exercise explores the causal chain that is expected to result in impact; that is, to make a difference to the ultimate outcomes. Once undertaking this exercise, the assignments tend to become broader in scope than only measuring outcomes, with attention to other (‘minor’?) evaluation questions with (‘major’?) learning potential.

Our preferred intervention logic set-up is found in Figure 5.1, as visualised by Mayne (2001); Mayne refined the traditional outputs-outcomes-impact logic to pay greater attention to the specification of outputs and expected outcomes in the different stakeholder groups involved, and, especially the disaggregation of outcomes into three categories: immediate, intermedi-



Source: Mayne (2001), p. 9

Figure 5.1 Logical framework to capture an intervention logic (Logic Chart)

ate and ultimate/final outcomes. In practice, the different outcome categories are used in a flexible way: what is considered an immediate outcome of an activity can also be considered an intermediate output or final outcome of others. The tandem output/reach and their link to immediate outcomes of activities helps us find key indicators of progress that are closer to the control span of the project/intervention, and more informative for learning and results-based management; they then become performance indicators and not just indicators of impact (Armytage, 2011).

This first step leads to a list of ‘areas’ where research on causality and attribution may help to reflect on progress and/or generate information on key assumptions. This long list is then reviewed in an interactive process to select only a subset of key questions and related key outcome indicators on which to focus the research. Each key question may demand a specific research approach and, in our experience, many questions tend to be more process-based (e.g. ‘how can we do it?’) rather than impact-related (e.g. ‘how did we do it?’). In addition, it is possible for the project staff to answer many of these key questions themselves through their internal monitoring and evaluation system. This process therefore narrows the questions to very few ‘difficult but important’ impact evaluation questions on which we, as external researchers, can focus.

## **Step 2: Anticipate the validity threats to the expected type of conclusions**

The second step is to critically examine the (mix of) methods used to generate evidence to answer each evaluation question. As explained above, we wish to avoid one-method research and design a mix of methods that generates ‘valid’ conclusions, within our constrained room for manoeuvre. Bamberger, Rao and Woolcock (2010) rightfully argue that ‘rigour is not determined solely by the use of a particular method as such, but rather the appropriateness of the “fit” between the nature of the problem being assessed and the particular methods deployed in response to it’. It is generally accepted that a combination of quantitative and qualitative methods in a ‘nested’ design brings synergy (as demonstrated by Garbarino and Holland (2009) and Bamberger et al. (2010)). Using transparent methods for case studies on impact (e.g. random selection, purposive sampling, contrasting cases, etc.) increases the generalisability of case study findings and helps us understand patterns and logic in survey data. However, mixed-method design is not only about combining qualitative and quantitative methods in one design. There is a whole range of qualitative and quantitative tools that have different strengths and weaknesses in generating evidence to underpin conclusions. Moreover, qualitative methods can result in quantitative data, while qualitative data can be used in statistical or econometric analysis. Mixed-methods design is more generally about adding methods and design elements (whether they be qualitative or quantitative, or descriptive as opposed to analytical in nature) into a coherent package to achieve a more valid ‘fit’ with the evaluation question .

The concept of threats to validity, as developed by Cook and Campbell in ‘Experimental and Quasi-Experimental Designs for Generalised Causal Inference’, and refined and extended in the latest edition by Shadish (Shadish et al., 2002) proved very useful to get closer to this ‘fit’. These authors emphasise that validity applies to the conclusions, and is not intrinsic to any one method; in the end, it depends on how boldly you phrase the conclusions. The evidence

collected and analysed with the research methods has to be strong enough to back up the types of conclusions expected from the assignment. A helpful tool that has largely emerged from this book, a ‘threats to validity checklist’, was developed to help pinpoint major weaknesses in research methods (Bamberger et al., 2006; Bamberger and Rugh, 2008). Applied to the initial core method(s) of collecting information (e.g. a household survey, case studies, focus groups, (quasi-)experimental design, participatory observations, etc.), a review of the threats to validity highlights the need for additional research methods and/or design elements to underpin the conclusions, leading to a broader set of methods to collect data (evidence, information or observations). Checking validity threats associated with the evolving mix of methods is an iterative process. The final design will include procedures that (partially) control some of the limitations of the research methods used. Shadish et al. (2002) distinguish between four types of validity that have to be convincingly addressed in the design of evaluation research:

1. **Statistical conclusion validity:** how are inferences about correlations made in data set observations? This emphasises the need to comply with proven methods to estimate association or correlation between variables.
2. **Internal validity:** how is causality attributed in the evaluation? This type refers to the logic behind the observed correlations and explains why and how interventions contribute to the observed change.
3. **Construct validity:** how are generalisations made from the categories used in the evaluation to broader units of representation? This stresses the importance of precise definitions and concepts.
4. **External validity:** how are the findings generalisable to other persons, times and contexts? This requires being precise about conditions and requirements that define the generalisation domain.

By using these four types of validity in a workshop setting, researchers with different methodological traditions can come together to generate pro-active and creative thinking on appropriate methods in addition (or complementary) to the initial methodology that had been chosen.

There are other concepts of validity that can also be useful for checking research designs. In a recent paper, Chen (2010) refines the original ‘Campbellian’ distinction between internal and external validity by adding a third dimension that he calls ‘**Viable validity**’. He intended to open up research designs to take on a bottom-up perspective for checking the practical feasibility of implementation of interventions. In this dimension, stakeholder consultations are a key feature of the methods included in the research design to cope with threats to validity. More importantly, viable validity also draws attention to the viability of and risks associated with particular research designs (i.e., implementation failures) when put into practice. For example, control groups sometimes end up as (partially) treated groups; ‘random’ assignment of an intervention may become ‘almost random’; or measurements ‘before’ and ‘after’ may suffer from bias as a result of changing enumerators. The awareness of these types of viability risks in the research design tends to lead to the addition of other design elements that create

fall-back options and make the mix of methods more robust in the face of unexpected changes in the research conditions.

In addition, Bamberger and Rugh (2008) suggest to add three more dimensions on which threats to validity can be assessed. These come from the constructivist tradition of social research:

1. **Objectivity:** are the conclusions drawn from the available evidence, and is the research relatively free of researcher bias?
2. **Reliability:** is the process of the study consistent, coherent and reasonably stable over time and across researchers and methods? If emergent designs are used are the processes through which the design evolves clearly documented?
3. **Utilisation:** how useful are the findings to clients, researchers and the communities studied?

These additional dimensions relate to the process and outputs of the evaluation in particular. An important validity threat relates to 'objectivity' and implies that research results have to be considered as a subjective 'argument' from a specific research perspective - rather than as an objective 'truth'. We consider that, when properly used, the three-step process is in itself an operationalisation of the 'reliability dimension', and we consider that the aspects of 'objectivity' and 'utilisation' are also inherent to it. Although these additional types of validity are useful for checking evaluation designs and results, we have chosen to limit ourselves to the four dimensions of validity developed by Shadish et al. (2002).

### Step 3: Maximise the scope for comparative research

The third and final step in our approach is to fit each evaluation assignment into the wider research agenda on agricultural value chain development, to generate a body of evidence for a wider group of stakeholders regarding the effectiveness of interventions. To make evaluation outputs more useful, we try to include comparative case studies in our evaluation assignments. In our experience, learnings from case studies on different interventions are still quite limited, due to the overwhelming site-specific details. We expect that organising and analysing case studies in a structured way helps to expand the learning potential from case studies (of experiences).

We propose to use the concepts developed in the realist school of social research (Sayer, 1992) more, in order to develop 'headings' relating to contextual factors that make intended mechanisms and processes of change work (or not work) in generating outcomes. Almost all value chain interventions are intended to change attitudes and behaviour in persons or institutions. Realist evaluators (Pawson and Tilley, 1997; Stame, 2004; Pawson, 2006) propose to test key assumptions about the mechanisms that stimulate these behavioural changes through the 'Context-Mechanism-Outcome Configurations (CMOCs)' concept. The CMOC lens allows for cases to explain why and how interventions worked, under what conditions, in a wider

configuration of causal factors (Easton, 2000). In real world value chain dynamics, these case studies can suggest ‘good practices’ or promise ‘principles’ that may work in a comparable context with a similar configuration of conditions. They can also be used as ‘food for thought’ in a learning process with stakeholders from other contexts. An additional advantage of comparative case studies is the potential to use them at a later stage, to draw lessons out of a large range of similar descriptions of change processes that feed into the quest for more evidence-based policy making.

## 5.4 Application

This section shows how the three-step process described in the previous section was applied to two research assignments. The first example is a programme that promotes high value horticultural production in Africa and Asia through the facilitation of micro-irrigation. The causal link between the intervention of ‘more technology’ and the outcome of ‘increased household income’ was quite straightforward for this project. In the first phase, the team developed a comprehensive evaluation framework with learning questions and mixed methods to collect evidence on outcomes. In the second phase, the team was assigned to look at one of the project’s key evaluation questions, the tracking of income effects. In close coordination with our client, country project staff and several external reviewers, we decided for a core methodology of cohort studies and added design elements that anticipated the most salient validity threats.

The second experience took place within a multifaceted and complex change process that posed bigger challenges for the operationalisation of indicators of impact. This case is a small grants fund set up in Bolivia to co-finance business plans submitted by collective marketing groups to increase value-added activities. These activities were intended to be instrumental in strengthening organisational capacity and generating economic benefits to the membership. The research team collected essentially qualitative information on organisational dynamics before and after the business plan was put into practice, to then be recoded and classified for use in case-based methods of configurational comparative analysis (Rihoux and Ragin, 2009; Byrne and Ragin, 2009). Design elements were included that increased the usefulness of the impact research for stakeholder learning. Both of these experiences illustrate the potential of the three-step process to improve rigour and relevance in the research design of evaluation assignments.



## Example 1: Impact evaluation of micro-irrigation technology supply

### **Background**

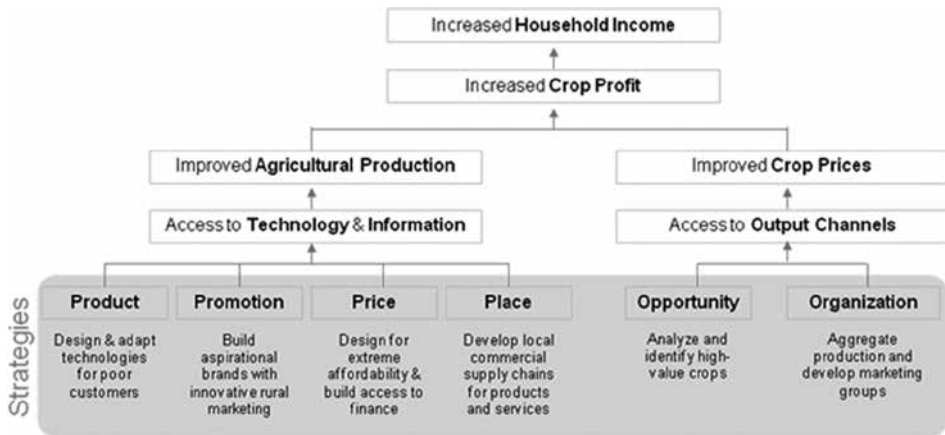
The Rural Prosperity Initiative (RPI) of our client, International Development Enterprises (IDE) based in Denver, USA, had the explicit goal of raising household incomes through micro-irrigation induced horticultural value chain development. The programme was funded by the Bill and Melinda Gates Foundation and the Dutch Development Cooperation, and was implemented in Nepal, Vietnam, Cambodia, Ethiopia, Zambia and Zimbabwe. IDE works on two interrelated issues: supply chains of low-cost micro-irrigation technology, and establishment of value chains ending in urban markets (wholesalers, supermarkets, etc.). Generally, low-cost and water efficient micro-irrigation technologies (e.g. plastic drip kits, human-powered treadle pumps,) are designed to meet the specific local conditions; manufacturers and dealers are supported to distribute the technology, and farmer households are trained in intensive horticulture to generate cash income from the additional area of irrigated fields. The donors asked for a rigorous impact evaluation of the RPI to explore conditions necessary for upscaling and replicating the micro-irrigation technology supply model in other countries. The piloting of the impact evaluation methodology took place in three countries: Nepal, Zambia and Ethiopia. The evaluation assignment started in 2006 and finished in 2010.

### **Mixed-methods design**

#### *Step 1: FOCUS - Refine the evaluation questions*

Our team started to interact with IDE in a kick-off workshop in December 2006. The workshop was attended by all the people who had been or would be involved in the programme's M&E activities. IDE staff had already developed a logic model of their approach (IDE, 2005); the workshop was meant to deepen the understanding of the impact evaluation requirements elaborated in the contract with the main funder, the Gates Foundation, and to link these to the model. The core of IDE's intervention theory is summarised in Figure 5.2

The first step in our workshop process involved a closer look at the project's intervention logic in order to refine the evaluation questions. Several causal links (represented in Figure 5.2 by arrows) in this intervention logic can be challenged. For example, the assumption that high-value crop systems based on low-cost micro-irrigation devices (drip-kits, treadle pumps) generate more income than the traditional crops, and thus increase the profitability of the agricultural/horticultural system requires further evidence. The assumption that access to low-cost irrigation equipment improves agricultural production (yields) is less contested, but a measure of impact is important for accountability and communication purposes. The assumption that marketing groups create better access to output markets with higher prices is also less contested. However, important questions related to this are whether output markets with better prices are a necessary condition for reaching crop profitability; or whether higher agricultural productivity alone is sufficient. This last question calls for an assessment of the impact of different kinds of support packages on farmer household profits.



Source: IDE (2011), personal communication

Figure 5.2 IDE-PRI Intervention theory

Our reflections on the logic and assumptions of the programme led IDE to develop a comprehensive M&E framework including several quantitative and qualitative methods and metrics. A critical assumption in the result chain regards attribution/contribution of the changes in the agricultural system to IDE's own efforts. Based on our analysis, we proposed to generate multiple streams of evidence to make attribution or contribution plausible, and place IDE support in the context of other development agencies working with the same households on horticultural development. For example, in Nepal, the Village Development Committees (VDC) work in close cooperation with IDE in supporting irrigation development and, exceptionally, provide grants to lower-caste groups to make the technology affordable or give access to horticultural seeds. But other projects also work on income-generating activities with or through the VDC with the same households. In Zambia, most intervention areas are likewise supported by the establishment of marketing groups that source for specific supermarkets in the cities. IDE is never the sole actor that triggers a change in the livelihood of the 'treated' households.

Our team assisted IDE in developing a toolkit with multiple methods to gather data on the outcomes that could help attribute change to their efforts (IDE, 2007). It triangulates data from surveys with in-depth livelihood case studies, and monitors the activities of partners or others working in the area (e.g. credit, buyers); it further complemented the survey-based research design with context information to support claims of attribution/contribution. The toolbox of data collection methods is summarised in Box 4.1. Seven distinct tools were selected, each with specific challenges in its design and implementation. In addition, to trigger a learning process around key issues from a series of national and international learning meetings, story harvesting was suggested as the main tool to engage field staff in discussions about 'good practices'.

#### BOX 4.1 IDE'S MIXED-METHOD MONITORING AND EVALUATION FRAMEWORK

1. **Rolling Baseline Households Surveys** (200 households annually per country) to monitor the impact of IDE interventions on the net income of participants;
2. **Farming System Case Studies** (25-50 households per country during the project period) , purposively sampled, to specific crops and farming practices to learn about the effectiveness and sustainability of IDE's products/services/technologies;
3. **Livelihood Impact Case Studies** (10-25 households per country during project period), in a random sub-sample of the rolling baseline survey, to assess changes in detail in intra-household relationships, household expenditures and wellbeing;
4. **Qualitative Studies** (5 per country during project period) to assess the enabling environment and support structures around value chain development to identify causal links and attribution of impact. Studies are part of or complementary to sub-sector and value chain analyses;
5. **Natural Resource Impact Assessments** (2-5 per country during the project period) to assess the impact of the major technologies and interventions;
6. **Focus Group Discussions with Smallholders** (4-10 per country per year) to understand primary stakeholder perspectives, changes in farming systems and perceptions of strengths and weaknesses. Focus groups studied specific types of interventions and technologies in each country;
7. **Customer Satisfaction Surveys** (annually in each country for marketed technologies) to receive feedback on the use and appropriateness of technologies and services.

Source: IDE (2007), pp. 53-63

In this first phase of our assignment, this framework was further refined by our team, in close coordination with IDE Headquarters and the respective country coordinators, in terms of survey questions and impact indicators to track progress over time. For several of these methods (focus group discussions, livelihood impact case studies, story harvesting), separate field manuals were developed in 2007-2008 and a training-for-trainers took place in Kathmandu in June 2008.

In a second phase, most of the activities in the M&E Framework were assigned to each IDE Country Office. A small team at IDE's Denver Headquarters worked on learning and knowledge management, and used information from the M&E Framework for their regular performance tracking and reporting to corporate management and donors. The credible measurement of income, the 'rolling baseline household surveys' (see first point in Box 4.1), was outsourced to our institute (i.e., LEI). The key question and design challenge that became the focus of our assignment was mainly related to measuring household income impact and collecting information on household characteristics through surveys. IDE contracted LEI to develop a software application to process survey results, with special attention to resolve the 'counterfactual', to be able to incorporate data collection into standard project management procedures in all countries and projects where it is active in value chain development. Additionally, the data from the surveys in the countries fed into a web-based portal, where

impact information was made accessible to external stakeholders and used for comparative GIS-based analysis.

*Step 2: DESIGN – Anticipate validity threats to the conclusions*

IDE wished to focus resources specifically on the measurement of impact through a design that met the standards for credible impact evaluation, such as those developed in the Donor Committee on Enterprise Development (DCED, 2010), in which IDE was a key participant. The choice for the core methodology was made during a workshop with the IDE country coordinators at the end of 2006. IDE had negative experiences with large one-off surveys that took a lot of effort and provided little learning. The organisation initially opted for a difference-in-difference approach, where treatment groups and control groups are compared before and after the project period. However, during this workshop several issues were raised that made us question the value of this approach. A major issue was the expectation that the intervention areas and customer groups might undergo important changes during the implementation period, which would result in baseline information that could prove to be no longer relevant when the programme unfolded. This did in fact happen in two of the three pilot countries: Zambia and Ethiopia. Another issue was that only a tiny fraction of the rural population in each project area was expected to make up the client base, making community-wide impact measurement averages unreliable, unless very large surveys were to be used. Instead of community-wide averages, client group-specific averages were considered to be a more cost-effective way to get an indication of impact. Clients have to pay for the technology and are therefore per definition self-selecting and not necessarily representative of the population of farmers in the area.

Therefore, to increase cost-effectiveness of the income monitoring system, we proposed a design that would not rely on extensive surveys to get community-wide averages, but opted for a pipeline design to track changes in customer households cohorts only; this provided 'case-based' estimates of impact in each of the client households included in the sample. Through yearly surveys, households were asked about their productive activities in both the current year (having adopted a micro-irrigation technology) and the previous year (prior to adoption). These data compared the starting position of cohort  $t+1$  with the income of the one-year client cohort  $t$ , which were expected to generate a correct estimate of the counterfactual situation. As mentioned before, it was clear to all in the team who designed the research that income changes over the two-year period for each household could not be directly attributed to the IDE-promoted technology adoption. Other exogenous variables –including weather, prices, inflation and other economic circumstances– may crucially influence household income in a year, invalidating the registered income change as a sole measure of impact. Therefore, the impact would instead have to be derived from the comparison of the successive cohort of clients.

The author, the IDE specialists and external reviewers (especially someone from the Institute of Development Studies in Sussex) discussed in a series of teleconferences the core design to identify important threats to validity to the expected estimates of attributable net income impact. Our team explored the survey instruments regarding the types of validity to think about other weaknesses and we translated these into more robust data-collection and analytical procedures. Table 5.1 summarises the main results of this iterative process. Most of the additions

were extra design elements and controls within the essentially quantitative research method of the household survey. In comparing consecutive cohorts, we introduced a form of matching of treatment and control farmers to limit selection bias and to reduce key differences in background characteristics. Conclusions were phrased cautiously, as a result of the awareness of the validity threats, and the reports explicitly referred to the need for complementary context and process information from the more qualitative methods used by the IDE Country teams (e.g. sector studies, focus group discussions, monitoring of other projects' activities in the area, etc.).

**Table 5.1** Summary of additional methods to counteract threats to validity

Type of validity threat	Main threat	Additional mixed methods	Result/Observation
Statistical conclusion	Selection bias	Tests on significant differences between client cohorts and a group of average households	In several intervention areas differences between cohorts proved highly influential. Data from these areas cannot be used for counterfactual calculation, except when a matching procedure is applied.
Internal	Attribution in complex systems	Qualitative case studies and sector studies	IDE applied several qualitative methods to triangulate data from the core surveys, and to monitor the activities of partners or others working in the area (e.g. credit, buyers). In presenting conclusions from our surveys, we emphasised that the impact relates to the changes in the farming system, not to IDE activities directly.
Construct	Estimation of income based on recall	Repetition of measurement in the same households with differing recall periods	Recall proved to be an issue in one of the countries (the resulting impact figures seem to be inflated), but not in the other two. The sources of recall bias are being explored.
	Definition of household income	Design software that generates a range of different income categories	Income from animal husbandry was contested; in particular, the values for yearly herd size changes are in some countries considered part of income and in others part of assets.
External	Agro-climatic specificity	Disaggregation of averages	The research was carried out in three countries with different agro-ecological zones in each of them. Sample sizes were chosen in a way that disaggregated averages of impact could be based on sufficient observations.

Another important validity threat related to the composition of the construct 'household income'. This income measure was derived by analysing flows in and out of the different components of the agricultural system - such as animal husbandry, dryland agriculture and irrigated horticulture, and the importance of non-agricultural income (e.g. from migration labour). Initially, the target increase was defined as the rise in total family income, converted to US\$ (IDE, 2006); at a later stage, this was complemented by their equivalents in 'Purchase Power Parity'(PPP\$). However, when analysing the data of the first survey, family income very much proved to be influenced by changes in off-farm employment (e.g. migration) and by the changing stock of animals held in the households. For example, changes in herds and stocks of small husbandry like chickens or goats, were considered part of income by the Nepal and Zambia teams, while in Ethiopia, they were considered to be part of the asset base that functions as a risk-insurance strategy and must not be included in household income. Although

tracking family income was preferable to monitoring poverty levels, and to trace eventual changes in the use of labour and resources between horticulture and other activities, it proved necessary to include another income category that was more directly related to the intervention. We therefore started to report not on one, but on several different income categories (e.g. 'family income', 'agricultural income' and 'target crop income') and to distinguish between gross income (including home-consumption) and cash income. By means of the database application MonQI (<http://www.monqi.org>), alternative 'constructs' for income estimates could be generated quite easily from the data based on input and output flows in the smallholder farming household. IDE had different uses for these different constructs and decided upon 'agricultural household cash income' as its preferred construct for 'income impact', because attribution to the irrigation and horticultural value chain development activities was much more direct.

Another assumption in the proposed core design proved to suffer highly from validity threats: a potential recall bias. Both observations in each household (before/after) are based on recalling activities realised, but with important time differences: the 'before situation' recalls information from between 12 and 24 months ago, while the 'after' situation was based on at most a 12-month recall period. The difference in recall accuracy between both moments could inflate or deflate the resulting income impact calculations. Based on this validity check, we included an additional test: we repeated the before-after measurements in a follow-up survey of a sub-sample of households, and compared answers from the two years: two assessments of income/production in the same year with varying recall periods.

### *Step 3: COMPARE - Maximise learning*

Most of the implementation of the qualitative research carried out for the evaluation system lay outside our assignment. The team helped IDE to generate training material for project staff on focus group discussions and focussed story harvesting. However, implementation and refinement was out of our span of control, as well as access to the information generated by them; our institute's assignment became focussed on the survey measurement. However, our data offered potential for comparative learning because the method generated before/after estimates for each of the respondents. The data proved useful for making rough estimates, even when no counterfactual could be calculated yet. They offered insights into the poverty status of clients, described their key characteristics and made it possible to differentiate between areas, types of farmers and types of intervention 'packages' (e.g. drip-kit irrigation, treadle pumps, new technology combined with crop training or combined with marketing support, etc.). Target groups could be adjusted based on, *inter alia*, data on the poverty status of clients in the different intervention pockets. A very important potential learning inherent to the design was that contradictory findings on income changes in (groups of) clients could be traced down to the source, which stimulated discussions and reflection by field staff on the logic of these patterns of change. Understanding the reasons behind income changes in relation to certain characteristics of the household generates additional knowledge on geographical, climatic and economic influences shaping the context of the intervention.

## Example 2: Impact evaluation of small grants for collective processing

### **Background**

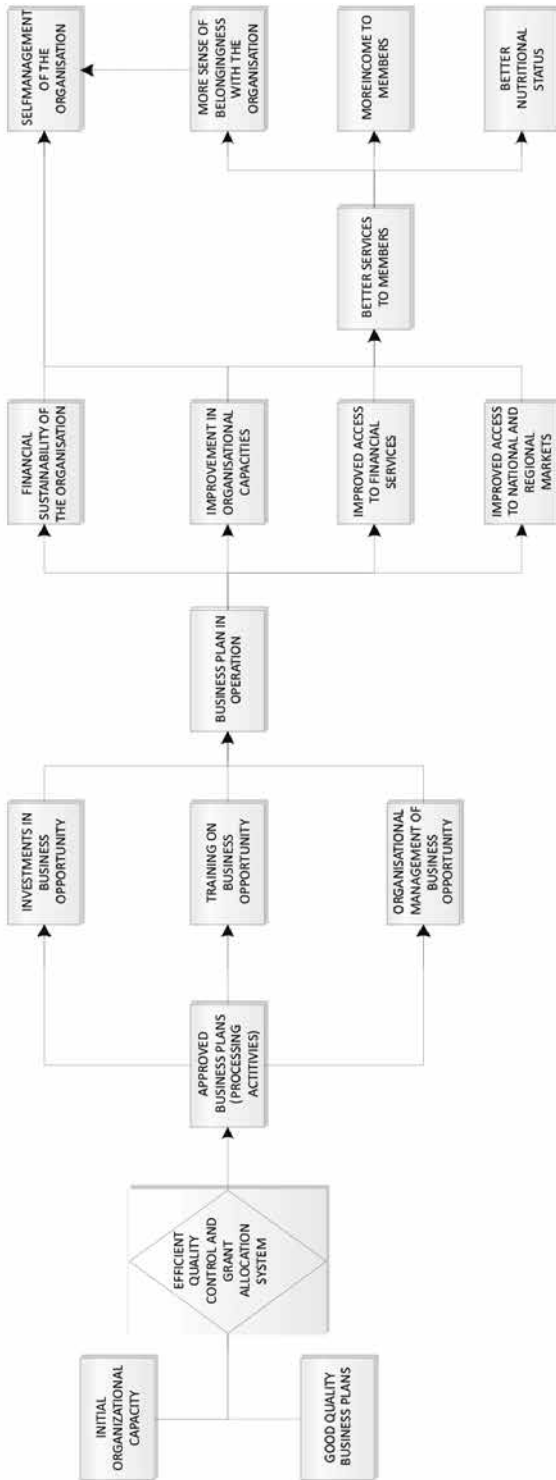
FONDOECAS is a semi-autonomous entity managing a small grants programme in Bolivia that caters to smallholder economic associations. The small grants fund goes toward investments in processing equipment and related capacity building for organised farmers' groups. These organisations need to already have had several years of experience in collective marketing activities. The grants amount to around US\$10,000 for each group. Group size differs widely - from 20 members to several thousand. Between 2007 and 2010, FONDOECA reached around ninety organisations, with an estimated total membership of 18,000 families, all members of the national platform CIOEC-Bolivia. The total number of similar organisations in Bolivia is estimated to be 778 (CIOEC-Bolivia, 2009), with one-third of them being a member of CIOEC-Bolivia.

FONDOECAS received funds from a group of European development organisations. A major funder, ICCO, the Dutch interchurch organisation for development cooperation, was in particular interested in novel ways to track impact in complicated or complex change processes that would comply to the 'emergent' standards of quality and rigour discussed in the Dutch evaluation community. The author already had in-depth knowledge about CIOEC-Bolivia, for whom he worked as policy analyst between 1999 and 2004. ICCO and LEI agreed to carry out an impact evaluation of FONDOECAS to develop tools and methods to capture outcome patterns in value chain development; of particular interest was strengthening of social capital in the chain (Ton et al., 2011) as well as the learning potential for the stakeholders involved. The research started in 2010 with a baseline survey and is expected to be completed in 2014.

### **Mixed-methods design**

#### *Step 1: FOCUS – Refine the evaluation questions*

We started by reflecting with the FONDOECAS team on the logic model behind the intervention, and the evaluation questions that should be central to the assignment. The farmers' group beneficiaries all build on distinct trajectories of organisational change, work in a wide range of economic sectors, and all propose to seize different niche-market business opportunities in highly dynamic and contested markets; this complexity poses huge challenges to the tools and methods proposed to identify and attribute impact. This diversity *in extremis* is typical for many value chain support interventions in the financial sphere. The two main requirements of beneficiaries were formal existence as a legal entity and a minimum of two years of collective marketing experiences, together with a lean and transparent fund management based on an impartial and knowledgeable analysis of the viability of the business plan submitted. These are assumed to provide a sufficient guarantee that a small grant can really make a difference in the organisations. By generating or expanding the activities of processing of agricultural products, both income and organisational capabilities are expected to be enhanced by experiential learning. Increased organisational performance makes it possible to improve service delivery to members, which eventually translates into increased wellbeing. The essence of the impact logic of this intervention is summarised in the concise result chain shown in Figure 5.3.



Source: Result of discussions with fund managers, May 2011

Figure 5-3 Intervention logic of the FONDOECAS small grants fund



The fund managers expect the research to generate evidence that may attract major new donors to FONDOECAS and/or to induce the government or donors to replicate or upscale this lean and transparent small grant allocation model for smallholder farmers' economic organisations. The core evaluation question in the assignment therefore became: For what kinds of organisations, and in what kinds of contexts does a small grant result in positive outcomes in the economic and organisational sphere? Attribution and generalisation are the key challenges that needed to be resolved by the core methodology, next to the inevitable measurement challenge.

A key assumption in the intervention logic of FONDOECAS was also prioritised. FONDOECAS managed the crucial assumption that the FONDOECAS business proposal evaluation team, the Technical Committee – comprising a group of six experienced professionals from micro-finance institutions, consultancy firms and NGOs – come up with information that reflects the 'real' quality differences of the business proposals on the ground, not just on paper. For the grant system as an intervention to be scalable, it must ensure that its positive or negative decisions about the business plans submitted do not merely reflect the abilities of the person who has written the plan on behalf of the smallholder organisation (often without computer literate staff or members!) in the required format.

*Step 2: DESIGN – Anticipating validity threats to the expected conclusions*

The design challenges particularly related to measuring impacts in organisational strength and needed outcome indicators that would enable cross-site comparisons. To cover the major threats to validity related to the attribution issue, a difference-in-difference design was chosen as the core methodology. Difference-in-difference is a quasi-experimental design that implies a comparison between a baseline and a future situation for a group of fund beneficiaries (i.e., the 'treated') and a comparison group. A sample of 30 beneficiary organisations and 20 'could-be' beneficiaries were selected randomly, from the list of beneficiaries or the national platform membership list, respectively, to facilitate extrapolation of findings with representative group averages.

As an indicator of outcomes in the organisational capacities of the farmers' organisations, the evaluators decided to focus on their capacity to contain the disintegrative tendencies of collective action. In spite of the great diversity among these organisations, these kinds of agency dilemmas (i.e., tensions between the individual members' interest and the collective interest of the group) prove to be 'inherent' to most collective marketing experiences, especially when they mature and scale-up (Ton, 2010). For example, dairy producers starting to make cheese seem very different from handicraft producers. Nevertheless, when designing systems to make their group members comply with quality requirements in delivering their produce and in establishing pre-financing mechanisms that do not pose too much of a financial burden to the group, these two groups have much to learn from each other. Rules and routines to control these kinds of agency dilemmas are the backbone of every farmers' economic organisation. The agility and detail of these rules and their effective application distinguishes the 'weak' from the 'strong' organisations.

We decided, therefore, to gather time series data on economic and organisational variables and to use these in a before-after assessment of the capacities of the organisations. This re-

sulted in a research tool to map the pertinence of common agency dilemmas, or disintegrative tendencies in collective marketing. The research (carried out by local researchers) produced descriptions of the way the organisations coped with these dilemmas in the last three to five years, and how this learning became codified and institutionalised in internal regulations and dispute settlement procedures. The farmers' organisations are expected to adjust and improve their internal organisation as a result of new business experiences. The baseline dynamics will be compared with the organisational dynamics in the three years after the new business plan has been put into practice.

**Table 5.2** Summary of additional methods to counteract threats to validity

Type of validity threat	Main threat	Additional mixed methods	Result/Observation
Statistical conclusion	Selection bias between treatment and comparison group	Case-based statistics to maintain case integrity in group comparisons	Instead of measuring and comparing averages of impact, we identify types of responses related to types of contexts and types of constellations of factors. These typologies are refined/validated in focus group discussions with key stakeholders.
Internal	Attribution in complex systems	Process tracing based on significant experiences in resolving agency dilemmas in collective action	Evidence of ways that organisations change their organisational capabilities by collective marketing activities is gathered, with thick descriptions of key moments to do so. The evidence underpins claims that experience with value-added activities translates into learning and refined internal regulations and incentive structures.
Construct	Measurement of organisational capabilities	Repetition of measurement of the self-assessment procedure with differing panel composition in the same organisation	The self-assessment procedure for qualifying the strength of farmers' organisations is cross-checked before assuming that it can be used as a monitoring device.
External	Diversity <i>in extremis</i>	Structured case studies, with due attention to incentive structures (mechanisms) that limit opportunistic behaviour	By focusing on behavioural incentives for internal control, instead of functional diversity in economic activities, common challenges of organisations are explored and solutions presented with a defined generalisation domain

As a next step, the evaluation design team explored major threats to the validity of the expected evaluation conclusion derived from this core methodology. This demanded critical scrutiny of the measurement tools and the triangulation of results with panel scoring and focus group discussions. The validity check resulted in the addition of methods as part of a mixed method design. Table 5.2 summarises this exercise. The core design already embraces the notion of diversity in contexts and development trajectories; it also collects information on a range of context variables, especially the mechanisms that are active in the baseline situation. Nevertheless, the description at the baseline level is necessarily exploratory. Additional and systematic mapping of context variables will have to be done after the analysis of all 50 baseline reports, to enable case-based analyses (e.g. cluster analysis, qualitative comparative analysis, etc.). A major threat to validity related to the proposed comparison of the measures of organisational strength: even when ordinal measures could be implemented (based on systematic data extraction from the qualitative baseline reports, cross-checked with panel

interviews), the scoring will be sensitive to differences between local researchers who prepare the report and the composition of the panel. To check the robustness of the measurement tool in a sub-set of the sample, interviews are being cross-checked between researchers and sometimes the scoring tool will be applied to similar panels from within the same organisation.

### *Step 3: COMPARE - Maximise learning*

The clients of the impact evaluation were positive about the methodological innovation in the pilot study but stressed that emerging results must be modelled to fit internal learning on organisational development and business strategies. An extra tool added to the mixed-methods design is focus group discussions that reflect on the ‘good practices’ reported from the comparative case studies. These focus groups consist of farmer leaders enrolled in the existing training programmes of the two national platform organisations that steer FONDOECAS. Over ten ‘illustrative events’, where internal organisational mechanisms have been defined or refined in response to an urgent situation, will be re-worked into short texts to be discussed in these courses. Material from the study will also appear in a searchable database on the website of the ESFIM (Empowering Smallholder Farmers in Markets) network, as part of a wider process of story harvesting on incentive structures in collective marketing.

## **5.5 Concluding remarks**

Impact evaluation assignments suffer from a range of constraints that ask for focused, lean and flexible research designs to collect evidence to support evaluation conclusions. Time and budget constraints are the most obvious reasons to decide for single-method or ‘quick-and-dirty’ research instead of more rigorous ways of evaluating impacts. Rigorous designs are necessary especially when there are high expectations of the stakeholders for an evidence base that supports conclusions on causality/attribution and an evaluation of effectiveness. However, this is not necessarily the case for all key questions in an evaluation. Many evaluation questions that were prioritised after reflecting on the result chain relate more to qualitative lessons learned in the implementation of the intervention and the possibilities for replication. Not all these questions need comprehensive research designs. Strong and comprehensive mixed-methods designs are necessary and budgetary feasible for only a small sub-set of specific research questions and for key indicators of progress of the intervention strategy under evaluation.

Over time, we have developed an internal peer-to-peer process to distil these key questions, and to improve the rigour of the methods used for tackling them, respecting budgetary and time constraints (our ‘room for manoeuvre’). As shown in this paper, the process consists of three fairly straightforward steps.

The first step is to refine the evaluation questions to the point that resources can be concentrated on the questions and assumptions that are most relevant for the stakeholders involved. Stakeholder discussions about the implicit and explicit logic behind their interventions proved to be very useful for narrowing the assignments down to key indicators of progress and key assumptions in the intervention logic. Our experiences lead us to suggest that it is good

to focus resources on top priority learning questions for ‘engaged insiders’, and on the critical points where ‘cynical outsiders’ may attack the claim of attribution of an intervention.

The second step consists of a process of critically challenging the chosen methodology for providing evidence to answer each of the questions. The use of the four dimensions of validity threats have proven useful to spur creative thinking on the possible addition or combination of methods or internal checks on methods. This resulted in additional data collection tools or the use of additional statistical tests. The result is not necessarily a mix of qualitative and quantitative methods (though often it will be) but may also result in the mixing of different qualitative methods (e.g. cross-checking interview findings in a focus group discussion; adding case studies to more general panel interviews) or the mixing of quantitative methods (e.g. check for recall-bias with a complementary survey; disaggregate units of analysis in statistical analysis).

The third step is the use of the data in a format that maximises stakeholder learning. The realist approach to unravelling mechanisms in social systems that have made an intervention work is promising for structuring comparative case studies and for documenting ‘illustrative events’. Comparative reporting with due attention to context helps to distil good practices while taking care not to make over-generalisations.

We experienced that this simple, straightforward and structured process helped us to open methodological discussions between researchers within different research traditions. The sensitivity to certain types of validity threats and the creativity to find ways for handling them was enhanced by the critical reflection in the team.

## Acknowledgements

*The author would like to thank the commissioners of the impact evaluation assignments and his colleagues at LEI Wageningen UR, who allowed this process of methodological refinement. A disclaimer is necessary, as the design process is a voluntary endeavour within our institute and is an abstraction from our ‘bumpy’ daily practice. The arguments in this paper, therefore, remain the sole responsibility of the author. The author acknowledges the support of the Knowledge Base Programme of the Dutch Ministry of Economic Affairs, Agriculture and Innovation (KB-11-004-003).*





## 6 | Tension Containment Capacity: Towards a measure of organisational social capital in economic farmer organisations

Giel Ton

Agricultural Economics Research Institute, LEI Wageningen UR

### Abstract

*This paper documents the field testing of a data-collection tool to assess farmer organisations involved in collective processing and marketing. It maps the relevance of common agency dilemmas, areas in which the interests of group and individual diverge, and the extent to which the group has found internal rules to contain these inherent tensions. We develop a construct to measure an organisation's tension containment capacity (TCC). We identify five core areas that proved most relevant for economic farmer organisations in Bolivia: quality-assurance systems, payment systems, rules that anticipate side-selling, delegation of commercial tasks, and the management of party-political aspirations. Using panel data from two rounds of measurement in 2011 and 2013, we show that the construct is suited for cross-sectional analysis of collective marketing groups.*

## 6.1 Introduction

Farmer organisations exist in various legal formats, such as informal community groups, civil associations, cooperatives, or farm-led enterprises. They also serve a diversity of functions, be they in the social, cultural, political or economic sphere. We focus on farmer organisations that collectively market their products. This type of market-oriented farmer organisations are considered to be key agents in agricultural innovation (Bebbington, 1997; Ashby et al., 2000; Bijman et al., 2011; Bernard et al., 2010). Donors such as the World Bank (World Bank, 2007; Bosc et al., 2001; IFAD-UNEP, 2013) and a large number of development NGOs propose to increase the development support to strengthen their capacities (Penrose-Buckley, 2007; Flores et al., 2007; Lundy et al., 2002; Blokland, 1992; Helmsing, 2003), or to provide an enabling environment for their development and strengthening (Markelova et al., 2009; Lyon, 2003; Paumgarten et al., 2012). The World Bank Report 2008 warns that methods of empowering farmer organisations need further experimentation and solid impact analysis to become more effective (World Bank, 2007: 157). Evidence on the effectiveness of strengthening approaches is still scarce (Shiferaw et al., 2011; Ton et al., 2015). This is partly because of methodological limitations to measure key outcomes, especially to capture the impact on organisational capacities.

We propose a new measure for the quality of social relations in the group to advance collective interest, as a proxy-indicator for ‘organisational social capital’ (Leana and Van Buren, 1999). We argue that the essence of organisational strength in collective marketing lies in the containment of inherent contradictions between group and member in ‘agency dilemmas’ with effective rules and regulations (Shapiro, 2005; Ostrom and Ahn, 2009). Some comparative measures have been applied, e.g. in community forestry (Poteete and Ostrom, 2004), non-timber forest product extraction (Donovan et al., 2008) and coffee growing (Ortiz-Marcos et al., 2011). However, the indicators and constructs used in these studies tend to be quite sector-specific. Many institutions and development programmes work with organisations in different sectors and need to compare organisational strength across a more heterogeneous group of organisations. With a common indicator for organisational strength, it could become better possible to compare alternative approaches to farmer-group strengthening, such as Fair Trade certification, preferential government procurement, innovation grant funds, or business training.

Our measure - Tension Containment Capacity (TCC) - fills a void. The measure is derived from semi-structured interviews about the presence and effectiveness of rules and regulations in the group in order to contain the inherent tensions in the group, for example as a result of opportunistic behaviour and free-riding (Hellin et al., 2009; Lyon, 2003). In these interviews, ten areas are reviewed in which agency dilemmas tend to become relevant in groups that collectively market their products. Out of these ten, we distilled five core agency dilemmas: quality-assurance systems, payment systems, ways of discouraging side-selling, commercial task delegation and management of political aspirations of leaders. We used these to derive a quantitative measure of organisational social capital. We field-tested this measure in a panel of 38 economic farmer organisations in Bolivia in 2011 and 2013. We show that the measure had face validity when used in cross-sectional research, and was correlated with economic performance indicators. The field test pointed to some changes in the tool that would increase the reliability of the quantitative measure for longitudinal research.

The paper is organised as follows. First, we specify the domain of the construct, and describe the specific challenges of farmer organisations in managing collective marketing activities. Second, we explore the literature for methods used to assess these capacities. Third, we present our data-collection tool and identify the core agency dilemmas for deriving a quantitative comparative measure of organisational strength. Fourth, we check the reliability of this measure for cross-sectional and longitudinal research, and use alternative ways to compute the TCC-score in order to make the measure more robust. In the final section, we discuss these results and sketch the relevance of the tool for impact evaluation.

## 6.2 Farmer organisations in collective marketing

Smallholder farmers need institutional arrangements that facilitate their access to markets. Because smallholder farmers are scattered, they need to aggregate (“bulk”) their produce in order to transport it cost-efficiently to urban or regional markets, or to the processing industry. The form of coordination varies, from contract-farming and trader-agent-networks to collective marketing by farmer organisations. The essence of these various forms is their effectiveness in performing key logistic functions with acceptable financial and managerial costs. We focus on the modality of collective marketing. Through collective marketing, farmers expect lower marketing costs and higher on-farm prices than in the traditional spot market or agent-trader network.

Collective marketing is never easy: structures for coordinated action, in whatever form, result in tensions due to the threat of opportunistic behaviour by individual farmers towards the group, and/or opportunistic behaviour by the group towards the individual members (Hellin et al., 2009; Lyon, 2003). Any group needs rules, trust and discipline to contain these agency dilemmas. The essence of the agency dilemmas in farmer groups is that the members want the organisation to do some things for them (e.g., sell their produce for a good price), and need some assurance that the organisation will do this well. At the same time, the organisation wants the members to do something (e.g., provide good-quality products) and needs to prevent disloyal behaviour on the part of members. In these situations, a workable middle road has to be found to make the deal acceptable to both the members and the organisation: the farmers have to trust the organisation to do a good job, while the organisation has to prevent disloyal behaviour by members. To contain the agency dilemmas in the group, a successful governance structure is needed that gives both the member and the organisation enough confidence to accept the collective marketing ‘deal’.

Farmer organisations deal with multiple agency dilemmas (Shapiro, 2005). A group needs to balance these inherent tensions and management ‘paradoxes’ (Cornforth, 2004). Some agency dilemmas may not be problematic at a certain point in time but become relevant in response to changes in the socio-economic environment. The organisation must be able to cope with these disturbances and show resilience: it should have the capacity to experience shocks while retaining essentially the same function, structure, feedbacks, and therefore identity (Holling, 1973). This capacity to adjust internal regulations is the main manifestation of organisational social capital. The development of effective rules and regulations is a learning processes that takes time and resources (Ostrom and Ahn, 2009).



*“Self-governing systems in many areas of social interaction tend to be more efficient and stable not because of any magical effects of grassroots participation itself but because of the social capital in the form of effective working rules (...) Simply agreeing on an initial set of rules (...) is rarely enough. Working out exactly what these rules mean in practice takes time. (...) Part of learning through experience is what happens when things go wrong” (Ostrom and Ahn, 2009, p.29-30)*

Agency dilemmas in collective marketing are the central focus in the field of cooperative studies, but scholars mostly focus on larger and professionalised cooperatives (Cook et al., 2008; Royer, 1999; Soboh et al., 2009; Bijman et al., 2012). Most empirical research in which smaller farmer groups are compared on the basis of their organisational social capital relate to common pool resource extraction, especially in forest management (Rustagi et al., 2010; Ostrom, 1990), and not to collective marketing activities. To fill this void, we propose a measure to quantify the organisational capacities of small groups active in collective marketing activities across different sectors, which can be used in impact evaluation and comparative research. Such a measure of organisational social capital may help to make organisational strengthening a more prominent indicator of development support, in addition to economic performance indicators such as group sales, profits or farmer income. Economic effects often need more time to mature and are influenced by many other actors and factors than only the support intervention. Organisational strengthening is likely to be a more direct effect of these support programmes, or a key moderating factor for farm-level impact (Elder et al., 2012; Verhofstadt and Maertens, 2014; Wollni and Fischer, 2015). An indicator that can measure this (intermediate) outcome may help to reflect on the effectiveness of the support, and compare alternative modalities of support (Ton et al., 2014c; Meinzen-Dick et al., 2004).

### 6.3 Comparing organisational strength

Economic farmer organisations face specific challenges that are difficult to resolve through traditional forms of governance. For example, Bernard et al. (2010) showed how market-oriented groups emerged within villages in Burkina Faso. At the start, these market-oriented groups were more similar in their objectives to the village-oriented organisations, whereas later they had more room to deviate from this pre-existing organisational structure in order to prioritise their economic and commercial objectives. With less influence from village-oriented organisations, the economic performance of the economic groups improved. They explain the negative influence of traditional village authorities on the performance of economic farmer groups as being a result of their fear of economic differentiation. Emerging market-oriented groups are perceived as a threat to the reproduction of the traditional social structure and the solidarity system. This mirrors the observation of Woolcock (1998), that the homogeneity and ‘closure’ characterizing traditional village structures may at some point stifle the members’ personal and business development. Serra (2011) also stresses this ‘downside’ of traditional systems of local governance and dispute resolution, and the need for market-oriented farmer groups to develop a different, group-specific governance system. She points to the importance of subgroup-specific rules and highlights the influence of outside factors that may constrain the effective implementation of agreements in the group.

Agency dilemmas and related decision-making processes have received much attention in cooperative studies. For example, Cechin et al. (2013) point to the importance of mechanisms that create loyalty and commitment in the membership to support managerial decisions on a group's economic strategies. This need for mechanisms and procedures to convince the members of fair decision-making by the group is a reflection of cooperatives' relatively high "social capital dependency" (Valentinov, 2004). The literature stresses the distinctive feature of cooperatives compared with conventional firms: members are both owners and investors. In developing countries the 'ownership' by members is often less clear, because financial investment by members in group assets tends to be low or absent, and start-up capital is often provided by the government or development cooperation. There are only a few studies that tried to measure and compare the capacity of smaller groups to handle these agency dilemmas.

To measure the organisational strength of smaller groups and relate this to development support, many authors follow Uphoff (2000), who makes an analytical distinction between structural and cognitive elements in social capital. Structural elements are roles, rules, procedures, and precedents as well as networks that facilitate collective action, whereas the cognitive elements are the norms, values, attitudes and beliefs that create and reinforce collective action. For example, Barham and Chitemi (2009) show for Tanzania that market performance was significantly correlated with social capital. For example, the maturity of the group, its organisational age and history of activities proved important for economic success. "*A set of rules must be followed in order to run successful group activities (.). Unlike new groups, mature groups had a set of institutions to guide group behavior.*" (Barham and Chitemi, 2009: 58) This points to the organisational intelligence embodied in the group's governance system. Barham and Chitemi did not develop a refined measure to assess this organisational social capital. They only differentiated whether a group already existed or whether it was newly formed for the purposes of the intervention. In this way, they bypassed differences in organisational strength within these new and old organisations.

To measure differences in organisational social capital, Bernard et al. (2010) use two governance indicators, 'Leadership/Participation' and 'Rules'. The first factor maps whether a president is taking decisions alone or whether decision-making is channelled through organisational structures, such as the committee (board) or general assembly; the second factor maps the functioning of the committee and controlling body. These governance arrangements only cover a narrow part of the group's organisational social capital to carry out economic activities. Likewise, in India, Adhikari and Goldey (2010) explored the difference in governance capacity. These researchers showed that the capacity to sanction appeared to be important for the performance of groups. To quantify this quality, they asked in a household survey whether rules were breached. Their scope was limited; they reduced it in their questionnaire to the issue of transparent fund management.

Lyon (2003; 2000) had a more comprehensive overview of the subject. In an analysis of Ghanaian farmer groups, he points to the importance of trust between the members and the price negotiator in collective marketing. Especially in bigger groups, this trust depends on rules that enforce transparency and facilitate dispute-settling strategies. He indicates that the process of group building may require self-selection of members and flexible rules set by these members. Johnson et al. (2002) provide more examples of relevant rules and regulations such

as the importance of complying with quality requirements in markets, adjusting payment systems to members, reducing side-selling, etc. They note that trust is an essential component of social capital and that history matters. They tried to quantify the 'supply of social capital' with the proxy-indicator 'number of groups to which the owner/manager belongs in his or her personal life'. For farmer organisations having multiple board members, however, this proxy-indicator seems inappropriate to compare levels of organisational social capital over time.

Donovan et al. (2008), in a study on forest-related rural community enterprises, compare economic farmer organisations on their levels of asset building, consolidation of management, internal structure, positioning in markets, and impacts on members and community. They developed a framework (Donovan and Stoian, 2012) to assess the impact of value-chain development on changes in the five capitals that form their asset endowments – human, social, physical, financial and natural capital. Among others, it covers the knowledge and skills in business administration to meet the service needs of affiliated producers (human capital) and the fostering of linkages between farmers, buyers and service providers (social capital). The tool evaluates the internal management procedures, management skills and business development plans, member satisfaction with payment systems and credit transactions, participation in decision making, and effective quality control. The rich qualitative information provided in the case studies uncovered several agency dilemmas in collective marketing, such as side-selling, payment systems and quality control, and ways that organisations coped with them. The results of the tool were, however, somewhat limited when used in cross-sectional analysis. They summarised the status of each organisation on human and social capital in three overall values: sufficient, somewhat insufficient or insufficient.

Similarly, to assess the impacts of Fair Trade, Ortiz-Marcos et al. (2011) adapted a management tool used by companies in developed countries (EFQM, 2010). Their tool systematically assessed the presence or absence of internal processes and administrative procedures. They applied the tool on second-tier coffee cooperatives, where it proved a useful diagnostic device. Nevertheless, the emphasis on administrative procedures and professional management routines makes it less appropriate for capturing (changes in) capabilities in smaller, less professionalised farmer organisations. Recent tools, such as ScopeInsight (2013), also tend to focus on the existence of management procedures, human resource management and financial transparency issues, which are less relevant for most smaller farmer groups.

Other scholars have tried to assess the organisational strength of a group by the individual attributes of its leader. For example, based on data from behavioural economic field-experiments in Ethiopia with small forest management groups, Kosfeld and Rustagi (2012) showed that the punishment behaviour of leaders mattered for the performance of the groups. Especially those leaders who punish members based on motives other than equality and effectiveness, e.g., because of ethnic (clan) reasons, cause groups to perform worse than average. Their research took place in a context of fairly homogenous organisational forms, with small groups and personalised authority. In membership organisations, where members elect a board, several persons work together to enforce rules, which makes it more difficult to relate performance to personal behavioural traits. The punishment capacity of an organisation will result from a more complex interaction between multiple members of the board and, sometimes, contracted staff.

## 6.4 Tension Containment Capacity (TCC)

### Ten agency dilemmas

The above overview of the literature underlines the importance of internal institutions in the sanction/prevention of deviant behaviour in collective marketing. Moreover, the overview points to the diversity of indicators used to measure organisational social capital, and the absence of common measures to assess organisational social capital in cross-sectional and longitudinal research. We propose to develop a comparative measure for organisational strength in collective marketing that would focus on agency dilemmas and the way that a group develops effective ways to resolve inherent tensions between the interest of the individual and the group. From the literature as well as in discussions with farmer organisations, we identified ten agency dilemmas that we consider inherent to most farmer groups engaging in collective marketing activities in developing countries (Ton, 2010b). We consider organisations that have rules and regulations to resolve many of these agency dilemmas as having more organisational social capital. In Table 6.1 we present these ten agency dilemmas and the starting questions used during the interviews, when reflecting with the interviewed board members about their rules and regulations to cope with these agency dilemmas. We describe the agency dilemmas briefly:

- T1 ‘Regulating member supply’ - When output markets are constrained, tensions can emerge when individual members would like to increase their supply to the marketing organisation, and, by doing so, negatively affect the opportunity for other members to supply. For example, in many dairy cooperatives the farmers are given delivery rights based on their membership and can trade these delivery rights. Handicraft associations also tend to have a system in place to distribute production among members.
- T2 ‘Quality assurance systems’ - Increasingly, modern value chains have strict rules and mechanisms for quality control (Vorley et al., 2007; Poulton et al., 2010), for which member interests must be balanced with the demands of downstream buyers, especially the retail sector and the processing industry. When a deal is made, the promised/contracted quality has to be checked: individual members may tend to deposit lower quality and the organisation needs a system to maintain minimum quality requirements.
- T3 ‘Ways to reduce the need for working capital’ - Many smallholder farmers tend to face cash constraints and want quick payment, while the organisation needs time to complete transactions with the buyer. The group needs working capital to resolve this tension. The legal form, patrimony and scale of a group will influence their access to bank loans (Von Pischke and Rouse, 2004). Internal agreements, such as delayed payment systems, are often used to limit the need for trade capital.
- T4 ‘Prevention of disloyal behaviour in sales’ - The organisation might provide a credit service or advance payment system to enable production. However, there is a risk that farmers “side-sell” their product to competing buyers to whom they have no repayment obligation (Barrett et al., 2012).

- T5 'Define ways to distribute profits'- When the organisation makes profit, the board will prefer to invest or increase their financial reserves, whereas the members tend to prefer more short-term benefits, e.g., better prices. In many cooperatives, the final price paid to the farmer is defined at the end of the year, creating a direct trade-off between short term gains for members and accumulation of financial buffers for the group.
- T6 'Differentiating benefits and services to members and non-members' - Most economic organisations need contributions from members to realise their business opportunities. However, members face a number of disincentives to do so, when benefits accrue to members and non-members alike. This problem is also relevant when new members have the same rights as older members (Staatz, 1983), and between active and passive members.
- T7 'Decision making on investments that do not benefit all' - Subgroups of members (e.g. with specific type of crops) may have different objectives than the group in general and a need to negotiate a compromise with the other members. Investment decisions that may seem economically optimal from the perspective of the group, are not necessarily agreed upon in the group's decision-making process (Staatz, 1983). This leads to a portfolio problem, which is a tendency to invest only in activities that provide benefits to all members, not to strategic subgroups and sectors. Also, a member can opt out of the cooperative and therefore prefer lower but short-term benefits above higher mid-term benefits.
- T8 'Delegating and supervising marketing tasks'- Organisations may have board members or professional staff that negotiate prices for them. They need rules to do be assured that these people are doing the job well, while giving them sufficient room for effective commercial decision making (Lyon, 2003). Decision rights on economic transactions need to be balanced between members, board and management staff (Cook, 1994; Hendrikse, 2005; Henehan and Anderson, 1994).
- T9 'Assuming liability in contracts and loans - The board contracts on behalf of the group, which creates liability towards the contracting partners. The board wants to transfer this risk to the members, while the members want to limit their exposure. Access to conventional (bank) credit may be constrained when the liability of boards and members in case of default are ill-defined. For example, in Uruguay board members of cooperatives need to guarantee group loans with personal assets, which creates risk aversion in decision making on group investments (Samson, 2010).
- T10 'Managing political aspirations of board and staff' - Members delegate their political voice to board members, who may use their position in the group to pursue individual political interest not aligned with the group interest. Especially in rural areas, where human capital is scarce – e.g., limited availability of literate leaders –, board members may participate in various community organisations, interest groups and political movements simultaneously, and create the need to define rules that contain possible tensions between members and their elected representatives .

**Table 6.1** Ten agency dilemmas in collective marketing

	Agency dilemma	Starting question in the interview
$T_1$	Regulating member supply	Did some members complain when the group decided not to buy all of their product?
$T_2$	Quality assurance systems	Did some members try to deliver products that were below the required quality?
$T_3$	Ways to reduce the need for working capital	Did members press for cash payment or did they have to wait until the organisation finished selling the product?
$T_4$	Prevention of disloyal behaviour	Did some members sell part of their product to other buyers, although they had promised to sell to the organisation?
$T_5$	Define ways to distribute profits	Did members accept that the organisation did not distribute all surplus/profit?
$T_6$	Differentiating benefits and services to members and non-members	Was there preferential treatment for members compared to non-members?
$T_7$	Decision making on investments and activities that do not benefit all	Has the organisation made investments or had projects that only benefited a sub-group of the organisation?
$T_8$	Delegating and supervising marketing tasks	Did members accept that others take decisions on the price of products that were sold by the organisation?
$T_9$	Assuming liability in contracts and loans	Did members take responsibility for penalties or sanctions for default on contracts negotiated by the board?
$T_{10}$	Managing political aspirations of board and staff	Did members accept that board members or staff take up party-political responsibilities?

We are aware that the above ten agency dilemmas do not cover all organisational challenges in groups. For example, financial transparency and accountability will influence organisational strength. The same holds true for other aspects of good governance in farmer groups, such as cooperative values and principles of democracy, equality, equity and solidarity (ICA, 1995), or administrative requirements, such as complying with tax regulations and having audited financial statements (Mendoza and Ton, 2003). We restricted our measure to the capacity of governing collective marketing functions.

### Field test in Bolivia

The field test of the data-collection tool was embedded in an impact evaluation of FONDOECAS, *Fondo para el Fortalecimiento de las Organizaciones Económicas Campesinas*, a decentralised administrative entity that manages a small-grants programme and is legally part of the apex organisation CIOEC-Bolivia. This small grant fund finances investments in processing equipment to economic farmer organisations. Economic farmer organisations in Bolivia have a peculiar identity, which is different than the traditional village organisations. While all households in rural Bolivia belong to so-called territorial grassroots organisations (*Organizaciones Territoriales de Base* – OTBs), which are often rural unions (*sindicatos*) created during land reform in 1952, some of them are at the same time members of an economic farmer organisation. These have a self-selected membership and tend to work on a larger geographical scale than the villages. The governance structure of economic farmer organisations, therefore, is different from that of the traditional community *sindicatos*, and they need to develop their own internal system of checks and balances.

We formed a team of local researchers that all had in-depth knowledge of the rural sector. One local researcher had been executive in CIOEC-Bolivia, another a long-standing consultant in a capacity-building programme directed to economic farmer organisations, and a third local

researcher had worked with indigenous organisations and local governments. In 2011, interviews with each of the organisations were conducted by these three local researchers, using a semi-structured questionnaire to describe the history of the organisations, the rules and regulations related to each of the ten agency dilemmas in collective marketing, and the changes made in these internal rules and regulations. Complementary to these interviews, secondary data was extracted from the membership lists of CIOEC-Bolivia, from the business plan proposals submitted to FONDOECAS and earlier studies on these organisations (Camacho et al., 2005). The interviews resulted in a detailed qualitative description of the dynamics around agency dilemmas, and captured how groups managed to resolve them.

In line with the tool developed by Donovan and Stoian (2012), the reports permitted insight into the learning process that characterised each of these organisations. Each interview was condensed into thick descriptions, reports of approximately 10 pages<sup>5</sup>.

In addition to these thick descriptions, we summarised the information with a view to computing a quantitative measure of organisational social capital. Therefore, each interview report ended with a one-page summary sheet (Annex 3) with two assessment questions, each with three answer options to be filled in by the researcher after the interview. The first question captures the ‘presence’ of each of the agency dilemmas in the practice of the organisation, and the second the ‘effectiveness of the organisational solution’ (see Table 6.2). To harmonise the interpretation of the qualitative information, a quality check on the classification of each of the researcher’s summary sheets was made, when the first set of 18 summary sheets was reviewed and discussed in the research team. Some corrections were made during this session to align criteria.

**Table 6.2** Composition of the measure of Tension Containment Capacity (TCC)

	Question 1 (Q <sub>1</sub> )			Question 2 (Q <sub>2</sub> )			
Agency dilemmas	The tension comes up in the activities of our organisation	The tension hardly comes up	The tension never comes up	We managed to resolve it with agreements and organisational arrangements	We are looking for a way to resolve it	We do not need to resolve anything	
	3 points	2 points	0 points	3 points	2 points	1 point	Score
T <sub>1</sub>							TC <sub>1</sub> =Q <sub>1</sub> *Q <sub>2</sub>
T <sub>2</sub>							TC <sub>2</sub>
...							...
T <sub>10</sub>							TC <sub>10</sub>
Tension containment capacity – TCC(10C) =							$\sum_{i=1}^{10} TC_i$

- 5 Some organisational solutions were described more in-depth, following the format of realist case-studies (Ton, Vellema and De Ruyter de Wildt, 2011), which described the problem, the various options that were considered and how the selected solution worked out in practice. Some of these examples served as topics for reflection in CIOEC’s leadership courses. We launched the method as a crowd-sourcing initiative in the professional journal *Farming Matters* (Ton G. (2012a) Crowd-sourcing Organisational Intelligence: capturing the rich experiences of farmers’ organisations. *Farming Matters* 28: 20-21.), and the website [www.collectivemarketing.org](http://www.collectivemarketing.org).

The summary sheet was used to derive a measure of the strength of these internal organisational mechanisms – the Tension Containment Capacity (TCC). Organisations that faced more agency dilemmas are considered to have a higher organisational social capital than organisations that only faced a few of them, and organisations that have resolved these agency dilemmas are considered to be stronger than organisations that are still looking for solutions. To derive a quantitative measure for the tension-containment capacity, we used weighting factors. The first question assesses the presence of the agency dilemma in the organisations ('never comes up' = 0 points; 'hardly comes up' = 2 points; 'comes up' = 3 points). The second questions assesses the extent to which the organisations resolved the agency dilemma with effective rules and regulations ('no need to resolve' = 1 point; 'looking for a way to resolve' = 2 points; 'resolved' = 3 points). To calculate a quantitative proxy indicator of the capability to contain each of the agency dilemmas, we multiplied the values of both questions (see Table 6.3). This resulted in a total score, summing the scores on each of the ten agency dilemmas. Some scoring possibilities are logically impossible: 'comes up, and no need to resolve', 'never comes up, and resolved' and 'never comes up, and looking for ways to resolve' are the three categories that have no empirical cases.

**Table 6.3** Weighting factors used for calculating each tension-containment variable

	This tension comes up in their activities	This tension hardly comes up	This tension never comes up / not relevant
They have managed to resolve	9	6	no empirical cases
They are looking for ways to resolve it	6	4	no empirical cases
They don't need to resolve it	no empirical cases	2	0

## Calibrating the measure

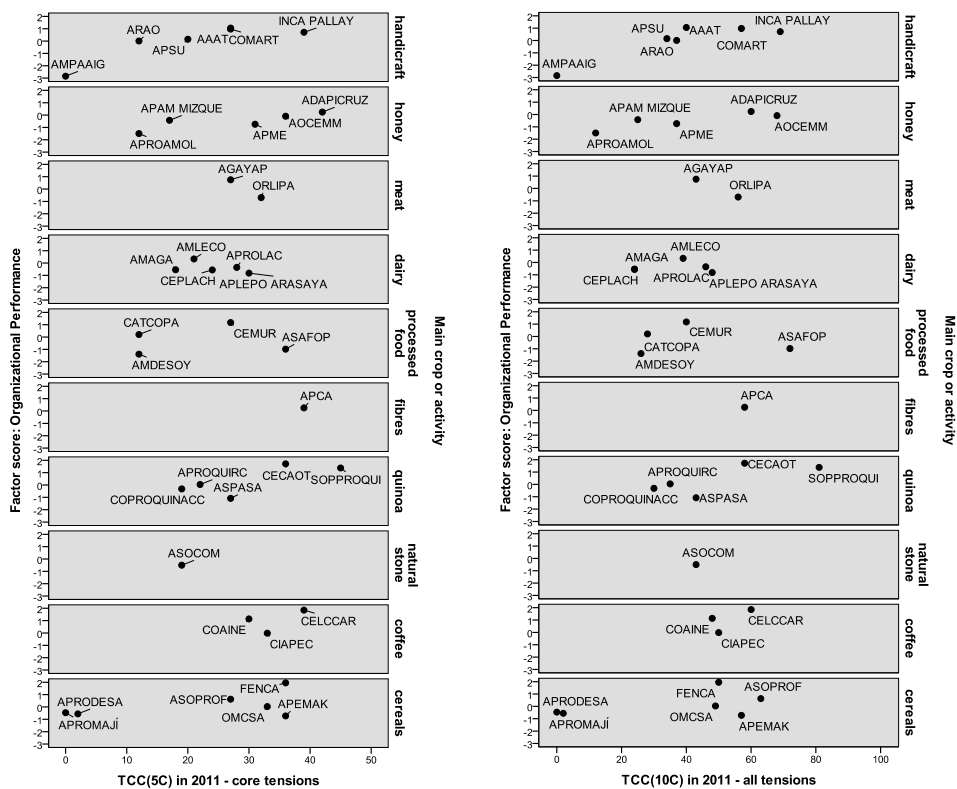
In 2011 we collected information on 38 organisations. Initially, we used all ten agency dilemmas to compute the TCC score and ranked the organisations accordingly (Ton et al., 2014a). The TCC-scores correlate quite well with a factor scores of organisational performance. The ranking of organisations per sector (Figure 6.1) seemed valid according to the researchers and the staff of FONDOECAS, CIOEC and AOPEB. However, we aspired to have a measure that would be applicable to organisations active in different sectors. And, not all the ten agency dilemmas proved to be relevant, often due to sector characteristics. Including all ten agency dilemmas in a quantitative measure could result in a sectoral bias and distort cross-case comparison of tensions containment capacity. Therefore, we decided that it was necessary to distil a subset of agency dilemmas. Table 6.4 shows that three agency dilemmas proved little relevant in the 2011 interviews: T<sub>1</sub>, 'Regulating member supply', T<sub>7</sub>, 'Decision making on investments and activities that do not benefit all', and T<sub>9</sub>, 'Assuming liability in contracts and loans'. Therefore, we used the seven other agency dilemmas, instead of all ten, in other applications of the tool, and for our crowd-sourcing initiative (Ton, 2012a).

Because we could apply the tool twice in the same organisations, as part of an impact evaluation of the FONDOECAS grant fund, we had the opportunity to further fine-tune the tool. Though some organisations experienced major changes, in most organisations the relevance of many of the agency dilemmas had remained unchanged between 2011 and 2013. Therefore,



we could apply a further check on the validity of the measure by comparing the results of both measurements.

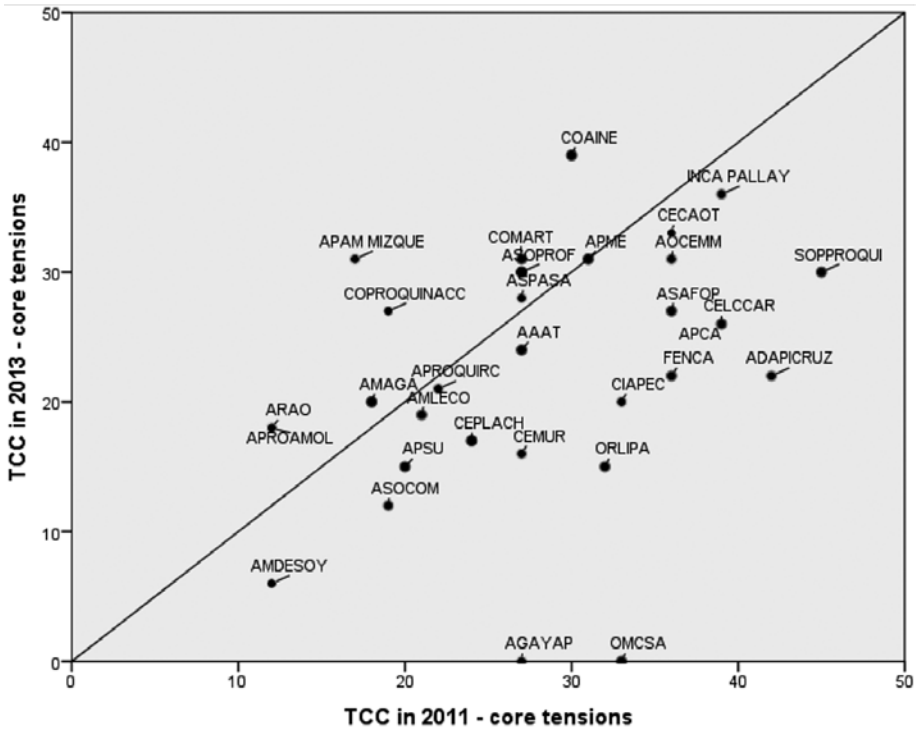
In 2013, we had complete information on 31 organisations<sup>6</sup>. The second round of interviews showed a slightly different pattern, with a low relevance of the agency dilemma T6 'Differentiating benefits and services to members and non-members'. The answers on the agency dilemma T5 'Ways to use profits' proved to be inconsistently coded, suggesting divergent interpretations of the issue in both rounds of interviews. Also the three agency dilemmas that were left out due to a likely sector bias proved very inconsistently coded. The major inconsistency was in the agency dilemmas that were considered resolved in 2011 and not needing to be resolved in the 2013 measurement.



Source: Author's own elaboration

Figure 6.1 Scatter plot of organisations on TCC and organisational performance, as used in Ton et al. (2014a)

6 Of the initial sample of 38 organisations for which baseline data was collected and reported on in Ton et al. (2014), one organisation could not be considered to be a collective marketing group but an NGO, two organisations had provided information on the rules and regulations related to past activities and, as later became clear, had no collective marketing activities anymore in 2011; two other organisations could not be contacted again in 2013 and had allegedly stopped their economic activities. Two others functioned as second-tier organisations that supported activities in their membership but did not manage any economic activities themselves.



Source: Author's own elaboration

Figure 6.2 Scatterplot of the changes in valuation of the tension containment capacities between 2011 and 2013

To limit this likely source of researcher bias in the longitudinal analysis for the impact evaluation, we decided to compute our measure of organisational social capital only on those agency dilemmas which appeared to be most consistently relevant in the organisations, and especially when the same researcher had conducted both interviews. Therefore, we ended up with five core agency dilemmas for which we computed the TCC-score for organisational strength in collective marketing. The difference in ranking of organisations on the ten, or the five agency dilemmas proved similar (for 2011: Kendall's tau<sub>b</sub>=.711; p<.01; for 2013: Kendall's tau<sub>b</sub>=.603; p<.01), as was the apparent validity of the resulting chart according to the experts involved in CIOEC, AOPEB and FONDOECAS (Figure 6.1). The five core agency dilemmas were Quality assurance systems (T<sub>2</sub>), Ways to reduce the need for working capital (T<sub>3</sub>), Prevention of disloyal behaviour in sales (T<sub>4</sub>), Delegating and supervising marketing tasks (T<sub>9</sub>) and Managing political aspirations of board and staff (T<sub>10</sub>).

Another step in fine-tuning the tool was the comparison between the measurements in 2011 and 2013. As expected, the TCC-scores based on the five core agency dilemmas are positively correlated (Pearson  $r=.364$ ;  $p=.044$ ) while the TCC-score based on the ten agency dilemmas was not (Pearson  $r=.291$ ;  $p=.112$ ). This significant correlation gave us confidence in the validity of the measure as a rough indicator of organisational social capital. The imperfect correlation shows that there are many other influencing factors and/or that the above-mentioned

**Table 6.4** Relevance and consistency of agency dilemmas

Sector	N	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>7</sub>	T <sub>8</sub>	T <sub>9</sub>	T <sub>10</sub>
		Supply systems	Quality systems	Work capital	Side selling	Use of profits	Member policy	Focussed investments	Task division	Liability	Political aspiration
handicrafts	6	100%	80%	80%	80%	40%	60%	80%	80%	80%	60%
coffee	3	0%	100%	100%	100%	67%	67%	67%	100%	67%	67%
fibres	1	0%	100%	100%	100%	100%	100%	0%	100%	100%	100%
meat	2	0%	50%	50%	100%	100%	100%	100%	100%	50%	100%
dairy	5	0%	100%	100%	33%	33%	100%	33%	0%	67%	100%
honey	5	20%	100%	100%	80%	60%	80%	20%	80%	20%	60%
quinoa	5	40%	100%	100%	60%	80%	80%	60%	100%	40%	100%
processed food	4	33%	67%	67%	33%	67%	100%	33%	67%	100%	100%
natural stone	1	100%	100%	100%	0%	0%	0%	0%	0%	100%	100%
cereals	6	33%	100%	100%	100%	100%	100%	67%	100%	67%	67%
Overall relevance 2011	38	34%	82%	89%	68%	63%	74%	47%	68%	53%	79%
Included in TCC-score after first round			X	X	X	X	X		X		X
handicrafts	5	80%	100%	100%	40%	20%	20%	40%	80%	40%	60%
coffee	3	33%	100%	100%	100%	100%	33%	33%	100%	33%	67%
fibres	1	0%	100%	100%	100%	0%	0%	100%	100%	0%	100%
meat	2	0%	0%	50%	0%	50%	50%	50%	50%	0%	50%
dairy	3	0%	100%	67%	0%	100%	67%	33%	67%	0%	100%
honey	5	60%	100%	100%	100%	80%	80%	40%	80%	60%	80%
quinoa	5	40%	100%	100%	80%	60%	40%	40%	100%	60%	60%
processed food	3	0%	67%	33%	0%	100%	33%	33%	67%	67%	100%
natural stone	1	0%	100%	100%	0%	100%	0%	0%	0%	0%	0%
cereals	3	33%	67%	67%	67%	67%	67%	67%	67%	67%	67%
Overall relevance 2013	31	35%	87%	84%	55%	68%	45%	42%	77%	42%	71%
Consistently coded as relevant or irrelevant*	29	65%	87%	84%	71%	58%	65%	45%	74%	55%	69%
Consistently coded by same researcher	14	79%	93%	86%	93%	79%	79%	57%	93%	79%	93%
Included in TCC-score after second round			X	X	X				X		X

\* Excluding two organisations that stopped collective marketing activities (AGAYAP, OMCSA)

measurement errors still influenced the results. All organisations had experienced changing conditions and emerging problems in the two years that separated both measurements. For example, OMCSA and AGAYAP are clear outliers (Figure 6.2), because both organisations stopped their collective marketing activities and changed their focus and membership.

In the initial design (Ton, 2012b; Ton et al., 2011b), in order to address construct validity threats, we had planned to compare the results with another measure of organisational strength, a guided self-evaluation, which was to become part of the formal intake procedures for bank loans to economic farmer organisations. Unfortunately, the funding of these self-evaluation exercises by APSA-DANIDA was discontinued (PROFIN, 2009).

As an alternative, we asked a knowledgeable practitioner to rank the organisations in the sample based on his knowledge and insights. This person worked as a monitoring officer for a grant fund (FONDOECAS) that stimulates collective marketing activities. Therefore, his ranking can be considered as an alternative measure for organisational strength in collective marketing. We used an Excel application for pairwise comparison of organisations in the sample. The monitoring officer had substantive knowledge on 8 of the 38 organisations in the 2011 sample. The ranking on the TCC-score on the five core agency dilemmas (TCC5C) proved to be significantly correlated with this alternative ranking by the field officer ( $N=8$ ; Kendall's  $\tau=.668$ ;  $p<.05$ ), while the TCC-ranking based on all ten (TCC10C) was not (Table 6.5).

**Table 6.5** Comparison of relative ranking

			TCC5C in 2011	TCC10C in 2011
Kendall's tau_b	Ranking by FONDOECAS expert	Correlation Coefficient	.668*	.327
		Sig. (2-tailed)	.028	.262
		N	8	8

## Correlation with economic performance

The TCC measures the capacity to manage collective marketing activities. Therefore, we expected to see a positive correlation between the TCC-score and economic performance indicators. We could test this correlation with the high quality time-series data on group sales after the 2013 interviews. We used as proxy-indicators the total value of 'group sales' and 'group sales per member'. To make these variables suitable for correlation analysis, they were transformed using their natural log.<sup>7</sup>

We used a linear regression model (Model 1a in Table 6.6) to test the hypothesis that baseline TCC-scores predict the level of group sales in 2012. The results in Table 6.6 show that the TCC-score predicts the increase in group sales ( $R\text{-squared} = .525$ ;  $P<.05$ ;  $\beta=.286$ ).

<sup>7</sup> Based on the 2011 data, we had used another performance construct to explore the correlation in Ton et al. (2014a). This construct reflected organisational performance, and was based on a principal component of membership, patrimony and organisation age.

We also expected that the intensity of transactions, the average value of transactions between the group and the member, would be a predictor of the group's organisational capacities (Soboh et al., 2009). Indeed, Model 2 show that the increase in tension containment capacity is also predicted by the baseline level of group sales per member (R-squared = .293;  $p < .05$ ;  $\beta = .421$ ). Both results provide support for the validity of the construct.

**Table 6.6** Relationship between tension containment capacity and performance indicators

	Model 1a	Model 1b	Model 2a	Model 2b
R-squared	.121	.525	.111	.293
Significance	$p = .064$	$p = .000$	$p = .067$	$p = .009$
Outcome:	Group sales (Ln) in 2012	Group sales (Ln) in 2012	Tension containment capacity in 2013	Tension containment capacity in 2013
Beta-values of predictors <sup>a</sup>				
Tension containment capacity in 2011	.348*	.286**	.334*	.352**
Group sales (Ln) in 2010	--	.630***	--	--
Group sales per member (Ln) in 2010	--	--	--	.421**

a. A constant was included in all regressions.

\* Statistical significant at  $p < .10$ ; \*\*  $p < .05$ ; \*\*\*  $p < .01$

## 6.5 Reliability in longitudinal analysis

In the former chapter, we used the TCC-score as a measure of the 'stock' of tension containment capacity in 2011 and 2013, a proxy of organisational social capital. We will now explore the reliability of the change between two measurements of the TCC-score ( $\Delta TCC$ ), the 'flow' of the tension containment capacity in time, as an indicator of organisational strengthening. The requirements for precision of measurements are much higher when net-effects are the goal of the research (Ton et al., 2011b), especially when used to assess changes on a case by case basis. In larger samples, computing group averages, random measurement error would disappear, as upward and downward changes would offset each other.

To be useful as an indicator of organisational strengthening in longitudinal research, the registered changes between the 2011 and 2013 measurements need to have empirical relevance and not be the result of measurement error. A potential threat to the reliability of  $\Delta TCC$  as an indicator of change that we need to refute is the phenomenon called 'Regression Towards the Mean' (RTM). RTM occurs when, by chance, the observed TCC-score of an organisation fluctuates randomly around a stable 'true' score (Barnett et al., 2005). Logically, organisations with higher TCC-scores in the 2011 would tend to have lower scores in the 2013 measurement, and vice versa. If all  $\Delta TCC$ -scores in the sample would suffer from RTM and only reflect this measurement error, the average  $\Delta TCC$ -score would be distributed normally, and fluctuate round an average value of zero. We tested for both aspects (Table 6.7). The Shapiro-Wilk test

for normality indeed showed that the  $\Delta$ TCC was normally distributed ( $p=.778$ ), but the mean was significantly different from zero (mean =-5.9;  $T=-3.1$ ;  $p<.01$ ). The latter does not completely dispel the threat of RTM, because the difference from zero could reflect a systematic bias between the 2011 and 2013 measurements.

**Table 6.7** Correlation of base-line values and changes between base-line and follow-up of core tensions (N=31)

Agency dilemmas	TCC 2011	TCC 2013	2011-2013 score change	Correlation $\Delta$ TCC and TCC 2011
	Mean ( $\pm$ S.E.)	Mean ( $\pm$ S.E.)	Mean ( $\pm$ S.E.)	Pearson
T <sub>2</sub> – Quality assurance	6.58 (.51)	5.03 (.41)	-1.55 (.51)***	-.675***
T <sub>3</sub> – Working capital	6.26 (.29)	5.81 (.55)	-0.45 (.59)	-.387**
T <sub>4</sub> – Disloyal side-selling	4.65 (.59)	3.06 (.54)	-1.58 (.51)***	-.534***
T <sub>8</sub> – Task delegation	5.39 (.61)	4.58 (.52)	-0.81 (.73)	-.712**
T <sub>10</sub> – Managing politics	5.32 (.63)	3.81 (.52)	-1.52 (.77)*	-.742***
TCC - Tension Containment Capacity	28.2 (1.61)	22.3 (1.71)	-5.90 (1.98) ***	

\* Statistical significant at  $p<.10$ ; \*\*  $p<.05$ ; \*\*\*  $p<.01$

We already showed that a certain level of researcher bias was present. For example, in Table 6.4 we presented the consistency of the presence of agency dilemmas. We showed that the consistency improved when the same researcher interviewed the organisations in both rounds. Precisely for this reason, we excluded the agency dilemmas that appeared to be the least consistently measured from the computation of the TCC-score. Nevertheless, some systematic bias may still be present in the five core agency dilemmas. A test for researcher bias of the  $\Delta$ TCC, using ANOVA, showed that having different researchers in the two measurement rounds resulted in a slightly higher  $\Delta$ TCC, but that this bias was only weakly statistical significant ( $N=31$ ;  $F=3.02$ ;  $p=.093$ ). When we exclude AGAYAP and OMCSA, which stopped collective marketing activities, from the average  $\Delta$ TCC computation, this systematic bias becomes unlikely ( $N=29$ ;  $F=1.06$   $p=.311$ ). However, systematic bias may also result from other sources, such as increased substantive knowledge on the case, and the possibility of cross-checking information during the interview, which may lead to a deeper, more nuanced insight about the effectiveness of rules and regulations in the second interview round. Furthermore, election processes in member-based organisations imply that board composition may have changed between two observations of the TCC. A new board may have a different perspectives on group dynamics and effectiveness of rules and regulations, while in reality these may not have changed substantially.

Acknowledging these sources of bias, we continued to verify if  $\Delta$ TCC is a measure that reflects the organisational dynamics in each organisation. To do so, we looked more closely at the ‘real’ events that had taken place in each organisation.

## Reality check

For this reality check, we had two detailed interview reports, ‘thick descriptions’ on each organisation, one made in 2011 and another in 2013. Therefore, we could check if the  $\Delta$ TCC indeed reflected ‘real’ events, as described in these reports. We reviewed the thick descriptions based on the qualitative interviews to look for explanations of the registered changes. When the interviews would not provide an explanation of the changed TCC-score, this indicated a possible measurement errors in one or both rounds of data collection.

To select the sample, we plotted the TCC-scores for 2011 and 2013 (Figure 6.2). We selected those organisations that changed most<sup>8</sup>, and those that changed least in their 2011 and 2013 scores. These cases were expected to have a ‘story’ to tell that would explain the relatively large change in their tension containment capacity. We analysed four organisations with the most positively changed ones (APAM-MIZQUE, COPROQUINACC, ARAO and APROAMOL), the four most negatively changed ones (AMDESOY, ADAPICRUZ, ORLIPA and ASOCOM), and four organisations with the most stable TCC-score (APROQUIRC, ASPASA, APME and INCA PALLAY).

**Table 6.8** Reality check on organisations with most positive change in tension containment capacity ( $\Delta$ TCC)

	ARAO	APROAMOL	COPROQUINACC-T	APAM MIZQUE
<i>TCC-score (2013-2011)</i>	<i>(18-12)</i>	<i>(18-12)</i>	<i>(27-19)</i>	<i>(31-17)</i>
<i>Change in TCC between 2011 and 2013</i>	<i>+6 (50%)</i>	<i>+6 (50%)</i>	<i>+ 8 (42%)</i>	<i>+ 14 (82%)</i>
T <sub>2</sub> – Quality assurance	6 (6-0)	0 (4-4)	2 (4-2)	0 (6-6)
T <sub>3</sub> – Working capital	0 (6-6)	0 (4-4)	5 (9-4)	-5 (4-9)
T <sub>4</sub> – Disloyal side-selling	-6 (0-6)	2 (6-4)	-3 (6-9)	6 (6-0)
T <sub>8</sub> – Task delegation	6 (6-0)	4 (4-0)	2 (4-2)	4 (6-2)
T <sub>10</sub> – Managing politics	0 (0-0)	0 (0-0)	2 (4-2)	9 (9-0)
<i>Researchers in 2011 and 2013s</i>	<i>Different</i>	<i>Same</i>	<i>Different</i>	<i>Different</i>
<i>Results of the reality check</i>	Supported with evidence on all core agency dilemmas	Supported with evidence on all core agency dilemmas	Supported with evidences on all core agency dilemmas	Supported with evidence on core agency dilemmas, except managing politics

We started the reality check with the organisations that registered a relative large increase in tension containment capacity between 2011 and 2013; details on the TCC-scores are given in Table 6.8.

<sup>8</sup> We excluded AGAYAP and OMCSA because these stopped all their collective marketing activities. For analysing the reliability of the  $\Delta$ TCC-construct, they are, thus, not very informative.

**ARAO** – This organisation, specialised in the production of carpets, sweaters, shawls and ponchos, has its own shop in Oruro for distribution and has started to sell through other shops as well (e.g., COMART and INCA PALLAY), which explains the increase in task delegation ( $\Delta T_8 +6$ ). ARAO uses a system for planning of production, and cash payment for products covered by this plan, with the possibility to deposit any additional handicraft products in the shop under a system of consignment ( $\Delta T_3 0$ ). In 2011, the board mentioned that they had solicited a loan to start paying in cash also for the additional production of members. However, in 2013 they considered side-selling not an issue because ARAO could not sell all products made by the weavers to consumers. Resulting from an investment in new weaving equipment, production has increased in volume and quality ( $\Delta T_4 -6$ ). This made it feasible to apply quality criteria in procurement ( $\Delta T_2: +6$ ).

**APROAMOL** – This small honey association (20 members) provides the service of harvesting honey. They started selling honey directly to institutional clients with delayed payments. In 2013, they changed this and started selling through dedicated distributors to consumers, resulting in faster payment ( $\Delta T_3 0$ ). The issue of side selling is given more importance in the 2013 interview ( $\Delta T_4: +2$ ). The president and vice-president are the ones that do the marketing, on a voluntary basis. They have started to report quarterly in meetings at which the majority of members participate ( $\Delta T_8: +4$ ).

**COPROQUINACC-T** - The quinoa organisation (130 members) operates in a booming market with increasing prices. They experience more problems of side-selling ( $\Delta T_4: -3$ ). They have access to trade capital through their second-tier organisation ANAPQUI, which also manages the marketing ( $\Delta T_8 +2$ ). ANAPQUI used a delayed payment systems. The delay in final payment to the farmers had been reduced from several months to only two weeks ( $\Delta T_3 +5$ ). In 2013, they mentioned that the issue of political representation had become more relevant, and there was a need to define rules on this issue ( $\Delta T_{10} +2$ ). The issue of quality assurance had gained importance ( $\Delta T_2 +2$ ) due to stricter control of organic quinoa, and increased risk of contamination from neighbouring fields.

**APAM MIZQUE** - This honey organisation (decreasing from 135 in 2009 to 50 members in 2013) introduced more rigid member obligations, to purge the association of certain members that directly competed with the organisation as intermediaries. These intermediaries were paying cash, while the organisation did not have enough funds to do so ( $\Delta T_3: -5$ ). In 2011, they considered the issue of side-selling irrelevant but it became more problematic due to these competing member-traders ( $\Delta T_4: +6$ ). In 2011 they had commercial staff paid with external donor support. This support project ended, however, and in 2013, the sales work was resolved 'within' the organisation, which the interviewed board members considered to be a positive change ( $\Delta T_8: +4$ ). While in 2011 they considered the issue of political aspiration irrelevant, in 2013 they assessed it as resolved ( $\Delta T_{10}: +9$ ). This large change in score was however not supported in the interviews, in which there was no mention of changes in rules and regulations.



**Table 6.9** Reality check on organisations with most negative change in tension containment capacity ( $\Delta$ TCC)

	ADAPICRUZ	ORLIPA	ASOCOM	AMDESOY
<i>TCC-score (2013-2011)</i>	(22-42)	(15-32)	(12-19)	(6-12)
<i>Change in TCC between 2011 and 2013</i>	-20 (48%)	-17 (53%)	-7 (37%)	-6 (50%)
T <sub>2</sub> – Quality assurance	-3 (6-9)	-9 (0-9)	-3 (6-9)	0 (0-0)
T <sub>3</sub> – Working capital	0 (6-6)	3 (9-6)	2 (6-4)	-6 (0-6)
T <sub>4</sub> – Disloyal side-selling	-5 (4-9)	-2 (0-2)	0 (0-0)	0 (0-0)
T <sub>8</sub> – Task delegation	-9 (0-9)	-2 (4-6)	0 (0-0)	0 (0-0)
T <sub>10</sub> – Managing politics	-3 (6-9)	-7 (2-9)	-6 (0-6)	0 (6-6)
<i>Researchers in 2011 and 2013</i>	<i>Same</i>	<i>Different</i>	<i>Different</i>	<i>Same</i>
<i>Results of the reality check</i>	Supported with evidence, except on task delegation	Supported with evidence, except on quality assurance	Supported with evidence on all agency dilemmas	Supported with evidence, except on the issue of working capital

The organisations with a relative large negative change in TCC-score between 2011 and 2013 are covered below and summarised in Table 6.9.

**ADAPICRUZ** - This honey processor (300 members) grew quickly in a (niche) market that is constrained. In addition to public procurement, lucrative other markets are explored but only piloted with piecemeal deliveries (organic exports). Quality control has improved, although the organic market poses new demands to traceability which means adjusting the quality assurance systems ( $\Delta$ T<sub>2</sub>: -3). Due to an oversupply, ADAPICRUZ reported more problems due to disloyal behaviour ( $\Delta$ T<sub>4</sub> -5), especially from members that sell honey from non-members as their own. ADAPICRUZ assumed its role in representing the honey sector, rethinking a way to manage party-political interference in this role ( $\Delta$ T<sub>10</sub>: -3). The registered change in task delegation ( $\Delta$ T<sub>8</sub>: -9) was not supported in the interviews, as no changes have been made in these regulations according to the interview reports. This error is likely due to the peculiar business organisation of ADAPICRUZ. It manages its commercial activities as a separate legal identity, Apícola del Bosque S.A., in which ADAPICRUZ and some (large) individual producers have shares.

**ORLIPA** - This meat-processing group (52 members) delivered its products to the school meal programme. In 2011 they were prepared to deliver a processed product made from dried meat and broad beans, sourced from members. However, in the final contract this product was removed from the list of required food items, and ORLIPA became a mere intermediary for grocery products. Without this contract with the municipality, the issue of side-selling became irrelevant ( $\Delta$ T<sub>4</sub>: -2). They also changed leadership in 2013. The new president was also elected as secretary general of the village organisation, and has not dedicated himself to the re-launching of the commercial activities of ORLIPA. This was criticised by some, and explains the changed score on the issue of

managing political aspirations ( $\Delta T_{10}$ : -7). After the first interview, ORLIPA received a working-capital loan from FONDOECAS ( $\Delta T_3$ : +3), which resolved their working-capital constraints. The change in quality assurance procedures ( $\Delta T_2$ : -9) was inconsistent, but can be explained with reference to the interviews as a result of a faulty interpretation by the 2011 researcher, who focused on quality in animal husbandry for which capacity building had taken place, instead of the procedures around quality assurance in their processing activities, which were lacking.

**ASOCOM** – This group of stone miners (72 members) was created in 2003, when the community took over the stone mine owned by the former (expelled) president Sanchez de Losada. In 2011, the organisation changed its legal character and converted into a cooperative (COCACOM). They do not source from members, but work as a production cooperative with central negotiation of contract and work assignments to mining teams. Managing political aspirations ( $\Delta T_{10}$ : -6) was considered irrelevant in 2013, while in 2011 they mentioned a satisfactory gentlemen’s agreement on this issue.

**AMDESOY** - This is a women’s group that makes food products with soya-meal. They do not purchase from members but function only as a micro-enterprise. The member-workers sell the products directly to consumers in Santa Cruz, with a system of door-to-door sales. They use bank loans for working capital. In 2013, they experienced more constraints in working capital but in fact they had no tensions related with payment systems to members. The local researcher qualified the agency dilemma as not relevant (‘the tension never comes up’), while in the valuation in 2011 she had indicated that the agency dilemma was considered relevant. This change has no grounding in the interview reports and has to be considered a measurement error.

**Table 6.10** Reality check on organisations with relatively unchanged tension containment capacity ( $\Delta TCC$ )

	INCA PALLAY	APROQUIRC	ASPASA	APME
<i>TCC-score (2013 – 2011)</i>	<i>(36-39)</i>	<i>(21-22)</i>	<i>(28-27)</i>	<i>(31-31)</i>
<i>Change in TCC between 2011 and 2013</i>	<i>-3 (8%)</i>	<i>-1 (5%)</i>	<i>+ 1 (4%)</i>	<i>0 (0%)</i>
T <sub>2</sub> – Quality assurance	-3 (6-9)	0 (6-6)	0 (6-6)	3 (9-6)
T <sub>3</sub> – Working capital	0 (9-9)	3 (9-6)	0 (9-9)	0 (9-6)
T <sub>4</sub> – Disloyal side-selling	-3 (6-9)	0 (0-0)	3 (9-6)	-2 (4-6)
T <sub>8</sub> – Task delegation	0 (6-6)	2 (6-4)	0 (4-4)	2 (6-4)
T <sub>10</sub> – Managing politics	3 (9-6)	-6 (0-6)	-2 (0-2)	-3 (6-9)
<i>Researchers in 2011 and 2013</i>	<i>Same</i>	<i>Different</i>	<i>Different</i>	<i>Same</i>
<i>Results of the reality check</i>	Supported with evidence on all core agency dilemmas	Supported with evidence on core agency dilemmas, except managing politics.	Supported with evidence on all core agency dilemmas	Supported with evidence on all core agency dilemmas

We finish the reality check by examining the changes in tension containment capacities in three organisations that have changed less in their aggregate TCC-scores (Table 6.10).

**APROQUIRC** – This rapidly growing regional branch of the quinoa federation ANAPQUI has 210 members. APROQUIRC managed to get a loan from the development bank, *Banco de Desarrollo Productivo*, to pay cash to supplying farmers ( $\Delta T_3$ : +3). The issue of disloyal side-selling is present but did not change much between 2011 and 2013 ( $\Delta T_4$ : 0). Marketing is delegated to ANAPQUI staff. Improvement in this area ( $\Delta T_8$ : +2) is due to the more extensive information provided by ANAPQUI on market issues. The issue of political representation has come to the forefront, as ANAPQUI is more explicit in its support for the ruling party. Erroneously, the 2013 researcher had marked this political issue as irrelevant ( $\Delta T_{10}$ : -6). On this issue the difference between the answer ‘hardly comes up, and resolved’ and ‘never comes up, and no need to resolve’ is even more difficult to see than on other agency dilemmas.

**ASPASA** – This relatively new and small quinoa organisation (52 members) is close to obtaining organic certification. This niche market is expected to be a good way to increase member loyalty in their sales, due to the higher price ( $\Delta T_4$ : +3). In the 2011 interview they mentioned that they had convinced a leader not to take part in the elections, but in 2013 this agency dilemma was considered not relevant ( $\Delta T_{10}$ : -2). The other agency dilemmas had not changed in rules and regulation.

**APME** – This honey organisation (104 members) is increasingly independent from the NGO that helped them during start-up and they have a steady increase in production and group sales. They sell most of the honey to the government nutrition programme. In 2012, they opened a new processing facility, which explains the need to address the issue of quality assurance ( $\Delta T_2$ : +3) and task delegation ( $\Delta T_8$ : +2). The issue of side-selling was less of a problem in 2013 ( $\Delta T_4$ : -2) than it was in 2011, mainly as a result of the better harvest, which caused an oversupply. The issue of managing political relations was considered to be less relevant in 2013 than in 2011, when one of their members was elected as city mayor ( $\Delta T_{10}$ : -3).

**INCA PALLAY** – This organisation is specialised in weaving and targets the high-end market of international tourists. It has shops in Sucre and La Paz, and a museum-shop in Tarabuco. On the issue of quality assurance, they experienced more problems in 2013 compared to 2011. They consider their current system of grading insufficient to address this problem ( $\Delta T_2$ : -3). They have sufficient funds to buy products but need to compete with the higher prices offered by a new governmental artisanal centre in Sucre that started operations in 2012. This has increased the problem of side-selling ( $\Delta T_4$ : -3). The coordinator of INCA PALLAY also mentions that the new president has a background in an opposition party, which negatively affected relations with the local governments. To address this they have decided to give the vice-president a more prominent role ( $\Delta T_{10}$ : +3).

## Towards a more robust measure

The reality check showed that most changes in TCC are reflected in the information provided by the interviewed board members. The main sources of error are inconsistencies in the evaluation of the relevance of an agency dilemma in the organisation's practice, especially when valued as '(hardly) relevant and resolved' in 2011 and 'not relevant and no need to resolve' in 2013. However, of the 60 changes that were reviewed, only six were not supported by any information in the interview reports. This relatively large number of 'grounded' changes makes it implausible that the phenomenon Regression Towards the Mean (RTM) explains the differences in TCC score. Most changes in the disaggregated agency dilemmas reflected 'real' processes. The measurement errors cause problems in longitudinal analysis ( $\Delta$ TCC). However, the influence of these errors on the absolute value of the TCC-score (the 'stock' of organisational social capital) seems less of a problem. The relative ranking of the organisations is fairly stable, when we apply the corrections.

In the process of developing the TCC-score, we focused only on the five agency dilemmas that were most evident and having the least researcher bias. Therefore, the fine-graded assessment of relevance in Question 1 about the relevance of the agency dilemmas became less relevant. We experimented with two simpler alternatives to compute the TCC score (Table 6.11). In Alternative 1, we gave an equal score to an agency dilemma that 'comes up' or 'hardly comes up', and in Alternative 2 we excluded Question 1 and only considered the answers on Question 2 in the computation of the TCC-score. Table 6.12 shows that the TCC-scores calculated with these alternative weighting factors have similar patterns of correlation with economic performance indicators. This suggests that Question 1 may well be eliminated in deriving the TCC-score. The comparative ranking of organisations based on their TCC-scores and their correlation with the economic performance indicators seems quite robust to these simpler weighting alternatives.

**Table 6.11** Alternative weighting factors for test of robustness

	Question 1: This tension [ <i>comes up</i> ] in the activities that we realise.			Question 2: We [ <i>managed to resolve</i> ] with organisational agreements and arrangements.		
	Comes up	Hardly comes up	Never comes up	Managed to resolve	Looking for ways to resolve	Don't need to resolve it
Original weighting	3	2	0	3	2	1
Alternative A	1	1	0	3	2	1
Alternative B	-	--	--	3	2	1

Note: See Annex 1 for Spanish wordings used in the field-test

The effects of the different weighting factors have understandably more effect on the measure of change ( $\Delta$ TCC). This simplification of the TCC computation in Alternative 1 and Alternative 2 are not the solution for the problem in longitudinal analysis, where sometimes an agency dilemma is marked as 'resolved in 2011' with 'no need to resolve in 2013', while we did not register any change in an organisation's reality. The  $\Delta$ TCC-score alone might not yet be a reliable indicator for organisational strengthening. However, we see two ways make  $\Delta$ TCC less prone to measurement error.

First, a more intensive preparation and more detailed guidelines for the interviews could help to harmonise interpretations between researchers. Differences in interpretation of interviews and differences in perspectives on issues in time are unavoidable, but can be reduced when the researchers better explain and discuss their assessments. For example, if the researchers would have conducted the initial interviews together, they could have discussed interpretation issues and aligned criteria.

A second way to reduce the error would be to provide the interviewers with the first summary sheet, and require that all changes between the two measurements be documented with reference to real processes and dynamics. A change from 'relevant' to 'irrelevant' would than always need clarification in the interview report. The researchers would have been helped in the interview process when they could have reflected on the summary sheet made two years earlier for the same organisation. This could help their interpretation of the current situation around each agency dilemma within an organisation, and/or help to correct apparently deficient assessments in these earlier measurement, based on the increased information and insights.

Because we wanted to use the results for an impact evaluation of the FONDOECAS grant fund, and were afraid of a potential positive bias, we had deliberately removed the summary sheet from the first interview to reduce an eventual positive bias to favour the results of the impact evaluation. It generated two relatively independent measures that made our reliability test stronger, but, likely, negatively influenced the coherence between the two interview reports, and the interview dynamics, which could have benefitted from this information to ask probing questions and clarifications about the apparent changes. In the end, we had no reason to suspect a positive bias on the part of the individual researchers to influence the outcomes of the impact evaluation, whereas the measurement errors made negatively influenced the use that we could make of the  $\Delta$ TCC as proxy-indicator of organisational strengthening between 2011 and 2013.

**Table 6.12** Results of robustness check (Pearson correlations)

	Alternative 1	Alternative 2	Group sales 2012 (Ln)	Group sales per member 2012 (Ln)	Patrimony 2012 (Ln)	Size of the membership 2012 (Ln)
TCC - 2011 – Original	.914**	.911**	.326	.050	.265	.520**
TCC - 2011 – Alternative 1			.382*	.112	.246	.473**
TCC - 2011 – Alternative 2			.313	.066	.279	.459*
	N	31	31	30	29	30
TCC - 2013 – Original	.947**	.908**	.542**	.427*	-.033	.168
TCC - 2013 – Alternative 1			.520**	.443*	-.027	-.088
TCC - 2013 – Alternative 2			.471**	.398*	.031	.161
	N	31	31	30	29	30
$\Delta$ TCC - Original	.853**	.895**	.146	.088	-.211	-.283
$\Delta$ TCC – Alternative 1			.074	-.020	-.167	-.192
$\Delta$ TCC – Alternative 2			.132	.264	-.210	-.276
	N	31	31	30	29	30

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\*. Correlation is significant at the 0.01 level (2-tailed).

## 6.6 Discussion and conclusions

There is a need for methods to compare the organisational strength of farmer groups (Donovan and Stoian, 2012; Meinzen-Dick et al., 2004). Most authors use case-specific proxy-indicators to assess organisational strength, which limit cross-case comparisons (Poteete and Ostrom, 2004). Existing benchmarking tools (SCOPEInsight, 2013; EFQM, 2010) focus more on organisational appearance (staffing, infrastructure, management practices, financial management, etc.) than on inherent organisational strength (trust, commitment, internal governance). This makes them less suited for assessing organisational strength of smaller groups with limited assets and funding, but which might well be ‘stronger’ than larger better-endowed ones (Olson, 1965).

We developed a tool that makes a ‘radiography’ of the capacities of organisations to address the inherent tensions between the members and the group in collective marketing. The tool is appropriate for smaller, less professionalised groups that work in a wide diversity of economic sectors. We identified ten areas in which these tensions tend to occur when farmer groups are active in collective marketing activities. Through in-depth interviews with board members and staff, we made detailed descriptions of the rules and regulations developed by each group to resolve their agency dilemmas. The rich qualitative information shows that farmer groups are learning organisations that adapt their internal rules and regulations to emergent dynamics (Flores and Ton, 2015).

We developed and tested a quantitative measure to compare organisational capacities between organisations and to track organisational strengthening within them. To obtain this quantitative measure, qualitative information from in-depth interviews was used to derive a measure of organisational social capital, called Tension Containment Capacity (TCC). The tool was field-tested in Bolivia, in economic farmer organisations active in a diversity of sectors, with varying activities and organisational forms. We showed that in the Bolivian context, five of these ten agency dilemmas proved to be the most important. These five core agency dilemmas are: quality assurance, working capital and payment systems; prevention of disloyal behaviour (side-selling), task division in commercial decision making, and the management of the political aspirations of board and staff. We showed that this measure of organisational social capital predicts economic performance (group sales). At the same time, the intensity of transactions with members (sales per member) predicts to a large extent their tension containment capacity.

The use of  $\Delta$ TCC as a longitudinal measure of organisational strengthening proved to be more challenging, due to higher demands on measurement accuracy. Independent measurements improve the validity of average-oriented research designs, but, in our case, limited the possibilities to trace the change processes in each of the cases. The field test showed that some changes in the application of the tool are needed. This may be possible with minor modifications in the instrument, mainly by requiring the researchers to give explicit reasons for any registered change with the, earlier, (baseline) scores. This would create room for probing questions that deepen the understanding of the process of change, and, eventually, the detection and correction of measurement errors or faulty interpretations.

The radiography instrument used to review the organisational capacities on ten agency dilemmas seems appropriate for use in other countries and settings where farmer groups are organised around collective marketing activities. It is likely that in other countries the TCC-score needs to be computed with different sets of agency dilemmas that have cross-sectional relevance.

Organisational strengthening is a key objective of agricultural development but is overlooked as an outcome area in many commissioned impact evaluations of agricultural support, who stress impact measurements in farmer households (Ton et al., 2015; Nelson and Martin, 2012). We consider the measure of tension containment capacity (TCC) a promising 'common indicator' of organisational social capital of collective marketing groups for use in impact evaluation of development interventions. This would help to compare the effectiveness of various approaches and support interventions that explicitly aim to strengthen collective action of smallholders in markets, and to highlight the influence of collective marketing groups as important moderating factors for farmer-level impacts of development support.

### **Acknowledgements**

The Netherlands' Ministry of Economic Affairs (KB-11-004 and BO-0-010-129) supported this work as part of ESFIM Comparative Research ([www.esfim.org](http://www.esfim.org)). The fieldwork was co-funded by the Dutch Interchurch Development Organizations, ICCO, as part of a baseline for an impact evaluation of the innovation grant fund FONDOECAS. The first round of data collection, in 2010-2011, was done by Lithzy Flores, Evaristo Yana and Rubén Monasterios. The second round of interviews, in 2013, was done exclusively by Lithzy Flores. I would like to thank board and staff of CIOEC-Bolivia and its regional branches, who facilitated the contacts for the interviews, and, especially, Freddy Ticona, Oscar Chambi and Richard Arguedas in FONDOECAS, who provided background information on most of the organizations in the sample.



# 7 | Using Qualitative Comparative Analysis to Explore Outcome Patterns of Grant Support to Farmer Organisations in Bolivia<sup>9</sup>

Giel Ton

Agricultural Economics Research Institute, LEI Wageningen UR

## ABSTRACT

*We used Qualitative Comparative Analysis (QCA) to study the combinations of factors that are consistently related to success or failure of grants given to farmer groups. Using data from a sample of 26 grant beneficiaries, we explored whether baseline characteristics of the organisations related to group sales, organisational scale and organisational strength could predict the intended outcomes of the grant system: improved access to markets for member products, increased organisational capacity, and more income to pay organisational expenses. We explain the calibration process used to assign each organisation to (fuzzy-set) conditions, and the iterative process of QCA to explore the resulting truth-table for plausible causal configurations that may help to target grant funds. We use the ambiguities in the evaluation of success or failure of certain organisations to verify the robustness of the analysis under real-world conditions of measurement error. We detected some single conditions consistently related with success, especially if they were sourcing raw material from members or the spot market, and could triangulate these patterns with logistic regression. The grants to the older, larger and stronger organisations were consistently unsuccessful, because the grant resulted in under-scaled investments in secondary activities that were discontinued after pilot experiences. Finally, we discuss the strengths and weaknesses of QCA as a method for explorative research and causal inference.*

---

9 Published as: Ton, G. (2015) Using Qualitative Comparative Analysis to Explore Outcome Patterns of Grant Support to Farmer Organisations in Bolivia. COMPASS Working Paper 2015-82.



## 7.1 Introduction

Many development projects want to know how they can improve the effectiveness of their support, but have information on only a limited number of cases from which to draw conclusions. There is a need for approaches that maximize synergy between qualitative and quantitative research traditions, between the need for sufficient data set observations and sufficiently informative causal-process observations (Brady and Collier, 2004; Brady et al., 2006).

The reality of having to draw causal inferences from a small sample is not uncommon in other areas of science, e.g., political science or management studies, where data sets relate to a limited ‘population’ of countries or companies. Case-based comparative methods (Byrne and Ragin, 2009) are presented as tools to bridge the qualitative and quantitative divide. Case-based comparative methods use the empirical diversity/heterogeneity of cases in a data set to propose, modify or test theories of causal explanation. Variable-based analytical tools included under this label are Cluster Analysis, Structural modelling and Scatter Plots, which detect causal relations in data sets exploring for correlations between (combinations of) variables and outcomes. Conversely, configurational comparative methods (Rihoux and Ragin, 2009) use Boolean algebra to make inferences on causality and detect (configurations of) causal conditions in data sets that are related with outcomes (Thiem et al., 2015).

Qualitative Comparative Analysis, developed by Charles Ragin (1987; 2000; 2008) is a prominent configurational comparative method. QCA is explicitly explorative in nature and geared to detect, and reflect on, the combinations of conditions consistently related to an outcome. QCA is especially useful when we expect multiple causal pathways that are conducive to producing a certain outcome. Instead of one single causal model that fits the data best, QCA “explores multiple causal models that exist among comparable cases” (Ragin, 1987; Rihoux and Marx, 2013).

Having been developed in political science, QCA has become popular in organisation research (Fiss et al., 2013a) and evaluation (Befani, 2013), as it can be applied on small samples sizes. QCA is criticised for being susceptible to changes in parameters and model specifications. Korgslund et al. (2015) argue that by searching for patterns within the data set, the QCA analysis has a marked confirmation bias and may find patterns in randomly generated data that may mistakenly be interpreted as causal configurations. Lucas and Szatrowski (2014) and Thiem (2013) point to the sensitivity to different specifications and consistency thresholds used in QCA.

In this paper we apply QCA to a real-world evaluation challenge. The research took place in Bolivia between 2010 and 2014 and concerned a small-grant fund that catered to economic farmer organisations. We used QCA to explore whether organisations’ start conditions predict success or failure of the grant. The insights are used to better target grants in the future and increase the effectiveness of the grant system.

The research provided a natural setting in which to check the robustness of QCA results under real-world conditions of measurement error. Two researchers analysed whether grants given to 26 different organisations had been successful or not. Because they had different sources of information, they differed in the evaluation of success or failure on some of these cases. We could use the evaluation before and after the reconciliation as real-world measurement error

to reflect on the stability of the QCA solutions. We expected that the effectiveness of the grant would differ for configurations of conditions, for example smaller and economically stronger groups versus larger and economically weaker ones. We detected some single conditions consistently related with success, especially the characteristic of being organisations that source their raw material from members or from spot markets, and could triangulate these with logistic regression. The grants to the best-endowed organisations appeared to have been consistently unsuccessful, most likely because the grant amount was limited and caused under-scaled investments in secondary business activities.

The paper proceeds as follows. First, we discuss the method of Qualitative Comparative Analysis in more detail. Second, we describe the context, background and rationale of the FONDOECAS small-grant support fund. We describe the construction of the data set (26 grant beneficiaries), the way that we calibrated the fuzzy-set conditions and the outcomes of the grants. Fourth, we present the results of the QCA analysis and discuss the suggested causal configurations of conditions that could explain/predict success or failure of the grant. We check the stability of the QCA results adapting consistency thresholds, and using the ambiguity in outcome evaluation. Fifth, we triangulate the results of the QCA with those of binary logistic regression. We finish with a discussion on the results and the usefulness of QCA as a method in impact evaluation.

## 7.2 Qualitative Comparative Analysis

Qualitative Comparative Analysis (QCA) has been developed by the political scientist Charles Ragin (1987; 2000; 2008). It is used to explore configurations of factors within a data set, that are related to the presence or absence of an outcome condition. Cases share certain attributes, called conditions, and each case is successful or not according to an outcome condition. Conditions can be ‘crisp-sets’, with the value 1 to denote presence of the condition and 0 to denote absence, or ‘fuzzy-sets’, with scores between 0 and 1, which denote partial membership of the case in the condition. The data set of observations is a matrix, with the cases in rows and the conditions in columns, similar to the data set used in statistical software.

The conditions used in the QCA analysis are expected to have explanatory power, be they as single conditions or as part of a configuration of conditions. QCA searches the data set for possible causal relations. This process of explorative analysis, looking at the outcomes to derive hypotheses about explanatory models, is called abduction (Reichertz, 2004; Minnameier, 2010) or ‘retroduction’ (Ragin, 2008; Rihoux and Lobe, 2009).

Abduction is a strategy that seeks satisfactory explanations of observed phenomena that can be adopted as new hypotheses and worthy candidates for further investigation (Douven, 2011; Peirce et al., 1935). By revealing patterns of associations across cases in a data set, QCA generates hypotheses about possible causal relations (Schneider and Wagemann, 2012; Legewie, 2013). QCA is an exploratory tool, and in many respects similar to the iterative use of the technique of cluster analysis in statistics, where cases are grouped according to their similarities in a set of variables. In econometrics, it is similar to the practice of model fitting, looking for possible explanations for observed patterns in data. And, just like model fitting, it bears the

risk of data fishing and ‘harking’ - hypothesis testing after the results are known (Kerr, 1998). Shadish rightfully warns that ‘many different models can fit a data set, so our confidence in any given model may be small.’ (Shadish et al., 2002). Therefore, QCA results need to be examined critically to prevent spurious causal explanations.

The QCA exploration for causal conditions starts with the data set of observations. This data set has an identifier (e.g., name of the organisation), several conditions and an outcome variable. Several cases may share the same set of conditions. Therefore, QCA creates an overview of all possible combinations of conditions (configurations) and the number of cases that share the same combination. This matrix is called a ‘truth table’. The origin of the truth-table matrix as a device to show all combinations of conditions involved in the causal explanation is ascribed to the 19th century American philosopher and logician Charles Sanders Peirce (Anellis, 2012). Each row of a truth table represents a logically possible combination of conditions. A complete truth table will possess  $2^k$  rows, where  $k$  equals the number of conditions. Not all rows in the truth table are necessarily covered by empirical cases. The rows that are not covered are called ‘logical remainders’.

The truth table is a ‘revealing data-display’ (Collier, 2014) that helps us to reflect about explanations for outcomes. One of the key assumptions in the interpretation of truth table rows is that cases with similar conditions behave in a similar manner. When the row is consistently related to the same outcome (presence or absence of success), it is a sufficient causal configuration. Each consistent row in the truth table is considered as a statement of causal sufficiency, called a ‘term’.

A QCA term can be written as:

$$\text{CONDITION}_1 * \text{condition}_2 * \text{CONDITION}_3 * \dots \rightarrow \text{Outcome} \quad (1)$$

where,  $\rightarrow$  denotes consistency,  $*$  is the Boolean logical operator AND, while UPPERCASE font indicates presence of the condition, while lowercase font indicates absence.

Before searching for the more parsimonious terms with the Quine-McCluskey algorithm, we first need to check for the presence of necessary conditions (Legewie, 2013; Rihoux and Ragin, 2009). Necessary conditions are ‘always’ present in all cases that share a certain outcome (be it success or failure) and therefore tend to be excluded in the Boolean minimisation as being redundant (Schneider and Wagemann, 2012: 221-225).

While inspection of a truth table is recommended practice (Byrne and Ragin, 2009; Collier, 2014), a much more contested feature of QCA is the use of the minimisations algorithm, which, using Boolean logic, distils causal ‘recipes’ from each configuration of conditions (rows) in the truth table. Consistency is the key criterion in QCA to decide on the strength of patterns in the data set. Consistency is the degree to which the empirical evidence supports the claim that the relation between conditions and outcome exists (Rihoux and Ragin, 2009: 183). In fuzzy-set QCA, some cases can be partial member of a condition. This implies that the causal relation might not be totally consistent, as the group of cases might have one or more cases that do not

share the conditions entirely. To define a causal statement as being ‘consistent’, most authors recommend to use a consistency score of at least 0.75. The consistency score is computed as the lowest of the membership score of the term in the set of conditions, or the membership score of the term in the set of the outcome (Ragin, 2008; Smithson and Verkuilen, 2006).

The consistency score ( $C$ ) of a group of  $n$  cases is:

$$C = \frac{\sum_1^n \min(\text{membership score in set of conditions}, \text{membership score in set of outcome})}{\sum_1^n (\text{membership score in set of conditions})} \quad (2)$$

As explained above, QCA considers each row with a proper threshold consistency score as a case-as-configuration, the bearer of a set of conditions that are sufficient for the outcome to occur. However, not all conditions and configurations are necessarily relevant for the causal explanation. Some may be trivial or redundant, while others may provide the clue for explaining a causal relation. Using Boolean logic, QCA searches for ‘simplest’ combinations of conditions that are still consistent with the outcome. This minimisation algorithm used in QCA is the Quine-McCluskey algorithm (McCluskey, 1956). It reduces the complex Boolean expressions of the rows in the truth-table into more parsimonious terms. We can write the resulting causal statements (the QCA solution) as a Boolean expression of these terms.

The QCA solution is written as:

$$\text{Term}_1 + \text{Term}_2 + \dots \rightarrow \text{Outcome} \quad (3)$$

where,  $\rightarrow$  denotes consistency, and  $+$  is the Boolean logical operator OR.

Various software packages provide the Quine-McCluskey algorithm (McCluskey, 1956) to do this truth-table minimisation, such as fsQCA 2.5 (Ragin and Davey, 2009), TOSMANA (Cronqvist, 2009), Kirq 2.1.12 (Reichert and Rubinson, 2014), the *fuzzy* command in STATA (Longest and Vaisey, 2008) and QCA for R (Duşa and Thiem, 2014). However, there are fierce debates about the appropriateness (Collier, 2014), validity (Lucas and Szatrowski, 2014) and robustness (Thiem, 2013; Krogslund et al., 2015) of ‘automatic’ truth table minimisation as a method of causal inference. Slight changes in data, e.g., leaving out one or more of the cases (Lieberson, 2004), the use of different consistency thresholds and varying assumptions about ‘logical remainders’ have implications for the results of minimisation. Krogslund et al. (2015) replicated the QCA analysis of three prominent studies and showed that their results were not stable. They showed that QCA may feed a preliminary analysis about possible causal configurations but that consistent causal terms derived with QCA minimisation alone cannot be considered strong evidence to prove the existence of these causal factors. Lucas and Szatrowski (2014) and Krogslund et al. (2015) show with simulation data that the QCA minimisation can easily produce Type I errors: finding causal configurations while these are just random patterns in the data. Thiem (2013) stresses the need for more extensive robustness checks to accompany a QCA analysis.

Where Thiem and Krogslund et al. have constructive criticism, with a view to improving the procedure of case-based comparative analysis, the critique from Lucas and Szatrowski is more devastating. They attack the dominant positive image of QCA as a *useful case-oriented method*

*for comparative analysis of asymmetrical causal relations*, and reverse this wording into “QCA is actually a self-contradictory, cell-oriented, non-comparative, non-analytic means to identify asymmetric causal illusions” (Lucas and Szatrowski, 2014). Their critique is valid when QCA is used mechanically, but we do not agree with their verdict on QCA as a research approach. Most QCA analysts will not accept the results of the QCA solution without a thorough reflection on the process and mechanisms of change. Ragin (2008; Ragin, 2014) and other scholars who developed QCA (Schneider and Wagemann, 2012; Rihoux and Ragin, 2009) have always warned against an uncritical interpretation of the results of QCA. They situate QCA in a cautious process of qualitative analysis, and an iterative process of analysis and interpretation of results in view of refining the understanding of the cases and the potential causal process suggested in the solutions.

The fsQCA-software application (Ragin and Davey, 2009) is partly to blame for uncritical and mechanical use. As results of the Boolean minimisations, fsQCA 2.5 automatically generates three different solutions, which differ according to the inclusion or not of the information in the logical remainder rows of the truth table. The complex solution only uses the rows of the truth table that have empirical cases, the parsimonious and intermediate solutions also include the ‘empty’ rows (logical remainders). The parsimonious and intermediate solutions differ in the simplifying assumptions used, assumptions about the causal direction of the conditions involved in these logical remainders. Rubinson points to a shortcoming in fsQCA 2.5 which makes that the choice for inclusion or not of a row has to be made by the researcher without direct/explicit reference to the ‘names’ of the empirical cases that are affected by these decisions. As a result, Rubinson indicates that the software fsQCA lost a key feature, present in the earlier crisp-set version of the software, and necessary for a proper QCA, as emphasised by Ragin himself (1987: 113). “To follow the case-oriented approach, then, is to treat any specification of relevant causal conditions as tentative and to use theoretical and substantive knowledge to achieve a proper specification of causal conditions before reducing the truth table”. The new QCA application Kirq (Reichert and Rubinson, 2014) facilitates this necessary reflection on consistent and inconsistent cases in the truth table rows. Researchers can, therefore, better apply their substantive knowledge about the cases to interpret the consistency of a row. For each truth table row and term in the solution, Kirq 2.1.12 gives the identifier of the case that is consistent or inconsistent in the outcome condition.

### 7.3 FONDOECAS

We applied QCA on data from an impact evaluation of a Bolivian grant fund. The grant fund FONDOECAS was started in 2006 and by 2010, it had allocated 130 grants to the same number of farmer groups, from a population of 400 possible grantees organisations. In 2010, we started a research project to capture and assess the outcomes of the fund (Ton, 2010a). In 2010, we designed our study in a way that would create a promising context for applying QCA.

FONDOECAS provided a relatively uniform and replicable ‘treatment’, a grant of USD 10,000 to invest in processing or collective marketing activities. Moreover, we had a group of organisations that, in spite of their specificity, shared distinctive characteristics, all of them being member-based rural organisations with a legal status. The diversity of baseline characteristics of

the beneficiaries made it clear that statistical power of any sample, even when doing a census, would be too low to result in quantitative impact estimates with a quasi-experimental design.

The preparation of the impact evaluation coincided with a discussion on the findings of an external evaluation of the pilot phase of the fund (Prudencio, 2010), which suggested that FONDOECAS should focus more on smaller and less-developed organisations, and introduce credit as an additional support service, alongside grants. The suggestion by the external evaluator to focus on small and new organisations was understandably contentious, as the larger and stronger organisations did not want to lose this facility. The idea was also contested by leaders in CIOEC, who preferred to work with organisations that had already proven capable of organising their economic activities, rather than inexperienced newcomers (Pardo, 2010). It was also contrary to the spirit of discussions held during the design phase in 2005/2006 (Ton, 2005), in which the grant was presented as a novel institutional arrangement to resolve specific bottlenecks in relatively strong, well-functioning organisations to access new markets, for example, to help them with the investments needed to comply with the quality requirements in government procurement programmes.

This discussion on the external evaluation influenced our research design. The core question on effectiveness became: For what type of organisations, under what type of conditions, did FONDOECAS result in positive outcomes? The research aimed to generate recommendations that would be useful to the managers of the FONDOECAS small-grant fund, to be worded as follows: “When your objective is to create [intended outcome], based on the available monitoring information, we would suggest that you focus the support [allocation mechanism] on these types of organisations [eligibility criteria] with these characteristics [baseline conditions].” This answer implied the need for an explorative analysis to detect multiple impact pathways, typically the strength of QCA.

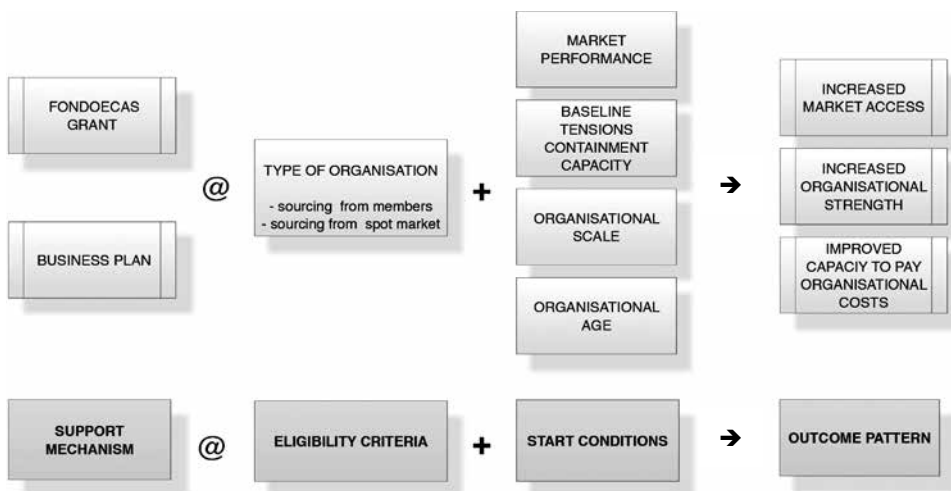
## 7.4 Conditions used in QCA

The organisations receiving the grant and the context in which these grants are used are diverse. They work in different sectors, vary in size, in age, in patrimony, baseline turnover, gender composition, geographical location, legal format, etc. The number of conditions used to describe the characteristics of each case is large. For a meaningful analysis of causal pathways, however, we needed to restrict the number of conditions. As explained above, the central feature of QCA is the truth table, which consists of  $2^k$  rows, all possible combinations of conditions. An analysis with two conditions generates four rows. With three variables, this increases to 8 and with six variables it is 64 rows. With only 26 cases, a high number of conditions in the QCA model would lead to a situation in which most rows are unpopulated. Using simulated data sets with random data, Axel Marx (2006; 2010) shows that when the proportion of variables on cases in a crisp set QCA analysis is low, the results of the Boolean minimisation process may become unstable and trivial solutions are likely to appear. This means that, in crisp set QCA, it is advised, as good practice, to constrain an analysis of 26 cases to five or six conditions only. Fuzzy set QCA is likely to be at least as vulnerable as crisp set QCA. We opted therefore to use only four or five conditions in our analysis.

Based on the documents and discussions around FONDOECAS, we can distil some conditions that are mentioned as possible moderators (see Figure 7.1). First, there have been discussions about the eligibility of some of the grant recipients, because they had characteristics that deviated from the ‘ideal type’ economic farmer organisation. Economic farmer organisations are framed by CIOEC as rural membership organisations that sell or process member products. Some beneficiaries, however, do not sell products that they buy from members, but work more as micro-enterprise, processing inputs bought in the market, or from non-members. The effectiveness of the grant can be expected to vary according to this different relation between group and members. Therefore, we distinguished two types: groups that process member products and organisations that process non-member products.

We selected three start conditions for which organisations could be compared at the moment of deciding on the grant proposal, and that could be potential predictors of success. In Chapter 6, we describe our tool to assess the organisational social capital of collective marketing groups, focussing on their capabilities to manage the inherent governance challenges in collective marketing: ‘tension containment capacity’. Two other conditions are related to the economic and organisational performance of the organisations at the moment of granting: ‘market performance’ and ‘organisational scale’.

In Figure 7.1, we present the conceptual model with these constructs. The grant-supported business plan interacts (@) with the type of organisation (+) and contextual conditions present at baseline to cause (→) certain outcome patterns. Our quest is for configurations of conditions that may ‘predict’ success or failure of the grant. If we find them, FONDOECAS may use them to target the grant to organisations that seem more likely to become a success, adjusting the eligibility criteria of the grant fund. None of the conditions is expected to be necessary or sufficient on its own, but they may be part of a causal configuration of conditions.



Source: Author's own elaboration

Figure 7.1 Conditions used in the QCA model

This type of causal conditions is often called INUS-condition (Mackie, 1965; Mahoney, 2008; Shadish et al., 2002): each condition is insufficient but a non-redundant part of a larger configuration of conditions that is unnecessary but sufficient to cause an outcome.

## Fuzzy-set calibration

In QCA, each condition takes a value between 0 and 1, describing the (partial) membership of a case in the group of cases that shares a specific condition. Fuzzy-sets make it possible to have cases that are ‘nor completely in, nor completely out’ of the set-condition (see Table 7.1). Fuzzy-sets often better represent the reality in the field with some cases that are difficult to classify under a specific condition. In our model, Market performance, Baseline tension containment capacity and Organisational scale are three constructs that are better represented in fuzzy-sets than in crisp-sets.

**Table 7.1** Verbal description of fuzzy-membership scores

Fuzzy value	The case is....
1	Fully in
0.9	Almost fully in
0.8	Mostly in
0.6	More in than out
0.5	Crossover: neither in nor out
0.4	More out than in
0.2	Mostly out
0.1	Almost fully out
0	Fully out

Source: Schneider and Wagemann (2012: 29)

There are several variables that could represent market performance of an organisation, be it total sales or sales per member. Moreover, group sales, patrimony and membership are informative for an assessment of the scale of an organisation. As explained above, however, we needed a reduced number of conditions for the QCA analysis. To limit the number of conditions in the QCA model, we used a Principal Component Analysis to distil two factors that could proxy for Market performance and Organisational scale. These factors were derived from the four performance variables: group sales, turnover per member, patrimony and membership (see Table 7.2). We normalised each variable using a natural log transformation. The Principal Component Analysis with Varimax rotation resulted in two factors, which together explained 81% of the variance. The first factor was defined by group sales and group sales per member. The second factor is dominated by patrimony and membership.

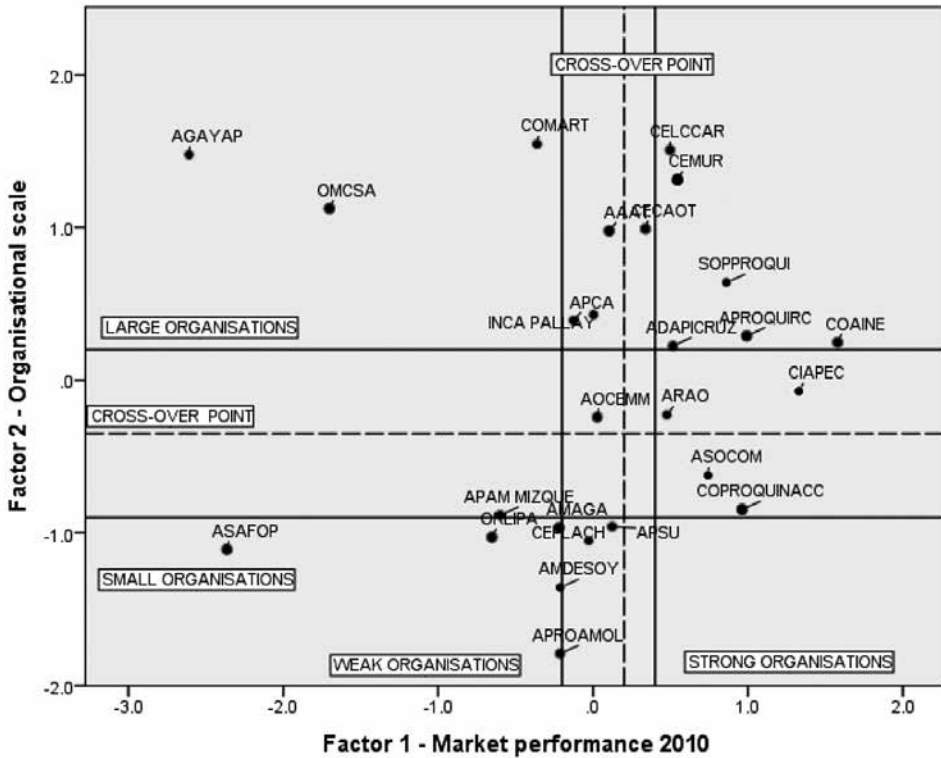
**Table 7.2** Rotated component matrix of Principal Component Analysis on four performance indicators (2010 data)

Component	Factor 1 “Market performance”	Factor 2 “Organisational scale”
Proportion of variance explained	53%	28%
Group sales (Ln)	.930	.289
Turnover per member (Ln)	.990	.004
Patrimony (Ln)	.166	.741
Membership (Ln)	.049	.874

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.



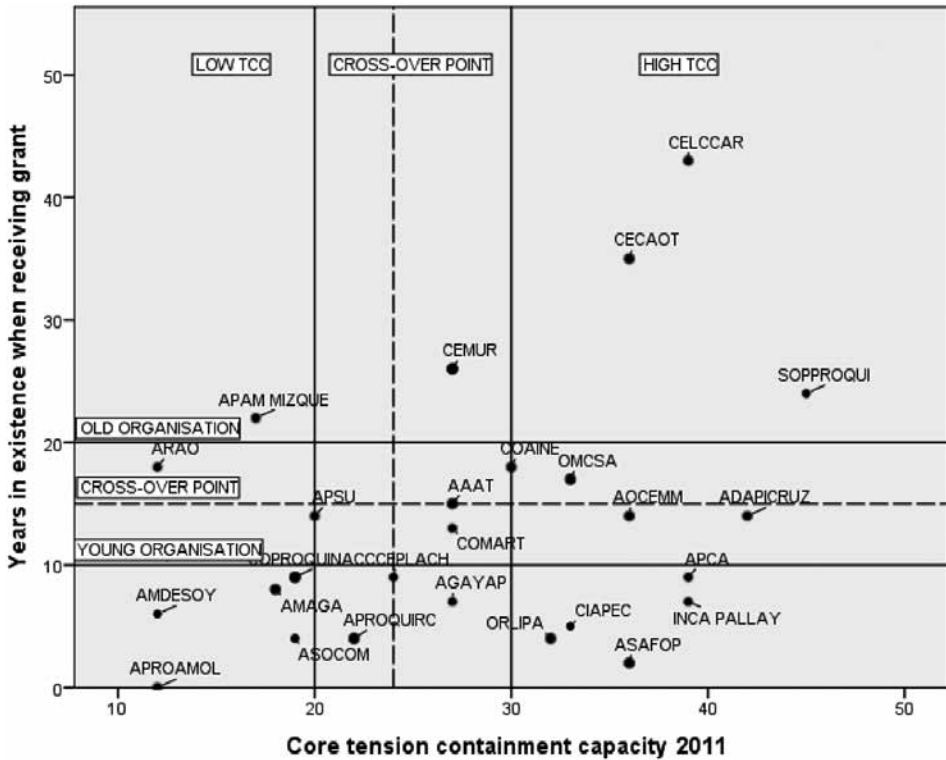


Source: Author's own elaboration

Figure 7.2 Scatter plot for the calibration of the fuzzy sets 'Market performance' and 'Organisational scale'

To make the variable suitable for QCA analysis, these principal components need to be transformed into fuzzy sets. The fuzzy-set score of a case represents the membership score of that case in the set. To calibrate the set, we used the 'direct method of calibration', as recommended by Rihoux and Ragin (2009) and provided in the software that is market leader for this analysis, fsQCA 2.5 (Ragin and Davey, 2009). Based on three thresholds, or 'qualitative anchors' (Schneider and Wagemann, 2012), a fuzzy-set variable can be computed with values between 0 and 1, with 0.5 as the cross-over point. The fsQCA 2.5 software applies a logistic function to calculate the continuous fuzzy-set scores between the cross-over point and the threshold cases that are definitely in or definitely out.

The direct method of fuzzy-set calibration (Ragin, 2008) is transparent but explicitly normative. For each set, the organisation that is considered definitely in and out had to be defined based on substantive knowledge and/or existing theory. We used scatter plots to identify the organisations that could function as appropriate qualitative anchors for defining the fuzzy-set scores. These qualitative anchors were selected primarily based on the substantive knowledge of the author on a fair number of organisations in the sample, having worked with many of them during 1999 and 2004 when employed by the national platform of economic farmer organisations CIOEC-Bolivia.



Source: Author's own elaboration

**Figure 7.3** Qualitative anchors used to calibrate the fuzzy set 'Organisations with high baseline tension containment capacity' and 'Old organisations'

Figure 7.2 shows the qualitative anchors used to define the fuzzy-set scores of two conditions. We selected CEMUR as an organisation that was definitely a member of the set of organisations with 'Strong market performance'. We considered APSU as closest to the cross-over point and AMDESOLY as the organisations that had to be considered as 'definitely weak'.

On the second condition, 'Large organisational scale', we considered ADAPICRUZ as definitely large, ARAO as the cross-over point and COPROQUINACC as definitely small scale. Table 7.3 describes the main characteristics that motivated the selection of these cases as qualitative anchors for calibration of the fuzzy sets.

The construct Baseline tension containment capacity was based on information collected through interviews with board members of the farmer groups. The interviewed board members and the local researchers reviewed the organisational dynamics in each of the organisations in the last few years for ten areas in which agency dilemmas in collective marketing tend to be present (Ton, 2010b). The interviews were summarised using a fixed format, and a 'core tension-containment capacity score' was derived from this information (see Chapter 6 of this thesis). For each agency dilemma, the status of the rules and regulations to resolve the inherent tensions between group and member was assessed on two aspects: the relevance of

**Table 7.3** Motivation to select cases as qualitative anchors for the calibration of fuzzy-sets

Fuzzy-set condition	Name	Scale score	Fuzzy score	Motivation
<b>Organisations with high baseline tension containment capacities (HIGHTCCBASE)</b>				
- Definitely high (HIGHTCC)	COAINE	30	0.95	COAINE is one of the oldest cooperatives in Bolivia. It manages several coffee processing facilities for pulping and removal of mucilage and parchment. COAINE employs four permanent staff and around 30 persons who work in the drying and parchment centres
- Cross-over	CEPLACH	24	0.50	CEPLACH is a small women's association of dairy processors, legally founded in 2001. It specialises in the production of yoghurt and cheese, when there is a sales opportunity for its products. Its main objective is the generation of part-time employment and complementary cash income for the female worker-members. However, CEPLACH also creates market access for several of the members who supply milk.
- Definitely low (high tcc)	APSU	20	0.05	APSU is a small handicraft organisation located near the border of Chile, specialised in alpaca weavings with a membership that declined from 60 in 2010 to 32 households in 2012. It sells most of its member products in an alliance with the federation COMART, which manages a shop in La Paz.
<b>Strong market performance (STRONGSALES)</b>				
- Definitely strong (STRONGSALES) CEMUR 0.4			0.95	CEMUR is an association of women's groups organised around capacity building and business development. It manages collective production units to sell the products to their members. In 2010 they had a turnover of US\$ 142,857, which means US\$952/member.
- Cross-over	APSU	0.2	0.50	APSU is a small handicraft organisation located near the border of Chile, specialised in alpaca weavings. It sells its products in an alliance with the federation COMART, which manages a shop in La Paz. Their sales in 2010 totalled US\$18,571, or US\$413/member.
- Definitely weak (strongsales)	AMDESOY	-0.2	0.05	AMDESOY started in 2005 and is a women's group that creates products from soy-meal. The member-workers sell the products directly to consumers in Santa Cruz with a system of door-to-door sales. Their annual sales in 2010 were less than US\$3,749, which is US\$187/member.
<b>Large organisational scale (LARGESCALE)</b>				
- Definitely large (LARGESCALE)	ADAPICRUZ	0.4	0.95	In 2010, the honey processor ADAPICRUZ had 150 members, based in the city of Santa Cruz de la Sierra. They manage a processing unit in Sta. Cruz worth around US\$140,000.
- Cross-over	ARAO	-0.2	0.50	ARAO, formed in 1983 with legal recognition in 1990, is specialised in the production of carpets, sweaters, shawls and ponchos, and has its own shop in Oruro for distribution. In 2010, ARAO had a membership of 90 members and patrimony of US\$92,000
- Definitely small (largescale)	COPROQUINACC	-0.9	0.05	COPROQUINACC-T is the smallest of 12 regional organisations that form the national quinoa federation ANAPQUI. It started in 1998 and by 2010, it grouped 60 producers. Only in 2006 did they manage to obtain legal status with the 60 members. They had a patrimony of US\$44,000.
<b>Organisational age (OLDAGE)</b>				
- Definitely old (OLDAGE)	APAM MIZQUE	22	0.99	Since 1988, APAM MIZQUE started to produce honey in the area, with help from an NGO. They sell the honey in the consumer market in Cochabamba.
- Cross-over	AAAT	15	0.5	AAAT is a handicraft organisation that uses wool from its 150 members living in a remote area, to produce products for the high-end market. The organisation started in 1992 supported by an NGO. In 2000 they became independent from the NGO but continued to receive support from development cooperation.
- Definitely young (oldage)	CEPLACH	9	0.03	CEPLACH is a small women's association of dairy processors, founded in 2001. It specialises in the production of yoghurt and cheese, whenever there is a sales opportunity for the products. Its main objective is the generation of part-time employment and complementary cash income for the female worker-members.

the issue in the organisations ('the tension comes up / hardly comes up / never comes up') and if the issue was resolved or not (resolved, trying to resolve, no need to resolve). Five of these agency dilemmas proved to be core to most organisations in Bolivia, which resulted in a so-called 'core tensions containment score' as the construct to measure organisation strength.

To select the qualitative anchors for the fuzzy-set condition 'High tension containment capacity', we plotted their respective core tensions containment scores for 2011 (Figure 7.3). We qualified COAINE as definitely in, CEPLECH as cross-over point and APSU as definitely out of the group of organisations with high baseline tension containment capacities (see Table 7.3 for more detail). In the same figure, we added the age of the organisation, because this variable is used in some of our QCA analysis as a complementary start condition.

### Crisp-set outcomes

To derive the crisp-set scores of cases in the outcome condition, that is, their 'membership in the group of organisations with a successful outcome of the grant', we used the time-series data and tension containment scores plus the qualitative information about dynamics and processes related to the grant and grant-supported business plan. A case-by-case interpretation of the information was necessary, because the performance indicators between 2008 and 2012 could not be interpreted directly/mechanically to assess success or failure of the grant. Proxy-indicators such as changes in turnover or membership had changed largely in response to dynamics in the traditional business activities of the organisation, whereas the grant only tackled a constraint in a specific business segment, often only one of multiple activities. Therefore, we could not determine the success or failure of a grant by simply subtracting these overall latter indicators of performance from those before the reception of the grant. For example, the contribution of the grant to increased market access was different for large coffee exporters that invested the grant in a pilot roasting machine, as a future complementary activity, than in small dairy plants or honey processors that invested in equipment that improved uniformity and quality of their product, often a mandatory requisite for access to government procurement markets. Therefore, to define if an organisation attained a certain outcome as a result of the grant, we had to infer the effectiveness of the grant through a case-by-case analysis of the processes and dynamics that were set in motion by the grant.

In this interpretative analysis of the change dynamics in each organisation, we applied counterfactual thinking (Vellema et al., 2013). We asked ourselves the question: Would the outcome have been achieved even without the grant? In Annex 1 we describe, for each of the 26 cases, the dynamics in the organisation and the reasons for classifying a case as successful or unsuccessful in outcome.

This classification of success or failure implies interpretation involving normative decisions based on the available information. We classified a case as successful or unsuccessful after reconciling the independent evaluation by the two main researchers, GT, the author of this paper, and LF, the local researcher who had done all 2013 interviews. GT had more knowledge on the performance indicators and the differences between the baseline and end-line interview reports. LF was more knowledgeable on the organisational dynamics, because she had additional information and impressions resulting from the actual interviews. Both quasi-independent

judgements were reconciled in a discussion in January 2015. The final, reconciled outcome classification was agreed upon between the two researchers. Table 7.4 shows the agreement in these evaluations. Agreement between both researchers was ‘moderate’. Cohen’s kappa scores are ‘substantial’ for the outcomes enhanced market access and improved organisational capacities, and ‘fair’ for the increased capacity to pay organisational expenses (Landis and Koch, 1977).

**Table 7.4** Cohen’s kappa scores of agreement in valuations<sup>a,b</sup>

Outcome	LF original versus reconciled	GT original versus reconciled	LF original versus GT original	Arguments used for reconciliation
Grant enhanced market access	0.75	0.70	0.48	One difference (AAAT) was due to an erroneous interpretation of the use of the investment in the shop. Another was due to considering different time intervals and grant investments (CECAOT). Also, a dairy plant (CEPLACH) appeared to have several milk-producing members who sold to the group but the grant served to build and relocate the place of operation, which negatively affected sales. ORLIPA accessed the local school food programme but appeared to have done so without the products from the grant-supported business plan.
Grant improved organisational capacities	0.83	0.59	0.41	Two cases (CELCCAR, CIAPEC) did not use the grant investment, though they continued with the supported business plan. In two cases the interviewees expressed the importance of the decision-making process around the grant to discuss internal group pressures, even though the grant did not contribute to production.
Grant increased capacity to pay organisational expenses	0.26	0.92	0.22	We noted a difference in interpretation of the question between the two researchers. During reconciliation, it was agreed that the capacity to pay expenses will increase when the level of sales increases due to the grant, even though in most organisations the total amount of expenses or member income did not change.

<sup>a</sup>. Interpretation of Cohen’s kappa: <0.00=Poor agreement; 0.00–0.20=Slight agreement; 0.21–0.40=Fair agreement; 0.41–0.60= Moderate agreement; 0.61–0.80=Substantial agreement; 0.81–1.00=Almost perfect agreement (Landis and Koch, 1977)

<sup>b</sup>. GT used fuzzy scores for some of the valuations. These were converted to crisp scores before calculating the Cohen’s kappa

The differences in the evaluation of the outcomes by the two researchers (Table 7.5) provided an opportunity to reflect on the stability of the QCA-solution under real-world conditions of measurement error. Implicitly, this reconciliation implied that one of the researchers had had a ‘measurement error’ when evaluating the respective organisation. The source of this measurement error differed according to the case, but was mostly due to missing information on the way the grant had been invested or additional, non-recorded information on organisational dynamics. Without this reconciliation process, the QCA would have proceeded with the evaluations of GT only. This makes it possible to assess the robustness of the QCA analysis in function of GT’s initial measurement error. Therefore, we will use the differences in the authors’ evaluation of the outcomes in each organisation (‘GT original’ versus ‘reconciled’) to verify the stability of the terms in the QCA solution.

**Table 7.5** Evaluation of the contributory role of the grant in three outcome areas

	Positive outcome market access			Positive outcome organisational strengthened			Positive outcome capacity to pay organisational expenses		
	LF original	GT original	Reconciled	LF original	GT original	Reconciled	LF original	GT original	GT + LF Reconciled
AAAT	0	1	0	1	0	1	0	1	1
ADAPICRUZ	1	1	1	1	0	1	0	1	1
AGAYAP	0	0	0	0	0	0	0	0	0
AMAGA	0	0	0	0	1	1	0	1	1
AMDESOY	0	0	0	0	0	0	0	1	1
AOCEMM	1	1	1	1	1	1	1	1	1
APAM MIZQUE	0	0	0	0	0	0	0	0	0
APCA	.	1	1	.	0	0	.	0	0
APROAMOL	0	1	0	1	1	1	0	1	1
APROQUIRC	1	1	1	1	1	1	0	1	1
APSU	0	0	0	0	0	0	0	0	0
ARAO	1	1	1	1	1	1	0	1	1
ASAFOP	0	0	0	1	0	1	0	1	1
ASOCOM	0	0	0	0	0	0	0	0	0
CECAOT	1	0	0	1	0	0	0	0	0
CELCCAR	0	0	0	0	1	0	0	0	0
CEMUR	0	0	0	0	0	0	0	1	1
CEPLACH	1	0	0	0	0	0	1	1	1
CIAPEC	0	0	0	0	1	0	0	0	0
COAINE	0	0	0	0	0	0	0	0	0
COMART	0	0	0	1	1	1	0	1	0
COPROQUINACC	0	0	0	0	0	0	0	0	0
INCA PALLAY	0	0	0	1	1	1	0	0	0
OMCSA	0	0	0	0	0	0	1	1	1
ORLIPA	0	1	0	0	0	0	0	0	0
SOPPROQUI	0	0	0	0	0	0	0	0	0
Success Rate	6/25	8/26	5/26	10/25	9/25	10/26	3/25	13/26	12/26

**Table 7.6** Organisations with consistent membership in each combination of conditions

row	Start conditions			Type of org.	N	Organisations that are member in each row
	HIGH TCC	STRONG SALES	LARGE SCALE	SOURCING		
1	TRUE	TRUE	TRUE	TRUE	6	ADAPICRUZ;CECAOT;CELCCAR;CIAPEC;COAINE;SOPPROQUI
2	TRUE	TRUE	TRUE	FALSE	1	CEMUR
3	TRUE	TRUE	FALSE	TRUE	0	-
4	TRUE	TRUE	FALSE	FALSE	0	-
5	TRUE	FALSE	TRUE	TRUE	6	AAAT;AGAYAP;AOCEMM;APCA;COMART;INCAPALLAY
6	TRUE	FALSE	TRUE	FALSE	1	OMCSA
7	TRUE	FALSE	FALSE	TRUE	1	ORLIPA
8	TRUE	FALSE	FALSE	FALSE	1	ASAFOP
9	FALSE	TRUE	TRUE	TRUE	2	APROQUIRC;ARAO
10	FALSE	TRUE	TRUE	FALSE	0	-
11	FALSE	TRUE	FALSE	TRUE	1	COPROQUINACC
12	FALSE	TRUE	FALSE	FALSE	1	ASOCOM
13	FALSE	FALSE	TRUE	TRUE	0	-
14	FALSE	FALSE	TRUE	FALSE	0	-
15	FALSE	FALSE	FALSE	TRUE	4	APAMMIZQUE;APROAMOL;APSU;CEPLACH
16	FALSE	FALSE	FALSE	FALSE	2	AMAGA;AMDESOY

Source: Original data, analysed using Kirq 2.1.12.

Cases are considered part of the configuration when consistency score of the case in the set  $>0.5$ .

## Description of the data set

Table 7.6 reflects the distribution of the 26 organisations according to the configuration of contextual conditions in which they have their highest membership score. The empty rows reflect the ‘limited diversity’, which is inherent to most social research: often, not all possible combinations of conditions can be observed in the real world. In QCA, these empty rows are called ‘logical remainders’.

Within the sample of grant beneficiaries, there is a group of six organisations that do not share any of the three start conditions. These organisations are relatively small, weak and had limited market performance around the time that they received the grant. Two of them are non-sourcing organisations (row 16), and the other four are classified as sourcing organisations (row 15), though CEPLACH has only partial membership with a fuzzy set score of 0.7. In contrast, seven organisations were strong on all three contextual conditions, one of them non-sourcing (the women’s organisation CEMUR), while five of the other six are quinoa- or coffee-bulking organisations.

The rows in the truth table are generalisations from more than one case. Some organisations have fuzzy-set scores between 0.25 and 0.75, which indicates ambiguity of membership in the row. With slightly changed qualitative anchors, these organisations would have changed

**Table 7-7** Data set of observations with membership scores of the cases in the fuzzy-set conditions

No	NAME	Outcomes										Conditions									
		Market access to 7.r member products		Improved organisational capacities		Increased income to pay organisational costs		SOURCING member products		HIGHTCC High tension containment capacity (2011)		STRONGSALES Strong marketing performance (2010)		LARGESCALE Large scale (2010)		ORGADAGE Organisational age when receiving grant					
		crisp set score	fuzzy set score	crisp set score	fuzzy set score	crisp set score	fuzzy set score	crisp set score	fuzzy set score	TCC score	fuzzy set score	PCA Factor 1 score	fuzzy set score	PCA Factor 2 score	fuzzy set score	scale	fuzzy set score				
1	AAAT	0	1	1	1	1	1	0	0	27	0.77	0.1	0.33	0.98	1	15	0.5				
2	ADAPICRUZ	1	1	1	1	1	1	1	1	42	0.77	0.51	0.99	0.23	0.96	14	0.35				
3	AGAYAP	0	0	0	0	0	0	1	1	27	0.77	-2.61	0	1.48	1	7	0.01				
4	AMAGA	0	1	1	1	0.3	0.3	1	1	18	0.01	-0.22	0.04	-0.97	0.03	8	0.01				
5	AMDESJOY	0	0	1	1	1	1	1	1	12	0	-0.21	0.04	-1.36	0	6	0				
6	AOCEMM	1	1	1	1	1	1	1	1	36	1	0.03	0.21	-0.24	0.64	14	0.35				
7	APAM MIZQUE	0	0	0	0	0	0	0	0	17	0.01	-0.6	0	-0.88	0.05	22	0.99				
8	APCA	1	0	0	0	0	0	0	0	39	1	-0.13	0.08	0.39	0.98	9	0.03				
9	APROAMOL	0	1	1	1	1	1	1	1	12	0	-0.21	0.04	-1.79	0	0	0				
10	APROQUIRC	1	1	1	1	1	1	1	1	22	0.14	0.99	1	0.29	0.97	4	0				
11	APSU	0	0	0	0	0	0	0	0	20	0.05	0.12	0.36	-0.96	0.03	14	0.35				
12	ARAO	1	1	1	1	0	0	0	0	12	0	0.48	0.98	-0.23	0.66	18	0.86				
13	ASAFOP	0	1	1	1	0	0	0	0	36	1	-2.36	0	-1.11	0.02	2	0				
14	ASOCOM	0	0	0	0	0	0	0	0	28	0.86	0.74	1	-0.62	0.18	4	0				
15	CECAOT	0	0	0	0	0	0	0	0	36	1	0.34	0.89	0.99	1	35	1				
16	CELCCAR	0	0	0	0	0	0	0	0	39	1	0.5	0.99	1.51	1	43	1				
17	CEMUR	0	0	0	0	1	1	1	1	27	0.77	0.55	0.99	1.31	1	26	1				
18	CEPLACH	0	0	0	0	1	1	0.7	0.7	24	0.35	-0.03	0.15	-1.05	0.02	9	0.03				
19	CIAPEC	0	0	0	0	0	0	0	0	33	0.99	1.33	1	-0.07	0.82	5	0				
20	COAINE	0	0	0	0	0	0	0	0	30	0.95	1.58	1	0.25	0.96	18	0.86				
21	COMART	0	1	0	0	0	0	0	0	27	0.77	-0.36	0.01	1.55	1	13	0.23				
22	COPROQUINACC	0	0	0	0	0	0	0	0	19	0.03	0.96	1	-0.85	0.06	9	0.03				
23	INCA PALLAY	0	1	0	0	0	0	0	0	39	1	0	0.19	0.43	0.99	7	0.01				
24	OMCSA	0	0	1	1	1	1	1	1	33	0.99	-1.7	0	1.12	1	17	0.77				
25	ORLIPA	0	0	0	0	0	0	0	0	32	0.99	-0.65	0	-1.03	0.02	4	0				
26	SOPPROQUI	0	0	0	0	0	0	0	0	45	1	0.86	1	0.64	1	24	1				



membership in a specific truth table row. We will pay special attention to these ambiguous organisations. As Table 7.7. shows, AMAGA and CEPLACH are ambiguous in sourcing from members; AAAT, ADAPICRUZ, CMUR, CEPLACH and COMART are ambiguous when looking at their tension containment capacities, AAAT and APSU when looking at sales performance, and AOCEMM and ARAO when considering their organisational scale. Annex 1 provides more qualitative detail about each of the organisations.

## 7.5 Truth table analysis

The data set that resulted after the set calibration of context and outcome conditions (Table 7.7) is the input for the Qualitative Comparative Analysis. We used of Kirq 2.1.12 (Reichert and Rubinson, 2014) as the preferred computer interface for truth table analysis, as it makes it easier to reflect on the empirical cases. Kirq 2.1.12 lists consistent and inconsistent observations, a feature to help decision making about the inclusion of each row in the QCA minimisation.

The consistency score is computed as the lowest score of the membership of the row in the set of conditions, or the membership of the row in the set of the outcome (Ragin, 2008; Smithson and Verkuilen, 2006). In our case, with a crisp outcome and fuzzy-set conditions, we can easily separate both aspects of consistency. Because the outcome variables are crisp sets, an organisation will always be fully consistent with either the negative or the positive outcome. The proportion of cases that are consistent or inconsistent can thus be computed. Whenever the consistency score is lower than the proportion of consistent cases, this is necessarily the result of the lower consistency of the configuration of conditions. We may use this additional information provided in Kirq when we need to determine if a configuration must or must not be considered as being a ‘sufficient explanation’ for the outcome, and used as input in the minimisation process. We accept higher inconsistencies of the configuration of conditions than inconsistencies in the outcome. Therefore, where the proportion of cases that shared the same outcome is higher than the consistency score computed by the software, but at least 75%, we included the row in the minimisation process.

The reasons for success may be different from the reasons for failure. Both analyses can give us insights into which factors to take care of when better targeting the scarce resources of the FONDOECAS grant system. QCA explores configurations of conditions that can predict a positive outcome separately from the configuration of conditions that might predict a negative outcome. Based on the computed consistency scores and a reflection on the number of consistent and inconsistent observations, we classify whether a contradictory row can be considered as ‘probabilistically sufficient’ (Mahoney, 2008) for the positive outcome or for the negative outcome. Often the verdict will be symmetrical, a row is TRUE in one and FALSE in the other. However, contradictory rows may well be considered FALSE for both: they may be inconsistent and therefore insufficient explanations of either the positive outcome or the negative outcome.

In the following, we present truth tables related with the three intended outcome areas of FONDOECAS and reflect on possible explanatory factors that explain why the positive outcome of the grant was present or absent in each of the organisations.

## Outcome 1: Increased market access of members

### Conditions that predict success

The number of organisations that created market access with the grant-supported business plan is very small. Only five cases are classified as such. When we review the rows in the truth table that contains these successful cases, we see that they all share the conditions of being large-scaled organisations that source from members. The truth table also shows that these conditions are not sufficient for success, as many organisations with a similar configuration of baseline conditions proved unsuccessful (rows 1 and 5 in Table 7.8).

**Table 7.8** OUTCOME: Grant contributed to market access for members

Row #	Start conditions				Type of group	Consist with success	Suff. for success	Suff. for failure	Observations consistent with successful outcome	Observations consistent with unsuccessful outcome
	high tcc	strong sales	large scale	sourcing						
1	TRUE	TRUE	TRUE	TRUE		0.21	Con (FALSE)	Con (TRUE)	ADAPICRUZ	CECAOT; CELCCAR; CIAPEC; COAINE; SOPPROQUI
2	TRUE	TRUE	TRUE	FALSE		0.00	FALSE	TRUE	-	CEMUR
3	TRUE	TRUE	FALSE	TRUE		n/a	Rem	Rem	-	-
4	TRUE	TRUE	FALSE	FALSE		n/a	Rem	Rem	-	-
5	TRUE	FALSE	TRUE	TRUE		0.31	Con (FALSE)	Con (FALSE)	AOCEMM; APCA	AAAT; AGAYAP; COMART; INCAPALLAY
6	TRUE	FALSE	TRUE	FALSE		0.00	FALSE	TRUE	-	OMCSA
7	TRUE	FALSE	FALSE	TRUE		0.22	FALSE	TRUE	-	ORLIPA
8	TRUE	FALSE	FALSE	FALSE		0.00	FALSE	TRUE	-	ASAFOP
9	FALSE	TRUE	TRUE	TRUE		0.78	TRUE	FALSE	APROQUIRC; ARAO	-
10	FALSE	TRUE	TRUE	FALSE		n/a	Rem	Rem	-	-
11	FALSE	TRUE	FALSE	TRUE		0.19	FALSE	TRUE	-	COPROQUINACC
12	FALSE	TRUE	FALSE	FALSE		0.00	FALSE	TRUE	-	ASOCOM
13	FALSE	FALSE	TRUE	TRUE		n/a	Rem	Rem	-	-
14	FALSE	FALSE	TRUE	FALSE		n/a	Rem	Rem	-	-
15	FALSE	FALSE	FALSE	TRUE		0.01	FALSE	TRUE	-	APAMMIZQUE; APROAMOL; APSU; CEPLACH
16	FALSE	FALSE	FALSE	FALSE		0.00	FALSE	TRUE	-	AMAGA; AMDESOY

Source: Original data, analysed with Kirq 2.1.12.

FALSE = inconsistent configuration; TRUE = consistent configuration; Con = contradictory (coded FALSE or TRUE after reflecting the degree of inconsistency); Rem = logical remainder; observations are considered consistent when consistency score of the case is >0.5.

As a next step in the QCA analysis, we checked whether these conditions were necessary conditions (INUS conditions). To qualify as such, all cases with a positive outcome need to be a subset of the set characterised by this necessary condition. The consistency threshold used to qualify as necessary condition needs to be high, at least 0.90 (Legewie, 2013). The QCA anal-

ysis shows that this is the case only for sourcing (see Table 7.9). The condition of large scale is not consistent enough to qualify as necessary condition (consistency score = 0.84). ARAO is ambiguous on the fuzzy-set large organisational scale (fuzzy-set score = 0.66). With a slight change in the fuzzy set qualitative anchors of this fuzzy-set, it could have been classified as having a small scale.

**Table 7.9** QCA analysis of necessary conditions for success in market access of members

Term	Consist	Cov	Obs
SOURCING	1.00	0.25	ADAPICRUZ;AOCEMM;APCA;APROQUIRC;ARAO
LARGESCALE	0.84	0.27	ADAPICRUZ; APCA;APROQUIRC;ARAO

The inconsistent cases in the truth-table rows indicate that there are almost no configurations that can be considered sufficient for a positive outcome. Only one row, with two cases, is consistently related to a positive outcome (APROQUIRC and ARAO, row 9). This group shares all start conditions except high tension containment capacity. With only one truth table row, Boolean minimisation is not possible. The coverage of cases is low and the configuration is, therefore, an unlikely predictor of success.

### ***Conditions that predict failure***

We applied a minimum consistency score of 0.75 when deciding on inclusion of the row in the subsequent minimisation to explore for sufficient causes for success. The complex solution that results from the Boolean minimisations of these nine rows results in five terms (Table 7.10). Four of these terms cover highly consistent groups. Some of these groups have partial overlap in membership.

One group is characterised by having both high tension containment capacities at baseline, with strong sales and being large in scale. It comprises the three coffee exporters and the two largest quinoa exporters. Another group of eight organisations is characterised as being weak considering their tensions containment capacities and being small scale. But there is also another possible group, also with eight cases, that is characterised as having weak sales and being small scale. Six of the eight organisations overlap in membership and are classified in both groups.

**Table 7.10** Complex solution for failure in generating market access for members

Term	Consist	RawCov	UniqCov	ObsConsist	ObsInconsist
HIGHTCCBASE*strongsales*sourcing+	1.00	0.10	0.00	ASAFOP; OMCSA	-
HIGHTCCBASE*LARGESCALE*sourcing+	1.00	0.07	0.00	CEMUR; OMCSA	-
strongsales*largescale+	0.95	0.35	0.06	AMAGA; AMDESOY; APAMMIZQUE; APROAMOL; APSU; CEPLACH; ASAFOP; ORLIPA	-
HIGHTCCBASE*STRONGSALES*LARGESCALE+	0.81	0.29	0.24	CECAOT; CELCCAR; CIAPEC; COAINE; SOPPROQUI; CEMUR	ADAPICRUZ
hightccbase*largescale	0.95	0.35	0.10	AMAGA; AMDESOY; APAMMIZQUE; APROAMOL; APSU; CEPLACH; ASOCOM; COPROQUINACC	-
Solution	0.90	0.77			

When we apply the Boolean minimisation of the truth table including the ‘empirically empty’ logical remainder rows, we reduce these five terms of the complex solution to only three terms in the parsimonious solution (Table 7.11). One term comprises a group of organisations that have in common that they do not source from members. This is consistent with our earlier finding that being a sourcing organisation is a necessary condition for success. There is another term with organisations that are characterised as being small scale. This points to a plausible explanation and predictor of success; when market access of members is the goal, it seems wise to hesitate in allocating grants to organisations that are small in scale.

**Table 7.11** Parsimonious solution for failure in generating market access for members

Term	Consist	RawCov	UniqCov	ObsConsist	ObsInconsist
sourcing+	1.00	0.27	0.05	AMAGA; AMDESOY; ASAFOP; ASOCOM; CEMUR; OMCSA	-
largescale+	0.93	0.47	0.27	AMAGA; AMDESOY; APAMMIZQUE; APROAMOL; APSU; CEPLACH; ASAFOP; ASOCOM; COPROQUINACC; ORLIPA	-
HIGHTCCBASE*STRONGSALES	0.82	0.30	0.24	CECAOT; CELCCAR; CIAPEC; COAINE; SOPPROQUI; CEMUR	ADAPICRUZ
Solution	0.90	0.80			

One term groups the organisations with high tension containment capacities at baseline and strong sales. The result that high tension containment capacities, large scale and strong sales, predict failure of the grant is somewhat counterintuitive: Why would organisations that were already both organisationally strong and good performers be especially unsuccessful in generating positive outcomes with the grant? This made us reflect on the specific situation of this subgroup. The literature shows that many organisations in our sample had a history of working with NGOs that had invested large sums in these organisations (Healy, 2001; Flores et

al., 2007; Bebbington, 1996). Would this ‘easy money’ possibly have led to lower importance being attached to the business plan presented to FONDOECAS?

This reflection induced us to experiment with the inclusion of the age of the organisation as an additional explanatory variable or condition, a modification of our initial conceptual model. The inclusion of OLDAGE as complementary condition in the QCA-model resolved the inconsistent row 1 that involved ADAPICRUZ. The terms and groups in the parsimonious solution change due to the inclusion of this additional condition. With the inclusion of OLDAGE, the non-sourcing organisations disappear as term in the solution; the conditions that were most consistently related to the well-endowed group of organisations change into being old and with high tension containment capacities. The consistency of the solution is high with 0.94, and the coverage is slightly better, but coverage is slightly lower with 0.70 (Table 7.12). The characteristic of being old seems to be a more consistent condition and a better characterisation of this group of organisations than having strong sales or being large scale.

**Table 7.12** Parsimonious solution for failure in generating market access for members and OLDAGE as additional condition

Term	Consist	RawCov	UniqCov	ObsConsist	ObsInconsist
largescale+	0.93	0.47	0.46	AMAGA; AMDESOY; APAMMIZQUE; APROAMOL; APSU; CEPLACH; ASAFP; ASOCOM; COPROQUINACC; ORLIPA	-
OLDAGE*HIGHTCCBASE	0.89	0.30	0.29	CECAOT; CELCCAR; CEMUR; COAINE; SOPPROQUI; OMCSA	-
Solution	0.93	0.76			

### **Robustness**

Applying a higher consistency threshold (0.80 instead of 0.75), leads to the exclusion of truth table row 1 and 7 as sufficient causal configuration for failure. The parsimonious solution would only have two consistent groups (terms). One group is defined by the condition of being non-sourcing organisations, and covers six (27%) of the unsuccessful cases. The other group is characterised by small scale and low tension containment capacities at baseline, and covers eight (35%) of the unsuccessful cases. They partially overlap in membership.

The application of the higher consistency threshold removes the large group of older and stronger organisation from the solution of the initial QCA model. However, the higher consistency threshold does not change the parsimonious solution of the expanded model (with the age condition included), and the term improves in consistency but has a lower coverage (0.35). In this expanded model the older organisations with high tension containment capacities are consistently related with failure of the grant to contribute to market access of members.

**Table 7.13** Ambiguity in evaluation of the grant's contribution to improved market access of members

			Reconciled evaluation		Total (original)
			Unsuccessful	Successful	
GT original evaluation	Unsuccessful	Count	18	0	18
	Successful	Count	3	5	8
	Total (reconciled)	Count	21	5	26

As shown in Table 7.13, the measurement error in the evaluation of success on the outcome market access of members was relatively high. Three out of eight cases were 'downgraded' during reconciliation from successful to unsuccessful. All cases rated as unsuccessful remained so after reconciliation. The three organisations that changed as a result of the reconciliation are AAAT, APROAMOL and ORLIPA. AAAT and APROAMOL are members of rows that were not included in the truth table minimisation, nor would they have been included in it had they been classified as successful. Row 5 would have remained excluded in the minimisation even if AAAT had been classified as a success. The measurement error related to APROAMOL would have changed the consistency score of row 15 to 0.72. If we leave that row out of the minimisation (applying 0.75 as consistency threshold), the parsimonious solution would still result in a similar grouping of organisations. Neither the reclassification of ORLIPA from unsuccessful to successful does not change the final solution substantially. The parsimonious solution is exactly the same, though with slightly different consistency and coverage. All in all, the QCA analysis on the outcome market access of members proved robust to measurement errors and varying consistency thresholds.

## Outcome 2: Improve organisational capacities

### *Conditions that predict success*

Ten out of the 26 organisations registered a positive outcome in this area. Most organisations that were successful, however, are covered by a contradictory truth table row (row 1, 5, 15, 16) with a high proportion of inconsistent cases. The truth table shows only two rows that are possible sufficient causes for a successful outcome (ASAFOP in row 8, APROQUIRC and ARAO in row 9). Interestingly, these two groups differ in all the four baseline conditions. It is, therefore, clear that it is not possible to further reduce the truth table.

**Table 7.14** OUTCOME: Grant contributed to organisational strengthening

row	Start conditions			Type of group	Consist with success	Sufficient for success	Sufficient for failure	Observations consistent with successful outcome	Observations consistent with unsuccessful outcome
	high tcc	strong sales	large scale	sourcing					
1	TRUE	TRUE	TRUE	TRUE	0.28	Con (FALSE)	Con (TRUE)	ADAPICRUZ	CECAOT; CELCCAR; CIAPEC; COAINE; SOPPROQUI
2	TRUE	TRUE	TRUE	FALSE	0.01	FALSE	TRUE	-	CEMUR
3	TRUE	TRUE	FALSE	TRUE	n/a	Rem	Rem	-	-
4	TRUE	TRUE	FALSE	FALSE	n/a	Rem	Rem	-	-
5	TRUE	FALSE	TRUE	TRUE	0.57	Con (FALSE)	Con (FALSE)	AAAT; AOCEMM; COMART; INCAPALLAY	AGAYAP; APCA
6	TRUE	FALSE	TRUE	FALSE	0.04	FALSE	TRUE	-	OMCSA
7	TRUE	FALSE	FALSE	TRUE	0.22	FALSE	TRUE	-	ORLIPA
8	TRUE	FALSE	FALSE	FALSE	0.77	TRUE	FALSE	ASAFOP	-
9	FALSE	TRUE	TRUE	TRUE	0.91	TRUE	FALSE	APROQUIRC; ARAO	-
10	FALSE	TRUE	TRUE	FALSE	n/a	Rem	Rem	-	-
11	FALSE	TRUE	FALSE	TRUE	0.23	FALSE	TRUE	-	COPROQUINACC
12	FALSE	TRUE	FALSE	FALSE	0.04	FALSE	TRUE	-	ASOCOM
13	FALSE	FALSE	TRUE	TRUE	n/a	Rem	Rem	-	-
14	FALSE	FALSE	TRUE	FALSE	n/a	Rem	Rem	-	-
15	FALSE	FALSE	FALSE	TRUE	0.36	Con (FALSE)	Con (FALSE)	APROAMOL	APAMMIZQUE; APSU; CEPLACH
16	FALSE	FALSE	FALSE	FALSE	0.36	Con (FALSE)	Con (FALSE)	AMAGA	AMDESROY

Source: Original data, analysed with Kirq 2.1.12.

FALSE = inconsistent configuration; TRUE = consistent configuration; Con = contradictory (coded FALSE or TRUE after reflecting the degree of inconsistency); Rem = logical remainder; observations are considered consistent when consistency score of the case is  $>0.5$ .

### **Conditions that predict failure**

In the truth table (Table 7.14), we see that row 1 has one successful and five unsuccessful cases. Grants to this group of well-endowed organisations proved rather unsuccessful. The proportion of unsuccessful cases is above our threshold of 0.75, but the consistency score is slightly below our threshold of 0.75. When we included row 1, we had five rows that feed the minimisation into the complex solution, which resulted in a solution with four terms. The inclusion of the logical remainder rows in the minimisations did not make much of a difference; the parsimonious solution still has four groups that cover 63% of the unsuccessful cases. The terms in the solutions are very diverse; neither the complex nor the parsimonious solution points to any plausible configuration of conditions related with failure (Table 7.15 and Table 7.16). When we excluded row 1 from the minimisation, three of the four groups remained, with a coverage of 26% of the unsuccessful cases.

**Table 7.15** Complex solution after Boolean minimisation of the truth Table for failure in organisational strengthening

Term	Consist	RawCov	UniqCov	ObsConsist	ObsInconsist
HIGHTCCBASE*LARGESCALE*sourcing+	0.98	0.10	0.04	CEMUR; OMCSA	-
HIGHTCCBASE*strongsales*largescale*SOURCING+	0.78	0.09	0.07	ORLIPA	-
HIGHTCCBASE*STRONGSALES*LARGESCALE+	0.75	0.35	0.29	CECAOT; CELCCAR; CIAPEC; COAINE; SOPPROQUI; CEMUR	ADAPICRUZ
hightccbase*STRONGSALES*largescale	0.84	0.15	0.13	ASOCOM; COPROQUINACC	-
Solution	0.87	0.62			

**Table 7.16** Parsimonious solution after Boolean minimisation of the truth Table for failure in organisational strengthening

Term	Consist	RawCov	UniqCov	ObsConsist	ObsInconsist
HIGHTCCBASE*largescale*SOURCING+	0.79	0.10	0.07	ORLIPA	-
STRONGSALES*largescale+	0.78	0.16	0.12	ASOCOM; COPROQUINACC	-
LARGESCALE*sourcing+	0.97	0.12	0.06	CEMUR; OMCSA	-
HIGHTCCBASE*STRONGSALES	0.76	0.37	0.29	CECAOT; CELCCAR; CIAPEC; COAINE; SOPPROQUI; CEMUR	ADAPICRUZ
Solution	0.80	0.63			

### **Robustness**

For ten organisations, the grant contributed to organisational strengthening. Some ambiguity in this classification of success was present. Table 7.17 shows that three cases (AAAT, ADAPICRUZ and CECAOT) were upgraded to successful, whereas two cases (CELCCAR, CIAPEC) were downgraded to unsuccessful after the exchange of information and opinions between the two researchers.

When we perform the analysis with the original outcome classifications of GT, thus including measurement error, the results of the QCA alter. All the rows in the truth table become inconsistent as configurations that may predict a successful outcome.

Also, only three rows remain consistent when analysing the conditions related to failure (row 2, 6 and 12). These rows cannot be reduced any further, and only represent three (15% coverage) of the unsuccessful cases. This shows that the QCA analysis is susceptible to measurement errors, especially when the results show many different terms that cover only a limited number of cases. This gives us even more reason to refrain from making strong inferences in the analysis of unsuccessful outcomes.



**Table 7.17** Ambiguity in evaluation of the grant's contribution to organisational strengthening

			Reconciled evaluation		
			Unsuccessful	Successful	Total (original)
GT original evaluation	Unsuccessful	Count	14	3	17
	Successful	Count	2	7	9
Total (reconciled)		Count	16	10	26

**Table 7.18** OUTCOME: Grant contributed to capacity to pay organisational costs

row	Start conditions			Type of group	Consist with success	Suff. for success	Suff. for failure	Observations consistent with successful outcome	Observations consistent with unsuccessful outcome
	high tcc	strong sales	large scale	sourcing					
1	TRUE	TRUE	TRUE	TRUE	0.25	Con (FALSE)	Con (TRUE)	ADAPICRUZ	CECAOT; CELCCAR; CIAPEC; COAINE; SOPPROQUI
2	TRUE	TRUE	TRUE	FALSE	0.96	TRUE	FALSE	CEMUR	-
3	TRUE	TRUE	FALSE	TRUE	n/a	Rem	Rem	-	-
4	TRUE	TRUE	FALSE	FALSE	n/a	Rem	Rem	-	-
5	TRUE	FALSE	TRUE	TRUE	0.32	Con FALSE	Con (FALSE)	AAAT; AOCEMM	AGAYAP; APCA; COMART; INCAPALLAY
6	TRUE	FALSE	TRUE	FALSE	1.00	TRUE	FALSE	OMCSA	-
7	TRUE	FALSE	FALSE	TRUE	0.41	FALSE	FALSE	-	ORLIPA
8	TRUE	FALSE	FALSE	FALSE	1.00	TRUE	FALSE	ASAFOP	-
9	FALSE	TRUE	TRUE	TRUE	0.92	TRUE	FALSE	APROQUIRC; ARAO	-
10	FALSE	TRUE	TRUE	FALSE	n/a	Rem	Rem	-	-
11	FALSE	TRUE	FALSE	TRUE	0.31	FALSE	FALSE	-	COPROQUINACC
12	FALSE	TRUE	FALSE	FALSE	0.22	FALSE	TRUE	-	ASOCOM
13	FALSE	FALSE	TRUE	TRUE	n/a	Rem	Rem	-	-
14	FALSE	FALSE	TRUE	FALSE	n/a	Rem	Rem	-	-
15	FALSE	FALSE	FALSE	TRUE	0.55	Con (FALSE)	Con (FALSE)	APROAMOL; CEPLACH	APAMMIZQUE; APSU
16	FALSE	FALSE	FALSE	FALSE	1.00	TRUE	FALSE	AMAGA; AMDESOY	-

Source: Original data, analysed with Kirq 2.1.12.

FALSE = inconsistent configuration; TRUE = consistent configuration; Con = contradictory (coded FALSE or TRUE after reflecting the degree of inconsistency); Rem = logical remainder; observations are considered consistent when consistency score of the case is  $>0.5$ .

### Outcome 3: Increase capacity to pay organisational expenses

#### *Conditions that predict success*

Twelve organisations registered a positive outcome on their ability to source income to pay for organisational expenses. The truth table (Table 7.18) shows that row 2, 6, 8, 9 and 16 are consistently related to a positive outcome, in which the grant had contributed to increased capacity to pay organisational expenses. The complex solution (Table 7.19) reduces these five rows into three groups that cover 7 out of 12 successful cases, without overlapping members. Five of the successful cases are non-sourcing organisations. The two others (APROQUIRC and ARAO) are large and strong, but with weak tension containment capacities. The parsimonious solution (Table 7.20) presents more terms than the complex solution. It suggests various alternative ways to reduce the number of conditions in the term, depending on the other prime characteristic attached to ASAFOP, CEMUR and OMCS along with being a non-sourcing organisation.

The sole condition of being a non-sourcing organisation is not enough to explain success in generating income to pay for organisational expenses. However, it is clear that the group of non-sourcing organisations is a group that is likely to be successful in raising the capacity to pay organisational expenses.

**Table 7.19** Complex solution after Boolean minimisation of the truth Table for success to increase the capacity to pay organisational costs

Term	Consist	RawCov	UniqCov	ObsConsist	ObsInconsistent
strongsales*largescale*sourcing+	1.00	0.24	0.24	AMAGA; AMDESOY; ASAFOP	-
hightccbase*STRONGSALES*LARGESCALE*SOURCING+	0.92	0.15	0.15	APROQUIRC; ARAO	
HIGHTCCBASE*LARGESCALE*sourcing	0.98	0.13	0.12	CEMUR; OMCSA	-
Solution	0.97	0.51			

**Table 7.20** Parsimonious solution after Boolean minimisation of the truth Table for success to increase the capacity to pay organisational costs

Term	Consist	RawCov	UniqCov	ObsConsist	ObsInconsistent
strongsales*sourcing+	1.00	0.31	0.24	AMAGA; AMDESOY; ASAFOP; OMCSA	-
hightccbase*LARGESCALE+	0.71	0.17	0.15	APROQUIRC; ARAO	
LARGESCALE*sourcing+	0.91	0.15	0.02	CEMUR; OMCSA	-
HIGHTCCBASE*sourcing	0.99	0.23	0.00	ASAFOP; CEMUR; OMCSA	
Solution	0.88	0.53			

#### *Conditions that predict failure*

The truth table only has two rows with a consistency score above 0.75. Rows 7, 11 and 12 have only cases with a negative outcome but the configuration of conditions fails to be consistent

enough. This is a result of partial membership of these cases (and other cases that have fuzzy set scores between 0 and 1) in one or more of these conditions.

**Table 7.21** QCA analysis of necessary conditions for success in market access of members.

Term	Consist	Cov	Obs
SOURCING*	0.93	0.64	AGAYAP; APAMMIZQUE; APCA; APSU; CECAOT; CELCCAR; CIAPEC; COAINE; COMART; COPROQUINACC; INCAPALLAY; ORLIPA; SOPPROQUI

The analysis (Table 7.21) shows that being a sourcing organisation is a consistent necessary condition for failure. With a consistency score of 0.93, covering 13 out of 14 unsuccessful organisations, this characteristic seems a good predictor of failure to raise the organisation's capacity to pay organisational expenses. There are plausible explanations for this result. Compared to organisations that do not source their products from members, this group faces more agency dilemmas. For example, tensions with members on input price determination or quality control of products supplied to the group are absent in the group of smaller non-sourcing organisations. The grant is generally used to start an additional processing activity, next to bulking of unprocessed member products. The impact of the sales of processed products with this new investment is likely very small compared to the already existing income flows to pay organisational expenses. Thus, we infer that the commitment of the group to use the grant to gain money is higher in organisations that buy inputs on the spot market (similar to micro-enterprises) than in the larger organisations that source from members.

The minimisation of the rows that are related to unsuccessful outcomes will therefore not result in much additional insight. The complex and parsimonious solutions (Table 7.22 and Table 7.23) are identical and show two terms that are identical with the specification of these two truth table rows. ASOCOM has a unique term in the solution, being a non-sourcing organisation with low baseline tension containment capacities, strong sales and small scale. The other term defines a large group of strong and large sourcing organisations that had high tension containment capacities. The large group covers five unsuccessful cases (37% of all unsuccessful cases), and one inconsistent case (ADAPICRUZ). The consistency of the solution is slightly higher than the threshold (0.75), with an overall consistency score of 0.76 that precludes strong inferences.

**Table 7.22** Complex solution after Boolean minimisation of the truth Table for failure to increase the capacity to pay organisational costs

Term	Consist	RawCov	UniqCov	ObsConsist	ObsInconsist
HIGHTCCBASE*STRONGSALES*LARGESCALE*SOURCING+	0.75	0.36	0.36	CECAOT; CELCCAR; CIAPEC; COAINE; SOPPROQUI	ADAPICRUZ
hightccbase*STRONGSALES*largescale	0.78	0.06	0.06	ASOCOM	-
Solution	0.76	0.42			

### **Robustness**

The use of a higher consistency threshold does not alter the results of the QCA analysis of conditions related to success. A higher consistency threshold of 0.80 would make all rows insufficient, precluding any causal analysis of sufficiency.

There was little ambiguity in the evaluation of success in this outcome area, when considering the classification before and after reconciliation (Table 7.23). Only one case (COMART) changed from successful to unsuccessful after the exchange of information between the two researchers during the reconciliation. This had no influence on the results of the analysis, except slightly different values for consistency and coverage.

**Table 7.23** Ambiguity in evaluation of the grant's contribution to increased capacity to pay organisational expenses

			Reconciled evaluation		Total (original)
			Unsuccessful	Successful	
GT original evaluation	Unsuccessful	Count	13	0	13
	Successful	Count	1	12	13
Total (reconciled)		Count	14	12	26

## **7.6 Why are grants to well-endowed organisations unsuccessful?**

Five of the six organisations with high tension containment capacity, strong sales and a large scale, all sourcing member products, failed on all three outcomes. This begs an explanation of the causal mechanisms that may have been involved. The available data is too limited to make strong inferences. Nevertheless, some plausible hypotheses emerge, when we review the grant-related dynamics in this group of organisations (Table 7.24). All five unsuccessful organisations mention as a reason for failure of the grant-supported business plan that they had invested the grant in under-scales equipment. Also, most of them mentioned the abortion of the grant-supported business plan as a rational business decision, after the pilot experience. For these organisations, the grant investment of USD 10,000 apparently was too small in relation to their existing collective marketing activities.

They could stop the new business plan without disintegrating as a group, because they did not depend for their organisational and economic survival on its success. ADAPICRUZ, who shares the characteristics of this group of well-endowed organisations, was successful, but this does not contradict this inference. ADAPICRUZ used a first grant to enter into a new market with a new product, but they needed complementary investments to make the diversified product portfolio commercially viable. Reviewing the grant-related dynamics in the other 20 beneficiary organisations, we see that only the young and small organisation ASOCOM mentioned under-scaled investment as a reason for failure (see Annex 1 for details).

**Table 7.24** Grant dynamics in the well-endowed organisations that proved consistently unsuccessful

Name	Grant dynamics
CECAOT	CECAOT used their grant, in 2009, to repair an optical quality control unit in their plant, to limit labour costs in the plant. The maximum amount available from FONDOECAS (US\$10,000) motivated them to repair the equipment instead of buying a completely new machine (US\$40,000). However, the equipment broke down again in 2010, partly due to improper handling. The optic sensor has not been repaired anymore due to the high costs. Instead, CECAOT considered buying a completely new optical sensor, which they did not do, however, partly due to the crisis and resulting internal organisational problems in 2011 which resulted from the failure to obtain a pre-harvest sales contract.
CELCCAR	CELCCAR channelled the FONDOECAS grant to one of its member cooperatives. They experimented with fruit processing on a pilot scale. They mention internal organisational problems and lack of complementary equipment as the major factors that negatively affected the business plan. The capacity of the equipment was considered by the 2013 interviewees to be too low to seriously create market access. An expansion of production capacity is needed to obtain real access to the market.
CIAPEC	CIAPEC wanted to develop a production line for roasted coffee for the national market in La Paz, and had expectations for export. It started to experiment with roasting and packaging but the production capacity was lower than expected and they experienced technical problems with the equipment after only one year of operation. They consider the equipment not suitable for processing on an industrial scale.
COAINE	The equipment bought with the grant was far too small for the use that COAINE projected. Additional access to markets has not been created, nor has COAINE visibility in the market been enhanced by the grant. The average yearly turnover of processed coffee was an insignificant amount when compared to the size of the total turnover and size of membership.
SOPROQUI	SOPROQUI wanted to invest the grant in processing and packaging equipment to supply processed quinoa products (quinoa popcorn, quinoa soup) to the market, including the school meal programmes. However, the equipment was never properly delivered and installed and the project never took off. The current board members do consider quinoa processing still to be an interesting business opportunity but indicate that other machinery and skilled personnel is needed to start doing so.

Source: For details, see Annex 1

## 7.7 Triangulating the results of QCA with logistic regression

In spite of the fundamentally different notion of causality implied in regressional-analytical and configuration-comparative methods, detailed in Thiem et al. (2015), and treated in Chapter 1 of this thesis, in our research logistic regression could help to triangulate part of the results of QCA because most of the conditions identified as potential predictors of effectiveness in the QCA resulted to be single conditions, not configurations.

We applied logistic regression to the scale variables that were underlying the fuzzy-sets. We included five variables in our model with a data set of only 26 observations. In Table 7.25 we present the results with the statistical significance level of the model, according to the Omnibus Test, and we use the Nagelkerke r-squared as the indicator of the capacity of the model to explain the total variance. We only present models for which the Hosmer and Lemeshow test rejects the null-hypothesis that observed and predicted values have a similar distribution (Meyers et al., 2006). Model 1 suffers from ‘quasi-complete separation’, a situation in which an explanatory variable perfectly predicts one of the two values for the outcome variable (Rainey, 2014). Grants to non-sourcing organisations will, per definition, always be unsuccessful in improving market access for group members. Quasi-complete separation causes the odd ratio to be infinitely high and the constant infinitely small. To address this problem, we included two alternative models. In Model 2 we omitted the constant from the model and

in Model 3 we omitted the sourcing-variable. The Hosner and Lemeshow test indicates there is predictive value in each model. The direction of the odd ratios is stable. Model 1 appears as the best interpretable. It points to the causal necessity of sourcing for success on this objective, and can be read both as ‘the condition of being a non-sourcing organisation reduces the chance of the grant to be successful to zero’.

Table 7.25 Logistic regression for success of grant to increase market access for members with scale variables

	Scale variable			Fuzzy sets		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Omnibus test	p=.432	p=.029	p=.671	p=.235	p=.068	p=.293
Nagelkerke R-squared	0.273	0.509	0.139	0.369	0.435	0.277
Hosner and Lemeshow test	0.952	0.318	0.741	0.620	0.526	0.574
Correct classification:						
• Overall percentage	77%	81%	77%	85%	81%	85%
• Prediction of success	0%	20%	0%	20%	20%	20%
• Prediction of failure	95%	95%	95%	100%	95%	100%
Odds ratio of predictors:						
• Sourcing organisation	8.66E+23	2.67	--	5.49E+21	0.43	--
• Tension containment capacity	1.02	0.96	1.04	0.14	0.08	0.15
• Market performance	1.63	1.99	2.01	2.62	0.78	2.82
• Organisational scale	1.13	2.16	1.34	23.56	16.93	53.26
• Organisational age	0.94	0.91	0.94	0.23	0.20	0.18
• Constant	0.00	--	0.15	0.00	--	0.05

Note: None of the odds ratio is statistically significantly at the 0.10 level

Table 7.26 Logistic regression for success of grant on organisational strength and capacity to pay organisational expenses

	Grant contributed to organisational strengthening (yes=1, no=0)		Grant contributed to capacity to pay org. expenses (yes=1, no=0)	
	Scale variables Model 7	Fuzzy-sets Model 8	Scale variables Model 9	Fuzzy-sets Model 10
Omnibus test	p=.571	p=.521	p=.296	p=.261
Nagelkerke R-squared	0.187	0.203	0.279	0.295
Hosner and Lemeshow test	0.770	0.237	0.394	0.130
Correct classification:				
• Overall percentage	62%	73%	77%	69%
• Prediction of success	40%	50%	58%	58%
• Prediction of failure	75%	88%	93%	79%
Odds ratio of predictors:				
• Sourcing organisation	2.63	1.68	0.11	0.07*
• Tension containment capacity	0.99	0.26	0.97	0.19
• Market performance	0.93	0.37	0.89	0.38
• Organisational scale	1.36	7.05	0.91	4.06
• Organisational age	0.90	0.19	0.98	0.96
• Constant	1.38	0.80	17.94	13.51*

\*. Odds ratio is statistically significantly at the 0.10 level

The logistic regressions using the fuzzy-set variables as predictors (Model 4, 5 and 6) yield similar results but with more accentuated odd-ratios. The fuzzy-set condition 'large organisational scale' seems to be a strong predictor of success, and being a non-sourcing organisation a strong predictor of failure. Organisations with a large organisational scale have a likelihood of being successful that is approximately 20 times higher than organisations that are small scale. In the logistic regression, having high tension containment capacities appears to lower the likelihood of success, and older organisations are likely to be less successful than young organisations. As shown above, QCA detected the combination of both conditions, old age and high tension containment capacity, as a causal configuration consistently related with failure. Overall, the chances that a grant contributes to increased market access for members are very low, which is reflected in the extremely low odd ratio of the constant.

The logistic regressions on the other two intended outcomes of the grant fund, organisational strengthening (Model 7 and 8) and capacity to pay organisational costs (Model 9 and 10) show weaker patterns. The characteristic of being a sourcing or non-sourcing organisations is important as a predictor of effectiveness of the grants on these objectives. Being a sourcing organisation seems to predict success of the grant in organisational strengthening, while being a non-sourcing organisation seems to predict success in raising income to pay organisational expenses.

**Table 7.27** Results of the configurational comparative and the regression-analytic exploration for predictors of success and failure

Outcome Area	Outcome pattern detected with QCA	Outcome pattern detected with logistic regression
Market access for members	<p>Being a sourcing organisation is a necessary condition for success on this outcome.</p> <p>Being small scale and being a non-sourcing organisation is consistently related with failure.</p> <p>Older, stronger and larger organisations with high tension containment capacities at baseline, are consistently unsuccessful on this outcome.</p>	<p>Being a sourcing organisation seems to increase the likelihood of success; being a non-sourcing organisation seems to lower the likelihood of success.</p> <p>Organisational scale seems the factor that most increases the likelihood of success.</p>
Increased organisational strength	<p>No conditions are consistently related with success.</p> <p>No conditions that are consistently related with failure on this outcome.</p>	<p>Being a sourcing organisation seem to increase the likelihood of success.</p>
Improved income to pay for organisational expenses	<p>No conditions are consistently related with success on this outcome.</p> <p>Being a sourcing organisation is consistently related with failure.</p> <p>Strong and large organisations with high tension containment capacities at baseline are consistently unsuccessful on this outcome.</p>	<p>Being a sourcing organisation seems to decrease the likelihood of success.</p>

Table 7.27 presents the causal inferences derived from both data-analytical approaches. The QCA results show that for a non-sourcing organisation, it is impossible to create market access for members. Likewise, the logistic regressions suggest that the likelihood of success for non-sourcing organisations is almost zero. Both data-analytical approaches also detected the importance of organisational scale. Grants to small scale organisations are likely to fail.

The logistic regressions suggest that being a sourcing organisation seems to increase the likelihood of success on the outcome organisational strengthening. For this outcome QCA did not identify any plausible configuration of conditions that could explain effectiveness. QCA identified that being a sourcing organisation was consistently related with failure to increase the capacity to pay for organisational expenses, and, likewise, the logistic regressions suggest that being a sourcing organisation decreases the likelihood of success.

## 7.8 Discussion and conclusions

The FONDOECAS grant facility wants to allocate grants to organisations that are most likely to be successful, or less likely to be unsuccessful. We used 26 case studies of grant beneficiaries to explore for baseline conditions that could predict success or failure of the grant, in order to gain insights that could help FONDOECAS to improve the targeting of the grants.

The results of the Qualitative Comparative Analysis show that being a sourcing organisation is a necessary condition for success on the outcome of increased market access for members. Having a small organisational scale was consistently related with failure. Also, being an older organisation combined with high tension containment capacities at baseline, was consistently related with an unsuccessful outcome. These results suggests that, when the intention of a grant fund is market access for member farmers, it could increase its effectiveness by targeting the grants to younger organisations that are not too small in scale (in patrimony and/or membership). This is not to be seen as a guarantee for success, but as a strategy to limit the risk of failure. Grants to non-sourcing organisations, however, are more likely to be successful in improving the capacity to pay for organisational expenses. Because some conditions were directly linked with success or failure, and not as part of a configuration, we could triangulate them with logistic regression.

The main difference between the two ways of data-set analysis is in the identification, through QCA, of a pattern of organisations that are older, stronger, larger and with high tension containment capacities but where the grants were, nevertheless, consistently unsuccessful. Revising the history and dynamics in these organisations, we found a plausible causal explanation for this pattern. The small size of the grant available for the investment and the presence of important other collective marketing activities resulted in most grant-supported processing activities not being commercially viable. Therefore, these well-endowed organisations decided to discontinue the new business after the pilot experiences.

We verified the stability of the QCA solutions by applying different consistency thresholds and by repeating the analysis using ‘real’ measurement error. We show that, overall, the results from the analysis with and without measurement error proved to be similar. The results of the



QCA seem also robust to changes in fuzzy-set calibration. The qualitative anchors used for the fuzzy-set calibration resulted in only a small number of organisations with fuzzy-set scores between 0.25 and 0.75. Slightly different qualitative anchors would, therefore, result in a fairly similar truth tables and QCA solutions.

QCA helps to address the explorative question, What works, for whom and under what conditions? However, to do so, it needs additional reasoning on the causal mechanisms that may explain the patterns detected, for example, by the method of process tracing which we used (Collier, 2011; Beach and Pedersen, 2013). We showed that logistic regression can be used to strengthen the validity of causal inferences made with QCA, especially when the QCA solution refers to single conditions with a fair coverage of cases. The Boolean logic in the minimisation of the truth table helps us to think, but does not think in itself (Rihoux and Lobe, 2009). Empirical knowledge on the specificities of each case is needed to interpret the detected data patterns.

### **Acknowledgements**

*The Netherlands' Ministry of Economic Affairs (KB-11-004) financially supported this research. The fieldwork was co-funded by the Dutch Interchurch Development Organizations, ICCO, as part of a baseline for an impact evaluation of the innovation grant fund FONDOECAS. Most of the field data was collected by Lithzy Flores, who made a tremendous effort travelling to all those remote places. Oscar Chambi and Richard Arguedas in FONDOECAS, provided additional data and background information. Erwin Bulte, Lan Ge, Barbara Befani provided useful advice on the data-analytical methods used.*



## 8 | Innovation Grants to Farmer Groups for Collective Marketing: A contribution analysis of the FONDOECAS small-grant fund in Bolivia

Giel Ton

Agricultural Economics Research Institute, LEI Wageningen UR

### Abstract

*Innovation grants help to reduce the risk faced by innovators when starting a new business process. We used Contribution Analysis to assess the relevance, effectiveness and efficiency of an innovation fund in Bolivia, FONDOECAS, directed to economic farmer organisations that develop value-added activities and collectively market their products. We select three key assumptions in the intervention logic on which we focussed the research: relevance of the farmer groups for local economic development, effectiveness of the fund in strengthening these group, and efficiency of the grant allocation mechanism. A large household survey showed that the majority of farmers was interested in collective marketing of their products and considers economic farmer organisations as important actors in local development. We used a case-based comparative analysis to assess the effectiveness of the grant. Most economic farmer organisations registered a high growth, however, the contribution of FONDOECAS to this growth had been modest. To assess the efficiency of FONDOECAS' grant allocation system, we used administrative data. The efficiency of the technical committee to target grants to the most feasible business plans appeared to be quite low.*

## 8.1 Introduction

Innovation grants help to reduce the risk faced by innovators when starting a new business process (Kessler, 2013). Public support for these competitive funds is, however, contentious. There must be evident public interest to legitimise governments investments in private ventures. That is, grant funds should support business activities that have a positive impact on society. Moreover, the grants need to facilitate investments that would otherwise not have been made (Heinrich, 2014). Poverty alleviation is a common policy objective for public funding of these funds in developing countries. Farmer organisations are considered to be key actors and hotspots for innovation in rural areas (World Bank, 2007; World Bank, 2010; Ashby et al., 2009; Bebbington, 1997; Berdegúé, 2001). And, farmer groups that handle economic activities are especially attractive partners for rural development initiatives, when poverty alleviation and local economic development are key policy objectives (World Bank, 2007; Bernard et al., 2010). Competitive grant systems specifically targeted to economic farmer organisations are part of agricultural innovation policies in several countries (Berdegúé, 2001; Toro and Espinosa, 2003).

Evidence on the effectiveness of this type of grant systems on local economic development is still scarce. An explorative systematic literature review on the effectiveness of innovation grants to smallholders (Ton et al., 2013b) found only a few peer-reviewed impact studies of business development grants. Several factors explain this relative neglect of impact evaluation. First, attributing outcomes to grants is challenging, as the business plans emerge from a complex process in which the grant covers only part of the investments needed. Second, the economic benefits of grant-funded investments often materialize only after some years of starting-up and gradual market penetration. Third, scale and activities of farmer groups vary a lot. This diversity results in a large variance in performance indicators such as turnover, membership, and patrimony. This implies the need of large samples to detect statistically significant effects, whereas the limited number of farmer organisations that exists in a country or region functions as a hard cap on sample size. Therefore, the possibilities for experimental or quasi-experimental designs to measure the impact of grants on farmer organisations are constrained. Alternative ways to make use of smaller data-sets are needed to verify the assumption that these funds generate development impact.

Contribution analysis has been developed to design monitoring and evaluation systems for complex interventions where experimental and quasi-experimental designs are impossible (Mayne, 2001; 2012). Despite substantial attention to contribution analysis in the field of evaluation methodology, few studies have applied it in practice (Dybdal et al., 2010). Our research is, to our knowledge, the first academic paper in which contribution analysis is used to assess the effectiveness of an agricultural support intervention in a developing country. We used this approach to assess the relevance, effectiveness and efficiency of an innovation fund in Bolivia, FONDOECAS (*Fondo para el Desarrollo de Organizaciones Económicas Campesina*), directed to economic farmer organisations that develop value-added activities and collectively market their products.

The paper proceeds as follows. First, we describe in more detail the context of the grant fund, the importance of economic farmer organisations, and the institutional set-up of FONDOECAS.

Second, we discuss the intervention logic and theory of change of FONDOECAS and select three key assumptions on which we focussed the research: relevance of the farmer groups for local economic development, effectiveness of the fund in strengthening these group, and efficiency of the grant allocation mechanism. Third, we present the design and results of research to verify these assumptions. Support to economic farmer organisations proved a relevant activity, but the effectiveness of the grants to improve market access for farmers is low. Effects on organisational capacities and income to pay organisational expenses were higher, though the larger and stronger organisations proved particularly unsuccessful. We finish with a reflection on contribution analysis and the methodology used.

## 8.2 Context

### Bolivia's rural structure

The international debt crisis in the 1980s struck Bolivia hard and resulted in a breakdown of the domestic economy due to hyperinflation. Bolivia was the first country in which structural adjustment programmes were implemented (in 1985) and agricultural policies have been shaped accordingly for two decades. The landslide economic reform package of 1985, known in Bolivia as *Decreto Supremo 21060*, changed the role of the state, privatising many state-owned enterprises, closing down the agricultural credit and extension services, and reducing protective border tariffs (Prudencio and Ton, 2004; Loza, 2002).

Domestic food supply was increasingly influenced by imports from distant areas with lower net production costs (Prudencio and Ton, 2004). This process has led to a more profound specialisation of the agricultural sector in Bolivia within the world market, and to an accentuated differentiation of agricultural regions inside the country (Pérez Luna, 2003): the Andean region with a predominantly smallholder production for domestic markets and the lowlands with predominantly large-scale production for export markets. In the mountainous Andean region, geographical conditions influence production costs in such a way that bulk production for large-scale processing industries is not economically feasible. Production for niche markets (e.g., organic, fair trade, quality seeds, local branding) is a tempting strategy for organised producers to reach more remunerative markets but is a marginal proportion of total marketed output, except for coffee, cocoa and quinoa. The comparative advantages of the Lowlands for low-cost protein production (meat and soybeans) and credit policies for agro-industrial development have resulted in an impressive growth, especially from 1990 onwards. After sugar and cotton in the 1970s, oil-seed production has been the main engine of agricultural growth. In 2012, one-third of the agricultural area in the department of Santa Cruz was dedicated to soybean production (ANAPO 2012).

Between 1985 and 2005, three political parties (MNR, MIR and AND) supported successive coalition governments without major changes in economic policy. These coalition governments gave Bolivia a relative stability but did not resolve the basic contradictions in the Bolivian agrarian structure, which resulted in large parts of the population being excluded from development (World Bank, 2005). The growing contradiction between a stagnating domestic econ-

omy<sup>10</sup> and increasing foreign domination of former state enterprises in mining and the service industry caused a series of outbursts of popular discontent (2000-2004) and accentuated the dichotomy between the state as being abusive and corrupt and the populist mass movement as being democratic and reliable<sup>11</sup>. A sequence of popular rebellions forced three presidents out of office. The political instability led to elections in 2006, followed by a new government that was supported by a coalition of popular movements (bundled in the ‘political instrument’ *Movimiento Al Socialismo* - MAS) led by the farmer Evo Morales. The Morales government re-introduced several rural support instruments that had characterized the pre-1985 period, especially soft loans to communal organisations and direct state interventions in agricultural markets.

### Economic farmer organisations

Until 2005, the large commercial farmers (represented by the *Cámara de Agrícola del Oriente* – CAO) had dominated agricultural policy making in Bolivia. After 2006, policy making has become much more focussed on the interests of the smallholder sector. In the lowlands, these were the indigenous people and the smallholders who had settled in land-reform areas (*colonizadores*). In the Andean Region, the main political organisations were the village-based unions - *sindicatos campesinos*- and *ayllus*. The unions are organised by the *Confederación Sindical Única de Trabajadores del Campo de Bolivia* (CSUTCB). They have a dense network of branches at village, provincial, departmental and national levels. *Ayllus* are (the remains of) indigenous territories that comprise various villages and are represented by the *Consejo Nacional de Ayllus y Markas del Qullasuyu* (CONAMAQ).

Since 1999, the *OECAs* (*organizaciones económicas campesinas* - economic peasant organisations) have entered the national policy arena as a new smallholder interest group, led by the *Coordinadora de Integración de Organizaciones Económicas Campesinas de Bolivia* (CIOEC-Bolivia)<sup>12</sup>. During the brief government of Carlos Mesa between 2003-2005 after president Sanchez de Losada fled the country, CIOEC influenced the orientation of agricultural policies. By 2005, they were involved in the drafting of the agricultural policy *Estrategia Nacional de Desarrollo Agrícola y Rural* (ENDAR) and played a leading role in the design of the policy “I Buy Bolivian” (*Compro Boliviano*), which gave preference to small-scale producers in government

---

10 The World Bank (2005) indicated that the per capita income in Bolivia fell slightly between 1950 and 2000 while, in comparison per capita income rose by 75% in Argentina, 200% in Chile and 350% in Brazil in the same period.

11 The urban economy grew, with the informal sector being its main engine of (precarious) employment generation. This informal sector of petty traders became dominated by an indigenous bourgeoisie. This challenged the traditional dichotomy of the urban ‘white and *mestizo*’ versus the rural ‘*indios*’ (Klein, 2003). This indigenous bourgeoisie fuelled the protests against the ‘corrupt white politicians’ by reconstructing and emphasizing their pre-colonial indigenous identity.

12 CIOEC had been founded in 1990 to support its members with capacity development and legal-administrative support. It gradually developed into an advocacy platform for public policies to enable economic farmer organisations in general. In 1999, when Bolivia organised a public consultation around its poverty reduction strategy, which was a prerequisite of the World Bank and the International Monetary Fund (IMF) for debt reduction, CIOEC gained political visibility and broadened its membership. With funding of British Embassy and the Inter-American Foundations (IAF), and in coordination with other organisations of small-scale producers which were grouped into the advocacy platform *Comité de Enlace* (Liaison Committee), CIOEC organised a series of 11 commodity-specific national meetings of economic farmer organisations to generate proposals for more enabling policies (CIOEC-Bolivia, 2000, 2004).

procurement (Mesa Gisbert et al., 2012; Elías, 2007). The sector of economic farmer organisations was increasingly mentioned as a key actor in agricultural innovation policies. This recognition was facilitated by international donors such as the United Nations Food and Agriculture Organisation (FAO), the International Fund for Agricultural Development (IFAD) and World Bank, which had put agricultural development by family farmers back on the agenda (World Bank, 2006; World Bank, 2007) and who had a crucial influence on public policy making (Rodríguez-Carmona, 2009). During the Morales government, the discourse on agricultural development in Bolivia shifted from a free-trade, export-orientation to a smallholder-oriented policy of food sovereignty. Village organisations and state enterprises became key actors in the ‘Communitarian Agricultural Revolution in Production’ (Government of Bolivia, 2011; Cordoba and Jansen, 2013). Economic farmer organisations were ‘caught in the middle’. Economic smallholder organisations enabled smallholder inclusion in markets. Nevertheless, they remained excluded from many new preferential policies for smallholder farmers because they overlapped in membership with the territorial grassroots organisations (OTBs - *Organizaciones Territoriales de Base*), such as *sindicatos*, *ayllus* and indigenous territories<sup>13</sup>.

According to the census by CIOEC (CIOEC-Bolivia, 2009), the number of first-tier economic farmer organisations in Bolivia is 712 (see Table 8.1). The census also registered 48 second-tier (federal) organisations and 16 representative organisations (e.g., the CIOEC departmental branches). One-third of the organisations registered in this 2008 Census of OECAs are informal groups having no legal status. Organisations with direct membership (first tier) have an average of 78 members, with a high variance S.D.=154). The constituency of second tier organisations varies even more, with an average of 1,057 and standard deviation of 3,148. The coca federation is a clear outlier, with a constituency of 19,505 registered members. This federation manages a large part of the highly regulated internal marketing of coca leaves for traditional use (chewing or tea). Without taking the coca sector into consideration, the second-tier organisations represent on average 584 farmers, with a standard deviation of 993.

---

13 In response, CIOEC launched a proposal to repair this omission with a special law (Antequera Guerra, 2013; Ton et al., 2013a). President Evo Morales proclaimed this law on 26 January 2013 (Government of Bolivia, 2013). This legal recognition of OECAs makes it easier for the government and international donors, such as the World Bank, IFAD and Swiss Development Cooperation, to target programmes and funding to this more innovative and market-oriented rural sector (Ton et al., 2013a).

**Table 8.1** Number and group size of economic farmer organisations registered in the 2008 Census of OECAs

Sector	Number of first-tier organisations (count)	First tier legalised (count)	First-tier group size (average)	Number of second-tier organisations (count)	Second-tier group size (average)
Agriculture	160	123	84	10	1,177
Processed foods	26	16	54	2	475
Honey	51	28	65	3	166
Handicrafts	82	38	60	4	444
Chicken	12	10	45	0	
Sugar	4	3	298	0	
Cocoa	2	2	64	2	767
Coffee	28	24	134	3	123
Coca	0	0	-	1	(19,505)
Fruits	72	60	59	1	650
VegeTables	30	21	65	0	
Dairy	109	44	70	7	380
Wood	5	1	20	0	
Nuts	5	2	133	0	
Cattle	53	37	147	4	472
Fish	7	4	29	0	
Quinoa	26	24	72	2	744
Seeds	19	17	51	1	75
Stones and tiles	2	2	40	0	
Tourism	11	9	125	1	300
Tree nurseries	8	6	40	0	
Total	712	471		48	
	100%	66%			
Mean			78		594 (1,057)

Source: Original data of the 2008 Census of OECAs (CIOEC-Bolivia, 2009)

## FONDOECAS

Before 2006, Bolivia had experimented with several grant facilities to stimulate collective marketing by smallholders. For example, Bolivia had piloted a grant fund called LIL/INDÍGENA (LIL stands for 'Learning and Innovation Loans') with World Bank funds between 2001 and 2005. Its objective was to demonstrate that initiatives based on indigenous identity, economic innovation and entrepreneurship, generated and self-managed by indigenous organisations, could help to reduce poverty (Llorenti Barrientos et al., 2005). The fund was promoted through the village organisations described above, and had as one of its eligibility requirements a minimum group size of 20 persons. LIL/INDÍGENA generated 1,800 project ideas, 172 of which were selected for project formulation and 51 were implemented. That LIL/INDÍGENA did not meet their expectations is an understatement. Many groups that formulated project ideas were formed with the sole objective of accessing funds from LIL/INDÍGENA and were unable to implement their business plans. The village organisations had been successful in generating interest and ideas for economic initiatives, but the groups that emerged from these processes

often lacked the capacities needed to implement and manage the projects. Economic farmer organisations, formally constituted as associations, proved better suited for this, especially organisations that had existed for some years (Llorenti Barrientos et al., 2005). Organisations that submitted proposals to the fund experienced long administrative delays (usually several years), which resulted in several business opportunities no longer existing at the time the grants were finally approved.

These, and other experiences showed the need for a grant fund that would specifically cater to functioning economic farmer organisations, and having a lean grant-allocation mechanism in order to limit the time between proposal and implementation of the business plan. FONDOECAS was designed by CIOEC-Bolivia in consultation with three Dutch development organisations (ICCO, CORDAID, and OXFAM NOVIB). CIOEC presented a proposal (Condori, 2005) for a competitive small-grant fund to strengthen economic farmer organisations through co-funding of some organisational costs, technical assistance, and productive investment. Initially, the donors were more inclined to support a credit facility (revolving fund) instead of a grant fund, but they agreed to the grant modality, because it would pose less organisational stress on CIOEC than a credit system, in which a higher capacity to sanction is needed in case of credit default. But the grant fund threatened to overstretch the organisational capacity of CIOEC. Grant allocation in member-based organisations is risky because board and staff can be subjected to member pressure that can affect transparency and effectiveness. For example, member pressure on grant allocation could influence the election and/or re-election possibilities of CIOEC's board members. CIOEC and ICCO decided to have a semi-autonomous grant fund with an anonymous technical committee of external experts that would make the decisions to approve or reject proposals, called FONDOECAS<sup>14</sup>. The national platform of biological producers AOPEB and the financial branch of the coffee producers (FINCAFE) were invited to join CIOEC and ICCO in the board of this grant fund to strengthen the semi-autonomous character and impartiality of the fund.

FONDOECAS hoped to develop a grant allocation system that could be replicated within other agricultural development programmes, such as the World Bank-funded programme *Programa de Alianzas Rurales* (PAR) and the public development bank BDP (*Banco de Desarrollo Productivo*). The strength of FONDOECAS, compared to PAR, would be in its less-rigid eligibility requirements, to also allow smaller organisations to get access to grant support. CIOEC and AOPEB's knowledge on the trajectories of the applicants could prevent investments in organisations that only existed in name, which were reactivated only when soliciting external support but had no real economic life of their own. In earlier grant funds, the selection of 'real' organisations had been a recurring challenge (Hartwich et al., 2007; Ton, 2007b).

---

14 The author was contracted to find a suitable institutional set-up for the grant fund: Ton G. (2005) Fondo para el Fortalecimiento Económico de las Organizaciones Económicas Campesinas (FONDOECA). La Paz: CIOEC-ICCO, 37.



**Table 8.2** Criteria used by the technical committee to assess the feasibility of business plans

Criteria	Relevant aspects	Points
Technical feasibility	Concerns the access to raw materials, processing, installed/spare capacity, etc.	max 30
Commercial feasibility	Considers whether the marketing strategy is coherent with previous experiences (market and organisation) and human, technical and financial possibilities.	max 30
Financial and economic feasibility	Considers the profitability of the investment in terms of cost/benefit, break-even point, or other financial indicators. Additionally, it considers the capacity of the organisation to pay for its organisational expenses, so that they do not affect the results of the business.	max 20
Funding structure	Co-funding by the organisation above the required percentage (30%) is considered for a higher score.	max 10
Organisational development	Considers the way that the organisation can lead the implementation of the business plan before it is operational. Additionally, it considers the operational structure around the business opportunity and the positive experiences of the organisation with earlier plans and investments. It values the participation of the different organisational instances in the implementation of the project, according to their role.	max 10
<b>TOTAL</b>		<b>max 100</b>

Source: FONDOECAS, 'Propuesta consolidada - criterios de calificación de proyectos', 26 August (2007)

Like in most grant systems and public procurement procedures, two types of documentation must be submitted by the applicant to FONDOECAS. One set of documents ('Envelope A') is needed to judge the eligibility of an organisation for the grant. A second set ('Envelope B') is needed to assess the quality of the business proposal. In the case of FONDOECAS, the eligibility requirements are: a legal status; a documented participatory consultation process having approved the business proposal; annual statements for the last two years; and being affiliated to one of the two national representative organisations CIOEC or AOPEB. The second set of documents reviewed the presence of a series of enabling conditions (see Table 8.2): raw materials and complementary processing equipment must be available, the group must show a credible marketing strategy, the break-even point needs to be reached in a time span within which the organisation is able to pay for the start-up costs, beneficiaries need to co-finance the investment as a guarantee of their commitment, and the organisational structure should be such that the business operations can be handled effectively. Thus, FONDOECAS has explicitly identified a list of assumptions about the supporting factors that need to be in place for a grant to be successful and 'cause' positive impact.

FONDOECAS started to allocate grants in 2007. The FONDOECAS Strategic Plan 2011-2013 (FONDOECAS, 2010) envisioned a process to scale up its activities with a view to attracting funds from the government or major donors. FONDOECAS started a procedure to become a separate legal institution, and as of 2010 it supported this research project to evaluate the impact of the grants.

## 8.3 Study design

### Objectives

The objective of the research was to develop a monitoring and evaluation system in FONDOECAS that could identify and evaluate the impact of grant investments on economic farmer organisations (Ton, 2010a). The end users of these research results, CIOEC, AOPEB and ICCO, expected to use the results to attract additional donors for this type of value-added grants to farmer organisations (FONDOECAS, 2010). Moreover, the funders of the field research, ICCO-CIOEC- ESFIM, requested a research process that would produce research outputs that could be used in capacity-building activities to strengthen the management capacities of the farmer organisations. ICCO had keen interest in the evaluation approach in view of replication on similar private-sector interventions in other countries. The fact that the author had been working with CIOEC between 1999 and 2004 and had had a role in the design of FONDOECAS in 2005 was considered an advantage, even though it meant that he was not a truly independent evaluator.

### Contribution Analysis

We used Contribution Analysis (Mayne, 2001; 2012) as our approach to address the challenges in impact evaluation of complex interventions (Ton et al., 2011b; Ton, 2012b). Mayne's principal motivation to develop contribution analysis was precisely to find a systematic way to collect and use monitoring information to reflect on the relevance and effectiveness of policy interventions, even when it is impossible to attribute the societal outcomes unambiguously to these policy interventions. Contribution Analysis is:

*“... a reasonable way to make evidence-based causal claims rather than being unable to say anything about causality – or worse, leaving readers to make their own assumptions. (...) From an evaluation perspective, the issue was what could be done to make credible causal claims in the absence of experimental approaches. Many evaluations seemed either to be silent on causality or, perhaps worse, made causal claims based solely on the views of interviewees.” (Mayne, 2012)*

Contribution Analysis is a form of theory-driven evaluation, collecting data to verify and strengthen the intervention's 'performance story'. In this process, multiple information sources are used to build, test and improve the understanding of the intervention logic. Contribution Analysis was well in line with our objective to build FONDOECAS' monitoring system, in which key information is collected on clients/beneficiaries to monitor progress and outcomes. Data collection in the monitoring and evaluation system needs to be able to capture the outcomes that result from the intervention and collect information to discard rival explanations for these outcomes, that is, *“build a compelling case with evidence from which it is reasonable to conclude with confidence that the intervention has made a contribution and why”* (Mayne, 2012).

Mayne (2001; 2012) describes Contribution Analysis as a logical sequence of six steps to obtain a convincing 'performance story'. These steps (see Box 8.1) describe an iterative process of building and refining the intervention logic. It identifies the key assumptions of impact that

**BOX 8.1 KEY STEPS IN CONTRIBUTION ANALYSIS****Step 1** Set out the cause–effect issue to be addressed

- Acknowledge the causal problem for the intervention in question
- Scope the problem: determine the specific causal question being addressed; determine the level of confidence needed in answering the question
- Explore the nature and extent of the contribution expected from the intervention
- Determine the other key factors that might influence the realisation of the results
- Assess the plausibility of the expected contribution given the intervention size and reach.

**Step 2** Develop the postulated intervention logic<sup>a</sup> and risks to it, including other influencing factors

- From intervention documents, interviews and relevant prior research, develop the postulated intervention logic of the intervention, including identifying the assumptions and risks for the causal links in the intervention logic
- Identify the roles other key influencing factors may play in the intervention logic
- Determine how contested the postulated intervention logic is to better understand the strength of evidence needed.

**Step 3** Gather the existing evidence on the intervention logic

- Gather the evidence that exists from previous measurements, past evaluations and relevant research to assess the likelihood: (1) of the expected results, assumptions and risk being realised; (2) of each of the causal links in the results chain occurring; and (3) of the other influencing factors making a significant difference.

**Step 4** Assemble and assess the contribution claim and challenges to it

- Set out the contribution ‘story’ on the likelihood that the intervention ‘worked’: the causal claim based on the analysis of logic and evidence so far
- Assess the strengths and weaknesses in the postulated intervention logic in light of the available evidence, and the relevance of the other influencing factors – which links seem reasonable and which look weak and need more evidence. If needed, refine or update the intervention logic.

**Step 5** Gather new evidence from the implementation of the intervention

- With a focus on the identified weaknesses, gather data on the intervention logic results that occurred, the assumptions and risks associated with the causal links and the other identified influencing factors.

**Step 6** Revise and strengthen the contribution story

- Build a more credible contribution claim based on the new data gathered
- Reassess its strengths and weaknesses, i.e., the extent to which the results, assumptions/risks and other influencing factors occurred
- Conclude on the strength of the intervention logic and the role played by other influencing factors and hence on the contribution claim
- If the evidence is still weak, revisit Step 5.

Source: Befani and Mayne (2014: :20) <sup>a</sup>. Befani and Mayne use the term Theory of Change. We changed this term to harmonise the terminology used in this thesis.

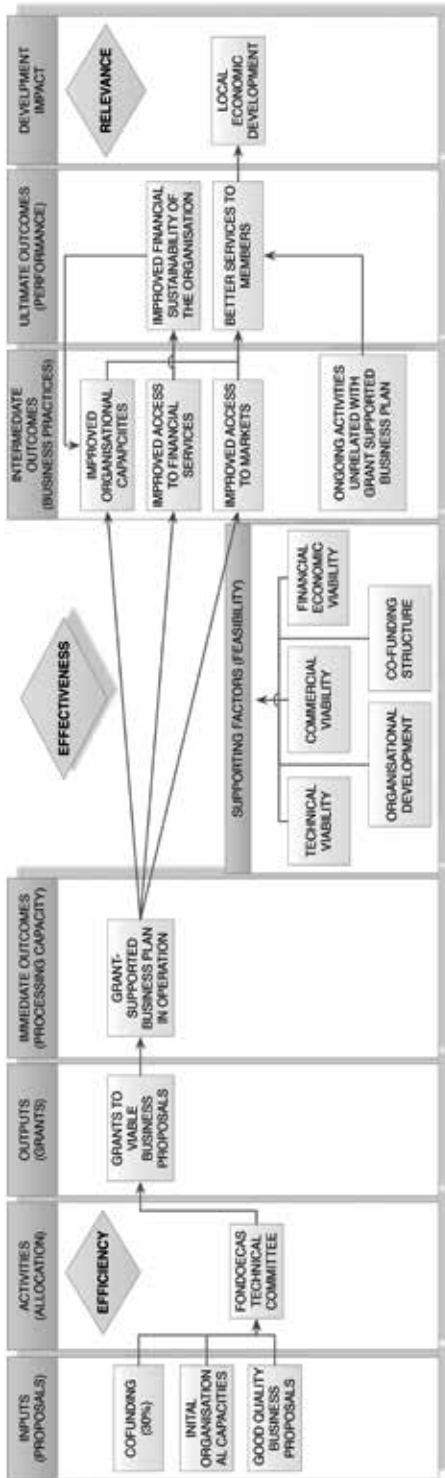
need to be verified and bolstered, gathers evidence to verify these; and reflects on the results. It is an exercise in which deductive and inductive research paradigms meet. Rohlfinger (2012) points to three essentially different types of research that are associated with different types of causal inferences: theory generating, theory testing, and theory modifying. In Contribution Analysis, all three types of research are combined. Contribution analysis checks multiple causal links along the intervention logic through a combination of research questions. It may include surveys, e.g., for net-effect estimates on outcomes that are directly influenced by the intervention, while using monitoring data, expert panels or other information sources to reflect on the contribution to outcomes that are outside the span of direct influence (Ton et al., 2014c).

### Intervention logic and key assumptions

We started by reflecting with the involved stakeholders (CIOEC, AOPEB, donors) on the rationale behind the grant fund in order to focus the evaluation questions. The grants are meant to help farmer groups to seize business opportunities. The two main eligibility requirements for beneficiaries were their status as a legal entity and a minimum of two years of collective marketing experience. This check on initial organisational strength, together with a lean and transparent fund management and an impartial analysis of the viability of the submitted business plan, were the key activities of FONDOECAS. By generating or expanding activities to process agricultural products, both income and organisational capabilities were expected to be enhanced. This would make it possible to improve service delivery and contribute to local economic development (Ton, 2012b).

Figure 8.1 depicts the grant-supported business plans as ‘immediate outcomes’, the technical and economic feasibility as ‘supporting factors’, the changed business practices as ‘intermediate outcomes’ and the changes in business performance as ‘ultimate outcomes’. We identified three interlinked assumptions of impact in the intervention logic of FONDOECAS, to guide our research. Each of these assumptions in the intervention logic needed a specific research design.

- **RELEVANCE** - The first assumption related to the expected impact on development of the collective marketing groups. They are assumed to contribute to a better income and nutritional status of the members – the development impact in the intervention logic.
- **EFFECTIVENESS** - The second assumption was that the grants would have a positive impact on organisational capacities, market access of members, the ability to pay for the costs of collective action, and improved access to financial service providers.
- **EFFICIENCY** - The third assumption was that the (lean) system of grant allocation would be efficient in selecting feasible business proposals to viable organisations. The feasibility analysis in the technical committee of experts was seen as the key institutional arrangement to target the grants.



Source: Author's own elaboration

Figure 8.1 FONDOECAS intervention logic and key assumptions about efficiency, effectiveness and contribution to development.

### Sample selection

Organisations will change in time, irrespective of whether they receive a grant. Therefore, it is useful to have a comparison group that can inform us about the changes that would most likely have taken place without the FONDOECAS grant. The literature (Khandker et al., 2009; Bamberger et al., 2004; Shadish et al., 2002) offers several quasi-experimental designs to estimate net-effects of interventions, where change is attributed to the intervention. Because of the diversity *in extremis* that characterises the sector of economic farmer organisations, most of these impact evaluation designs were not possible in our case<sup>15</sup>. The variance in most performance indicators registered was very high (see Table 8.3), which makes it impossible to implement a design with a comparison group that would have sufficient statistical power to detect net-effects. For example, even if we were to include all 200 CIOEC members in our sample, detectable effect-size in group sales would still lie far above expectations<sup>16</sup>.

**Table 8.3** Minimal effect sizes that could be detected, based on variance at baseline

Indicator	Mean	Std. Deviation	DetecTable effect size* n=200
Sample size			
Organisational capacities (TCC-score)	28.2	9.0	3.58
Annual group sales	US\$ 186,135	US\$ 300,634	US\$ 119,700
- unprocessed	US\$ 278,134	US\$ 386,523	US\$ 153,890
- processed	US\$ 36,642	US\$ 60,381	US\$ 24,040

\*.  $p < 0.50$ ; statistical power=0.8;

Therefore, we opted for a sample that focussed principally on the beneficiaries, complemented with a smaller group of unsupported organisations, also members of CIOEC, to feed counterfactual thinking. Moreover, instead of relying on group averages, we proposed comparative case studies with process tracing (Ton, 2012b; Vellema et al., 2013). When we started to implement the study, end 2010, we soon experienced problems of ‘contagion’, which reduced the number of organisations we could use in the comparison group. Several of these unsupported organisations had become grant solicitors in late 2010, early 2011. Initially, we were able to replace these with other organisations. However, later in 2011, after local researchers had been

15 Random assignment of grants, a requirement in a randomised control trial, would seriously compromise the legitimacy of the fund. Limiting the grants to a randomly selected sample of municipalities would negatively affect CIOEC’s image as an inclusive national organisation. Thus, an experimental impact evaluation design proved impossible. As a second option, we explored the possibility of selecting a comparison group based on the evaluation scores of the organisations that submitted proposals but were not awarded a grant. However, in our case, this design proved impossible due to the fact that most of the rejected applications never reached the Technical Committee for the scoring of the feasibility of proposals. Moreover, the few organisations that did get through this administrative selection and were rejected by the FONDOECAS evaluation committee qualified in their second or third attempt, having resolved the issues related to their business plan. Thus, a regression discontinuity design using the threshold scores on of the Evaluation Committee proved impossible. A panel study, comparing a grant beneficiaries with a comparison group, seemed the most feasible option.

16 The proponents needed to submit an estimate of the cash flow in the five subsequent years after being granted. The estimated increase of annual group sales was on average US\$58,509 (N=19; S.D.=108,988) with a median of US\$16,740.

contracted and were operational, this became more difficult. We report on the reasons for the exclusion of certain organisations from the sample in Annex 2. In 2013, the sample decreased by seven organisations. Three of these were beneficiary organisations on which we could not collect information because they had suffered an organisational collapse. We ended up with 26 grant beneficiaries and five organisations that had not (yet) received a grant.

Moderate sample sizes are not uncommon in other areas of social science, e.g., political science or organisation studies. Case-based comparative methods (Byrne and Ragin, 2009) and configurational comparative methods (Rihoux and Ragin, 2009) are tools used to explore patterns in a data-set. Cases are compared on characteristics and context conditions. We used Qualitative Comparative Analysis (QCA) to make (cautious) inferences about factors that may explain under what conditions and for what type of beneficiaries the FONDOECAS grants were more effective. In QCA, the researcher needs a good understanding of the characteristics and dynamics in each case-study to permit the required 'dialogue with the cases' (Ragin, 1987). Therefore, it is even recommended that the number of cases not be too high, in order to allow the researcher to obtain in-depth knowledge of each case (Rihoux and Ragin, 2009). We explain the method and results in detail in another chapter of this thesis.

## 8.4 Verifying the assumption of relevance

To understand the relevance of FONDOECAS, it is important to consider the unusual rural political constellation in Bolivia after 2006. There was a political-ideological debate about the role of economic farmer organisations versus traditional village organisations. Several preferential policies, such as co-funding of local business initiatives, were accessible only for village organisations, and not to economic farmer organisations with a more selective membership (Ton et al., 2013a). The explicit focus on economic farmer organisations reflected the conviction of the initiators and funders of FONDOECAS, who considered that economic farmer organisations were better suited for empowering smallholder farmers in markets than the traditional village organisations, and necessary for local economic development.

### Household survey

The academic literature is largely supportive of the assumption that independent economic farmer organisations are important for local economic development. For example, Bernard, De Janvry and Sadoulet (2010) show that the influence of village organisations negatively affects the performance of economic groups. This is in line with Woolcock (1998) who writes that the homogeneity and 'closure' characterizing ethnic communities may at some point stifle members' personal and business development. Many development programmes in Bolivia have learned by experience that farmer organisations need to be independent from the traditional village authorities in order to survive competition (Swen and Both, 1999; Healy, 2001; Flores et al., 2007; Bebbington, 1996; Llorenti Barrientos et al., 2005; Toornstra, 2000).

As a check whether economic farmer organisations were indeed considered by the rural population themselves as an important instrument for local economic development, we opted for a survey with Likert scale statements. Data was collected by piggybacking on a household

survey made by the consultancy firm *AnálisisReal-Latinoamérica* – AR-LAT, which managed a web-based training platform for village-based enumerators (many of them school teachers) to collect data on local economic development. We added a survey sheet with statements about the relevance of economic farmer organisations for local economic development and the characteristics of economic farmer organisations versus the village organisations. AR-LAT conducted the household survey between October 2012 and January 2013 (Aramayo, 2013). The survey could be conducted at low cost, as the results could be combined with data from a previous survey on 20 municipalities in 2008, in view of AR-LAT’s interest in producing an atlas on local economic development potential in Bolivia. The sampling strategy needed to be similar to that of their earlier surveys<sup>17</sup>. The survey resulted in a database of 1,945 observations from 39 municipalities (Aramayo, 2013). Without necessarily being representative of all producers in the municipalities, the data permits comparisons between households having a member in an economic farmer organisation or not (Table 8.4 and 8.5).

### Do economic farmer organisations contribute to local economic development?

Our direct question, “Do you consider the association/OECA an important instrument for economic development?”, was answered positively by two-thirds of the respondents. Surprisingly, there are no significant differences in opinion about this statement between respondents that had and those did not have membership in an economic farmer organisation. Sixty-five percent of the households that are members of an association consider this organisation to be more supportive to them than the traditional village organisation, the rural syndicate and/or the *ayllu*. Most of these members also consider that economic farmer organisations are among the best-functioning farmer organisations in the area. However, the answers to the question whether they would recommend their neighbours becoming a member reflects the mixed expectations towards economic farmer organisations in Bolivia: many households hold strong opinions on this statement. Non-members are a bit more negative than the households with a member in an economic farmer organisation. However, the majority of them agreed with the statement that they would recommend others to become a member.

On average, the characteristics of members of organisations are considered to be similar to average households. We asked their opinion on five statements to explore if land size, assets or political influence were different for members of economic farmer organisations. Two-thirds of the respondents disagreed with the statements expressing that these difference existed (Table 8.4), with a similar distribution of responses for households that have and those that do not have persons that participate in this type of organisation. Although most respondents indicated that the leaders of the association are highly committed to the community, the political tensions between village authorities and economic farmer organisations (Ton et al., 2013a) may well be reflected in the relatively high proportion (14%) of non-members that strongly disagreed with this statement. However, they are even more negative about the commitment

---

17 AR-LAT had used a peculiar sampling strategy for these surveys, see Aramayo R. (2008) Encuesta Fundamental para el Desarrollo Económico y Social: Informe 01. La Paz: INE-World Bank - CAN. They purposefully sampled three categories of respondents: households units, economic units and local government units. The economic units were selected in clusters of villages within the municipality, in areas with good agricultural production conditions, based on information provided by the local government.



Table 8.4 Survey results

	Statements	No-one member (N=1,007)				Someone member (N=938)				ANOVA		
		No opinion	Strongly disagree	Disagree	Agree	Strongly agree	No opinion	Strongly disagree	Disagree	Agree	Strongly agree	p-value
Relevance	The association/OECA is an important instrument for economic development in the area	224	10%	30%	47%	14%	166	7%	27%	55%	12%	
	The association/OECA is one of the farmer organisations in the area that functions best	933	100%	0%	0%	0%	166	7%	27%	55%	12%	p<.001
	The association/OECA supports me more than the <i>sindicato</i> or <i>ayllu</i>	1007	0%	0%	0%	0%	209	10%	25%	52%	13%	
	Not member, but someone in the household would be interested to become member of an association/OECA	268	0%	30%	70%	0%	789	0%	100%	0%	0%	
	I would recommend my neighbours to become member of the association/OECA	419	18%	25%	55%	2%	311	14%	22%	55%	10%	p<.001
Member characteristics	The association/OECA consists of a group of families that have different characteristics than the average in the area	248	7%	57%	26%	10%	250	8%	57%	24%	11%	
	The members of the association/OECA have more land or working capital than average in the area	317	8%	55%	26%	11%	277	9%	56%	22%	12%	
	The association/OECA have better houses than average	303	9%	64%	19%	8%	297	7%	71%	18%	4%	p<.050
	The members of the association/OECA have more political influence than average	332	10%	53%	28%	10%	380	12%	51%	31%	7%	
Social performance	The leaders of the association/OECA have ample commitment to the village	233	14%	36%	50%	0%	187	9%	30%	58%	4%	p<.001
	The leaders of the association/OECA have ample commitment to their members	1007	0%	0%	0%	0%	169	0%	33%	52%	14%	p<.001
	The leaders of the <i>sindicato/ayllu</i> have ample commitment to the village	200	11%	34%	56%	0%	176	11%	37%	52%	0%	
	The leaders of the <i>sindicato/ayllu</i> have ample commitment to the members of the association	774	0%	100%	0%	0%	204	0%	42%	48%	10%	p<.001
	The association/OECA has a good relation with the <i>sindicato/ayllu</i>	482	0%	53%	47%	0%	398	0%	47%	53%	0%	

Note: For Spanish original questions, see Annex 5

**Table 8.5** Opinion of farmers on functions of economic farmer organisations

FUNCTIONS	No-one member (N=1,007)					Someone member (N=938)					ANOVA Stat. sign.
	No opinion	Strongly disagree	Disagree	Agree	Strongly agree	No opinion	Strongly disagree	Disagree	Agree	Strongly agree	
The association/OECA is a mean that helps me to get support (aid)	1007	0%	0%	0%	0%	0*	0%	39%	54%	8%	p<.001
The association/OECA is a mean that helps to access markets	1007	0%	0%	0%	0%	0*	0%	60%	30%	10%	p<.001
The association/OECA is a mean that helps me in the social life of the village	195	3%	42%	46%	9%	178	6%	36%	47%	11%	
The association/OECA is a mean that helps me to improve production	1007	0%	0%	0%	0%	190	0%	26%	64%	11%	p<.001
I like to sell my products in an organised way together with others	287	7%	27%	59%	7%	271	7%	25%	61%	8%	

\*We cannot discard a data entry error on this statement. 'No opinion' had been coded as 'disagree'. This was corrected for households 'no-one member' but not for 'someone as member'. However, when using the two subsequent statements as control (leaving out the 'no opinion' respondents on both questions from the analysis) we see a similar distribution of opinions. Note: For Spanish original questions, see Annex 5

of leaders of village organisations. When we only compare respondents that had an opinion on both statements, we see that, overall, the opinion is slightly more positive about leadership in economic farmer organisations ( $p < 0.05$ , using a paired sample T-test). Interesting is also that most farmers, also the households without current membership in an economic farmer organisation, indicated that they would like to sell their products in an organised way. Table 8.5 indicates that only a minority considered the association as a means to gain access to markets. More prominent was the association's function in improving production and accessing outside support. Moreover, they are considered valuable aspects of social life, with 55% of the responding households that did not participate in an economic farmer organisation considering them positively in this respect.

### Conclusions on relevance

The survey data is supportive of the key assumption in the FONDOECAS intervention logic that farmer organisations are an important component of the social capital in rural areas and are relevant for local economic development. The finding shows that even though their identity as economic farmer organisations (CIOEC-Bolivia, 2010) is framed around their pivotal role in creating access to markets, they are only considered instrumental in doing so only by a minority of the households. Nevertheless, the willingness to engage in some sort of collective marketing is high, which supports the assumption that grant support to facilitate the emergence and development of economic farmer organisations is indeed a relevant activity.

## 8.5 Verifying the assumption of effectiveness

### Case-based comparative analysis

To assess the effectiveness of the grants, we proposed to apply case-based analysis on 50 organisations, a group of 30 beneficiary organisations and a comparison group of 20 non-supported organisations. The case-based analysis does justice to the complex change processes related with the grant, and the comparison group would help to reflect on the casual relationship between these changes and the grant (Ton, 2012b). The only performance variable with which we expected to detect a net effect in the observational study, was the 'tension containment capacity' (TCC) used to measure organisational strength in managing collective marketing activities (Ton, this thesis). We expected that the challenges needing to be resolved by rules and regulations within the group would increase due to the new, additional activities of the group: this would be a learning process during which the groups would improve their capacity to manage collective marketing activities.

Our analysis is based on information from 26 beneficiary organisations and five (relatively) unsupported organisations. We used the time-series data (see Annex 4) and qualitative interviews to assess, for each intended outcome, if the grant could be considered as a contributory factor for success. To triangulate this interpretative analysis, we included self-assessment questions (see Annex 3 and 4), in which we asked if changes in performance were related to the grant-supported business plan.

We assessed the effectiveness of FONDOECAS grants on four outcomes: organisational strengthening, increased market access for members, improved capacity to pay organisational expenses, and new access to financial service providers. These are the intermediate outcomes in the intervention logic presented in the introduction, and depicted in Figure 8.1.

Being aware of the subjective and normative element in the assessment of success and failure, we used a process to check and bolster the evaluation. Two researchers, the author of this paper (GT) and the local researcher (LF), assessed each case independently (see Chapter 7 for details) on three of the four intended outcomes. The outcome, ‘access to financial service providers’ was left out due to limited information provided in the interviews on the process of taking loans. The two relatively independent assessments of the change processes in each of the organisations were compared and reconciled. Table 8.6 shows the Cohen’s kappa scores, as an indicator of inter-rater agreement. Most differences were due to differences in access to information (GT had more detailed knowledge about the time-series data) or substantive knowledge about the cases (LF had conducted all the interviews and knew more about the context and process of the changes). The agreement between both researchers, according the common interpretation of the kappa-score (Landis and Koch, 1977) was ‘moderate’ for the outcomes enhanced market access and improved organisational capacities, and ‘fair’ for the increased capacity to pay organisational expenses. The latter low kappa-score was due to different interpretations of ‘success’ on this outcome, which was resolved during reconciliation.

**Table 8.6** Cohen’s kappa scores of agreement in valuations<sup>a,b</sup>

Outcome	LF original versus reconciled	GT original versus reconciled	LF original versus GT original	Arguments used for reconciliation
Grant enhanced market access	0.752	0.698	0.481	One difference (AAAT) was due to an erroneous interpretation of the use of the investment in the shop. Another due to considering different time intervals and grant investments (CECAOT). And a dairy plant (CEPLACH) appeared to have several milk producing members who sold to the group but the grant served to build and relocate the place of operation, which negatively affected sales. ORLIPA accessed the local school feeding programme but appeared to have done so without the products from the grant supported business plan.
Grant improved organisational capacities	0.833	0.586	0.407	Two cases (CELCCAR, CIAPEC) did not use the grant investment, though they continued with the supported business plan. In two cases the interviewees expressed the importance of the decision making process around the grant to discuss internal group pressures, even though the grant did not contribute to production.
Grant increased capacity to pay organisational expenses	0.257	0.923	0.224	We noted a difference in interpretation of the question between the two researchers. During reconciliation, it was agreed that the capacity to pay expenses will increase when the level of sales increases due to the grant, even though in most organisations the total amount of expenses or member income did not change.

a. Interpretation of Cohen’s kappa: <0.00 = Poor agreement; 0.00–0.20 = Slight agreement; 0.21–0.40 = Fair agreement; 0.41–0.60 = Moderate agreement; 0.61–0.80 = Substantial agreement; 0.81–1.00 = Almost perfect agreement (Landis and Koch, 1977)

b. Any fuzzy scores were converted to crisp scores before calculating the Cohen’s kappa

## Did the grants improve organisational capacities?

To verify the assumption that grants are effective in improving organisational capabilities, we collected panel data in two rounds, 2011 and 2013. In each round, we conducted in-depth interviews on the dynamics within each organisation surrounding 10 agency dilemmas that are common in collective marketing arrangements (Ton, 2010b). For each organisation, this resulted in two interview reports, with thick descriptions of the organisational dynamics and the internal rules and regulations to address each agency dilemma.

We summarized this qualitative information in a summary sheet, called an 'organisational radiography' (Annex 3). We mapped the relevance of each agency dilemma, in order to monitor the 'breadth' of organisational competencies. Based on the five agency dilemmas (inherent tensions) that were consistently present, we calculated a score to assess the depth of the organisational capacities, labelled as Tension Containment Capacity (TCC). The TCC-score is used to monitor change in time and for cross-case comparative analysis. The design and validation of the construct is described in chapter 6 of this thesis.

We expected that as a result of the investment in value-added production, in addition to their current activities, the number of relevant agency dilemmas would increase. To our surprise, we found a decrease of the average number of agency dilemmas that were relevant. Whereas in 2011 the grant beneficiaries had an average of 7.0 agency dilemmas, in 2013 this was only 5.8. The average change was significantly different from zero (paired sample test  $p=0.05$ ). However, the interviewed board members indicated that many of these changes were considered to be unrelated to the grant. They may have changed as a result of a wide range of possible intervening factors and actors, such as market dynamics, context-specific political dynamics, or support from other agencies.

In Table 8.7 we distil only the organisations that indicated that the change in their capacity to handle an agency dilemma was related to the grant-supported business plan. Eight different organisations mentioned a grant-related change in one or more agency dilemmas. In six organisations, the grant was related to an agency dilemma that had become less problematic. Five of these six organisations reported that they had fewer problems with quality assurance ( $T_2$ ). This is in line with FONDOECAS' focus on supporting added value through processing, and paying attention to the niche markets of public-procurement programmes. Processing creates the need to improve the quality of the products supplied by members and develop rules and regulations that are effective in doing so.

Other agency dilemmas that changed due to the grant-supported activities were the issue of working capital constraints ( $T_3$ ) by ASPASA<sup>18</sup>, the issue of differentiating services between

---

18 ASPASA indicated that they had fewer problems surrounding the working capital constraints. They related this change to FONDOECAS. However, they never received a grant but received a loan from FONDOECAS' new credit window, which started to operate in 2012 alongside the grant fund. This explains the improvement. They sold their quinoa through the national association of quinoa farmers, ANAPQUI, who paid an advance payment to the farmer, to be completed after the export process had been finalised and the product had been paid by the international client. With this trade capital accessed through FONDOECAS, ASPASA was able to sell to a private exporter based in La Paz, who offered a better price than ANAPQUI, and, also paid them in cash on the day of purchase. Curiously, thus, the loan increased the capacities of ASPASA to manage collective marketing operations directly as a grassroots organisation, but it reduced the amount of quinoa that was exported through collective marketing in their federation, ANAPQUI.

**Table 8.7** Changes in organisational capacities related to grant-supported business plan

Agency dilemma	Outcome	Agency dilemma is relevant	Change is unrelated to grant support	Change is related to grant support	Name of the organisation that reported a relation with the grant-supported business plan.
T <sub>1</sub> - 'Regulating Member Supply'	More problems	4	4	0	
	No change	4	4	0	
	Fewer problems	3	1	2	AOCEMM, ARAO
T <sub>2</sub> - 'Quality Assurance Systems'	More problems	4	3	1	ADAPICRUZ
	No change	13	9	4	COMART, AAAT, ASAFOP, CEMUR
	Fewer problems	9	4	5	AOCEMM, ARAO, APROQUIRC, APROAMOL, CEPLACH
T <sub>3</sub> - 'Coping with Working Capital Constraints'	More problems	3	2	1	AMAGA
	No change	18	16	2	ARAO, CEPLACH
	Fewer problems	4	3	1	ASPASA
T <sub>4</sub> - 'Anticipating Side-Selling'	More problems	6	6	0	
	No change	10	9	1	ADAPICRUZ
	Fewer problems	0	0	0	
T <sub>5</sub> - 'Ways to Use Profits'	More problems	3	2	1	AMAGA
	No change	15	15	0	
	Fewer problems	2	2	0	
T <sub>6</sub> - 'Differentiating Services to Members and Non-Members'	More problems	2	1	1	AMAGA
	No change	10	8	2	AOCEMM, ADAPICRUZ
	Fewer problems	1	1	0	
T <sub>7</sub> - 'Decision Making on Activities that Benefit Only a Sub-group'	More problems	2	1	1	AMAGA
	No change	8	4	4	AMLECO, ASAFOP, CIAPEC, ORLIPA
	Fewer problems	2	1	1	AOCEMM
T <sub>8</sub> - 'Task Delegation and Supervision of Professional Staff'	More problems	4	4	0	
	No change	14	14	0	
	Fewer problems	5	5	0	
T <sub>9</sub> - 'Liability in Contracts and Loans'	More problems	1	1	0	
	No change	10	8	2	ASPASA, ARAO
	Fewer problems	1	0	1	COMART
T <sub>10</sub> - 'Managing Political Aspirations'	More problems	8	8	0	
	No change	11	11	0	
	Fewer problems	2	1	1	COMART

Source: Based on 2013 interview data. Valid N=28 (Missing data for FENCA, AGAYAP and OMCSA)

members and non-members ( $T_7$ ) by AOCEMM<sup>19</sup>, the issue of distribution of legal responsibilities in contracts and loans ( $T_9$ ) and the issue of managing political aspirations ( $T_{10}$ ) by COMART<sup>20</sup>. Two organisations indicated that the grant was related to an area in which they experienced more problems. ADAPICRUZ mentioned that there were more problems with quality assurance due to the implemented business plan. They were in the process of organic certification of their honey products and experienced that the requirements in this market forced them to be more demanding. In AMAGA, problems around various agency dilemmas had increased between 2011 and 2013<sup>21</sup>, after the processing plant started to operate. Interestingly, an application for a second FONDOECAS grant in 2012 had catalysed the awareness of these problems within AMAGA and provided a context within which to address them.

*“[Recently] we presented ourselves again to FONDOECAS, but they rejected us because we did not have the formal resolution of conformity, signed by all members. We will have to define our membership, see who clearly is a member, and what members need to comply with. It can’t be that all believe themselves to be a member (..) The issue is that those that want to work are in, and not those that only want benefits, and harm the organisation.”* (Interview with AMAGA, 2013)

As explained above, we computed a quantitative measure of organisational strength, the TCC-score, based on the information on the five core agency dilemmas that proved to be relevant across different sectors and were consistently interpreted by different researchers in 2011 and 2013. Table 8.8 shows that the average change ( $\Delta TCC$ ) in grant beneficiaries is -21%, with a change of 5.9 points less than the average TCC-score of 27.7 in 2011 ( $p < 0.01$ ). Most organisations appear weaker in 2013 than in 2011. Acknowledging the limits in accuracy of the TCC-construct, we applied a margin of 15% up or down to classify the change between 2011 and 2013 as ‘the same’. Six grant recipients increased their tension containment capacities, six stayed the same, and 14 had a lower score.

Table 8.8 summarizes the quantitative information used to assess the effectiveness of FONDOECAS in strengthening organisational capacities, using the information from the two interview reports. Independently, the two researchers evaluated the outcome in each case, reflecting on the quantitative data and the qualitative information obtained in the interviews on the organisational dynamics and grant implementation process. In the column ‘reconciled verdict’ we present the result of this case-by-case analysis. This verdict was based not only on

---

19 AOCEMM upgraded their honey processing units with the grant received from FONDOECAS. This made it possible to satisfy the food safety requirements of the buyer, the governmental food programme. They needed to buy more honey than the members produced and, therefore, they introduced a pricing system that differentiated between members and non-members.

20 COMART is a second tier handicraft organisation that received two FONDOECAS grants. One in 2007 to equip their shop, and a second grant to support 20 of its member organisations with small investments (e.g., lighting equipment). The first grant was invested in (re)furnishing their shops in down-town La Paz, renewing products and revising stock level, which explains the improvement in the quality assurance system. The second grant, ‘gifts’ to member organisations, apparently facilitated the political aspirations of some leaders. And, as such, it is an unintended outcome from the perspective of (the back-donors of) FONDOECAS, which tried to prevent the use of the grant for political patronage.

21 AMAGA manages a diary processing plant. After the processing plant started to operate, it faced increasing problems in procuring raw milk due to lack of working capital, had an on-going discussion with its members who demanded the distribution of profits and caused a deadlock in decision-making processes related with key investment decisions.

Table 8.8 Changes in organisational capacities between 2011 and 2013

NAME OF ORGANISATION	Relevant agency dilemmas (out of 10)		Relevant core tensions		TCC score	Relevant agency dilemmas (out of 10)		Relevant core tensions		TCC score	ΔTCC score	ΔTCC from 2011 level	Significant change in TCC	Grant is related to agency dilemma	Reconciled verdict: Grant contributed to organisational strengthening?
	Mean (S.D.)	7.0 (1.7)	4.2 (.97)	28 (9.6)	2011	2013	5.8 (2.6)	3.6 (1.6)	2013	22 (10.0)	-6.6 (11.3)	Positive: 6 Neutral: 6 Negative: 14			
GRANTEES	Mean (S.D.)	7.0 (1.7)	4.2 (.97)	28 (9.6)	2011	2013	5.8 (2.6)	3.6 (1.6)	22 (10.0)	-6.6 (11.3)	Positive: 6 Neutral: 6 Negative: 14			Yes: 10 No: 16	
AAAT	6	4	27	3	24						+/-	-11%	T <sub>2</sub>	yes	
ADAPICRUZ	7	5	42	4	22						-	-48%	T <sub>2</sub> , T <sub>4</sub> , T <sub>6</sub>	yes	
AGAYAP	8	4	27	0	0						-	-100%		no	
AMAGA	5	3	18	4	20						+/-	11%	T <sub>3</sub> , T <sub>5</sub> , T <sub>6</sub> , T <sub>7</sub>	yes	
AMDESQY	5	2	12	3	6						-	-50%		no	
AOCEMM	10	5	36	5	31						+/-	-14%	T <sub>1</sub> , T <sub>2</sub> , T <sub>6</sub> , T <sub>7</sub>	yes	
APAM MIZQUE	5	3	17	5	31						+	82%		no	
APCA	8	5	39	5	26						-	-33%		no	
APROAMOL	3	3	12	4	18						+	50%	T <sub>2</sub>	yes	
APROQUIRC	6	4	22	3	21						+/-	-5%	T <sub>2</sub>	yes	
APSU	8	5	20	2	15						-	-25%		no	
ARAO	6	2	12	3	18						+	50%	T <sub>1</sub> , T <sub>2</sub> , T <sub>3</sub> , T <sub>9</sub>	yes	
ASAFOP	8	4	36	5	27						-	-25%	T <sub>2</sub> , T <sub>7</sub>	yes	
ASOCOM	6	4	28	3	12						-	-16%		no	
CECAOT	8	5	36	5	33						+/-	-8%		no	
CELCCAR	8	5	39	5	26						-	-33%		no	
CEMUR	8	5	27	3	16						-	-41%	T <sub>2</sub>	no	
CEPLACH	4	3	24	3	17						-	-29%	T <sub>2</sub> , T <sub>3</sub>	no	
CIAPEC	8	5	33	4	20						-	-39%	T <sub>7</sub>	no	
COAINE	6	4	30	5	39						+	30%		no	
COMART	8	4	27	5	31						+	15%	T <sub>2</sub> , T <sub>9</sub> , T <sub>10</sub>	yes	
COPROQUINACC-T	6	4	19	5	27						+	42%		no	
INCA PALLAY	10	5	39	5	36						+/-	-8%		yes	
OMCSA	8	5	33	0	0						-	-100%		no	
ORLIPA	8	5	32	3	15						-	-53%	T <sub>7</sub>	no	
SOPPROQUI	9	5	45	5	30						-	-33%		no	
NON-GRANTEES	Mean (S.D.)	8.0 (1.2)	4.6 (.54)	28 (5.6)	2011	2013	7.4 (2.1)	4.4 (.89)	26 (5.2)	-2.4 (6.8)	Positive: 0 Neutral: 4 Negative: 1				
AMLECO	8	4	21	3	19						+/-	-10%	T <sub>7</sub>	loan 2012 grant 2012	
APME	6	5	31	5	31						+/-	0%		no grant	
ASOPROF	9	4	27	5	30						+/-	11%		loan 2012	
ASPASA	9	5	27	4	28						+/-	4%	T <sub>3</sub> , T <sub>9</sub>	loan 2012	
FENCA	8	5	36	5	22						-	-39%		no grant	



quantitative information, but also the thick descriptions of the change process, as documented in more detail in Chapter 7. Except for two organisations (CECAOT<sup>22</sup>, INCA PALLAY<sup>23</sup>), the final verdict is in line with the panel data collected on tension containment capacities as well as the self-assessments of board members.

The data show that, on average, the organisations had more problems in resolving agency dilemmas in collective marketing with effective rules and regulations in 2013 than in 2011. However, most of these changes were not related to the grant-supported business plans. The grants did trigger improvements in organisational capacities to manage collective marketing, but only in 10 of the supported organisations. The effects were largely related to the issue of quality assurance. Other improvements were on the issue of differential pricing between members and non-members, and the ability to focus investments on subgroups within their membership.

Four of the five non-beneficiaries were organisationally stable, considering their tension containment capacity in 2011 and 2013. Only non-grantee FENCA showed a decline, because it suffered a breakdown of collective marketing activities as a result of the emergence of parallel rice-procuring farmer associations initiated by the state-company EMAPA (see Annex 1).

### Did the grants improve access to markets?

The average group sales presented a tendency of growth in most of the organisations studied. The differences in growth rate per organisation varied greatly (see Table 8.9). The strongest growth of sales occurred in the quinoa exporters (COPROQUINACC, ASPASA, SOPROQUI, CECAOT and APROQUIRC), especially in the 2011-2012 period, when quinoa prices more than doubled. We see that the strong growth in the coffee cooperatives between 2008-2010 period slowed down in the period 2011-2012. This change is due to the implosion of the international coffee price during that period. Also, the growth rates of the honey processor ADAPICRUZ levelled out in 2011-2012<sup>24</sup>. Growth rates were extremely high for AMDESOY, a small soy processing enterprise, more than doubling their sales each year. However, the absolute value of their sales remained still relatively low.

The average growth rate of total group sales seems a good proxy for commercial health of the organisations. However, it is a poor indicator for the effectiveness of the FONDOECAS grants, especially when the grant investments had been directed to new business activities which were complementary to existing, traditional economic activities. Therefore, in Table 11, we make a distinction between processed and unprocessed products. We also indicate if the interviewed board members considered the change in turnover to be related to the grant.

---

22 The grant in CECAOT was invested in an optic-sensor to control quinoa quality, but it broke down within one year.

23 INCA PALLAY indicated in the interviews that investment in a production unit in one of the villages did not increase their capacities to manage collective marketing (they are already among the strongest) but it helped to resolve a conflict with the local *sindicato*, which had threatened to seize the equipment of the handicraft group.

24 In 2013 ADAPICRUZ reported that they had too much honey in stock and operated in a constrained market, where the public food programme was an attractive but stagnant buyer. Alternative markets (e.g., organic exports to Argentina) were being explored but not yet accessed.

**Table 8.9** Group sales in the period 2008-2012

Sector	Name	Annual group sales (US\$)					Average yearly growth rate (%)	
		2,008	2,009	2,010	2,011	2,012	Benef.	Comp.
Agriculture	ASOPROF	--	--	--	1,000,000	1,700,000		70%
	OMCSA	--	2,500	2,971	3,800	1,500	1%	
	FENCA	--	--	142,857	142,857	0		-50%
Cattle	AGAYAP	0	0	0	17,423	14,286	-18%	
	APCA	--	--	45,714	51,429	41,429	-3%	
	ORLIPA	0	2,286	3,071	5,857	10,214	66%	
Coffee	CELCCAR	285,714	285,714	372,143	443,571	501,429	16%	
	CIAPEC	400,000	600,000	802,857	1,302,857	1,331,428	37%	
	COAINE	608,000	714,143	1,409,571	1,388,571	1,116,857	23%	
Dairy	AMAGA	3,214	3,143	4,600	3,551	4,371	11%	
	AMLECO	--	--	339,000	355,714	367,143		4%
	CEPLACH	11,429	11,429	11,429	10,000	4,286	-18%	
Handicrafts	AAAT	28,571	42,857	64,324	61,830	45,792	18%	
	APSU	--	--	18,571	18,571	8,857	-26%	
	ARAO	60,532	71,420	83,322	88,923	97,793	13%	
	COMART	214,286	228,571	207,143	192,857	171,429	-5%	
	INCA PALLAY	69,202	82,262	72,748	89,034	97,309	10%	
Honey	ADAPICRUZ	36,286	114,857	145,429	197,714	205,714	71%	
	AOCEMM	47,143	45,181	31,991	45,994	55,015	8%	
	APAM MIZQUE	10,257	11,643	8,786	17,357	20,857	27%	
	APME	--	30,649	34,987	37,653	45,050		14%
	APROAMOL	0	0	2,934	2,649	2,220	-13%	
Processed foods	AMDESOY	--	1,102	3,749	4,412	10,286	130%	
	ASAFOP	47,143	51,429	0	28,571	38,571	-19%	
	CEMUR	171,429	114,286	142,857	204,224	217,448	10%	
Stone mining	ASOCOM	185,714	167,143	204,286	214,286	242,857	8%	
Quinoa	APROQUIRC	457,143	514,286	571,429	720,000	1,300,000	33%	
	ASPASA	0	0	100,000	200,000	500,000		125%
	CECAOT	142,857	178,571	172,000	186,286	795,714	89%	
	COPROQUINACC	128,571	135,714	200,000	205,714	457,217	45%	
	SOPPROQUI	600,000	500,000	571,429	1,400,000	1,600,714	39%	
Average annual growth rate							21%	33%
Standard deviation							(36%)	(67%)

Note: -- = no data available

Interestingly, in many instances this access to new markets was mediated by preferential public procurement policies.<sup>25</sup> Government procurement programmes can stimulate economic farmer organisations, which is most notable in the honey producers, but it also creates dependency. We see large fluctuations in the yearly sales of ASAFOP (bakery) and CEPLACH (yoghurt and cheese) which are the result of the discontinued contracts for deliveries to the school meal programmes. The latter illustrates the vulnerability of this niche market, especially for small organisations, due to the inherent administrative and political dynamics of government procurement programmes (Prudencio and Elías, 2014)± .

When we look at the comparison group, we see that the average growth rate is similar but with marked differences between the steady growing ASOPROF and ASPASA, the stagnant APME and the collapsed FENCA. None of these had value-added processing activities (see Table 8.10).

As described above, for each case we ‘traced’ the process of change and evaluated if the grant could be considered a contributory factor to improved market access of members. Market performance is influenced by more factors than FONDOECAS. Therefore, there might be some discrepancy between the trend in sales of an organisations and our evaluation of the successfulness of the grant in improving market access for members. Table 8.10 presents the reconciled verdict of the researchers based on data and qualitative information about the change process in each organisation (see Annex 1 for further details). In six cases, the ‘final verdict’ differed markedly from the tendency of growth and the self-assessed relationship with the grant. In these organisations, the grant could not be considered as a contributory factor. For example, the three coffee organisations registered an increase in sales of processed products (roasted coffee). However, the amount of coffee used in this processing was insignificant when compared with the amount of coffee sourced from their members for export. And, in two of them (COAINE, CIAPEC) the equipment bought with the FONDOECAS grant was in fact never used in processing for external clients. Furthermore, three organisations active in food processing bought their inputs from non-members in the local market. They clearly gained market access as micro-enterprises, but did not create market access for their members<sup>26</sup>. Only for five grant recipients did the grant indeed improve market access for members (APCA, ARAO, ADAPICRUZ, AOCEMM and APROQUIRC).

---

25 This is especially relevant for the government procurement of honey, dairy products and quinoa in the national Subsidy for Lactating Mothers programme -*Subsidio de Lactancia*-, and dairy products and processed food in municipal School Meal programmes -*Desayuno Escolar*. CIOEC has a long advocacy trajectory around this issue (Elías, 2007; CIOEC-Bolivia, 2010; Prudencio and Elías, 2014), which explains the close contacts with those farmer organisations and rural micro-enterprises that are involved in this, even very small ones. In many regional offices, CIOEC participated in public-private partnerships to promote the inclusion of local producers in public nutritional programmes (Garafulic and Bredow, 2006; Sánchez, 2006). The CIOEC’s regional branches promoted the grant funding opportunities, while they were helping these organisations to comply with the administrative and qualitative requirements in the procurement process. In the inception document (Ton, 2005) there was explicit mention of the obstacles for economic farmer organisations that had tried to access government procurement markets.

26 ORLIPA was classified as unsuccessful on this outcome of market access. It had received the grant to enter the school meal programme. Indeed, they did, but not with the product for which they had bought the equipment. This product (based on dried meat and broad beans) had been removed from the school menu. ORLIPA worked only as an intermediary, delivering grocery products to the school meal programme, not member products.

**Table 8.10 Group sales (turnover) for unprocessed and processed products during 2008-2012a**

Sector	Name	Grant contract year	Group sales unprocessed products in USD*					Group sales processed products in USD*					Reconciled outcome		
			2,008	2,009	2,010	2,011	2,012 related?	2,008	2,009	2,010	2,011	2,012 related?	Improved market access of members?	Increased income to pay organ. costs?	
Agriculture	ASOPROF	no grant			1,000,000	1,700,000								no grant	no
	FENCA	no grant			142,857	142,857	0							no grant	no
	OMCSA	2009/2011	0	0	0	0	0	1,000	1,471	2,300	0			no	yes
Cattle	AGAYAP	2008	0	0	17,423	14,286	0							no	no
	APCA	2009**	0	1,429	1,571	4,286	0							no	no
	ORLIPA	2009**	0	1,429	1,571	4,286	Much	40,000	42,857	31,429	Much	no	no		
Coffee*	CELCCAR	2009	285,714	285,714	371,429	442,857	500,000	Much	0	857	1,286	1,429	5,714	Much	no
	CIAPCC	2009**	400,000	600,000	800,000	1,300,000	1,328,571	No	0	2,856	2,856	Much	Much	no	no
	COAINE	2009	590,000	695,000	1,390,000	1,371,429	1,097,143	No	0	2,000	2,429	0	1,143	Much	no
Dairy	AMAGA	2008	0	0	0	0	0							no	yes
	AMLECO	no grant**			0	0	0							no grant	no
	CEPLACH	2008	0	0	0	0	0							no	yes
Handicrafts	AAAT	2007/2009	0	0	0	0	0							no	yes
	APSU	2009	0	0	0	0	0							no	no
	ARAO	2011/2012	0	0	0	0	0							no	no
Honey	COMART	2009/2011	0	0	0	0	0							no	yes
	INCAPALLAY	2007	0	0	0	0	0							no	no
	ADAPICRUZ	2009	22,857	100,000	125,714	168,571	171,429	No	12,857	14,286	17,143	25,714	30,857	Much	yes
Proc. foods	AOCEMM	2008**	44,381	31,673	44,880	54,915	Much							yes	yes
	APAM IZQUE	2010	6,286	8,000	4,286	5,143	3,429	No	3,714	3,429	4,286	12,000	17,143	No	no
	APME	(2012)	30,004	34,380	37,275	44,286	No	0	0	0	0	0	0	no grant	no
Stones	APROAMOL	2010	0	0	2,857	2,571	2,143	Much	0	0	0	0	0	no	yes
	AMDESQY	2009/2011	0	0	0	0	0							no	yes
	ASAFOP	2009	0	0	0	714	0							no	yes
Quinoa*	CEMUR	2009**	0	0	0	0	0							no	yes
	ASOCOM***	2008	0	0	0	0	0							no	no
	APROQUIRC	2007	457,143	514,286	571,429	714,286	1,300,000	Much	0	0	5,714	0	0	yes	yes
Quinoa*	ASPASA	no grant**	0	100,000	200,000	500,000	Much							no grant	no
	CECAOT	2009/2012	142,857	178,571	171,429	180,000	757,143	No	0	0	571	571	28,571	Little	no
	COPROQUINACC	2007	128,571	135,714	200,000	205,714	457,217	No	0	0	0	0	0	no	no
SOPPROQUI	2008	600,000	500,000	571,429	1,400,000	1,600,000	No	0	0	0	0	286	No	no	no

a. Empty cells indicate missing information; zero indicates no sales in the category; shaded boxes indicate the year when the first grant was received.  
 \*. Coffee beans and quinoa prepared for bulk export are considered as unprocessed. Raw honey is also considered as being unprocessed, except when innovatively packaged.

\*\* These organisations received a FONDOECAS loan in 2012.

\*\*\* The self-assessed relationship refers to FONDOECAS credit facility not the grant.

\*\*\*\* The equipment (compressor) was never used, being too heavy to carry. The answer is inconsistent.

Yes: 5  
 No: 21

Yes: 12  
 No: 14

As we showed in Chapter 7, the successful outcomes were found only among organisations that sourced their raw material from members. Also, small organisations proved more likely to be unsuccessful. The old, large and strong organisations that already had high organisational capacities when awarded the grant proved particularly unsuccessful. In many of these organisations, the grant had been used for complementary processing, a secondary economic activity alongside their primary activities in bulking and exports (e.g., quinoa, coffee). Interesting, these organisations indicated the limited amount of the grant (USD 10,000) as a reason for the failure of the business plan. The grant had resulted in under-scaled investments. With the grant alone, the supported business activities could not get the production volume needed for a commercially viable market launch.

### Did the grants increase the capacity to pay organisational expenses?

Outcomes were more positive considering the grant's contribution to the capacity to pay organisational expenses, such as office supplies, expenses related with communication and travel, or personnel. Annual organisational expenses are modest, with an average of US\$2,411 (S.E. US\$448) in 2012. ARAO, a handicraft organisation, registered the highest amount with US\$8,962. Handicraft organisations show relatively high costs because they need to rent shop space. On average, payments to hired staff were 46% (S.E. 6%) of total organisational costs. Small non-sourcing microenterprises that sell processed food (AMDESOY, ASAFOP, CEPLACH) had low costs, because members do most of the work.

Table 8.10 shows that for 12 organisations the grant contributed to an increased ability to pay organisational costs. Next to the sourcing organisations, several non-sourcing organisations also benefited from the grant on this outcome area. Two of the 12 organisations need some explanation. CEPLACH registered falling sales, partly as a result of the suboptimal location of their new infrastructure (processing unit and shop). However, before acquiring this property, they had paid rent. The grant was valued very positively by the members, especially because it had reduced their recurrent organisational costs. In contrast, APCA had improved access to markets for members, increasing their bargaining power in price negotiations with procuring companies, with the computerised fibre measurement tool that they bought with the grant. However, in 2012, they suspended their collective marketing activities, and, therefore, lost the mechanism to raise group income to pay organisational costs<sup>27</sup>.

Almost two-thirds of the implemented business plans were successful in generating group income, representing 41% of the grant recipients. In Chapter 7 we showed that particularly the sourcing organisations were unsuccessful. This can be explained by the agency dilemmas in collective marketing: non-sourcing organisations have the advantage that they can use all benefits to pay organisational costs and salaries and do not have to negotiate the collective interests of the group with the short-term interest of the members in a higher price of the raw material that is sourced from them. Also, the Qualitative Comparative Analysis (QCA)

---

27 APCA suffered from increasing costs of processing. They subcontracted spinning and colouring of wool to third parties, first to another farmer organisation, COPROCA, and, later, a private Peruvian company. However, in 2012, they ceased to do so, and only sold unprocessed alpaca wool. In 2013, they even stopped the procurement of wool from their members. They are trying to raise funds from donors to install their own processing unit.

indicated that the grant had proven particularly unsuccessful for the larger and stronger organisations that already had high organisational capacities and a large scale of operations at the moment of receiving the FONDOECAS grant. If the increased capacity to pay organisational expenses were to be the main objective of the fund, targeting the grants to non-sourcing organisations would be likely to improve the effectiveness of the fund.

### Did the grants improve access to financial services?

Twelve grant recipients mentioned that they had received one or more loans in the following years. However, five of these loans were provided by FONDOECAS itself, through a new loan facility, started in 2012. Also, two of the non-beneficiaries gained new access through FONDOECAS' credit line, ASPASA and AMLECO.

The coffee cooperatives continued to get access to pre-financing through their contracts with upstream buyers. The offer of a pre-financing facility is mandatory for buyers within the Fair Trade system. Furthermore, many of the coffee cooperatives had access to a credit scheme provided by their own financial entities, *Asociación de Servicios Financieros Cafetaleros* (FINCAFÉ), an initiative of FECAFEB, and *Financiera Asociación Agropecuaria del Altiplano Sur* (FAAAS).

Two organisations in the sample, COAINE and CECAOT, accessed a loan from the government-led national development bank (*Banco de Desarrollo Productivo –BDP*)<sup>28</sup>, which introduced a special credit line dedicated to the support of farmer organisations. Access to BDP loans was mentioned as an aspiration in many of the interviews. In practice, however, the access is limited to organisations having a large patrimony that can be used as collateral for the loan. However, their access to formal credit was not related to the business plan supported by FONDOECAS. In both organisations, the BDP loan was used for sourcing raw material for exports, while the grant-supported processing for consumer-ready products for the domestic market.

Three grant recipients accessed micro-credit for the grant-supported business plan. Two of these were given access on an exceptional basis: one through the international development NGO HEIFFER which manages a revolving fund for project beneficiaries only, and the second through an unregulated micro-credit institution in Tarija (FONCASOL), which in this case made an exception to their general policy of lending only to individual farmers. A third grant beneficiary accessed a regulated micro-credit institution (ANED). Also one of the non-beneficiaries, ASOPROF, managed to get a loan for trade capital with the international financial institution OIKO CREDIT.

---

28 BDP is a state-owned second-level bank that, through first level banks like Banco Union, provides loans with relatively low interest rates. Initially, the allocation of these loans had been highly influenced by the support of these organisations in the 2006 elections of Evo Morales (Córdoba and Jansen, 2013). However, gradually this type of loan was incorporated in the normal BDP's credit portfolio with (more) stringent eligibility and repayment obligations.

**Table 8.11** Access to grants and credit (2010-2012)

FONDOECAS GRANTEES (2008-2010)	Access to new grants (2010-2012)			Access to credit (2010-2012)					Patrimony (2012) US\$
	New grant?	Second FONDOECAS grant	Other grant	New loan?	FONDOECAS loan	Trade loan	BDP loan	Other loan	
Name	Count	Count	Count	Count	Count	Count	Count	Count	
AAAT	Yes	0	1	No	0	0	0	0	523,945
ADAPICRUZ	Yes	1	0	No	0	0	0	0	142,856
AGAYAP	No	0	0	No	0	0	0	0	700,000
AMAGA	Yes	0	1	No	0	0	0*	0	75,000
AMDESOY	Yes	1	0	No	0	0	0	0	42,000
AOCEMM	No	0	0	Yes	1	0	0	FONCASOL	63,411
APAM MIZQUE	Yes	0	1	No	0	0	0	0	7,800
APCA	Yes	0	1	Yes	1	0	0	ANED	88,500
APROAMOL	No	0	0	No	0	0	0	0	22,000
APROQUIRC	No	0	0	Yes	0	1*	0	0	150,000
APSU	Yes	0	1	No	0	0	0	0	39,000
ARAO	Yes	0	1	No	0	0	0	0	97,000
ASAFOP	Yes	0	1	No	0	0	0	0	36,000
ASOCOM	No	0	0	No	0	0	0	0	30,000
CECAOT	Yes	0	1	Yes	0	0	1	0	327,157
CELCCAR	Yes	0	1	Yes	0	1	0	0	660,000
CEMUR	Yes	0	1	Yes	1	0	0	0	2,054,137
CEPLACH	Yes	0	1	Yes	0	0	0	HEIFER	35,000
CIAPEC	Yes	0	1	Yes	1	1	0	0	150,000
COAINE	No	0	0	Yes	0	1	1	0	375,000
COMART	Yes	0	1	No	0	0	0	0	56,000
COPROQUINACC	No	0	0	Yes	0	1*	0	0	44,285
INCA PALLAY	Yes	0	1	No	0	0	0	0	82,192
OMCSA	Yes	1	0	No	0	0	0	0	50,000
ORLIPA	Yes	0	1	Yes	1	0	0	0	21,000
SOPPROQUI	No	0	0	Yes	0	1*	0	0	200,000
SUBTOTAL GRANTEES	18 73%	3 12%	15 58%	12 46%	5 19%	6 23%	2 8%	3 12%	
AMLECO****	Yes	1	1	Yes	1	0	0	0	880,500
APME	Yes	1	0	No	0	0	0	0***	32,580
ASOPROF	Yes	0	1	Yes	0	0	0	OIKO CREDIT	20,000
ASPASA****	Yes	0	1	Yes	1	0	0	0	15,000
FENCA	No	0	0	No	0	0	0	0	50,000
SUBTOTAL NON-GRANTEES	4 80%	2 40%	3 60%	3 60%	2 40%	0 0%	0 0%	1 20%	

\*. Client of BDP before 2010, with pending loan.

\*\*. Trade finance through the second-tier organisation ANAPQUI.

\*\*\*. Informal loans to the group by individual board members are considered as no increase in access

\*\*\*\*. Accessed a FONDOECAS loan in 2012

These results show that access of economic farmer organisations to commercial banks or micro-credit institutions is still extremely limited. Access to the BDP and other banks is an exclusive privilege of larger and richer organisations. It seems unlikely that an average FONDOECAS' beneficiary will gain access to bank credit in the near future. FONDOECAS provided some leverage in resolving the working capital constraints of economic farmer organisations, through their new loan facility. However, overall, there is no evidence that the FONDOECAS grant did influence the possibilities of the organisations to get a loan from financial institutions<sup>29</sup>. The assumption that the grant, through an increase in infrastructure, group patrimony and turnover, would increase access to formal credit for the grant beneficiaries is not supported.

We observed that the organisations continued to be supported by the development cooperation, apart from FONDOECAS. Table 8.11 shows that many organisations benefited from additional, new grants (not loans) to support their collective marketing activities during this period. Four out of five non-grantees received a follow-up grant, as well as 18 of the 26 grant beneficiaries. Except for one organisation (APCA) we have no reason to consider this to be a result of the FONDOECAS grant. Most likely, the organisations would have received this support in any case. This confirms the importance of the economic farmer organisation as a means to access outside support (Bebbington, 1996), and as a preferred interface for development NGOs and governments to reach innovative farmers (Bebbington, 1999; 2007; World Bank, 2012). In economic farmer organisations in Bolivia, the increase of patrimony, infrastructure and working capital, is largely based on grants. Organisations such as ANAPQUI and CECAOT could access commercial credit lines only after a long period of grant support, by North-American and European bilateral cooperation (see Laguna (2011) on ANAPQUI, and Healy (2001) on the trajectories of El CEIBO and CECAOT).

## Conclusions on effectiveness

Several of the organisations in the sample that had received a grant failed to implement their grant-supported business plan. Others did implement the grant and received a second grant after 2010. In Table 8.12 we give an overview of successful implementation by the grant beneficiaries in the sample. To prevent over-reporting due to attrition, we included the three organisations (APEMAK, APLEPO ARASAYA, APROLAC) for which we could not get complete data due to organisational break-down, and which were left out of the analysis in the preceding paragraphs (see Annex 2).

---

29 In 2014 FONDOECAS significantly increased their emphasis on loans instead of grants. On 13 November 2014, CIOEC communicated that FONDOECAS had received a two million dollar grant to expand the loan portfolio.



**Table 8.12** Implementation and outcomes of grant supported business plans

	Number of organisations	Success rate
Beneficiaries	29	100%
Grant implementers in 2012	23	79%
Business plans in operation in 2012	19	66%
Grants contributed to organisational capacities	10	35%
Grants contributed to market access of members	5	17%
Grants increased capacity to pay organisational expenses	12	41%
Grants improved access to financial services (FONDOECAS loans incl.)	8	28%
Grants improved access to financial services (FONDOECAS loans excl.)	3	10%

The implementation rate of grant-supported business plans in the sample is 66% (see Table 8.12). One third of the grants were not invested, or the business plan was aborted soon after the investments were made. Reflecting on the record of non-implementers, in 2011, FONDOECAS decided to strengthen the role of the CIOEC departmental branches. As an additional requisite, these branches were asked to certify that the organisation was active and eligible for support. The department branches also started to receive a small amount of money, approximately US\$80, upon presentation of a progress report on one of the grant beneficiaries. While in the initial proposal, in 2006, this direct interference of CIOEC in the selection process had intentionally been reduced to a minimum in order to prevent political clientelism or extortion, at the end of 2011 it was acknowledged by all stakeholders that the CIOEC branches needed to play a more active role in the ex-ante screening process to prevent failure (FONDOECAS, 2010; FONDOECAS MyE, 2012).

FONDOECAS improved organisational capacities in 10 of the 29 organisations. Half of the implemented business plans resulted in an improved capacity to contain the agency dilemmas in collective marketing, foremost being the capacity to resolve the issue of quality assurance. Although we cannot compare this with the outcomes of other grant funds, and a certain percentage of failure is inherent to innovative business plans, FONDOECAS has certainly been less effective in generating positive outcomes than was initially expected. This is even more evident when reviewing the effectiveness in increasing market access for member products. Only 17% of the grants were successful in this respect. The results on the capacity to generate income to pay organisational expenses were better. Almost two-thirds of the implemented business plans were successful in generating group income, representing 41% of the grant recipients. Finally, access to financial services has improved for only a minority of organisations, and mainly because FONDOECAS started to resolve the issue in 2012 on its own by opening a loan facility alongside the grant system.

The grants proved particularly unsuccessful for the larger and stronger organisations that already had high organisational capacities and a large scale of operations at the moment of receiving the FONDOECAS grant. If the increased capacity to pay organisational expenses were the main objective of the fund, targeting the grants to non-sourcing organisations would likely improve effectiveness. However, if increasing market access for members were the main objective, targeting the larger sourcing organisations would increase the likelihood of success.

## 8.6 Verifying the assumption of efficiency

### Analysis of monitoring data

The technical committee of external experts is the key institutional arrangement in FONDOECAS' grant allocation system. It had to target the grants at feasible proposals from viable organisations. We could analyse the available administrative data in FONDOECAS to verify if the committee's feasibility scores could predict progress in implementation. We had access to most of the evaluation scores by the technical committee on business proposals that had been granted between 2007 and 2013. The total number of observations was 604, concerning 150 different organisations. Next to this, we had access to the progress reports of the monitoring and evaluation officer in FONDOECAS, appointed in 2011. The officer reported his findings on the status of the funded business plans, which he had visited, on a yearly basis (FONDOECAS MyE, 2012; Prudencio, 2010). In 2011 and 2012, he visited 61 organisations to assess progress in implementation; 13 of these are part of our sample of comparative case studies.

### Did feasibility scores predict implementation progress?

The FONDOECAS technical committee reviewed each proposal on feasibility, innovation and social orientation (Table 8.2). During the period 2007-2013, a total of 10 different experts had taken part in the committee. Until 2011, the team had five members to evaluate each proposal. However, five out of the 10 experts were more consistently involved in these evaluations, of which three continuously so in the whole period 2007-2013. Two of the evaluators proved to be much more critical than the others. They rated the proposals on average about 10 points lower than the other members of the panel (ANOVA  $p < 0.05$ ; post hoc comparisons with Bonferroni correction). FONDOECAS respected the critical stance of these evaluators. When in 2011 FONDOECAS decided to reduce the committee from five to three members due to budgetary restrictions, they retained these two critical evaluators in the group.

We tested if their assessments converged in time. We used a linear regression using each evaluator's 'absolute difference with the mean score' as dependent variable. FONDOECAS organised 14 rounds of grant evaluations between 2007 and 2013, with an average of 12 proposals per round. We used the grant round number as the independent variable, and included the mean score as a covariate to control for collinearity. The differences in assessment between experts, proved to decline slowly, with a quarter point per round ( $F[2, 604]=17.48$ ,  $\beta: -0.240$ ,  $p < 0.01$ ;  $R\text{-squared}=0.06$ ), when the average difference was 7.5 points.

We could compare these evaluations with the results of the monitoring missions of the FONDOECAS M&E staff, who visited 61 organisations during 2011 and 2013 to check on progress of the business-plan implementation. Most of these organisations were visited one year after having had their grant approved. Other organisations were visited because of persistent delays in implementation. FONDOECAS' monitoring officer reported on progress in four areas: organisation, production, marketing strategy and financial investments. Each area had four aspects, coded as a crisp-variable (yes/no). For each area we computed a variable with five possible values. The inspection reports (Table 8.13) indicated that organisational progress

in business plans lagged. Only 13 of these 61 were part of our sample of comparative cases studies used to assess effectiveness. The progress indicators for these organisations showed a similar pattern. The involvement of members as suppliers or operators in the businesses was far lower than expected, and only sporadically were specialised persons operating the business. Also, progress in the area of marketing was less advanced than expected, especially on the issue of quality certification. Progress in financial investment and in production was relatively good, with 72 and 67%, respectively.

**Table 8.13** Monitoring scores on business plan implementation

Area	ASPECT	All visited (N=61)		Matched with evaluator scores (N=50)		Matched with case studies (N=13)	
		Average progress	Stand. dev.	Average progress	Stand. dev.	Average progress	Stand. dev.
Organisation	Increasing involvement of members	0.16	0.36	0.16	0.36	0.15	0.38
	Members know how to manage production	0.48	0.52	0.48	0.52	0.31	0.48
	There is professional staff available	0.36	0.48	0.4	0.48	0.54	0.52
	Business plan benefits all members	0.56	0.52	0.64	0.48	0.46	0.52
	SUBTOTAL	39%	25%	42%	24%	37%	32%
Production	Primary production	0.80	0.40	0.84	0.36	0.77	0.44
	Procurement from members	0.64	0.48	0.68	0.48	0.85	0.38
	Processing of product	0.52	0.52	0.56	0.52	0.77	0.44
	Collective marketing of product	0.92	0.24	0.96	0.2	1.00	0.00
	SUBTOTAL	72%	26%	76%	24%	85%	16%
marketing	Finished product	0.68	0.48	0.76	0.44	0.85	0.38
	Ready market	0.52	0.52	0.48	0.52	0.69	0.48
	Quality certification of product	0.24	0.44	0.28	0.44	0.23	0.44
	Dedicated buyer for product	0.52	0.52	0.52	0.52	0.38	0.51
	SUBTOTAL	48%	24%	50%	21%	54%	20%
Financial investment	Equipment in place	0.88	0.32	0.92	0.32	1.00	0.00
	Machinery in place	0.52	0.52	0.52	0.52	0.69	0.48
	Infrastructure in place	0.84	0.36	0.88	0.36	0.69	0.48
	Working capital available	0.44	0.52	0.48	0.52	0.46	0.52
	SUBTOTAL	67%	29%	69%	27%	71%	22%

We expected a positive relationship between the initial committee evaluation score and progress reported by FONDOECAS monitoring staff. We could pair 50 of the 61 monitoring observations with the average feasibility score of the committee. The unpaired observations differed in name or referred to the same organisation having been visited twice. We tested if the feasibility scores predicted the progress of successful implementation according to the monitoring report with a multivariate analysis of covariance (MANCOVA) to test for this predictive effect. The regression model can be written as:

$$Y_{ij} = [y_{i1}, y_{i2}, y_{i3}, y_{i4}] = \alpha + \beta X_i + \epsilon_i$$

with,

- $y_{i1}$  = progress in organisational commitment of organisation  $i$
- $y_{i2}$  = progress in production and logistics of organisation  $i$
- $y_{i3}$  = progress in marketing strategy of organisation  $i$
- $y_{i4}$  = progress in infrastructural investments of organisation  $i$
- $X_i$  = feasibility score of organisation  $i$
- $\epsilon_i$  = error term in regression of organisation  $i$

The model is statistically significant ( $F[4,50]=3.11, p=0.02$ , Pillai's Trace  $p=0.02$ , Partial eta-squared=0.22), which means that the feasibility scores predict the progress of implementation of the grant-supported business plan fairly well (see Table 8.14). The partial eta-squared is considered as a large effect according to Cohen's rules of thumb (Cohen, 1988) in multivariate analysis. Testing the Pearson correlation between the feasibility score and each area where progress was reported by the FONDOECAS monitoring expert, we see that the feasibility score predicts the progress in financial investments in the business plan ( $F[1,50]=7.54, p<0.01$ ), but the feasibility score is negatively related with progress in production and logistics ( $F[1,50]=4.16, p<0.05$ ). When we revised the data, we detected that the negative correlation was heavily influenced by the low scores of two dairy organisations, which had invested the grant in processing equipment but did not manage to get the business operational. These are the only organisations, of the 50, that had received zero points in both production and marketing. The negative correlation between high evaluation scores and progress in the area of production disappears when we exclude these two outliers ( $F[1,48]=0.03, p=0.87$ ), whereas the correlation with progress in the area of financial investment remains statistically significant ( $F[1,48]=5.01, p<0.05$ ).

**Table 8.14** Multivariate analysis with average feasibility score as predictor of field monitoring results (N=50)

Source	Area	Pearson correlation	F	Sig.	Partial Eta Squared
Feasibility score	Organisation	.049	.115	.736	.002
	Production	-.282	4.155	.047	.080
	Marketing	-.011	.006	.939	.000
	Financial investments	.368	7.539	.008	.136

## Conclusions on efficiency

We used administrative and monitoring data to assess the efficiency of the grant allocation system in FONDOECAS, where an independent, external committee of experts evaluates the feasibility of the business proposals.

The valuation of the feasibility of a proposal differed greatly between evaluators. Several proposals were approved, while one or more of the evaluators had serious doubts about the viability of the business proposal. This analysis of the feasibility scores proved useful for reflection on this undesirable aspect, and helped to improve the grant allocation system. In 2012, FONDOECAS (2012) decided to give individual evaluators a right to veto when they considered the feasibility very low (less than 40 points out of 100). Fortunately, we see that the differences between evaluator scores decreased in time, which suggest a learning process within the evaluation committee.

We compared the progress in implementation of the business plans with the feasibility scores received when awarded the grant. The field visits of the FONDOECAS M&E staff showed that the implementation of many business plans was slower than expected. The involvement of the members and the professionalisation of the business operators, especially, lagged behind. The evaluators' score on feasibility of the proposed business plan indeed predicted progress in implementation (financial investment). However, there is no positive correlation with the more substantial outcomes in the areas organisation, production and marketing (see Table 8.14). This suggests that the committee is not yet efficient at differentiating between feasible and less feasible projects.

## 8.7 Contribution story

FONDOECAS emerged as a pilot-project to address the specific problems of economic farmer organisations in collective marketing, in view of scaling up and replication. The fund provided grants (around US\$10,000) to business plans that were evaluated on feasibility by a committee of external experts. The donor community supported FONDOECAS because they consider stronger economic farmer organisations key actors for local economic development. Their relevance legitimises the use of public money to support them. To do so, the FONDOECAS grants need to be effective in strengthening the groups, and, for reasons of efficiency, the FONDOECAS evaluation committee needs to target the grants to feasible business plans. We used contribution analysis to reflect on these assumptions.

The household survey showed that there is majority support for economic farmer organisations among the rural population. Households that were participating in economic farmer organisations considered them even more supportive than the village organisations. Interestingly, the distinctive identity of economic farmer organisations is their role in creating market access, whereas this is reported as being the least evident benefit. There was more agreement on the statement that they are a means to support production, access outside support, and as a component of social life. There are, however, prospects for growth, considering the willingness of two-thirds of the farmers to market their products collectively.

Most economic farmer organisations registered a high growth of group sales between 2008 and 2013. Overall, they have become commercially stronger, especially due to higher agricultural prices and preferential procurement by public nutrition programmes. The contribution of FONDOECAS to this growth is modest. Due to FONDOECAS' focus on processing activities, often complementary activities, impact on market access of members was marginal. FONDOECAS did help to build organisational capacities in only a third of the beneficiary organisations, mainly the capacities needed for quality assurance. FONDOECAS contributed to the capacity to earn some money to pay organisational costs in almost half of the supported groups. There was no evidence of an improved access to formal credit as a result of the grant. When organisations are already well-endowed, small grants are less important and are likely to result in under-scaled investments in secondary activities. Access to trade finance may be a more effective strategy to strengthen these organisations.

The data suggests that the technical committee is not very efficient in its targeting of grants to feasible business plans. There were big differences in valuation between the experts in the committee, though these differences declined in time. FONDOECAS, based on our preliminary analyses, decided to retain the two most critical experts in the committee, which reflects awareness of the need to critically assess the feasibility of business proposals in order to better target the grants. However, although we showed a positive correlation between feasibility scores and progress in implementation of investments, there seems no predictive power of the viability scores of the technical committee for the organisation, production and marketing related aspects of the business plan.

## 8.8 Conclusion

The result of the research, the verification of the contribution story, helps the implementing agency to reflect on expectations and refine the intervention. But the findings have wider relevance. Grant funds may use these findings to refine their targeting policies. A grant fund that wants to support economic farmer organisations should distinguish between organisations that source and those that do not source their raw materials from members. Market access for members is only possible with sourcing organisations, and larger organisations are more likely to be successful in using the grant to do so. Instead, organisations that source their raw materials from spot markets are more likely to be successful in using the grant to raise group income to pay for organisational expenses.

Furthermore, the approach to impact evaluation used in this research has wider application. There is a need to reflect on the effectiveness of support interventions that, like FONDOECAS, want to trigger complex change processes and work in a restricted population of firms or organisations. The shift in development cooperation 'from aid to trade', with an increasing number of grant funds directed to companies, comes with an increasing need to verify relevance, effectiveness and efficiency of grants for business development (Kessler, 2013). We showed that through a creative mix of methods and data-analytical approaches, using fund-specific monitoring data, comparative cases studies on beneficiaries and surveys, we were able to critically review the key assumptions in the rationale behind grant funds.





## 9 | Synthesis and Discussion

*“... the best anyone can do is to understand deeply the specific problems that afflict the poor and identify the most effective ways to intervene. In some instances, no doubt, the best option will be to do nothing, but there is no general rule here, just as there is no general principle that spending money always works. It is the body of knowledge that grows out of each specific answer and the understanding that goes into those answers that give us the best shot at, one day, ending poverty.”*

(Banerjee and Duflo, 2011: 15)



## 9.1 Introduction

As described in the first chapter, I developed my research using methods from two approaches to impact evaluation. The ‘randomistas’ are characterised (or caricaturised) by their use of (quasi)experimental methods and surveys in order to detect net-effects between a treatment and a comparison group, which – through randomisation and matching – are similar enough to permit the assumption that the difference in outcomes is attributable to the intervention. The ‘realistas’, instead, are characterised by their focus on contextual conditions that allow an intervention to be successful, and prefer configurational comparative methods to detect causal patterns in data. ‘Randomistas’ and ‘realistas’ may tend to converge in their research focus on causal mechanisms of effectiveness, but they differ in the preferred methods to do so. In the introduction, I showed that these methods imply different logics of causal analysis. The regressional-analytics label a factor a cause when it *covaries* with the outcome: ‘the more/less of the cause, the more/less of the outcome’. Conversely, configurational comparativists label a factor a cause by *implication*: ‘if the cause is absent/present, then the outcome is absent/present’ (Thiem et al., 2015). I show, in Chapter 7, that certain causal patterns could only be detected by using the logic of implication of Qualitative Comparative Analysis (QCA).

The research for this thesis has provided me with an opportunity to learn more about the preferred methods used by both schools, and to experience their strengths, weaknesses and synergy in real-world impact evaluation. In this last chapter, I will reflect on the lessons learnt in order to identify some principles that have proven helpful in obtaining an appropriate research design for impact evaluation. The chapter proceeds as follows. First, I give an overview of the main findings and lessons learnt, following the sequence of chapters. I use these insights to distil seven design principles for impact evaluation design. Second, I reflect on the new tool to analyse and measure organisational strength in collective marketing. Tension Containment Capacity, and discuss its contribution to social theory. Finally, I reflect on the credibility and limitations of the research.

## 9.2 Main findings

In Chapter 2, my co-authors and I present the results of an explorative systematic review. This type of systematic review is located in the spectrum between two extreme approaches to synthesis, aggregative synthesis (meta-analysis), which verifies and tests the effectiveness of an intervention on the one hand, and configurative synthesis, which tries to make sense of the body of literature around a certain issue without a predetermined thematic framework, on the other (Gough and Thomas, 2012). We mapped studies on similar types of interventions and reflected on causal assumptions in the respective impact pathways. We did not intend to give a verdict on the effectiveness and effect size of innovation grants to smallholder farmers, but to develop a better understanding of three types of innovation grant systems: voucher systems, business development matching grants and innovation support funds. Our explorative systematic review may have been less restrictive in the inclusion of studies than is usual in a meta-analysis, because precise estimates of effect size were not our main interest. We experienced severe time/budget constraints due to the systematic electronic search and burdensome screening process required by the systematic review process. However, the analysis of previous

research on similar interventions and causal processes is a good starting point for the design of impact evaluation studies, and a learning experience on the ways in which others have mixed research methods in their own research designs. The limited number of quantitative impact studies that we ‘harvested’ in the systematic literature search, coupled with the high diversity of proxy-indicators used, signalled that the use of common proxy-indicators would have facilitated the comparison of findings. Several studies (Friis-Hansen, 2008; Gustafson, 2002) showed that the effectiveness of an innovation grant system (NAADS) depended to a large extent on the success of a previous intervention to stimulate group formation and experimentation (Farmer Field Schools). This indicates that a learning process is inherent to innovation, as are the capacities of farmers to evaluate new practices or business plans through experiential learning. The review also showed that common measures of human and social capital to sustain innovation processes would be helpful to better compare the effectiveness of interventions.

In Chapter 3, my co-authors and I analyse the challenges of measuring net-effects of certification. The number of certificates linked to agricultural commodities is increasing, and donors, companies and consumers want to be informed about their effectiveness. The discussions focus on the effects of certification on the income of certified farmers (Haight, 2011; Minten et al., 2015: 22; Méndez et al., 2010; Claar and Haight, 2015). Based on our experiences with impact research on cocoa certification in Côte d’Ivoire and Ghana, we showed that there are limits to the accuracy of measurement of net-effects on yields and income. The sample sizes needed to measure the expected effect-size on yields with sufficient statistical power would need to be approximately 2,000, whereas for intermediate effects on knowledge and practices, a sample size of less than 100 would be sufficient. We propose the delineation of a ‘span of direct influence’, inside of which net-effect estimates are possible. Outside the span of direct influence, the attribution of net-effect becomes impossible. The exact location of the boundary will depend on the complexity of the causal process and real-world constraints on research methods (Bamberger et al., 2004). This does not imply that measuring and monitoring outcomes outside the span of control is unimportant. However, it does imply that quasi-experimental research designs are ‘powerless’ or inappropriate. Outside the span of control, other methods are needed to verify the assumption of contribution of an intervention to processes of change. These other methods are not necessarily cheaper, but awareness of the impossibility to calculate net-effects in ultimate outcomes may result in smaller sample sizes or shorter surveys, thus freeing budget for additional research components. For example, to assess household poverty status, detailed questions on income components may be replaced with less precise but leaner instruments to assess the poverty position of beneficiaries, such as the PPI- Progress out of Poverty Index (Chen and Schreiner, 2009). We also plea for the use of common measurement tools to facilitate the comparison of various approaches to tackling similar types of problems. Common measures for intermediate outcomes could increase the learning about alternative training modalities, different approaches to group strengthening, or different pricing and incentive structures in contracts. It is precisely the quality of these ‘nitty-gritty’ institutional arrangements that triggers behavioural responses and new practices in persons and organisations: these are the ‘small institutional changes’ needed in order to fight poverty.

Evaluation design needs to ‘match’ the research methods to the evaluation question and to the context of the intervention. In Chapter 4, my co-authors and I present three challenges for doing so in agricultural value-chain interventions. These challenges are the measurement of outcome patterns, attribution of effects in open systems and the generalisation of findings from the particular research context to recommendations for replication or scaling of the intervention. The main point we stress in this chapter is the role of theory in impact evaluation design and data analysis. We show that ‘randomistas’ use theory in regression models and matching procedures, while ‘realistas’ explore theories on contextual conditions that explains why an intervention is effective. We point to the common ground between both approaches in so-called Theory-Based Evaluation (Weiss, 1997; White, 2009; Pawson, 2013; Stame, 2004), where impact evaluation is used to verify and refine assumptions about change processes that interventions are assumed to ‘cause’.

Chapter 5 presents a practical three-step process for improving rigour in impact evaluation. The three steps are: 1) Refine the evaluation questions based on the intervention logic; 2) Anticipate the validity threats to the expected type of conclusions; and 3) Maximize the scope for comparative research. The check on validity threats in a team setting proved a useful dynamic of research design. We started with a provisional core methodology to address a specific research question, and we added research design elements after we reflected on the threats to validity of the type of conclusions from this core method,. Often this resulted in adding quantitative components to qualitative core methodologies, or qualitative components to quantitative core methodologies. But it also resulted in adding qualitative methods to qualitative core methodologies, and quantitative design elements to quantitative core methodologies. In doing so, researchers having different methodological traditions came together, were stimulated in pro-active and creative thinking about methods to complement their own preferred tool-kits. This worked better than the usual ex-ante decision to have a quantitative and a qualitative component and budget, which often ends up with different methodological traditions working in isolation.

Chapter 6 documents the design and field test of a new way to assess and compare organisational capacities of collective marketing groups. I present Tension Containment Capacity as a new tool to assess the organisational strength of collective marketing groups. I focus on the quality of the organisational practices to respond to emergent tensions. The tool was designed to be lean enough to incorporate into the grant fund’s monitoring system, not only for impact evaluation. The tool is appropriate for both large and small groups, and relatively independent from the type of commodity handled by these groups. Because this is likely to be one of the main theoretical contributions of the thesis, I reflect on this in more detail in section 9.4 below.

In Chapter 7, I apply a relatively new method to detect patterns in data sets using Qualitative Comparative Analysis (QCA). It uses Boolean algebra to identify the most concise (parsimonious) recipe of conditions that are consistently related to success or failure. I identified patterns in data that point to predictors of effectiveness. And, because some of these predictors are single conditions, I could use logistic regression to triangulate and increase the validity of these predictors. I am not the first to combine regression-analytic and configuration comparative methods in one research design [see for example Vis ((2012))], but the application of a fuzzy-set QCA together with a logistic regression on real data from a real-world impact evaluation is

novel. This combined use of distinct methods of causal analysis in one research design is most likely the most innovative mix of methods in this thesis.

In Chapter 8, I used Contribution Analysis (Mayne, 2001; 2011; 2012) as the framework to present evidence on three key assumptions in the intervention logic of FONDOEAS, an innovation grant system that subsidises investment of economic farmer groups in processing and collective marketing. Reviewing the evidence generated with the mix of methods, I conclude that the results of the grants are somewhat disappointing. However, from the point of view of impact evaluation methods, I think that my methodological approach to impact research, respecting the different logics but combining the methods of ‘randomistas’ and ‘realistas’, has passed the proof of principle test.

### 9.3 Design principles

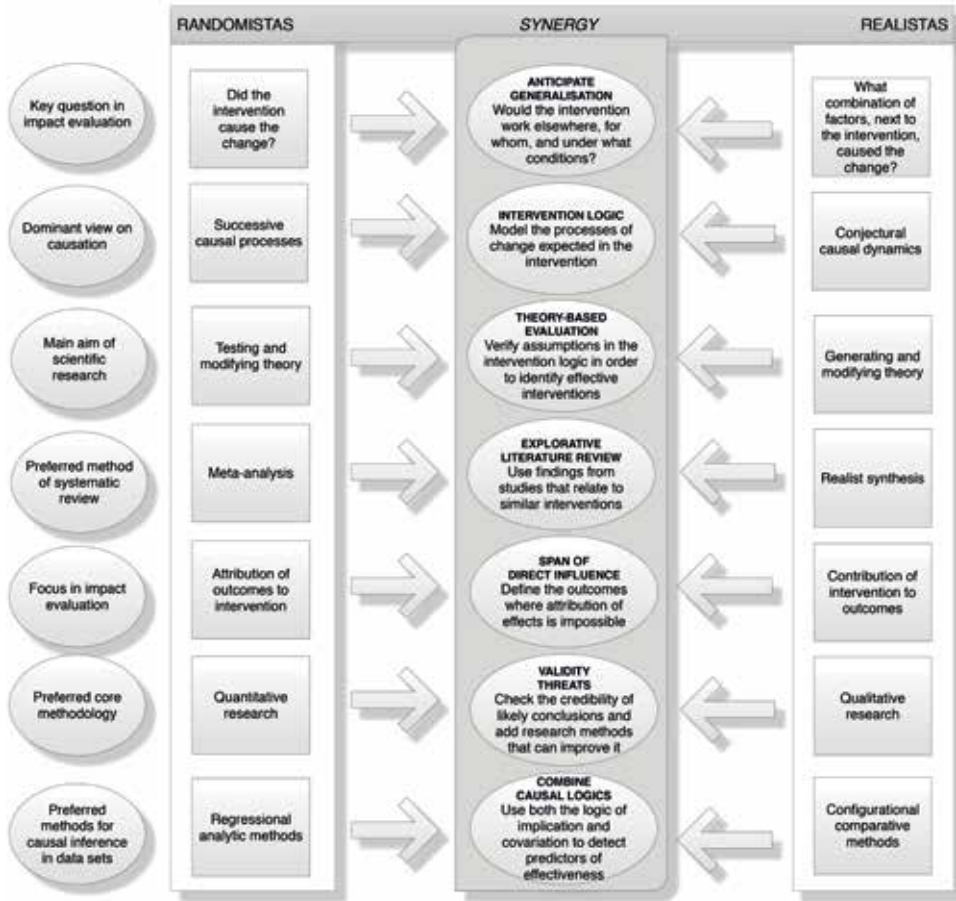
As shown in this thesis, the exact mix of methods used to verify assumptions in an intervention logic depends on the contextual conditions and logistical and budgetary constraints. Impact evaluation is applied research: research designs emerge from a process in which commissioners and researchers negotiate to balance budgets and expectations: a research design that is appropriate in one setting may not be appropriate in another. I will distil some ‘principles’ from the above experiences that helped me in research design and that have broader application. Figure 9.1 presents these principles. I locate them in between the two impact evaluation approaches which I discussed in the introduction chapter. These principles create room for synergy between both approaches, while respecting their differences.

#### ***Anticipate generalisation***

Impact evaluations examine the effectiveness of interventions. All these interventions are embedded in a specific context, which means that their conclusions can be only generalised to similar interventions in similar contexts. In order to increase the relevance of the findings, the research design should include methods that help to delineate the generalisation domain (Chen, 1994), and prepare an answer to the question ‘Would the intervention work elsewhere, for whom, and under what conditions?’. Figure 9.2 shows the three sub-questions implied in this overall question. The balance in commissioned impact evaluations and systematic reviews is too much in favour of the ‘What works?’ question only. They would do well to focus attention on heterogeneity in program design, context, and impacts (Pritchett and Sandefur, 2015). This will give way to research that tries to unpack the causal mechanisms that make interventions effective, in what ‘randomista’ Chris Blattman (2008) calls Impact Evaluation 2.0, and ‘realista’ Ray Pawson (Pawson, 2013) calls Evaluation Science.

#### ***Map the intervention logic***

Most development programmes try to tackle many issues at the same time, with multiple interventions and partly overlapping groups of beneficiaries. Impact evaluations of such comprehensive interventions are not very informative. Impact evaluations on potentially replicable programme components are much more informative. The mapping of an intervention logic



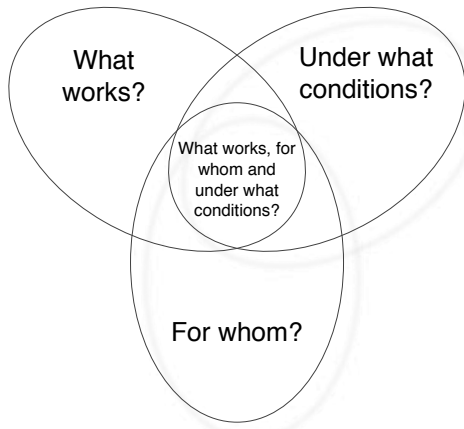
Source: Author’s own elaboration

Figure 9.1 Design principles that look for synergy between two approaches to impact evaluation

is a good way to differentiate the components of an intervention, in order to focus on those causal processes that seem most critical to the effectiveness of the intervention and most informative for future intervention design. Intervention logics model a successive process. There are expectations about social change processes due to the intervention in the minds of the people that implement a development intervention and in the minds of the ones that fund it. However, this does not imply that all causal links in the intervention logic are linear. Almost always, a mapping of an intervention logic will show linear as well as configurational causal processes, and include feed-back loops to model complexity (Funnell and Rogers, 2011).

**Use theory-based evaluation**

Theory-based evaluation conceptualises intervention logics as interlinked causal assumptions. These assumptions are the ‘theories’ behind the intervention. A theory-based impact



Source: Author's own elaboration

**Figure 9.2** The balance of three evaluation questions

evaluation collects data to verify whether these theorised causal processes indeed took place and reflects on the role of important contributing factors and necessary context conditions. Some assumptions are based on theories on the incentives for human behaviour (Westthorp, 2012; 2013). Others take a more pragmatic perspective and consider as ‘theory’ the stakeholders’ assumptions about how a programme will generate social change (Weiss, 1997). Not all assumptions in the intervention logic can be a focus of impact research. A decision will have to be made about the key causal assumption on which the research will focus. Some assumptions can best be verified through research methods that focus on the intervention as the ‘cause’ and try to reduce the influence of context (e.g., experimental and quasi-experimental methods). Other assumptions will need research that focuses on the presence of the conditions to make the intervention work. The verification of other assumptions may not need any primary data collection but can be addressed by reflecting on secondary data or findings in the literature.

### ***Explore the existing literature***

Reading previous studies is perhaps the most obvious but often the most neglected design principle. Systematic reviews may help to give an overview of the literature and condense the findings on certain interventions. But not many systematic reviews have yet been done on international agricultural development, and most of these review only quantitative studies on effectiveness for very specific types of interventions. Doing a systematic literature review is very labour intensive and, therefore, not an effective way to prepare an impact evaluation design. However, open access policies and platforms such as the Social Science Research network (SSRN) and ResearchGate, combined with fast search engines such as Google Scholar make it easy to find relevant literature that helps one to reflect on certain causal assumptions and find ways to verify them with appropriate methods.

### ***Define the span of direct influence***

The attribution of effects to a support intervention is only possible within a span of direct influence, and, of course, only when outcomes can be properly measured or observed. The impact pathway of each intervention will have a different boundary for this span. Moreover, of course, some research methods are better able to capture some of these ‘borderline’ outcomes than others. Within the span of control, quasi-experimental research and Randomised Controlled Trials can be informative research designs (Vaessen, 2011). While randomised assignment of support interventions such as grants, credit, training for certification, or technology supply is often logistically and politically unfeasible, randomisation of alternative implementation modalities of support may be possible and informative (Beekman et al., 2014; Bulte et al., 2014).

### ***Organise a threats to validity check***

The research design to answer an evaluation question needs to result in credible findings. Interdisciplinary research is enhanced by identifying the main threats to validity of the core methodologies preferred in each field. I propose, similar to Shadish et al. (2002), four types of threats to validity to check the rigour of the research design: data-set analytic validity, internal validity, construct validity and external validity. During the threats to validity check, researchers with different methodological traditions come together, which stimulates pro-active and creative thinking about methods, and learning about new methods that are complementary to their own preferred tool-kit.

### ***Combine different logics to detect causal patterns***

My last principle is most likely the most innovative. The regressional-analytic school uses ‘covariation’ to detect causal patterns in data, and configurational comparative methods use the logic of ‘implication’. I demonstrate that the combination of both the Qualitative Comparative Analysis (see Box 9.1) and the regressional analytic method of logistic regression, into one research helps to increase the validity of causal inferences. QCA is used especially in research with relatively small data sets. With a small number of cases, it is possible to go back to the individual cases to judge the plausibility of the detected pattern as causal explanations. However, QCA can also be applied for the analysis of large samples (Fiss et al., 2013b). In large samples, the issue of measurement error is likely to be more problematic, because it is more difficult to check a specific causal term with the information about each case. This implies that the strength of the evidence to label a condition as a cause needs to be higher. Single causal conditions are likely to appear as predictors in regressions. Causal configurations are less easy to detect with regressions. However, when the sample size is sufficiently large and the configuration has a fair coverage of cases, a regression with higher order interaction terms is likely to detect a causal configuration discovered by QCA (Fiss et al., 2013b).

**BOX 9.1 QUALITATIVE COMPARATIVE ANALYSIS**

Qualitative Comparative Analysis (QCA) was developed by the political scientist Charles Ragin (Ragin, 1987; 2000; 2008). It is used to explore configurations of factors within a data set, which are related to the presence or absence of an outcome condition. Cases share certain attributes, called conditions, and each case is successful or not according to an outcome condition. Conditions can be ‘crisp-sets’, with the value 1 to denote presence of the condition and 0 to denote absence, or ‘fuzzy-sets’, with scores between 0 and 1, which denote partial membership of the case in the set. The data set of observations is a matrix, with the cases in rows and the conditions in columns, similar to the data set used in statistical software. This data set has an identifier (e.g., name of the organisation), several conditions, and an outcome variable. Several cases may share the same set of conditions. Therefore, QCA creates an overview of all possible combinations of conditions (configurations) and lists the cases that share the same combination. This matrix is called a ‘truth table’. QCA considers each row in the truth table with a proper threshold consistency score as a case-as-configuration, the bearer of a set of conditions that are sufficient for the outcome to occur. However, not all conditions are necessarily relevant for the causal explanation. Some may be trivial or redundant, while others may provide the clue to explain a causal relationship. Using Boolean logic, QCA searches for ‘simplest’ combinations of conditions that are still consistent with the outcome, using the Quine-McCluskey algorithm (McCluskey, 1956). This algorithm reduces the complex Boolean expressions of the rows in the truth-table to more parsimonious terms. QCA reports these Boolean expressions and lists the cases that are covered by each term. QCA supports the findings with a consistency score and the coverage of cases, which are indicators that help to judge the importance of the term for causal explanation (see Chapter 7, for more detail and an empirical application).

## 9.4 Realist analysis of organisational social capital

Despite one of my own design principles, in which I suggest to use common indicators to facilitate comparative analysis, I needed to develop a new indicator of organisational strength in economic farmer organisations. This was because the literature provided no other construct that assessed the organisational strength of groups that were similar to the FONDOECAS beneficiaries. This indicator has the potential to become a common indicator. I argue, however, that this tool’s usefulness goes beyond that which I made of it in my research in Bolivia, and is not limited to impact evaluation.

The sector of collective marketing groups is very diverse. It includes organisations of all sizes, and economic sectors as diverse as honey processing, handicrafts and coffee exporters. Having worked for many years with this sector in Bolivia, I knew that these organisations had much in common in spite of their differences. I was also aware that small organisations could be very strong and resilient in their collective action without having an office, staff or large turnover, and that organisations with many assets and a large turnover could be in disarray. I wanted to have a construct that could measure this condition of organisational strength relatively independently from economic performance, and that could reflect the quality of social relations in the group so as to advance the collective interest.



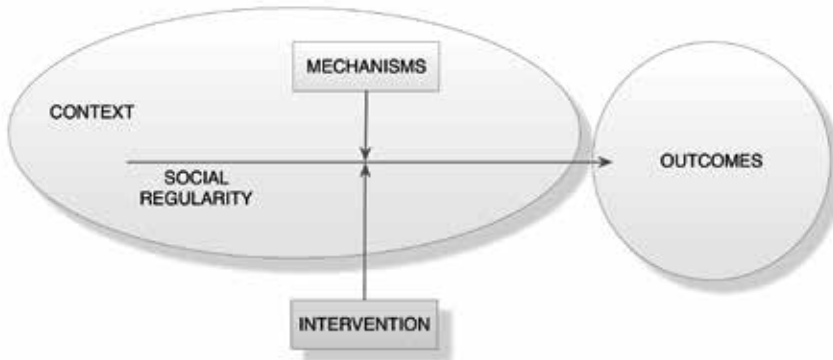
Similar concerns have been raised in the literature about the need for comparative indicators on collective action in natural resource management (Meinzen-Dick et al., 2004; Poteete and Ostrom, 2004). Most of these authors use the term social capital to refer to this quality of the group. The concept is used in many different ways. Woolcock (1998), for example, uses social capital to refer to the quality of the social linkages within communities, between communities and between the state and society. Most often, however, scholars use the term social capital to refer to an attribute of persons, not organisations, and the social networks in which these are involved or the informational resources that they have available. Collective marketing groups, typically with a self-selected membership that live in different villages, are in this approach considered as a part of the social capital of farmers (Grootaert and Narayan, 2004). They focus on the measurement of the density of a farmer's network, rather than on the quality of the organisations in the network. This focus on micro-level social capital is common for scholars and impact evaluators who use the sustainable-livelihood approach (Scoones, 1998), where livelihoods assets are analysed as human, social, financial, physical and natural capital.

In organisational studies, Leana and van Buren (1999) introduced the term 'organisational social capital', which I felt was a better name for the construct to describe the qualities of social relations in a group. This term is more easily understood as referring to the organisational quality of farmer organisations, as a subtype of social capital that is relatively independent from the social networks of members and the quality of social relations and institutions in a region or country. An economic farmer organisation can, thus, be conceptualised as a semi-autonomous field (Moore, 1973) in which the quality of the internal organisational agreements constitute its organisational social capital.

Farmer groups that are involved in collective marketing are a special form of organisation. They differ from farmer groups that have primarily social or political objectives. They are organised around economic transactions between the group and the member. These economic transaction and logistic operations within the group create problems of collective action that are qualitatively different than, for example, those in common resource management. Therefore, to differentiate it from other forms of organisational social capital, I labelled this collective marketing-specific form of organisational social capital as Tension Containment Capacity (TCC). In Chapter 6, I describe the TCC and data collection tool in more detail. In this concluding chapter I would like to present the TCC as a useful analytical framework and identify issues for further research.

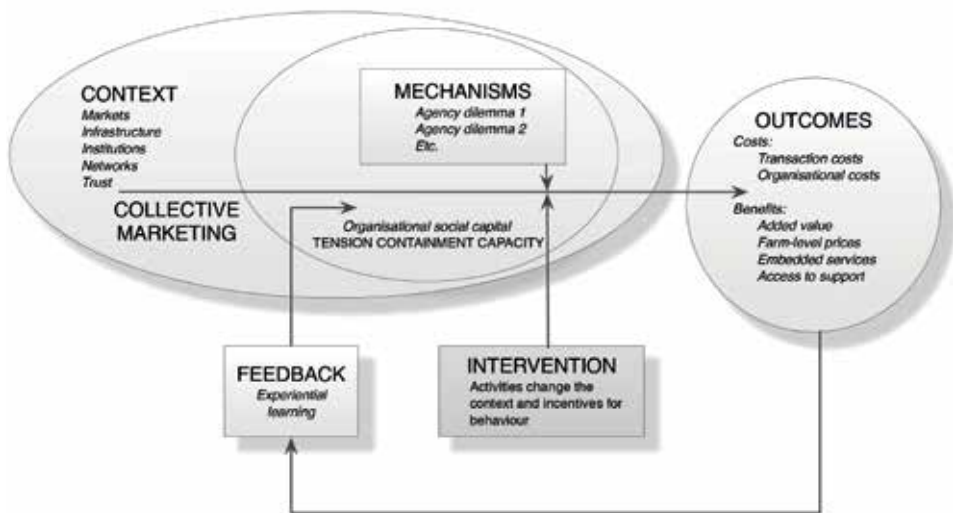
I conceptualised the support to economic farmer organisations using the realist framework of CMOC, Context-Mechanism-Outcome configuration. Pawson and Tilley (1997) use a simple figure to visualise the way that interventions cause social change. They focus on a specific social regularity, symbolised by the horizontal arrow, which as a result of some causal mechanism, present in the context, leads to certain outcomes. Support interventions influence the context, and change the way these mechanisms work, which (is expected to) result in different outcome patterns. Figure 9.3 presents this basic figure of realist analysis.

In Figure 9.4, I propose a similar format as a framework of analysis of collective marketing. In realist terms, collective marketing is the social 'regularity', symbolised by the central horizontal arrow. The causal mechanisms active in the context make this collective marketing result



Source: Author's own elaboration

Figure 9.3 Basic concepts in realist social explanation



Source: Author's own elaboration

Figure 9.4 Tension Containment Capacity as framework to study organisational social capital for collective marketing

in certain outcomes. I modified the basic figure, adding an arrow that represents the feedback process of experiential learning. It emphasises the fact that groups develop 'organisational intelligence' to resolve the disintegrative tendencies of various agency dilemmas (mechanisms). The organisation will learn from its own intents and experiments, and from the experiences of other organisations. Whenever the organisation conserves an institutional memory, through this experiential learning, older organisations will tend to grow stronger and become more resilient. This interdependency of context, agency dilemmas and tension containment capacity

needs more attention in social research. Collective marketing groups, especially cooperatives, are too easily suggested to be the panacea for smallholder market access, without giving due respect to the difficulties and costs that such an endeavour implies.

Leana and Van Buren (1999) stress that organisational social capital results in benefits and costs. The benefits of collective marketing are obvious, whereas the costs are often ignored. The costs of collective action add to the transaction costs, which determine to a large extent whether a collective marketing group can survive the competition with other market players (Williamson, 2002). I draw attention to organisational costs, e.g., for meetings, transportation and networking, which are generally only a tiny fraction of total turnover of a group but, nevertheless, prove to be very important in shaping organisational social capital. Because members tend to prefer high prices and low margins, board members and staff often have to operate with very low budgets for quality control, raising commitment of members or providing information about commercial transactions.

Figure 9.4 lists four key contextual elements. This is inspired by Ostrom and Ahn (2008), who distinguish trustworthiness, networks and institutions as the contextual conditions that support social capital in collective action. I added markets and infrastructure to emphasise these as important contextual influences on organisational social capital. For example, a common institutional constraint is provided by the Civil and Commercial Codes in a country, which define whether a group can distribute profit to its members, and the related tax regime. In many development countries, including Bolivia, the formal registering of a group automatically implies that they are at a competitive disadvantage relative to their direct competitors, the informal traders and intermediaries, because of rigid administrative and fiscal regulations (Mendoza and Ton, 2003). Other contextual factors may favour them, such as the access to niche markets through Fair Trade or preferential government procurement policies, like those in Brazil (Oldekop et al., 2015). But these also imply new organisational challenges and agency dilemmas. For example, preferential government procurement programmes require that groups have a system of food quality assurance in place. Furthermore, the time that governments delay the payment of their bills often implies that the group needs to introduce delayed payment systems. Certification schemes demand internal control systems that require costly third-party verification. I mention trust as another contextual element that defines the scope of an organisation to build organisational social capital. I did not find significant geographical differences in trust levels in Bolivia, when measuring it with questions derived from the World Value Survey (WVS, 2009) in the AR-LAT survey (Annex 5); however, it is an important contextual factor that may explain specific forms of resolving agency dilemmas, and the opportunity to build organisational social capital.

To 'capture' the tension containment capacity of an organisation, I developed a short semi-structured interview format for a 'realist interview' (Pawson, 1996). During the (approximately half-day) interviews, the researcher discussed ten agency dilemma (see Table 9.1) and used probing questions, discussing some organisational solutions that are commonly applied in other organisations, such as delayed payment instead of cash payment, or the need for an annual contribution in cash to better define the membership, etc. During this dialogue, both the interviewer and the interviewees 'discovered' the formal and informal rules and regulations that prevented the group from disintegrating or experiencing overt conflict.

**Table 9.1** Ten agency dilemmas where disintegrative tendencies in collective marketing are located

In short	Description
'Regulating Member Supply'	Tensions can emerge when individual members increase their supply to the marketing organisation, and, by doing so, negatively affect the opportunity for other members to supply.
'Quality Assurance Systems'	When a deal is made, the quality that the organisation has promised has to be checked: individual members may tend to deposit lower quality and the organisation needs a system to maintain minimum quality requirements.
'Coping with Working Capital Constraints'	Many smallholder farmers tend to face cash constraints and want quick payment, whereas the organisation needs time to complete transactions with the ultimate buyer.
'Anticipating Side-Selling'	The organisation might provide a credit service or advance payment system to enable production. However, there is a serious risk that farmers "side-sell" their product to competing traders or processors to whom they have no repayment obligation.
'Ways to Dispose of Profits'	When the organisation makes profit, the organisation prefers to invest or increase capital reserves, whereas the members prefer more short-term benefits, e.g., better prices.
'Differentiating Services to Members and Non-Members'	Most economic organisations need contributions from members to realize their business opportunities; however, members face a number of disincentives to do so when benefits that flow from investment accrue to investors and non-investors alike.
'Decision Making on Activities that Benefit Only a Sub-group'	When the type of investment is unlikely to benefit all members, investment decisions that seem economically optimal from the perspective of management are not necessarily desirable from the standpoint of (sub-groups of) members.
'Task Delegation and Supervision of Professional Staff'	Member-based organisations elect persons to supervise and support management; however, the limited technical knowledge of board members and the lack of transparency of information disclosed by management often limit the effectiveness of this governing structure.
'Liability in Contracts and Loans'	There is an inherent tension between members who want to limit their liability for group actions and the need of the group as a whole to generate as much collateral as possible. Organisations specify procedures for decision making when the board contracts on behalf of the group.
'Managing Political Aspirations'	Economic smallholders' organisations tend to take up a broader representative role in addition to their economic service provisioning to members. Members delegate their political voice to the organisation, whereas the political representatives of the organisation may never fully discuss all political decisions with them.

Source: Based on Ton (2010b), see Chapter 6 for more details

Each interview report provides a 'thick description' (Geertz, 1973) of the status of the organisational capabilities of an organisation to contain these agency dilemmas. The interview reports (approximately 10 pages) not only describe the rules and regulations but also reflect on the contextual influences and the apparent reasons for this particular way to contain the agency dilemma. The method, therefore, depends on the ability of the interviewer to discuss each of the agency dilemmas in depth in each organisation and write it up in the interview report<sup>30</sup>.

30 The complete set of interview reports (in Spanish) is available online at <http://www.esfim.org/case-studies-on-internal-governance-of-economic-farmer-organisations-in-bolivia>. An edited selection of these reports was published in: Flores L and Ton G. (2015) *Inteligencia Organizativa y Desempeño Económico de Organizaciones Económicas Campesinas: Dinámicas alrededor de planes de negocios para dar valor agregado a la producción campesina*, La Paz: CIOEC-Bolivia.

While this interview process and conceptualisation of the construct Tension Containment Capacity is firmly embedded in qualitative research, it is the conversion of the information from the interview into a summary sheet, and subsequently into a quantitative proxy-indicator of organisational social capital, that makes the link with quantitative research, be it configurational comparative or regression-analytic research. For example, Elders et al. (2012) and Wollni and Zeller (2007) recognise organisational strength of farmer groups as an important confounding factor when analysing farm-level impact of interventions, but they face problems when they try to take the organisational quality of the group into account. The quantitative TCC-score makes it possible to use the level of organisational social capital as an independent variable in regression-analytic models. I expect that this will improve the analysis of quasi-experimental survey data of farmer-level outcomes, and hope to show this in subsequent research.

## 9.5 Credibility of evidence

All impact evaluations covered in this thesis are related to grants and farmer groups. The intervention logics of these support programmes reflect the expectation of a positive effect of grants on the organisational strength and economic performance of these groups. The literature review on the effectiveness of innovation grants to smallholder farmers, in Chapter 2, showed that most authors believed in the positive effects of channelling innovation grants to farmer groups, even when the body of credible evidence to support those assumptions is rather small. This suggests a publication bias: studies on grants to farmer groups are likely to be done by scholars who sympathise with the logic behind the interventions, and, therefore, may be more motivated to publish positive than negative results. I am certainly not immune to this bias. I sympathise with interventions that aim to strengthen farmer organisations, and I was involved in the design of FONDOECAS, the object of the empirical research in this thesis. I hope to have reduced this validity threat to my findings by being transparent about my research methods and handling of data, and in drawing cautious conclusions.

Unlike the impact evaluations on certification and micro-irrigation, discussed in Chapter 3 and 5, the research on FONDOECAS did not result in quantitative estimates of effects. This is perhaps the main limitation of this study. Chapter 4 and 5, written at the initial stage of the research in Bolivia, describe my initial design, to implement the case-based comparative analysis in a quasi-experimental setting with a comparison group. However, I soon experienced that the sample was much more diverse than initially expected, with very small groups of less than 20 members and large groups with thousands of members. As a result, the variance in the indicators was prohibitively high for any quasi-experimental design to derive a meaningful group average. Nevertheless, I decided to keep a small comparison group to nurture counterfactual thinking. Contagion effects and logistic implementation problems eroded the size of the comparison group even further. The comparison group, therefore, played only a minor role in the analysis.

However, counterfactual thinking can be supported with other information than data from a comparison group. Counterfactual thinking implies a structured and transparent way of discarding alternative explanations for an effect (Vellema et al., 2013; Yin, 2013). This requires reflection on the question, What would have happened without the intervention? In the

FONDOECAS research this implies a case-by-case evaluation of the contributory role of the grant in generating certain outcomes, based on a close reading of the dynamics within the organisation, looking for ‘traces’ of grant-induced change and alternative explanations. This ‘process tracing’ as a method of counterfactual analysis (Perri 6, 2006; Beach and Pedersen, 2013) is feasible only when the sample size is not too large and when rich qualitative information on each case is available.

I showed that ‘real-world’ conditions constrain sample sizes and can make it difficult to obtain credible effect estimates. For many interventions, it is only possible to collect data on a relatively small number of beneficiaries, for example on organisations, firms, networks or countries, and the number of these units within a certain geographical area is small. Funding agencies, such as the International Initiative for Impact Evaluation (3ie), require that effect estimates have sufficient statistical power. Impact studies with a low statistical power can detect only large effects. But sample size is also a problem in household surveys. In Chapter 3, I showed that very large samples would be required to measure the effects of certification on yields and farmer income with sufficient statistical power, and I used this as an example of the ‘boundary’ of the span of direct influence. Each quasi-experimental impact evaluation design will need to acknowledge a span of direct influence, considering the complexity of the impact pathways that are being studied (Rogers, 2009), and the budgetary, time, data and political constraints for the impact research design (Bamberger et al., 2006).

In the case of FONDOECAS, I expected that the grant would have a measurable effect on the organisational capacities of the farmer groups. I also hoped that the effects would be reflected in the volume of group sales, although I anticipated that this would be less likely. The FONDOECAS research, in Chapter 8, shows that the monitoring of these outcomes provided useful information for analysis and reflection, but that both outcome areas appeared to be outside the span of direct influence. The time-series data on group sales, differentiated into processed and unprocessed products, proved to be measures that were too rough, because they often reflected other business activities of the group that had no relation with the grant-supported business plan. More detailed information and analysis of the economic effects of the business plans in each of the organisations would likely have permitted a rough estimate of ‘before-after’ effects, but these detailed analyses were not feasible within the budgetary and logistical constraints of the research for this thesis. Therefore, I focused the research only on the contributory role of the grant and the contextual conditions that increase the likelihood that the grants would be successful. In other impact evaluations in which I have been involved, however, the boundary of the span of control is such that quantitative estimates of net-effects on certain outcome areas are feasible, and also informative enough to deserve the research effort.

The seven principles that I distilled from these research experiences help to design impact evaluation that balance all three evaluation questions included in Figure 9.2: What works for whom under what conditions. The learning potential of impact evaluations would be greatly enhanced were this question to be more prominent in commissioned impact research. Not only ‘realistas’ like Pawson, who puts this at the centre of his manifesto for a ‘Science of Evaluation’ (Pawson, 2013), but also a ‘randomista’ like Blattman stresses the importance of this broader evaluation question in his plea for an ‘Impact Evaluation 2.0’ (Blattman, 2008). I look forward to applying these design principles in future impact evaluations. But, even

more, I hope that they may inspire others. Impact evaluations have a role to play in the quest for knowledge to address development problems. In the introduction I placed a quote from Ray Pawson (2013) from his book *The Science of Evaluation: a realist manifesto*, in which he argues for a social science that helps to discover partial truths about causal mechanisms that explain effectiveness. In this synthesis I quoted Abhijit Banerjee and Esther Duflo (2011), the champions of randomised controlled trials in development economics research, from their book *Poor Economics: a radical rethinking of the way to fight global poverty*. They point to the need for a growing body of knowledge to give us a ‘best shot at, one day, ending poverty’. I would be honoured if my design principles were to help others design creative and appropriate impact designs to generate *plausible evidence* on development interventions that are *likely to be effective* to help disfavoured groups, such as smallholder farmers in development countries. This could help to make the world a better place.

## References

- Adhikari KP and Goldey P. (2010) Social Capital and its “Downside”: The Impact on Sustainability of Induced Community-Based Organizations in Nepal. *World Development* 38: 184-194.
- Adjei-Nsiah S, Kuyper TW, Leeuwis C, et al. (2008) Farmers’ agronomic and social evaluation of productivity, yield and N<sub>2</sub>-fixation in different cowpea varieties and their subsequent residual N effects on a succeeding maize crop. *Nutrient Cycling in Agroecosystems* 80: 199-209.
- Alvarez G and von Hagen O. (2011) The Impacts of Private Standards on Producers in Developing Countries. *Literature Review Series on the Impacts of Private Standards, Part II*, Geneva: International Trade Centre.
- Alvarez G and von Hagen O. (2012) When Do Private Standards Work? *Literature Review Series on the Impacts of Private Standards; Part IV*. Geneva: International Trade Centre.
- Anellis IH. (2012) Peirce’s Truth-functional Analysis and the Origin of the Truth Table. *History and Philosophy of Logic* 33: 87-97.
- Aramayo R. (2008) Encuesta Fundamental para el Desarrollo Económico y Social: Informe 01. La Paz: INE-World Bank - CAN.
- Aramayo R. (2013) Encuesta para Levantamiento de Información Primaria en 40 Municipios del Programa FONDOECAS - Informe final (revisado). La Paz: FONDOECAS-LEI-AR-LAT-CIOEC, 76.
- Ashby J, Heinrich G, Burpee G, et al. (2009) What farmers want: collective capacity for sustainable entrepreneurship. *International Journal of Agricultural Sustainability* 7: 130-146.
- Ashby JA, Braun AR, Garcia T, et al. (2000) Investing in Farmers as Researchers: experience with local agricultural research committees in Latin America. Cali: Centro Internacional de Agricultura Tropical (CIAT).
- Ashby JA, Braun AR, Gracia T, et al. (2001) *La Comunidad se Organiza para Hacer Investigación: Experiencias de los Comités de Investigación*, Cali: Centro Internacional de Agricultura Tropical (CIAT).
- Bachmann R. (2003) Trust and power as means of co-ordinating the internal relations of the organization: a conceptual framework. *The trust process in organizations: Empirical studies of the determinants and the process of trust development*: 58-74.
- Bamberger M and Rugh J. (2008) A framework for assessing the quality, conclusion validity and utility of evaluations. Experience from international development and lessons for developed countries. *American Evaluation Association 2007*. Baltimore.
- Bamberger M, Rugh J, Church M, et al. (2004) Shoestring evaluation: Designing impact evaluations under budget, time and data constraints. *American Journal of Evaluation* 25: 5.
- Bamberger M, Rugh J and Mabry L. (2006) *RealWorld evaluation: working under budget, time, data, and political constraints*: Sage Publications Inc.
- Banerjee AV and Duflo E. (2011) *Poor economics: A radical rethinking of the way to fight global poverty*: PublicAffairs.
- Banful Afua B. (2011) Old Problems in the New Solutions? Politically Motivated Allocation of Program Benefits and the “New” Fertilizer Subsidies. *World Development* 39: 1166-1176.
- Barham J and Chitemi C. (2009) Collective action initiatives to improve marketing performance: Lessons from farmer groups in Tanzania. *Food Policy* 34: 53-59.



- Barnett AG, van der Pols JC and Dobson AJ. (2005) Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology* 34: 215-220.
- Barrett CB, Bachke ME, Bellemare MF, et al. (2012) Smallholder participation in contract farming: Comparative evidence from five countries. *World Development* 40: 715-730.
- Bashkar R. (1978) A realist theory of science. *A Realist Theory of Science*.
- Beach D and Pedersen RB. (2013) *Process-tracing methods: Foundations and guidelines*: University of Michigan Press.
- Bebbington A. (1996) Organizations and Intensifications: campesino federations, rural livelihoods and agricultural technology in the Andes and Amazonia. *World Development* 24: 1161-1177.
- Bebbington A. (1997) Social capital and rural intensification: Local organizations and islands of sustainability in the rural Andes. *Geographical Journal* 163: 189-197.
- Bebbington A. (1999) Capitals and capabilities: a framework for analyzing peasant viability, rural livelihoods and poverty. *World Development* 27: 2021-2044.
- Bebbington A and Sotomayor O. (1998a) Case study: agricultural extension in Chile. *Financing the Future*. Chili: Oxford Policy Management, 17.
- Bebbington A and Sotomayor O. (1998b) Demand-Led and Poverty-Oriented . . . Or Just Subcontracted and Efficient? Learning from (Semi-) Privatized Technology Transfer Programmes in Chile. *Journal of International Development*. Jan. Feb. 1998; 10: 17-34.
- Beekman G, Bulte E and Nillesen E. (2014) Corruption, investments and contributions to public goods: Experimental evidence from rural Liberia. *Journal of Public Economics* 115: 37-47.
- Befani B. (2012) Models of causality and causal inference. In: Stern E, Stame N, Mayne J, et al. (eds) *Broadening the Range of Designs and Methods for Impact Evaluations: report of a study commissioned by the Department for International Development*. London: DFID.
- Befani B. (2013) Between complexity and generalization: Addressing evaluation challenges with QCA. *Evaluation* 19: 269-283.
- Befani B and Mayne J. (2014) Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation. *IDS Bulletin* 45: 17-36.
- Begg C. (1994) Publication bias. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation, 399-409.
- Benin S, Nkonya E, Okecho G, et al. (2007) Assessing the Impact of the National Agricultural Advisory (NAADS) in the Uganda Rural Livelihoods. Uganda: IFPRI, 92.
- Benin S, Nkonya E, Okecho G, et al. (2008) Impact Evaluation and Returns to Investment of the National Agricultural Advisory Services (NAADS) Program of Uganda. IFPRI.
- Benin S, Nkonya E, Okecho G, et al. (2011) Returns to spending on agricultural extension: the case of the National Agricultural Advisory Services (NAADS) program of Uganda. *Agricultural Economics* 42: 249-267.
- Berdegúé JA. (2001) *Cooperating to Compete: Associative Peasant Business Firms in Chile*: Wageningen University.
- Bernard T, De Janvry A and Sadoulet E. (2010) When Does Community Conservatism Constrain Village Organizations? *Economic Development and Cultural Change* 58: 609-641.
- Bienabé E, Coronel C, Le Coq J-F, et al. (2004) Linking Small Holder Farmers to Markets: lessons learned from literature review and analytical review of selected projects. Washington: World Bank.

- Bijman J, Iliopoulos C, Poppe KJ, et al. (2012) Support for Farmers' Cooperatives. *Final Report European Commission, Brussels*.
- Bijman J, Muradian R and Cechin A. (2011) Agricultural Cooperatives and Value Chain Coordination. In: Helmsing AHJ and Vellema S (eds) *Value Chains, Inclusion and Endogenous Development: Contrasting Theories and Realities*. London: Routledge.
- Blackman A and Rivera J. (2010) The Evidence Base for Environmental and Socio-Economic Impacts of 'Sustainable' Certification. Washington DC: Resources for the Future.
- Blackmore E, Keeley J, Pyburn R, et al. (2012) Assessing the benefits of sustainability certification for small-scale farmers in Asia. In: Mayers J (ed) *IIED Natural Resource Issues*. London: IIED.
- Blaikie NWH. (1991) A critique of the use of triangulation in social research. *Quality and Quantity* 25: 115-136.
- Blaikie NWH. (2000) *Designing social research: The logic of anticipation*, Cambridge: Polity Press.
- Blatter J and Blume T. (2008) In search of co-variance, causal mechanisms or congruence? Towards a plural understanding of case studies. *Swiss Political Science Review* 14: 315-356.
- Blattman C. (2008) Impact Evaluation 2.0. *Presentation to the Department for International Development (DFID), London*.
- Blokland C. (1992) *Participacion campesina en el desarrollo economico: la Union Nacional de Agricultores y Ganaderos de Nicaragua durante la Revolucion sandinista*, Doetinchem: PFS.
- Borgen SO. (2004) Rethinking incentive problems in cooperative organizations. *Journal of Socio-Economics* 33: 383-393.
- Bosc P-M, Eychenne D, Hussein K, et al. (2001) *Reaching the Rural Poor: The Role of Rural Producers Organisations (RPOs) in the World Bank Rural Development Strategy - background study*, Washington: World Bank.
- Brady HE and Collier D. (2004) *Rethinking social inquiry: diverse tools, shared standards*: Rowman & Littlefield Publishers.
- Brady HE, Collier D and Seawright J. (2006) Toward a Pluralistic Vision of Methodology. *Political Analysis* 14: 353-368.
- Braun A, Thiele G and Fernandez M. (2000) Farmer Field Schools and Local Agricultural Research Committees: complementary platforms for integrated decision-making in sustainable agricultural: CIAL. *AGREN Network Paper No.105*.
- Bukenya C. (2010) Meeting Farmer Demand? An assessment of extension reform in Uganda. *Technology and Agricultural Development*. Wageningen: Wageningen University.
- Bulte E, Beekman G, Di Falco S, et al. (2014) Behavioral responses and the impact of new agricultural technologies: Evidence from a double-blind field experiment in Tanzania. *American Journal of Agricultural Economics* 96: 813-830.
- Byrne D and Ragin C. (2009) *The SAGE Handbook of Case-Based Methods*: Sage Publications Ltd.
- Cechin A, Bijman J, Pascucci S, et al. (2013) Decomposing the Member Relationship in Agricultural Cooperatives: Implications for Commitment. *Agribusiness* 29: 39-61.
- Chelimsky E and Shadish WR. (1997) *Evaluation for the 21st century: A handbook*: Sage.
- Chen HT. (1994) *Theory-Driven Evaluations*: Sage Publications Inc.

- Chen S and Schreiner M. (2009) Progress out of Poverty Index: a simple poverty scorecard for Indonesia. Progress out of Poverty - Grameen Foundation.
- Chocolate Working Group. (2010) Letter of Intent: sustainable cocoa consumption and cocoa production. In: IDH - Ministry of Agriculture NaFQ (ed).
- CIOEC-Bolivia. (2000, 2004) Agenda para el Desarrollo Estratégico de las Organizaciones Económicas Campesinas. La Paz: CIOEC-Bolivia.
- CIOEC-Bolivia. (2009) *1er Censo Nacional de Organizaciones Económicas Campesinas, Indígenas y Originarias*, La Paz: CIOEC-Bolivia.
- CIOEC-Bolivia. (2010) Las Organizaciones Economicas Campesinas, Indígenas y Originarias en el Desarrollo del País. La Paz: CIOEC.
- Claar VV and Haight CE. (2015) Is Fair Trade Worth Its Cost? *Faith & Economics* 65: 25-34.
- Coase RH. (1937) The nature of the firm. *economica* 4: 386-405.
- Cohen J. (1988) Statistical power for the social sciences. *Hillsdale, NJ: Laurence Erlbaum and Associates*.
- Collier D. (2011) Understanding Process Tracing. *Political Science and Politics* 4: 823-830.
- Collier D. (2014) Comment: QCA Should Set Aside the Algorithms. *Sociological Methodology*: 0081175014542568.
- Collier P and Dercon S. (2014) African Agriculture in 50 Years: Smallholders in a Rapidly Changing World? *World Development* 63: 92-101.
- Condori G. (2005) Proyecto Fortalecimiento de las OECAs para el Desarrollo Económico Rural - documento final. La Paz: CIOEC-Bolivia.
- Connell JP and Kubish AC. (1998) Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects, and Problems. In: Fulbright-Anderson K, Kubish AC and Connel JP (eds) *Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives Vol2: theory, measurement, and analysis*. The Aspen Institute.
- Cook ML. (1994) The role of management behavior in agricultural cooperatives. *Journal of Agricultural Cooperation* 9: 42-58.
- Cook ML, Burrell MJ and Iliopoulos C. (2008) New producer strategies: the emergence of patron-driven entrepreneurship. *12th EAAE Congress 'People, Food and Environments: Global Trends and European Strategies', 26-29 August 2008*. Gent (Belgium), , 24-27.
- Cordoba D. (2014) Participation, Politics and Technology: Agrarian development in post-liberal Bolivia. *Wageningen School of Social Sciences*. Wageningen: Wageningen University.
- Cordoba D and Jansen K. (2013) The Return of the State: Neocollectivism, Agrarian Politics and Images of Technological Progress in the MAS Era in Bolivia. *Journal of Agrarian Change*: n/a-n/a.
- Cornforth C. (2004) The Governance of cooperatives and mutual associations: a paradox perspective. *Annals of Public & Cooperative Economics* 75: 11-32.
- COSA. (2013) The COSA Measuring Sustainability Report: Coffee and Cocoa in 12 Countries. Philadelphia: The Committee on Sustainability Assessment.
- Creevey L and Ndiaye M. (2008) Common Problems in Impact Assessment Research. *Impact Assessment Primer Series #7*. Washington: USAID.
- Creevey L and Woller G. (2006) Methodological Issues in Conducting Impact Assessments of Private Sector Development Programs. *Impact Assessment Primer Series #2*. Washington: USAID.
- Cronqvist L. (2009) Tosmana - Tool for Small-N Analysis. Version 1.3.1 ed. Trier: Tosmana.

- Crosse W, Newsom D and Kennedy E. (2012) Appendix 1: Recommendations for Improving Research on Certification Impacts. In: Certification SCotS-o-K-AoS (ed) *Toward Sustainability: the roles and limitations of certification*. 169.
- DCED. (2010) The DCED Standard for Measuring Achievements in Private Sector Development: Control Points and Compliance Criteria. Version V.: Donor Committee for Enterprise Development.
- De Meur G, Bursens P and Gottscheiner A. (2006) MSDO/MDSO revisited for public policy analysis. *Innovative comparative methods for policy analysis*. Springer, 67-94.
- Deaton A. (2010) Instruments, randomization, and learning about development. *Journal of Economic Literature*: 424-455.
- Denning G, Kabambe P, Sanchez P, et al. (2009) Input Subsidies to Improve Smallholder Maize Productivity in Malawi: Toward an African Green Revolution. *Plos Biology*. Malawi, 9.
- Denzin NK. (1970) *The research act*: Aldine Chicago.
- DGIS. (2011) Protocol Resultaatsbereiking en Evalueerbaarheid in PSD 2011 e.v. . The Hague: DGIS-DDE - IOB.
- Donaldson S, Christie C and Mark M. (2008) *What counts as credible evidence in applied research and evaluation practice?*: Sage Publications, Inc.
- Donovan J and Stoian D. (2012) 5Capitals: a tool for assessing the poverty impacts of value chain development. *Technical Bulletin no.55*. Turrialba: CATIE.
- Donovan J, Stoian D and Poole N. (2008) A Global Review of Rural Community Enterprises: the long and winding road for creating viable businesses. *Technical Bulletin no.29*. Turrialba: CATIE.
- Dormon E, Van Huis A, Leeuwis C, et al. (2004) Causes of low productivity of cocoa in Ghana: farmers' perspectives and insights from research and the socio-political establishment. *NJAS-Wageningen Journal of Life Sciences* 52: 237-259.
- Dorward A and Chirwa E. (2011) The Malawi agricultural input subsidy programme: 2005/06 to 2008/09. *International Journal of Agricultural Sustainability* 9: 232-247.
- Dorward A, Chirwa E, Kelly VA, et al. (2008) Evaluation fo the 2006/2007 Agricultural Input Subsidy Programme, Malawi., 141.
- Douthwaite B. (2002) *Enabling innovation*: MIT Press.
- Douven I. (2011) *Abduction*. Available at: <http://stanford.library.usyd.edu.au/entries/abduction/>.
- Duflo E, Glennerster R and Kremer M. (2007) Using randomization in development economics research: A toolkit. *Handbook of development economics* 4: 3895-3962.
- Duflo E, Glennerster R, Kremer M, et al. (2006) Using Randomization in Development Economics Research: A Tool Kit. *MIT Department of Economics Working Paper No. 06-36*.
- Duşa A and Thiem A. (2014) QCA: A Package for Qualitative Comparative Analysis. R Package Version 1.1.4
- Dybdal L, Bohni Nielsen S and Lemire S. (2010) Contribution Analysis Applied: Reflections on scope and methodology. *CANADIAN JOURNAL OF PROGRAM EVALUATION* 25.
- EES. (2007) EES Statement: The importance of a methodologically diverse approach to impact evaluation—specifically with respect to development aid and development interventions. In: European Evaluation Society (ed).
- EFQM. (2010) *EFQM Excellence Model*, Brussels: EFQM.

- Ekwamu A and Brown M. (2005) Four years of NAADS implementation program outcomes and impact. *Proceedings of the Mid Term Review of National Agricultural Advisory Services*. 25-46.
- El Hage N. (2012) Guidelines for Sustainability Assessment in Food and Agriculture. *FAO* 2012.
- Elder SD, Zerriffi H and Le Billon P. (2012) Effects of Fair Trade Certification on Social Capital: The Case of Rwandan Coffee Producers. *World Development* 40: 2355-2367.
- Elías B. (2007) *Compro Boliviano: Evaluación de la aplicación del Decreto*. La Paz: Centro de Investigación y Promoción del Campesinado (CIPCA).
- European Commission. (2008) *Second Strategic Review of Better Regulation in the European Union* Brussels: European Commission.
- Fafchamps M. (2004) *Market institutions in sub-Saharan Africa: Theory and evidence*. MIT Press Books 1.
- Farrington DP. (2003) Methodological Quality Standards for Evaluation Research. *The ANNALS of the American Academy of Political and Social Science* 587: 49-68.
- Fiss PC, Cambré B and Marx A. (2013a) *Configurational theory and methods in organizational research*. Available at: <http://site.ebrary.com/id/10705877>.
- Fiss PC, Sharapov D and Cronqvist L. (2013b) Opposites attract? Opportunities and challenges for integrating large-N QCA and econometric analysis. *Political Research Quarterly*: 191-198.
- Flores L, Commandeur D, Ton G, et al. (2007) Coaching Organisational Transitions Towards Increased Market Orientation: SNV's experience with producer organisation support in the Bolivian valleys. In: Ton G, Bijman J and Oorthuizen J (eds) *Producer Organisations and Market Chains: facilitating trajectories of change in developing countries*. Wageningen: Wageningen Academic Publishers.
- Flores L and Ton G. (2015) *Inteligencia Organizativa y Desempeño Económico de Organizaciones Económicas Campesinas: Dinámicas alrededor de planes de negocios para dar valor agregado a la producción campesina*, La Paz: CIOEC-Bolivia.
- FONDOECAS. (2007) *Propuesta consolidada - criterios de calificación de proyectos*. La Paz: FONDOECAS.
- FONDOECAS. (2010) *Plan Estratégico Fondo de Fortalecimiento Económico para las OECAs 2011-2013*. La Paz: CIOEC-FONDOECAS-AOPEB, 60.
- FONDOECAS MyE. (2012) *Propuesta para Mejorar los Mecanismos de Otorgación de Fondos*. La Paz: FONDOECAS.
- Friis-Hansen E. (2008) Impact assessment of farmer institutional development and agricultural change: Soroti district, Uganda. *Development in Practice* 18.
- Friis-Hansen E and Egelyng H. (2006) Supporting local innovation for rural development: analysis and review of five innovation support funds. Copenhagen: DIIS, 88.
- Fundación Chile. (2009) *Impactos de los instrumentos de Transferencia tecnológica - Chile*. Santiago de Chile: Fundación Chile - Área Agroindustrias.
- Funnell SC and Rogers PJ. (2011) *Purposeful program theory: Effective use of theories of change and logic models*: John Wiley & Sons.
- Gandarillas A, Blajos J, Aguirre G, et al. (2007) Changing paradigms for organising R&D: Agricultural research and the creation of the PROINPA Foundation in Bolivia. *International journal of agricultural resources, governance and ecology* 6: 256-276.
- Garforth C, Angell B, Archer J, et al. (2003) Fragmentation or creative diversity? Options in the provision of land management advisory services. *Land Use Policy* 20: 323-333.

- Geertz C. (1973) Thick description: toward an interpretive theory of culture. *The interpretation of cultures*. New York: Basic Books, 3-30.
- Gereffi G, Humphrey J and Sturgeon T. (2005) The governance of global value chains. *Review of International Political Economy* 12: 78-104.
- Gerring J. (2007) The mechanistic worldview: thinking inside the box. *British Journal of Political Science* 38: 161-179.
- Giddens A. (1979) *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis*: University of California Press.
- Gough D, Oliver S and Thomas J. (2012) *An Introduction to Systematic Reviews*. , London: Sage Publications Ltd.
- Gough D and Thomas J. (2012) Commonality and diversity in reviews. In: Gough D, Oliver S and Thomas J (eds) *Introduction to systematic reviews*. London: EPPI-Centre, 35-65.
- Govere I, Foti R, Mutandwa E, et al. (2009) Policy perspectives on the role of government in the distribution of agricultural inputs to farmers: lessons from Zimbabwe. *International NGO Journal* 4: 470-479 478 ref.
- Government of Bolivia. (2011) Ley 144 - Ley de la Revolución Productiva Comunitaria Agropecuaria. La Paz.
- Grootaert C and Narayan D. (2004) Local institutions, poverty and household welfare in Bolivia. *World Development* 32: 1179-1198.
- Guijt I. (2008) Seeking Surprise: Rethinking Monitoring for Collective Learning in Rural Resource Management. *Communication Studies*. Wageningen, Wageningen University. PhD.
- Guijt I, Brouwers J, Kusters C, et al. (2011) Evaluation Revisited: improving the quality of evaluative practice by embracing complexity - Conference Report. Wageningen: Centre for Development Innovation, Wageningen UR - Context, international cooperation - Learning by Design.
- Gustafson DJ. (2002) Supporting the demand for change: recent experiences with farmer learning grants in Kenya. Case study for the workshop "Extension and Rural Development: Convergence of Views on International Approaches? : World Bank, Washington DC.
- Hagen-Zanker J. (2012) Making systematic reviews work for international development research [online]. Secure Livelihoods Research Consortium Briefing Paper 1.
- Haight C. (2011) The problem with fair trade coffee. *Stanford Social Innovation Review* 3: 74-79.
- Hall A, Clark N and Naik G. (2007) Technology supply chain or innovation capacity?: Contrasting experiences of promoting small scale irrigation technology in South Asia. *Technology*: 014.
- Hartwich F, Alexaki A and Baptista R. (2007) Innovation Systems Governance in Bolivia - lessons for agricultural innovation policies. IFPRI.
- Healy K. (2001) *Llamas, weavings, and organic chocolate*: University of Notre Dame Press.
- Heckman JJ. (2005) The scientific model of causality. *Sociological Methodology* 35: 1-98.
- Heckman JJ and Smith JA. (1995) Assessing the case for social experiments. *The Journal of Economic Perspectives* 9: 85-110.
- Heemskerck W and Wennink B. (2005) *Stakeholder-driven funding mechanisms for agricultural innovation : case studies from Sub-Saharan Africa*, Amsterdam: Royal Tropical Institute (KIT) KIT Development policy and practice.
- Heinrich M. (2014) Demonstrating Additionality in Private Sector Development Initiatives: A practical exploration of good practice for Challenge Funds and other cost-sharing mechanisms. London: DCED.

- Hellin J, Lundy M and Meijer M. (2009) Farmer organization, collective action and market access in Meso-America. *Food Policy* 34: 16-22.
- Helmsing AHJ. (2003) Local economic development: new generations of actors, policies and instruments for Africa. *Public Administration and Development* 23: 67-76.
- Helmsing AHJ and Vellema S. (2011) *Value Chains, Inclusion and Endogenous Development: Contrasting Theories and Realities*: Routledge.
- Hendrikse G and Bijman J. (2002) Ownership Structure in Agrifood Chains: The Marketing Cooperative. *American Journal of Agricultural Economics* 84: 104-119.
- Hendrikse GWJ. (2005) Contingent control rights in agricultural cooperatives. *Strategies for Cooperation*. Shaker Verlag: 385-394.
- Henehan B and Anderson B. (1994) Decision Making in Membership Organizations: A Study of Fourteen US Cooperatives. *ARME Research Bulletin* 94-5, Ithaca, NY: Cornell University.
- Henehan B and Anderson B. (2001) Considering Cooperation: A Guide for New Cooperative Development. *Department of Applied Economics and Management College of Agriculture and Life Sciences, Cornell University Ithaca, New York*: 14853-17801.
- Higgins J and Green S. (2011) Cochrane handbook for systematic reviews of interventions Version 5.1. o [www.cochrane-handbook.org](http://www.cochrane-handbook.org). The Cochrane Collaboration.
- Holden S and Lunduka R. (2010a) Impacts of the fertilizer subsidy programme in Malawi: targeting, household perceptions and preferences. Malawi, 38.
- Holden S and Lunduka R. (2010b) Too poor to be efficient? Impacts of the targeted fertilizer subsidy programme in Malawi on farm plot level input use, crop choice and land productivity. Malawi: NORAD, UMB, 59.
- Holden S and Lunduka R. (2012) Do fertilizer subsidies crowd out organic manures? The case of Malawi. *Agricultural Economics* 43: 303-314.
- Holling CS. (1973) Resilience and stability of ecological systems. *Annual review of ecology and systematics* 4: 1-23.
- Hospes O. (2008) Evaluation Evolution?- three approaches to evaluation. *The Broker* 2008: 24-26.
- Hounkonnou D, Kossou D, Kuyper TW, et al. (2012) An innovation systems approach to institutional change: Smallholder development in West Africa. *Agricultural Systems* 108: 74-83.
- Humphrey J and Navas-Aleman L. (2009) Multinational Value Chains, Small and Medium Enterprises, and 'Pro-Poor' Policies: A Review of Donor Practice. *IDS Research Report* 63.
- Humphries S, Gallardo O, Jimenez J, et al. (2005) Linking small farmers to the formal research sector: lessons from a participatory bean breeding programme in Honduras. *Agren Network Paper* 142.
- Humphries S, Gonzales J, Jimenez J, et al. (2000) Searching for sustainable land use practices in Honduras: lessons from a programme of participatory research with hillside farmers.
- IBM Corp. (2012) IBM SPSS Statistics for Windows, Version 21.0. Amonk, New York: IBM Corp.
- ICA. (1995) Co-operative Principles. Geneva: International Co-operative Alliance.
- IEG-World Bank. (2009) Agricultural research and competitive grant schemes: an IEG performance assessment of four projects in Latin America. World Bank.
- IFAD-UNEP. (2013) Smallholders, food security and the environment. *Rome: International Fund for Agricultural Development*.

- Ingram V, Waarts Y, Ge L, et al. (2014) Impact of UTZ Certification of Cocoa in Ivory Coast. The Hague: LEI Wageningen UR.
- ISEAL Alliance. (2013) Demonstrating and Improving Poverty Impacts: ISEAL Common Core Indicators. London: ISEAL Alliance.
- ISEAL Alliance. (2014) Code of Good Practice for Assessing the Impacts of Social and Environmental Standards Systems: revision 2.0. London: ISEAL Alliance.
- Johnson N, Suarez R and Lundy M. (2002) *The importance of social capital in Colombian rural agro-enterprises*, Washington: International Food Policy Research Institute.
- Kaaria S, Lilja N, Sandoval V, et al. (2006) Assessing Impacts of Farmer Participatory Research Approaches: A Case Study of Local Agricultural Research Committees in Colombia (CIALs). 21.
- Karkkainen B. (2005) Panarchy and Adaptive Change: Around the Loop and Back Again. *Minn. J. Sci. & Tech.* 7: 59.
- Kawai A. (2011) ¿Cómo las organizaciones de productores ecológicos (OPEs) tratan de mejorar la comercialización colectiva? La importancia de mecanismos organizacionales para mejorar la calidad en cuatro OPEs cafetaleras en Caranavi, Bolivia. *Farming Systems Ecology*. Wageningen: Wageningen University, 45.
- KENFAP. (2010) An Assessment of the Effects of Government Interventions in Input and Output Markets in Kenya: a case of maize seed, fertilizer and maize grain. KENFAP, 32.
- Kerr NL. (1998) HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review* 2: 196-217.
- Kessler A. (2013) Guidance on how to use the DCED Standard to measure results in Challenge Funds. London: DCED.
- Khandker S, Koolwal B and Samad H. (2009) *Handbook on Impact Evaluation*, Washington: World Bank.
- Kidd AD, Lamers JPA, Ficarelli PP, et al. (2000) Privatising agricultural extension: caveat emptor. *Journal of Rural Studies* 16: 95-102.
- Kilelu CW, Klerkx L and Leeuwis C. (2013) Unravelling the role of innovation platforms in supporting co-evolution of innovation: Contributions and tensions in a smallholder dairy development programme. *Agricultural Systems* 118: 65-77.
- Kilelu CW, Klerkx L, Leeuwis C, et al. (2011) Beyond knowledge brokering: an exploratory study on innovation intermediaries in an evolving smallholder agricultural system in Kenya. *Knowledge Management for Development Journal* 7: 84-108.
- Klerkx L, De Grip K and Leeuwis C. (2006) Hands off but Strings Attached: The Contradictions of Policy-induced Demand-driven Agricultural Extension. *Agriculture and Human Values* 23: 189-204.
- Klerkx L and Jansen J. (2010) Building knowledge systems for sustainable agriculture: supporting private advisors to adequately address sustainable farm management in regular service contacts. *International Journal of Agricultural Sustainability* 8: 148-163.
- Klerkx L and Leeuwis C. (2008) Matching demand and supply in the agricultural knowledge infrastructure: Experiences with innovation intermediaries. *Food Policy* 33: 260-276.
- Klerkx L and Leeuwis C. (2009) Operationalizing Demand-Driven Agricultural Research: Institutional Influences in a Public and Private System of Research Planning in The Netherlands. *The Journal of Agricultural Education and Extension* 15: 161 - 175.



- Klerkx L and Nettle R. (2013) Achievements and challenges of innovation co-production support initiatives in the Australian and Dutch dairy sectors: A comparative study. *Food Policy* 40: 74-89.
- Kosfeld M and Rustagi D. (2012) Leader Punishment and Cooperation in Groups: experimental field evidence from commons management in Ethiopia. Tilburg.
- Krogslund C, Choi DD and Poertner M. (2015) Fuzzy Sets on Shaky Ground: Parameter Sensitivity and Confirmation Bias in fsQCA. *Political Analysis* 23: 21-41.
- Laguna P. (2011) Mallas y flujos : acción colectiva, cambio social, quinua y desarrollo regional indígena en los Andes Bolivianos. Wageningen: Wageningen University.
- Landis JR and Koch GG. (1977) The measurement of observer agreement for categorical data. *biometrics*: 159-174.
- Lawson T. (2003) *Reorienting Economics*, London & New York: Routledge.
- Lawson T. (2009) Applied economics, contrast explanation and asymmetric information. *Cambridge Journal of Economics* 33: 405-419.
- Leana CR and Van Buren HJ. (1999) Organizational Social Capital and Employment Practices. *The Academy of Management review* 24: 538-555.
- Leeuw F and Vaessen J. (2009) Impact evaluations and development: NONIE guidance on impact evaluation. Network of networks for Impact Evaluation.
- Legewie N. (2013) An Introduction to Applied Data Analysis with Qualitative Comparative Analysis. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*.
- Lensink R. (2014) What Can We Learn from Impact Evaluations. *European Journal of Development Research* 26: 12-17.
- Liebersohn S. (2004) Comments on the Use and Utility of QCA. *Qualitative Methods* 2: 13-14.
- Llorenti Barrientos R, Espejo Uscamayta R, Cabezas Ortiz E, et al. (2005) Aprendizajes del proyecto LIL/INDIGENA. La Paz: Ministerio de Asuntos Indígenas y Pueblos Originarios - Proyecto LIL/INDÍGENA - Banco Mundial.
- Longest KC and Vaisey S. (2008) fuzzy: A program for performing qualitative comparative analyses (QCA) in Stata. *Stata Journal* 8: 79-104.
- Loza G. (2002) *Políticas y Estrategias de Comercio Exterior*, La Paz: MAGDR.
- Lucas SR and Szatrowski A. (2014) Qualitative comparative analysis in critical perspective. *Sociological Methodology* 44: 1-79.
- Lundy M, Ostertag C and Best R. (2002) Value adding, agroenterprise and poverty reduction: A territorial approach for rural business development. *Globalization of Agricultural Research conference*: 25-27.
- Lyon F. (2000) Trust, Networks and Norms: The Creation of Social Capital in Agricultural Economies in Ghana. *World Development* 28: 663-681.
- Lyon F. (2003) Community groups and livelihoods in remote rural areas of Ghana: How small-scale farmers sustain collective action. *Community Development Journal* 38: 323-331.
- Mackie JL. (1965) Causes and conditions. *American philosophical quarterly*: 245-264.
- Mahoney J. (2008) Toward a Unified Theory of Causality. *Comparative Political Studies* 41: 412-436.
- Maredia MK. (2009) *Improving the proof: Evolution of and emerging trends in impact assessment methods and approaches in agricultural development*: Intl Food Policy Res Inst.
- Markelova H, Meinzen-Dick R, Hellin J, et al. (2009) Collective action for smallholder market access. *Food Policy* 34: 1-7.

- Marx A. (2006) Towards more robust model specification in QCA: Results from a methodological experiment. *American Sociological Association, Philadelphia, PA*.
- Marx A. (2010) Crisp-set qualitative comparative analysis (csQCA) and model specification: Benchmarks for future csQCA applications. *International Journal of Multiple Research Approaches* 4: 138-158.
- Mason NM and Ricker-Gilbert J. (2012) Disrupting demand for commercial seed: Input subsidies in Malawi and Zambia. *World Development* 45: 75-91.
- Mayne J. (2001) Addressing attribution through contribution analysis: using performance measures sensibly. *Canadian Journal of Program Evaluation* 16: 1-24.
- Mayne J. (2011) Contribution analysis: Addressing cause and effect. *Evaluating the complex: Attribution, contribution, and beyond* 18: 53-96.
- Mayne J. (2012) Contribution analysis: Coming of age? *Evaluation* 18: 270-280.
- McCluskey EJ. (1956) Minimization of Boolean Functions\*. *Bell system technical Journal* 35: 1417-1444.
- Meinzen-Dick R, DiGregorio M and McCarthy N. (2004) Methods for studying collective action in rural development. *Agricultural Systems* 82: 197-214.
- Méndez VE, Bacon CM, Olson M, et al. (2010) Effects of Fair Trade and organic certifications on small-scale coffee farmer households in Central America and Mexico. *Renewable Agriculture and Food Systems* 25: 236-251.
- Mendoza M and Ton G. (2003) *Manual de Procedimientos Jurídicos y Tributarios para Organizaciones Económicas Campesinas*, Sucre: CIOEC-SNV.
- Mesa Gisbert CD, de Mesa Figueroa J and Gisbert T. (2012) *Historia de Bolivia - Octava edición*, La Paz: Editorial Gisbert.
- Meyers LS, Gamst G and Guarino AJ. (2006) *Applied multivariate research: Design and interpretation*: Sage.
- Mill JSS. (2009 (1843)) *System of Logic Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*: Cosimo Classics.
- Minh TT, Friederichsen R, Neef A, et al. (2014) Niche action and system harmonization for institutional change: Prospects for demand-driven agricultural extension in Vietnam. *Journal of Rural Studies* 36: 273-284.
- Minnameier G. (2010) The logicity of abduction, deduction, and induction. *Ideas in action: Proceedings of the Applying Peirce Conference*. Nordic Pragmatism Network Helsinki, 239-251.
- Minten B, Dereje M, Engeda E, et al. (2015) Who benefits from the rapidly increasing Voluntary Sustainability Standards? Evidence from Fairtrade and Organic certified coffee in Ethiopia. Working Paper 71. Addis Ababa: International Food Policy Research Institute (IFPRI).
- Moore SF. (1973) Law and social change: the semi-autonomous social field as an appropriate subject of study. *Law and Society Review*: 719-746.
- Nathaniels NQR. (2005) Cowpea, farmer field schools and farmer-to-farmer extension: a Benin case study. *Agricultural Research and Extension Network* 148.
- Neef A and Neubert D. (2011) Stakeholder participation in agricultural research projects: a conceptual framework for reflection and decision-making. *Agriculture and Human Values* 28: 179-194.
- Nelson V and Martin A. (2012) The impact of Fairtrade: Evidence, shaping factors, and future pathways. *Food Chain* 2: 42-63.

- North DC. (1990) *Institutions, Institutional Change and Economic Performance*.
- OECD. (2001) The DAC Guidelines:: Poverty Reduction. Paris: OECD.
- OECD. (2008) *Evaluating Development Co-operation: summary of key norms and standards*, Paris: OECD.
- Oldekop JA, Chappell MJ, Peixoto FEB, et al. (2015) Linking Brazil's food security policies to agricultural change. *Food Security*: 1-15.
- Olken BA. (2015) Promises and Perils of Pre-analysis Plans. *Journal of Economic Perspectives* 29: 61-80.
- Olson M. (1965) *The logic of collective action: Public goods and the theory of groups*, Cambridge (Massachusetts): Harvard University Press.
- Opondo C, German L, Stroud A, et al. (2006) Lessons from Using Participatory Action Research to Enhance Farmer-Led Research and Extension in Southwestern Uganda. Uganda: African Highlands Initiative - AHL, 14.
- Ortiz-Marcos I, Naranjo MS and Cobo JR. (2011) A diagnostic tool design for cooperatives and Fair Trade organisations. *Total Quality Management & Business Excellence* 22: 743-753.
- Ostrom E. (1990) Governing the commons: the evolution of institutions for collective action. *Governing the commons: the evolution of institutions for collective action*.
- Ostrom E and Ahn T. (2008) Social Capital and Collective Action. *The Handbook of Social Capital*: 70.
- Ostrom E and Ahn T. (2009) The meaning of social capital and its link to collective action. *Handbook of Research on Social Capital: The Troika of Sociology, Political Science and Economics*. Cheltenham: Edward Elgar Publishing Limited, 17.
- Ostrom E and Bloomington I. (2009) Building Trust to Solve Commons Dilemmas: Taking Small Steps to Test an Evolving Theory of Collective Action. *Games, Groups, and the Global Good*: 207.
- Pamuk H, Bulte E and Adekunle AA. (2014) Do decentralized innovation systems promote agricultural technology adoption- Experimental evidence from Africa. *Food Policy* 44: 227-236.
- Pan L and Christiaensen L. (2012) Who is vouching for the input voucher? decentralized targeting and elite capture in Tanzania. *World Development* 40: 1619-1633.
- Pardo J. (2010) Interview on FONDOECAS pilot experience in CIOEC-Cochabamba. In: Ton G (19-08-2010).
- Patton MQ. (2002) Utilization-focused evaluation. *Evaluation models*: 425-438.
- Patton MQ. (2008) *Utilization-Focused Evaluation. Fourth Edition*: SAGE.
- Paumgarten F, Kassa H, Zida M, et al. (2012) Benefits, Challenges, and Enabling Conditions of Collective Action to Promote Sustainable Production and Marketing of Products from Africa's Dry Forests. *Review of Policy Research* 29: 229-250.
- Pawson R. (1996) Theorizing the interview. *British Journal of Sociology*: 295-314.
- Pawson R. (2002) Evidence-based Policy: The Promise of 'Realist Synthesis'. *Evaluation* 8: 340-358.
- Pawson R. (2003) Nothing as practical as a good theory. *Evaluation* 9: 471-490.
- Pawson R. (2013) *The Science of Evaluation: A Realist Manifesto*: SAGE.
- Pawson R and Tilley N. (1997) *Realistic evaluation*: Sage Publications Inc.
- Pawson R and Tilley N. (2006) Realist Evaluation. *Realistic Explanation: linking explanatory and interpretive accounts in applied social research*. Utrecht & Wageningen: CERES.

- Peirce CS, Hartshorne C and Weiss P. (1935) *Collected papers of Charles Sanders Peirce*: Harvard University Press.
- Penrose-Buckley C. (2007) *Producer Organisations: a guide to developing collective rural enterprises*, London: OXFAM.
- Peppelenbos L. (2008) *Trading Up: building cooperation between farmers and traders in Africa*. Amsterdam: KIT/IIRR.
- Pérez Luna M. (2003) *Apertura comercial y sector agrícola campesino: la otra cara de la pobreza del campesino andino*: CEDLA.
- Perrett H. (2004) *Community Development Funds: emerging lessons for a project design - main report*. Rome: IFAD.
- Perri G. (2006) *Combining Styles of Research in One Research Project. Lecture presented at the University of Wageningen*. Nottingham-Wageningen.
- Poteete AR and Ostrom E. (2004) In pursuit of comparable concepts and data about collective action. *Agricultural Systems* 82: 215-232.
- Poulton C, Dorward A and Kydd J. (2010) The Future of Small Farms: New Directions for Services, Institutions, and Intermediation. *World Development* 38: 1413-1428.
- Pritchett L and Sandefur J. (2015) Learning from Experiments When Context Matters. *American Economic Review* 105: 471-475.
- PROFIN. (2009) *Sistema de Autoevaluación, Calificación y Acreditación de Entidades Asociativas Económicas Rurales: Documento de trabajo*. La Paz: APSA II - PROFIN.
- Prudencio J. (2010) *Evaluación Externa del FONDOECAS*. La Paz: CLAVE Consultores SRL.
- Prudencio J and Elías B. (2014) *Compras Públicas: ¿Una oportunidad para los pequeños productores? Publicaciones Regionales del Proyecto Mercados Campesinos*. La Paz: Afónomos y Veterinarios Sin Fronteras (AVSF).
- Prudencio J and Ton G. (2004) *Integración Regional y Producción Campesina: la urgencia de políticas de soberanía alimentaria*, La Paz: CIOEC-Bolivia.
- Ragin C. (1987) *The comparative method: Moving beyond qualitative and quantitative strategies*: Univ of California Pr.
- Ragin C. (2000) *Fuzzy-set social science*: University of Chicago Press.
- Ragin C. (2008) *Redesigning social inquiry: Fuzzy sets and beyond*: University of Chicago Press.
- Ragin C and Davey S. (2009) *fs/QCA, Version 2.5 (Computer program)*. Tucson: University of Arizona.
- Ragin CC. (2014) Comment Lucas and Szatrowski in Critical Perspective. *Sociological Methodology*: 0081175014542081.
- Rainey C. (2014) *Dealing with Separation in Logistic Regression Models. Working paper*. Buffalo NY: University of Buffalo.
- Ravallion M. (2009) Should the Randomistas Rule? *The Economists' Voice* 6: 6.
- Reardon T and Berdegue JA. (2002) The Rapid Rise of Supermarkets in Latin America: Challenges and Opportunities for Development. *Development Policy Review* 20: 371-388.
- Reichert C and Rubinson C. (2014) *Kirq, Version 2.1. 12 [Computer program]*. Houston, TX: University of Houston-Downtown.
- Reichert J. (2004) *Abduction, deduction and induction in qualitative research*. In: Flick U, Kardorff Ev and Steinke I (eds) *A Companion to Qualitative Research*. London: SAGE Publications Ltd, 159.

- Richards P. (2007) How does participation work? Deliberation and performance in African food security. In: Chataway J, Maat H and Waldman L (eds) *IDS Bulletin*. Brighton; UK: Institute of Development Studies, University of Sussex, 21-35.
- Ricker-Gilbert J and Jayne T. (2009) Do Fertilizer Subsidies Affect the Demand for Commercial Fertilizer? An Example from Malawi. *IAAE 2009*. Beijing.
- Ricker-Gilbert J, Jayne T and Shively G. (2013) Addressing the “Wicked Problem” of Input Subsidy Programs in Africa. *Applied Economic Perspectives and Policy* 35: 322-340.
- Rigby D, Woodhouse P, Young T, et al. (2001) Constructing a farm level indicator of sustainable agricultural practice. *Ecological Economics* 39: 463-478.
- Rihoux B and Lobe B. (2009) The case for qualitative comparative analysis (QCA): Adding leverage for thick cross-case comparison. *The SAGE Handbook of Case-Based Methods*: 222-242.
- Rihoux B and Marx A. (2013) QCA, 25 Years after “The Comparative Method”: Mapping, Challenges, and Innovations—Mini-Symposium. *Political Research Quarterly* 66: 167-235.
- Rihoux B and Ragin C. (2009) Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and related techniques. Thousand Oaks and London: Sage.
- Rivera WM. (2000) The changing Nature of Agricultural Information And The Conflictive Global Developments Shaping Extension. *The Journal of Agricultural Education and Extension* 7: 31-41.
- Roche CJR and Roche C. (1999) *Impact assessment for development agencies: Learning to value change*. London: Oxfam GB - NOVIB.
- Rodríguez-Carmona A. (2009) *El Proyección: Bolivia tras 20 años de ayuda externa*, La Paz: Plural Editores.
- Rogers PJ. (2008) Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 14: 29-48.
- Rogers PJ. (2009) Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation. *Journal of Development Effectiveness* 1: 217 - 226.
- Rohlfinger I. (2012) *Case Studies and Causal Inference: an integrative framework*. Palgrave Macmillan.
- Royer J. (1999) Cooperative organizational strategies: a neo-institutional digest. *Journal of Cooperatives* 14: 44-67.
- RSCE. (2009) Round Table on a Sustainable World Cocoa Economy - RSCE2/7 Draft Principles for a Sustainable Cocoa Economy. Accessed April 24, 2014 at: <http://tinyurl.com/lfyjls9>.
- Ruben R, Fort R and Zuniga-Arias G. (2009) Measuring the impact of fair trade on development. *Development in Practice* 19: 777-788.
- Ruben R and Heras J. (2012) Social capital, governance and performance of Ethiopian coffee cooperatives. *Annals of Public and Cooperative Economics* 83: 463-484.
- Russillo A and Pintér L. (2009) Linking Farm-Level Measurement Systems to Environmental Sustainability Outcomes: Challenges and ways forward. Winnipeg: International Institute for Sustainable Development (IISD).
- Rustagi D, Engel S and Kosfeld M. (2010) Conditional cooperation and costly monitoring explain success in forest commons management. *Science* 330: 961-965.
- Sachs JD and McArthur JW. (2005) The Millennium Project: a plan for meeting the Millennium Development Goals. *The Lancet* 365: 347-353.

- Samson S. (2010) Performance and Dynamics of Uruguayan Cooperatives. *International Development Studies*. Wageningen: Wageningen University, 139.
- Sandoval V, Kaaria S, Lilja N, et al. (2009) Impactos en términos del capital humano y social de los métodos de investigación participativa en agricultura: El caso de los Comités de Investigación Agrícola Local–CIAL, en el Cauca, Colombia.
- Savedoff WD, Levine R and Birdsall N. (2005) When Will We Ever Learn? Improving lives through impact evaluation. . *Report of the Evaluation Gap Working Group*. Washington DC: Center for Global Development.
- Sayer RA. (2000) *Realism and Social Science*: Sage Publications.
- Schneider CQ and Wagemann C. (2012) *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge, Cambridge University Press.
- Scoones I. (1998) Sustainable rural livelihoods: A framework for analysis. *Sustainable Rural Livelihoods: A Framework for Analysis*.
- SCOPEInsight. (2013) SCOPEc Report Scoring of Organizational Performance:: Anonymous Farmers Cooperative Union. SCOPEcInsight, 33.
- SCSSC. (2012) *Toward sustainability: The roles and limitations of certification*. Washington: Steering Committee of the State-of-Knowledge Assessment of Standards and Certification (SCSSC) - RESOLVE Inc.
- Serra R. (2011) The Promises of a New Social Capital Agenda. *The Journal of Development Studies* 47: 1109-1127.
- Shadish WR, Cook TD and Campbell DT. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*: Houghton Mifflin Co. Boston, MA.
- Shapiro S. (2005) Agency Theory. *Annual Review of Sociology* 31: 263-284.
- Shiferaw B, Hellin J and Muricho G. (2011) Improving market access and agricultural productivity growth in Africa: What role for producer organizations and collective action institutions? *Food Security* 3: 475-489.
- Shiferaw B, Obare G and Muricho G. (2006) Rural Institutions and Producer Organizations in Imperfect Markets: Experiences from Producer Marketing Groups in Semi-Arid Eastern Kenya. *Collective Action and Property Rights (CAPRI) Working Paper*.
- Shroff R, Martin J and Jersild A. (2012) Accelerating Innovation for Development: case studies: ETC Prolinnova. *Accelerating Innovation for Development: case studies*. Rockefeller Foundation, 115-136.
- Smith A, Fressoli M and Thomas H. (2014) Grassroots innovation movements: challenges and contributions. *Journal of Cleaner Production* 63: 114-124.
- Smithson M and Verkuilen J. (2006) *Fuzzy set theory: applications in the social sciences*: Sage.
- Snilstveit B. (2012) Systematic reviews: from 'bare bones' reviews to policy relevance. *Journal of Development Effectiveness* 4: 388-408.
- Snodgrass D. (2006) Assessing the Impact of New Generation Private Sector Development Programs. *Impact Assessment Primer Series #1*. Washington: USAID.
- Snowden D and Boone M. (2007) A leader's framework for decision making. *harvard business review* 85: 68.
- Soboh R, Lansink A, Giesen G, et al. (2009) Performance Measurement of the Agricultural Marketing Cooperatives: the gap between theory and practice. *Review of Agricultural Economics* 31: 446-469.
- Sotomayor C, Palma L, Díaz M, et al. (2008) Proyecto de Desarrollo Rural Corredor Puno - Cuzco: informe final. Peru.

- Staatz J. (1983) The cooperative as a coalition: a game-theoretic approach. *American Journal of Agricultural Economics*: 1084-1089.
- Stame N. (2004) Theory-Based Evaluation and Types of Complexity. *Evaluation* 10: 58-76.
- StataCorp. (2013) Stata Statistical Software: Release 13. . College Station, TX: StatCorp LP.
- Stern E, Stame N, Mayne J, et al. (2012) Broadening the Range of Designs and Methods for Impact Evaluations: report of a study commissioned by the Department for International Development. *Working Paper 38*. London: DFID.
- Swen H and Both F. (1999) *Tractores y auditores: una escuela para la cooperación técnica y las organizaciones económicas campesinas*: SNV.
- Sykuta ME and Cook ML. (2001) A New Institutional Economics Approach to Contracts and Cooperatives. *American Journal of Agricultural Economics* 83: 1273-1279.
- Tanburn J. (2008) The 2008 Reader on Private Sector Development: measuring and reporting results. Turin: International Training Centre - ILO.
- Thiem A. (2013) Membership function sensitivity of descriptive statistics in fuzzy-set relations. *International Journal of Social Research Methodology* 17: 625-642.
- Thiem A, Baumgartner M and Bol D. (2015) Still Lost in Translation! A Correction of Three Misunderstandings Between Configurational Comparativists and Regression Analysts. *Comparative Political Studies*, Prepublished March 16, 2015.
- Thomas J, Brunton J and Graziosi S. (2010) EPPI-Reviewer 4.0: software for research synthesis. *EPPI-Centre Software*. London: Social Science Research Unit, Institute of Education.
- Thomas J and Harden A. (2008) Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology* 8: 45.
- Thomas J, Harden A and Newman M. (2012) Synthesis: combining results systematically and appropriately. In: D. Gough D, Oliver S and Thomas J (eds) *An Introduction to Systematic Reviews*. London: Sage, 179–226.
- Ton G. (2005) Fondo para el Fortalecimiento Económico de las Organizaciones Económicas Campesinas (FONDOECA). La Paz: CIOEC-ICCO, 37.
- Ton G. (2007a) Farmers' Organisations in Agricultural Research and Development: governance issues in two competitive funding programs in Bolivia. *Producer Organisations and Market Chains: facilitating trajectories of change in developing countries*. Wageningen Wageningen Academic Publishers, 271-283.
- Ton G. (2007b) Farmers' organisations in agricultural research and development: governance issues in two competitive funding programs in Bolivia. In: Ton G, Bijman J and Oorthuizen J (eds) *Producer Organisations and Market Chains: facilitating trajectories of change in developing countries*. Wageningen: Wageningen Academic Press, 271-283.
- Ton G. (2008) Challenges for smallholder market access: A review of literature on institutional arrangements in collective marketing. *Stewart Postharvest Review* 4.
- Ton G. (2010a) Propuesta para el Proyecto de Monitoreo y Evaluación de Impactos del FONDOECAS. In: UR F-I-LW (ed). La Paz.
- Ton G. (2010b) Resolving the Challenges of Collective Marketing: incentive structures that reduce the tensions between members and their group. *ESFIM Policy Brief #4*. Wageningen: ESFIM.
- Ton G. (2012a) Crowd-sourcing Organisational Intelligence: capturing the rich experiences of farmers' organisations. *Farming Matters* 28: 20-21.
- Ton G. (2012b) The Mixing of Methods: a three-step process for improving rigour in impact evaluations. *Evaluation* 18: 5-25.

- Ton G. (2015) Using Qualitative Comparative Analysis to Explore Outcome Patterns of Grant Support to Farmer Organisations in Bolivia. *COMPASSS Working Paper 2015-82*.
- Ton G and Bijman J. (2006) The Role of Producer Organizations in the Process of developing an Integrated supply Chain: experiences from quinoa chain development in Bolivia. In: Bijman J, Omta SWF, Trienekens JH, et al. (eds) *International Agri-Food Chains and Networks: management and organization*. Wageningen: Wageningen Academic Publishers, 97-111.
- Ton G, Bijman J and Oorthuizen J. (2007) *Producer Organisations and Market Chains : facilitating trajectories of change in developing countries*, Wageningen: Wageningen Academic Publishers.
- Ton G, Flores L, Monasterios R, et al. (2014a) Capabilities and Performance in Collective Marketing: the importance of learning to cope with agency dilemmas. In: Christy R, da Silva C, Mhlanga N, et al. (eds) *Innovative Institutions, Public Policies and Private Strategies for Agro-Enterprise Development*. FAO / World Scientific Publishing Co. Inc., 113-149.
- Ton G, Gouet C and Gonzalez N. (2013a) CIOEC-Bolivia: successful advocacy for legal recognition of the sector. In: Ton G and Proctor F (eds) *Empowering Smallholders Farmers in Markets: experiences with farmer-led research for advocacy*. Wageningen: AGRINATURA-CTA.
- Ton G, Grip K, Klerkx L, et al. (2013b) Effectiveness of Innovation grants to smallholder agricultural producers: an explorative systematic review. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Ton G, Grip K, Lançon F, et al. (2014b) Empowering Smallholder Farmers in Markets: strengthening the advocacy capacities of national farmer organisations through collaborative research. *Food Security* 6: 261-273.
- Ton G, Klerkx L, de Grip K, et al. (2015) Innovation grants to smallholder farmers: Revisiting the key assumptions in the impact pathways. *Food Policy* 51: 9-23.
- Ton G, Opeero DM and Vellema S. (2010) 'How Do We Get it to the Mill?' A Study on Bulking Arrangements that Enable Sourcing from Smallholders in the Ugandan Vegetable Oil Chain. *VC4PD Research Paper #7* the Hague: LEI Wageningen UR.
- Ton G, Rau M-L, de Grip K, et al. (2011a) Effectiveness of Innovation Grants on Smallholder Agricultural Producers: protocol for an explorative systematic review. In: EPPI-Centre (ed). London: Social Science Research Unit - Institute of Education - University of London.
- Ton G, Taylor S, Mahapatra S, et al. (2012) Poverty Impacts of Irrigation Technology Supply and Horticulture Development: Income Impact of Treadle Pumps and Drip-Kits in Successive Client Cohorts in Nepal, Zambia and Ethiopia. Available at SSRN: <http://ssrn.com/abstract=2264157>.
- Ton G, Vellema S and de Ruyter de Wildt M. (2011b) Development impacts of value chain interventions: how to collect credible evidence and draw valid conclusions in impact evaluations? *Journal on Chain and Network Studies* 11: 69-84.
- Ton G, Vellema S and Ge L. (2014c) The Triviality of Measuring Ultimate Outcomes: Acknowledging the Span of Direct Influence. *IDS Bulletin* 45: 37-48.
- Toornstra C. (2000) Buscando una Identidad Empresarial Asociativa. Sucre: SNV, 20.
- Toro G and Espinosa N. (2003) Los Fondos Competitivos para la Agricultura y el Desarrollo Rural: Fundamentos, Aplicaciones y Lecciones Aprendidas. San José: Costa Rica: IICA.



- Triomphe B, Wongtschowski M, Krone A, et al. (2012) Module 5 - IAP4: Providing Farmers with Direct Access to Innovation Funds. *Agricultural Innovation Systems: an investment sourcebook*. 435-444.
- Uphoff N. (2000) Understanding social capital: learning from the analysis and experience of participation. In: Dasgupta P and Serageldin I (eds) *Social capital: A multifaceted perspective*. Washington: World Bank, 215-249.
- USAID. (2009) Glossary of Evaluation Terms. Washington: Planning and Performance Management Unit, Office of the Director of U.S. Foreign Assistance.
- UTZ Certified. (2014) UTZ Certified Impact Report January 2014: Combining results from 24 external impact studies and data from UTZ Certified. Amsterdam: UTZ Certified.
- Vaessen J. (2010) *Methodological and Conceptual Challenges in Evaluating the Impact of Development Interventions*, Maastricht: Datawyse / Universitaire Pers Maastricht.
- Vaessen J. (2011) Challenges in impact evaluation of development interventions: randomized experiments and complexity. In: Forss K, Marra M and Schwartz R (eds) *Evaluating the complex: Attribution, contribution, and beyond*. 283-313.
- Valentinov VL. (2004) Toward a social capital theory of cooperative organisation. *Journal of Cooperative Studies* 37: 5-20.
- van der Meer K and Noordam M. (2004) The Use of Grants to Address Market Failures - A Review of World Bank Rural Development Projects. Agriculture and Rural Development Discussion Paper.
- van Veldhuizen L, Wongtschowski M and Waters-Bayer A. (2005) Farmer Access to Innovation Resources (FAIR): findings from an international review of experiences. *ProInnova*, 15.
- Vellema S. (2010) Measuring impacts: About lean approaches, mechanisms, and critical control points. *ISEAL Alliance Conference*. London.
- Vellema S. (2011) *Transformation and Sustainability in Agriculture: connecting practice with social theory*, Wageningen: Wageningen Academic Publishers.
- Vellema S and Ton G. (2012) Is measuring ultimate impacts the way forward? *ISEAL Conference "Who's Who and What's On in research about the poverty reduction impacts of sustainability standards"*. London, 27-28 November 2012: ISEAL/NRI.
- Vellema S, Ton G, de Roo N, et al. (2013) Value chains, partnerships and development: Using case studies to refine programme theories. *Evaluation* 19: 304-320.
- Vellema S and van Wijk J. (2014) Partnerships intervening in global food chains: the emergence of co-creation in standard-setting and certification. *Journal of Cleaner Production*.
- Verhofstadt E and Maertens M. (2014) Smallholder cooperatives and agricultural performance in Rwanda: Do organizational differences matter?
- Vis B. (2012) The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N Analyses. *Sociological Methods & Research* 41: 168-198.
- Von Pischke J and Rouse J. (2004) New strategies for mobilizing capital in agricultural cooperatives. *FAO: Rome*.
- Vorley B, Fearn A and Ray D. (2007) *Regoverning Markets: a place for small-scale producers in modern agrifood chains?*, London: IIED - Gower Publishing Ltd.
- Waarts Y, Ge L and Ton G. (2013a) From Training to Practice: Mid-term evaluation of the UTZ-Solidaridad smallholder tea programme in Kenya. The Hague: LEI Wageningen UR, 171.

- Waarts Y, Ge L, Ton G, et al. (2012) Sustainable tea production in Kenya: Impact assessment of Rainforest Alliance and Farmer Field School training. *LEI Wageningen UR, The Hague*.
- Waarts Y, Ge L, Ton G, et al. (2013b) A Touch of Cocoa: Baseline study of six UTZ-Solidaridad cocoa projects in Ghana. The Hague: LEI Wageningen UR, 183.
- Weiss CH. (1997) How Can Theory-Based Evaluation Make Greater Headway? *Evaluation Review* 21: 501-524.
- Westhorp G. (2012) Using complexity-consistent theory for evaluating complex systems. *Evaluation* 18: 405-420.
- Westhorp G. (2013) Developing complexity-consistent theory in a realist investigation. *Evaluation* 19: 364-382.
- White H. (2009) Theory-based impact evaluation: principles and practice. *Journal of Development Effectiveness* 1: 271 - 284.
- White H. (2011) Achieving high-quality impact evaluation design through mixed methods: the case of infrastructure. *Journal of Development Effectiveness* 3.
- Williamson OE. (2000) The New Institutional Economics: Taking Stock, Looking Ahead. *Journal of Economic Literature* 38: 595-613.
- Williamson OE. (2002) The Lens of Contract: private ordering. *The American Economic Review* 92: 438-443.
- Wollni M and Fischer E. (2015) Member deliveries in collective marketing relationships: evidence from coffee cooperatives in Costa Rica. *European Review of Agricultural Economics* 42: 287-314.
- Wollni M and Zeller M. (2007) Do farmers benefit from participating in specialty markets and cooperatives? The case of coffee marketing in Costa Rica. *Agricultural Economics* 37: 243-248.
- Wong G, Greenhalgh T and Pawson R. (2010) Internet-based medical education: a realist review of what works, for whom and in what circumstances. *BMC Medical Education* 10: 12.
- Wong G, Greenhalgh T, Westhorp G, et al. (2013) RAMESES publication standards: realist syntheses. *BMC medicine* 11: 21.
- Wongtschowski M, Triomphe B, Krone A, et al. (2010) Towards a farmer-governed approach to agricultural research for development: lessons from international experiences with local innovation support funds. 11.
- Woolcock M. (1998) Social capital and economic development: Toward a theoretical synthesis and policy framework. *Theory and Society* 27: 151-208.
- Woolcock M. (2013) Using case studies to explore the external validity of 'complex' development interventions. *Evaluation* 19: 229-248.
- World Bank. (2005) *Bolivia - Country Economic Memorandum: policies to improve growth and employment*, Washington: World Bank.
- World Bank. (2006) *Enhancing agricultural innovation: how to go beyond the strengthening of research systems*. Washington: World Bank.
- World Bank. (2007) *World Development Report 2008: Agriculture for Development*, Washington: World Bank.
- World Bank. (2010) *Designing and implementing agricultural innovation funds: lessons from competitive research and matching grants*. The International Bank for Reconstruction and Development/The World Bank, 94.

- World Bank. (2012) *Agricultural Innovation Systems: An Investment Sourcebook*: World Bank.
- WVS. (2009) World values survey 1981-2008 official aggregate v. 20090901. Madrid: World Value Survey Association.
- Xu Z, Burke WJ, Jayne TS, et al. (2009) Do input subsidy programs “crowd in” or “crowd out” commercial market development? Modeling fertilizer demand in a two-channel marketing system. *Agricultural Economics* 40: 79-94.
- Yang H, Klerkx L and Leeuwis C. (2014) Functions and limitations of farmer cooperatives as innovation intermediaries: Findings from China. *Agricultural Systems* 127: 115-125.
- Yin RK. (2013) Validity and generalization in future case study evaluations. *Evaluation* 19: 321-332.
- Zandniapur L, Sebstad J and Snodgrass D. (2004) Review of Evaluations of Selected Enterprise development Projects. *MicroREPORT #3*. Washington: USAID.



# Annexes

<b>Annex 1. Case study summaries</b>	242
<b>Annex 2. Changes in the sample</b>	281
<b>Annex 3. Organisational radiography instrument</b>	282
<b>Annex 4. Time-series data sheet</b>	283
<b>Annex 5. Questionnaire included in the AR-LAT household survey</b>	284

## Annex 1 Case study summaries

### AAAT

#### Context

AAAT is a handicraft organisation that uses wool from its 150 members living in a remote area, to produce products for the high-end market. The organisation started in 1992 supported by the NGO *Centro de Capacitación Integral de la Mujer Campesina de Tarija* (CCIMCAT). In 2000 they became independent from the NGO but continued to receive support from development cooperation. The organisation owns a shop in the city centre of Tarija and is, therefore, a relatively well-capitalized artisanal organisation. AAAT manages 11 units of production with a decentralised governance system. Each unit of production has a committee that controls quality and decides whether the product is accepted for sale in the shop. Artisans get paid for products sold, every 5<sup>th</sup> day of the month.

#### Dynamics

After a surge in turn-over between 2008-2010, the group sales (alpaca weavings) declined recently. The interviews indicate that the main reason for this is the restriction of member supply. Many experienced (male) weavers migrated to Argentina to work as seasonal labourers.

Two of the eleven units of production stopped functioning in 2011 because they lost interest in weaving and pursued other income-generating activities (e.g., quinoa).

During the period 2010-2012, AAAT also participated in a European Union project to increase the quality of the weaving. The project finished in 2012, and therefore the two paid staff members had to be dismissed. All activities are now performed by board members.

#### Grant influence

AAAT used one grant in 2007 to equip the shop in Tarija, and a second grant in 2009 to invest in six production units. AAAT invested the 2009 grant in weaving equipment and industrial cookers for dyeing. The investment was meant to reduce the heterogeneity of the weaving material (sheep and lama wool), especially in the production of shawls.

The investments in the production units did not result in increased sales. The loss of EU-support induced the board to take up more responsibilities and created new internal rules and regulations, especially on quality assurance. The FONDOECAS grant contributed to this.

The interviewees in 2013 also indicated a long list of negative factors that affected the business plan, including the role of the NGO, competition of other shops, operating costs, and member commitment. They mention the quality of the products that resulted from the 2011 investments as a positive factor.

#### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: Yes
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

## ADAPICRUZ

### Context

The honey processor ADAPICRUZ (300 members), based in the city of Santa Cruz de la Sierra, has grown very fast in a (niche) market that is constrained. The organisation decided to separate its commercial activities into a separate legal identity, Apícola del Bosque S.A., in which both ADAPICRUZ and individual producers have shares. It is considered a front-runner in developing an adequate formal format for collective marketing that is able to mobilize investments from the membership and to attract credit from banks.

Apícola del Bosque sources honey from smallholders organised in local associations, and also from some larger, specialised individual beekeepers. They sell a fixed amount of honey, approximately US\$15,000 per month, to the governmental nutrition programme for lactating mothers. This generates approximately US\$2,000 net-income per month to pay the processing and organisational costs.

### Dynamics

ADAPICRUZ has grown steadily both in member numbers and turnover in the last few years. Due to the strong growth in production capacity, ADAPICRUZ encountered a limit in their ability to sell all honey offered by the farmers. As a result of the growing mismatch between supply and demand, the organisation had to introduce a more stringent system of supply management that, however, generated some frustration in members who had to sell part of the honey on the local market at lower prices. Although access to credit is good for ADAPICRUZ/Apicultores del Bosque, they increasingly face constraints in working capital, aggravated by an increasing passive stock of unsold honey.

Alongside the public procurement, new lucrative complementary niche markets are currently being explored but have only been piloted with piecemeal deliveries (organic honey exports). ADAPICRUZ increasingly emphasis its role as a representative body of beekeepers, leaving the production to the local associations and the processing to the company.

### Grant influence

In 2009, ADAPICRUZ invested in a packaging machine (*sachetadora*) for individual portions, demanded by the school meal programme in La Paz. In 2011 they received a second FONDOECAS grant for additional machinery to improve the quality of the processed products and to diversify their range of products, especially for making honey-sesame bars.

ADAPICRUZ increased their sales volume steadily. The first grant made it possible to access the market of school food in La Paz. The second grant is considered by all interviewees to have played an essential role in upgrading their processing capacity to access new markets. We do not have the information to specify how much of the additional sales are due to these new markets, but the positive effect of the grant on market access is clear. The 2013 interviewees only mention positive factors that helped the realisation of the business plan: government procurement and prices paid, price and quality of supply, equipment and role of the board.

### Outcome summary

- Increased access to markets for members: Yes
- Increased organisational capacities: Yes
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

## AGAYAP

### Context

AGAYAP, located in the Eastern Lowlands, was founded in 2001. AGAYAP owns and manages a slaughterhouse. When applying for the FONDOECAS grant, it had a membership of 202, largely dairy farmers that deliver milk to the dairy company PIL Andina in Santa Cruz. The members may use the slaughterhouse facilities and sell the meat themselves, paying AGAYAP a service fee, or they delegate the sales to AGAYAP and receive payment per kilo after AGAYAP has sold the meat.

The farm and slaughterhouse represents a huge patrimony. The property rights of this location (500 ha) have been disputed, however- a legal dispute that resulted partly from alleged participation of the former AGAYAP president in party politics, opposed to the current government.

### Dynamics

In 2010, the slaughterhouse was shut down, due to a failing supply of cattle. Many members transported their cattle to cities where the meat price was higher (e.g., Cochabamba). Due to this crisis, the economic activities of AGAYAP have been paralysed. According to the 2011 interview, the lack of income to pay for the expenses of the board members created a downward spiral, with the board remaining inactive while action was needed.

The association has a majority of members that in the past received a loan to establish dairy production, and who deliver their milk individually to a dairy company. Many members have defaulted on their loans (young heifers, to be paid back in kind), which limits access of AGAYAP to working capital and has created additional governance problems within the group.

Thanks to a bank loan, AGAYAP has maintained a small-scale dairy production in the area, using a dairy plant in the same locality. The dairy production is principally intended to secure the property title that is under dispute. It is an activity that is unrelated to the core functions of providing butchering services to their members.

### Grant influence

AGAYAP invested in a refrigerator van to take meat from the slaughterhouse to distant markets, to comply with the sanitary regulations. The business plan has not translated into any market access. Between 2010 and 2012, the grant-supported business plan was paralysed. They still had some sales turn-over related to dairy production by an external person on their property, and totally unrelated to the business plan supported by the FONDOECAS grant. In their proposal, AGAYAP projected to quadruple turn-over, from US\$70,525 (in 2006) to US\$324,000. We may well consider AGAYAP as an organisation on which the grant has had no positive effect on sales.

The break-down of AGAYAP seems unrelated to the grant investment, but a result of political and legal pressures that affect their property, and increased competition from slaughterhouses closer to the consumer market. In 2013, the interviewed board members identified the following negative factors influencing the business plan: the local government, the quantity of cattle supplied by members, the increased price competition with other slaughterhouses and the operational costs of the slaughterhouse.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## AMAGA

### Context

AMAGA started in 2000 as a small group of 60 dairy farmers that worked in projects to improve cattle farming, supported by funding from the Danish government agency DANIDA, which for many years had stimulated dairy production in the Highlands (PDLA – *Programa de Desarrollo Lechera del Altiplano*). In 2007, they formalised their organisation with 28 members and constructed a small dairy plant, with EU-funding through the Italian organisation *Progettomondo*. In 2010 they received a loan from the BDP to complete investments and for working capital. The municipality of Challapata supported them with the required co-funding to these projects.

### Dynamics

AMAGA sources its milk primarily from non-members that are located close to the plant for cost-efficiency reasons. Members still benefit from being member of AMAGA. AMAGA continued to be successful in obtaining support from governmental and non-governmental development programmes (e.g., DETI, APROSAR). However, their milk production is processed on-farm into artisanal cheese, and is not related to the processing activities.

The capacity of the plant is 1000 litres/day, but in 2011 they only used 10% of this capacity. One of the reasons for this was changes in the market conditions. The state enterprise LACTEOBOL started to procure raw milk at artificially high prices, which increased production costs, and several other small dairy plants had started to operate in the municipality that served the same market for quality yoghurt and pasteurised (non-artisanal) cheese.

AMAGA had made efforts to obtain a bank loan to increase working capital and increase the processing of milk but they did not succeed, as they had delayed the repayment of an earlier loan for investment in the plant, because they wanted to recover the investment made from an input provider that had not delivered a technical implement which they had bought with the loan.

In 2011 they got a contract to provide the school food programme in the municipality of Huanuni. However this (yearly) contract was discontinued by the municipality in 2012, and they hoped to be re-contracted again in 2013.

The supplying non-member farmers were willing to become members, but the existing membership decided on a membership fee of US\$1,000 for new members, which was considered too high in relation to the benefits of being a member.

### Grant influence

AMAGA invested in a packing machine for yellow cheese, to meet the sanitary regulations necessary to reach more remunerative markets. The effect of the grant on sales is difficult to assess. Packed cheese is only one of the products that they process in the plant. Even when we attribute the increased sales volume to the grant-supported business plan, the average yearly sales effect would be less than US\$3,000. This is very modest, even considering the small number of members.

AMAGA learned to face many new organisational challenges and agency dilemmas after they started processing milk. Many rules and regulations are under discussion. The grant supported this organisational development. Nevertheless it is clear that several issues still need to be resolved for the plant to survive.

The 2013 interviewees indicated the technical staff and the government procurement as a positive factors for the development of the business plan, and the role of board and members together with the increased competition for raw milk as the negative factors that constraint success.



The plant started up with grants from different sources, only part of the equipment was bought with the FONDOECAS grant. AMAGA had already good access to grants and loans when they received the grant, and they continue to have these support from other institutions. It is unlikely that the FONDOECAS grant had any influence on this.

### **Outcome summary**

- Increased access to markets for members: No
- Increased organisational capacities: Yes
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

## AMDESOY

### Context

AMDESOY started in 2005, and is a women's group that makes products from soya-meal, principally soya milk. They wanted to create value from the second-grade soybeans that the market did not accept. Initially, they were a group of 70 women. AMDESOY does not purchase from members but functions only as a processing unit. The member-workers sell the products directly to consumers in Santa Cruz with a system of door-to-door sales. Most members are not rural. For example, one male member has a lorry and two others are retired school teachers. Two female members emigrated to work in Spain.

They work daily in the soya processing, taking weekly turns with teams of three persons, only in the early morning. They can do the work aside from their other activities and work in the house. Each member sells approximately 50 litres of soy milk per day, with a margin estimated at US\$3,50/day/person. Sometimes they contract a worker to make the soy milk (100 US\$/month).

In 2010 they obtained a contract to provide to the school meal programme in Yapacaní. However, the contract was cancelled some days before the first deliveries, allegedly due to political pressure and corruption, however this was after that they had invested in the required inputs. The contract had been signed by an individual, and not as an organisation, which created internal organisational problems.

### Dynamics

The number of members further decreased from 20 in 2010 to only 13 in 2012. They started to make other food products based on soy, including bread, which proved to be a success and sales have been doubling each year between 2010 and 2012. The average income gained by each member-worker is estimated to be US\$120-150/month

### Grant influence

The president of AMDESOY had taken part in the municipal government when they received the grant. The proposal for the grant was written by an employee of the municipality. In 2009, AMDESOY invested in a processing infrastructure for food processing (based on soy), to comply with the sanitary requirements. They received a second grant from FONDOECAS, in 2011, to diversify their menu, including the production of bread, and invested in an industrial stove for the bakery.

Being a micro-enterprise that does not source from members, there was no increase in market access of member products. Nor did the grant have influence on their organisational capacities to manage collective activities.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

## AMLECO

### Context

AMLECO is a dairy association that manages a cooling tank, agricultural machinery and veterinary services, and mainly delivers milk to a large private dairy company, PIL Andina. This large and growing dairy association (353 members) has input provisioning to farmers and contract negotiation with the dairy plant PIL-Andina as their main functions. AMLECO has a steady turnover in the supply of feed and veterinary medicine for their members and manages the logistics around the cooling tanks. The collective processing and marketing of member produce is a new service of AMLECO, but this only uses a small part of the milk produced by the members.

### Dynamics

AMLECO started to develop its processing activities in 2012. It managed to equip the plant with funds from a development programme (Vida Campesina). This programme of the French NGO *Vétérinaires Sans Frontières* (AVSF), though co-implemented with CIOEC, initially functioned independently from FONDOECAS. In late 2012, however, they received a grant as well as a loan from FONDOECAS for working capital.

### Grant influence

AMLECO received a grant only in 2012 to invest in a cheese and yoghurt processing unit. The FONDOECAS grant had not had any impact as of 2013, as the investments made with it had not been completed by end 2012.

### Outcome summary

- Increased access to markets for members: NR
- Increased organisational capacities: NR
- Increased access to loans: NR
- Increased access to grants: NR
- Increased income to pay organisational expenses: NR

## AOCEMM

### Context

Founded in 1993, AOCEMM's collective marketing experience started with the production and marketing of quality wheat seed. They managed the supply of inputs (fertiliser) to their members.

AOCEMM received ample support from a diversity of European NGOs (MISEREOR, SNV, FOS, SOS FAIM, ACRA). In the late 1990s, AOCEMM concentrated its activities on honey processing and was actively involved in the setting up of the platform of honey associations ACPROABOL, which negotiated the contract with the national nutrition programme for lactating mothers in 2007.

AOCEMM had a professional staff of five persons that made decisions in close coordination with the supporting NGO. By 2010, AOCEMM had broken with the NGO (and coordinator) and the board became more important in coordinating the work. Three persons are employed in AOCEMM (marketing, half-time book-keeping, coordination); they are members of the organisation.

AOCEMM works in nine production units (villages), five of them specialised in honey. AOCEMM has some support activities on fruits and wheat with farmers in the units where honey production is not feasible. Membership fees are only collected from members that benefit from the honey business.

### Dynamics

AOCEMM has had a steady increase in associated processing units and group sales volume. More than half of the honey is procured from non-members at a 1% lower price. In 2010, 50% of the membership was actively involved in the production of honey. In 2013, the vast majority (90%) of members was involved in honey production, and AOCEMM is working to include as members eight local honey associations from which they buy. In 2013 they procured 70% of their honey from members. The four units of production where honey production was not feasible left the organisation.

They manage a micro-credit from FONCASOL for working capital. They tried to gain access to a BDP loan to increase trade capital to pay members cash at the moment of supply, but they had not managed to obtain it by June 2013. The loan was pending, to be discussed and approved in their general assembly of members.

They sell almost exclusively to the Programme for Lactating Mothers. However, they plan to diversify their market and coordinate with other economic farmer organisations to offer a diversity of products that could be provided to the local school meal programmes.

### Grant influence

FONDOECAS has been the only grant received for the honey processing unit. AOCEMM constructed the processing centre with the margin gained in the honey business. With it, in 2008, AOCEMM refurbished five buying centres for honey as well as purchasing 25 bee hives, ten centrifuges, a wax frame mould, and packaging material to transport honey to the central unit. They also invested in bar-scan technology, required to deliver to the big supermarkets in the five production centres.

The grant served to obtain certification of the Health department (registro sanitario), mandatory for selling to the government. The FONDOECAS grant invested in 2008 made this contract possible. The grant has had a direct effect on market access. The sales effect is estimated in US\$31,030/year, using the baseline volume of honey sales in 2007 of US\$4,150, reported in the grant proposal. The 2013 interviewees only indicate positive factors that influenced the business plan: government procurement at attractive price, quality of the equipment and the role of board and staff.

Though their tension containment capacities did remain stable between 2011 and 2013, the effects of the grant are considered to be positive.

### **Outcome summary**

- Increased access to markets for members: Yes
- Increased organisational capacities: Yes
- Increased access to loans: Yes
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

## APAM MIZQUE

### Context

Since 1988, APAM MIZQUE started to produce honey in the area, with help from the NGO CEDEAGRO. Later they obtained support from two other NGOs to improve apiculture and they received a revolving fund of US\$5,500 from the Swedish NGO SCC for trade capital. They sell the honey in the consumer market in Cochabamba. APAM has no paid staff. In 2010, APAM MIZQUE had 74 members located in 35 villages in the municipality of Mizque.

### Dynamics

The number of honey producers providing honey to APAM is stable (around 50), but the number of members has declined, especially individual producers that produce honey but sell individually.

Sales have increased sharply, principally because of the added value that they could create with better packaging and branding. APAM MIZQUE does not deliver to the nutrition programme for lactating mothers but developed their own local brand and package, gradually replacing sales of unprocessed honey. They see their role primarily as regional price-setters.

APAM introduced more rigid member obligations to purge the association of some members that directly competed with organisation as intermediaries. These intermediaries were paying cash, whereas the organisation had not enough funds to do so.

### Grant influence

In 2010 APAM MIZQUE invested the grant in a carpentry centre to produce beehives to expand its production and membership. The investment in machinery was made, but the business plan failed because of lack of skilled carpenters and low demand for hives from members. The 2013 interviews mention the role of board and members as the main factors that negatively influenced the grant-supported business plan.

Interestingly, the regional branch of CIOEC had voiced its concern when the business proposal was submitted, based on technical and ecological concerns (interview with the coordinator of CIOEC-Cochabamba, 19-08-2010). However, the anonymous process of evaluation in the committee did not permit them to influence the awarding process.

It is not sure if the business plan will ever be implemented, although the board indicated in 2013 that they considered the production capacity to be their major constraint and the production of low-cost hives an essential step to resolve this. We consider the sales effect of the grant to be zero.

The organisation increased its tension containment capacities but this has no relation to the FONDOECAS grant.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## APCA

### Context

APCA is an association of Andean camelid herders that covers the indigenous territory of Marka Antaquilla, the territory of eight *ayllus*, indigenous communities. APCA was established in 2000 as economic farmer organisation. APCA has a membership of 242 families. They sell alpaca wool that has been processed into high quality coloured fibres. The company that provided the spinning and colouring services was COPROCA, a company governed by the economic farmer organisation AIGACAA (*Asociación Integral de Ganaderos en Camélidos de los Andes Altos*), also a member of CIOEC. APCA manages a small shop in El Alto to sell the processed wool to consumers (*artisans*).

In the past, APCA provided other services to their members. For example, they worked with a fund from a Canadian NGO to influence the farm-gate price of wool in the area, offering to buy wool as a last-resort buyer at a minimum price. However, this project was discontinued.

The sales process is managed by a contracted coordinator. They adopted the governance system of the local indigenous form of organisations (*marka*), with the appointment of leadership (*malkus*) based on a two-year rotation between *ayllus*.

### Dynamics

APCA (242 members) ceased production at the end of 2012, due to competition from other buyers and the costs of subcontracting the processing to third parties. In 2010, COPROCA raised the price of its services. APCA therefore contracted a Peruvian company for spinning and dyeing the wool. However, in 2012, they ceased to do so, and thereafter only sold unprocessed alpaca wool. In 2013, they even stopped the procurement of wool from their members. In 2013 they started a project (supported by the NGO *Vétérinaires Sans Frontières*) to establish their own wool processing plant.

APCA's governance and membership overlaps completely with the indigenous organisation of the marka. While in 2011 this was considered a way to resolve many problems, in 2013 APCA considered it as a constraint for business development, considering their plan to set up their own processing plant. They started to reorganise and plan to have a reduced membership with more committed, and more alpaca-specialised members.

### Grant influence

APCA has invested in computerized fibre measurement equipment to better classify the wool quality in order to negotiate better prices with the companies that process the wool. They evaluate the grant investment as positive, and consider the acquired technology as a positive factor in the business plan. They indicated two factors that negatively influenced the performance of the business plan: the operating costs and the lack of complementary equipment.

The effects of the grant on APCA's performance are difficult to quantify. Most likely, the organisation has sold the fibre at a higher price because of the better classification system. Based on the estimates provided in the business plan, a 20% price difference due to the better classification of the wool is presented as a reasonable estimate. If so, this would have resulted in an estimated yearly sales effect of US\$7,619. This is a very rough estimate, but it is plausible to assume that the grant indeed improved the APCA's group sales turn-over during 2010 and 2012.

For the further development of the business, in 2013 APCA considered it necessary to start a process to better define who is member and who is not. The 'solution' found in 2000 to control several tensions in collective marketing by aligning APCA's board with the 'elders' that rule the indigenous organisation proved insufficient. Due to this reorganisation process, the tension containment capacity score in 2013 is lower than in 2011. When indeed the new venture, APCA's

own spinning centre, will have materialised (in some years), this decline may be interpreted as temporal, and part of a learning cycle to which the grant has contributed. However, in 2013, we cannot but consider APCA as one of the grant implementers that suffered a decline in tension containment capacity between 2010 and 2013, especially as a result of ceasing collective marketing activities.

### **Outcome summary**

- Increased access to markets for members: Yes
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: Yes
- Increased income to pay organisational expenses: No



## APME

### Context

The honey processor, APME, formally established in 2006, has operated since 1998 as a beneficiary group of a vocational training institute working on value chain development (CETHA). APME is a typical example of an NGO-initiated group that gradually gained autonomy, and is learning and experimenting with internal rules to manage its collective marketing operations. In 2009 they started to provide honey on a monthly basis to the national nutrition programme for lactating mothers, as part of the ANPROABOL consortium of honey associations. They have one person paid half-time to organise the logistics and one person half-time to do the processing and packaging.

### Dynamics

They deliver most of their honey to the Subsidy for Lactating Mothers programme. In 2011 as result of climatic conditions, they could not satisfy the contract with ANPROABOL completely. They had to complete the order by buying honey from other honey producers in other municipalities. In response they decided that a minimum amount per member had to be supplied each year (70 kg), a system that works fairly well.

Their sales show a steady growth. They also sell honey that is certified as being produced by ecological farmers, following an Bolivian government initiative to label organic production from smallholders for the internal market.

The local government supported APME with the construction of a new enhanced processing centre, open for use to all honey producers in the area, but in practice managed by APME.

APME had 210 registered members in 2010 and 104 in 2012. However, they indicate that the number of active members supplying honey has increased from 50 to 104. In the earlier years many members registered because through APME they could get access to NGO support (PUMA), which constructed bulking centres in each of the villages. However, many of them did not use the centres and did not start to produce honey. In 2011-12, the local government distributed equipment for bee-keeping. APME used that opportunity to visit the villages and check who really produced honey and paid their membership fee (US\$7/year). Only these are now considered members of APME.

Their relation with the municipality is very fluid (the mayor is the former president of APME) and they are looking for ways to supply to the local school food programme.

### Grant influence

They started the honey processing plant with support of other donors, including an investment by the local government. APME received a FONDOECAS grant only late in 2012 to invest in bins to comply with sanitary regulations, As the FONDOECAS grant has not yet been operational, no sales effect can be attributed to the grant.

### Outcome summary

- Increased access to markets for members: NR
- Increased organisational capacities: NR
- Increased access to loans: NR
- Increased access to grants: NR
- Increased income to pay organisational expenses: NR

## APROAMOL

### Context

The smaller honey organisation APROAMOL is part of the second-tier municipal honey association APRODAL. It started as a beneficiary group around technical assistance by a technician paid by the PROSAT programme. To access PROSAT funding, the co-funding had been paid by the technician through a reduction in his salary, a practice observed also in other places in Bolivia where co-funding in cash is required to get access to technical assistance (Ton, 2007a).

Most members are diversified farmers who have broad beans as main commercial crop (*habas*), with honey as a complementary activity, and have, therefore, also an affiliation in the farmer organisation ASOHABA that exports dried broad beans. Honey production started after a donation by the municipal government of bee hives, six for each family.

The proposal to FONDOECAS was elaborated and submitted by APRODAL, initially meant to benefit a women's group. They changed the intended beneficiaries during the process, when it became a project to benefit APROAMOL.

### Dynamics

In 2010, APROAMOL had a membership of only 15, falling from 49 in 2006. The number of members increased to 20 members in 2012.

The number of hives managed by each member has increased from six to an average of ten hives per member. APROAMOL manages 18 hives as collective production. Nevertheless, production has declined, principally due to climatic reasons.

APROAMOL provides the service of harvesting the honey to the members, operated by its president. The president and vice-president are the ones who do the marketing. They are not paid, but do this as voluntary work for the community. The organisation changed several internal regulations in the last year, e.g., distributing part of the profits made, and maintaining a stricter control on quality.

Half of the harvested honey is sold collectively, the other half is sold by each member individually; the uncommon 'black honey' is generally sold on local markets.

Instead of delivering honey to institutions, they now sell most of the honey through the shop in the CIOEC-Potosí office. Many other honey associations use this shop, which has made them aware of the differences in quality between associations and the importance of complying with the food safety regulations. The shop also provides the opportunity to sell various qualities, including the black honey, while the former institutional buyers needed a uniformly light-coloured honey.

### Grant influence

APROAMOL invested the grant in a processing unit for honey, constructed on a site that was bought by funds from the members (US\$100/person). Production level are still low and sales are stagnant. APROAMOL is looking for additional funding to finish their processing unit, comply with the sanitary regulations, raise production levels and enter more lucrative markets. They consider the quality of supply, the cost of processing and the role of the board as positive factors that facilitated the business plan.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: Yes
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

## APROQUIRC

### Context

APROQUIRC is a rapidly growing regional branch of ANAPQUI (210 members) producing organic quinoa. It manages one of the quinoa processing plants to prepare for export through ANAPQUI. It also manages agricultural machinery that is rented to members. APROQUIRC uses a system of delayed payment, paying 30% in cash and 70% after ANAPQUI has finished its export sales. The price for organic quinoa is sufficiently high to make sales to ANAPQUI attractive. Production is bought according to a production plan, linked to the organic certification. Quinoa from non-certified plots may be sold individually on the local markets.

APROQUIRC employed one person to assist the board in logistics and technical assistance, and one person to do the book keeping. At peak times, some labourers are contracted to work in the plant.

Through ANAPQUI, members have access to a micro-credit facility, technical assistance and reduced prices for organic agricultural inputs. These services are funded by FAUTAPO, the national quinoa support programme.

### Dynamics

Between 2010 and 2012, ANAPQUI managed to get a loan from the development bank, Banco de Desarrollo Productivo, to pay the full price of the quinoa to supplying farmers in cash. This helped resolve the problem of side-selling to other intermediaries. Next to the price effect, volumes of quinoa sourced from members increased sharply between 2010 and 2012.

### Grant influence

In 2007, APROQUIRCI invested the grant in two silos, which facilitated an increase in their capacity to source group sales. These silos added to the existing storage capacity of ANAPQUI, which falls short and limits its capacity to grow. ANAPQUI stimulates the construction of additional infrastructure in its departmental branches. According to the information provided in the 2010 interview, the silos would have doubled the processing capacity from 1,660 to 3,550 bushels/year. While it is likely that the silos indeed have increased the sales volume, we have insufficient information to respond to the asses net-effects, as most of the increase in turn-over in the years after 2007 is a result of the above-mentioned price and market effects that incentivised quinoa production in the area.

According to the 2010 interviewees, the silos increased the trust and commitment of the members to the organisation and the conviction that APROQUIRC could manage complementary commercial activities a group. In 2013, APROQUIRC considered starting a processing unit to sell added-value products on the local market.

### Outcome summary

- Increased access to markets for members: Yes
- Increased organisational capacities: Yes
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

# APSU

## Context

APSU is a small handicraft organisation located near the border of Chile, specialized in alpaca weavings with a membership that declined from 60 in 2010 to 32 household members in 2012. It sells part of its products in an alliance with the federation COMART, that manages a shop in La Paz. They also sell their products at a touristic centre/hostel in Livichuco, owned by APSU and intended for agro-tourism. They have some basic equipment. They sometimes obtain contracts, for example to make sportswear (using synthetic fibres).

APSU has a production centre in Challapata, constructed with support from the European Union, and they managed to renew their semi-industrial equipment with support of the PAR programme and co-funding by the municipality and the members (25%). They also procure wool from the members to sell collectively.

The board functions rotate every two years, similar to the governance of the indigenous village organisation (ayllu). APSU invested the grant in equipment for a shop in La Paz, opened in 2009. They hoped to double production and sales through this new outlet. However, they could not pay the rent and had to close down already in 2010. After this, they started to sell their weavings mainly through the second tier organisation COMART.

## Dynamics

Sales have declined, especially in 2012. This has led to a stock of unsold products, some of which were returned to the artisan that made them, to be sold independently. They also stopped procuring inputs collectively due to lack of working capital.

## Grant influence

APSU invested in furniture in a showroom for the direct sale of their handicraft products. It is clear from the interviews that the business plan to which the grant contributed failed completely.

## Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## ARAO

### Context

ARAO, formed in 1983, and with legal recognition in 1990, is specialized in the production of carpets, sweaters, shawls and ponchos, and has its own shop in Oruro. It uses a system of production planning, pays cash, and offers the possibility to deposit any additional handicraft products in the shop under a system of consignment: the artisan is paid only when the product has been sold. They sell part of their products through COMART (they are one of its members) and have a partnership with INCA PALLAY, which is interested in having a broader assortment of quality weavings in their shops in Sucre and La Paz.

ARAO coordinates the production in by the members according to projected sales. The artisans procure and pay inputs themselves in order to comply with this production plan.

### Dynamics

ARAO reduced its membership from 90 members in 2010 to 48 in 2012. The remaining members work more intensively in the production of handicrafts for ARAO, and generate more income from this activity than they did before, with a larger membership.

ARAO improved their turnover between 2010 and 2012. They invested the grant in improved weaving equipment for the production units, which positively affected the quality and marketability of the products. The yearly sales are estimated at US\$10,036, using the sales volume in 2010 as a baseline.

### Grant influence

ARAO received two FONDOECAS grants, invested in equipment for several local handicraft production centres (2011) and in furniture to increase the sales in their shop (2012). This made it possible for the individual members to produce more and generate more income from handicrafts than before. This effect is not totally attributable to the FONDOECAS grant but it is considered to have been a contributory factor..

The 2013 interviewees indicate the support of other NGOs and the local government in the success of the business plan, the quality of the equipment bought and the role of the board in managing the projects. As a negative factor they signal problems in the COMART shop, which affected the sales of ARAO's products.

### Outcome summary

- Increased access to markets for members: Yes
- Increased organisational capacities: Yes
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

## ASAFOP

### Context

ASAFOP is a small organisation with predominantly female members active in food processing. The group started as an informal women's group supported by charity NGOs, with a diversity of training programmes for 'women as mothers,' and was functional in distributing part of the food aid received, especially from the US programme PL-480. When the latter function disappeared in the 1990s, ASAFOP took part in a reforestation project. Based on these experiences, the group started to look for other ways of generating income. In 2004, as one of the first organisations in the country, it managed to get a contract with the municipal government to provision part of the school meals. ASAFOP started its processing activities on a small scale, mixing wheat flour with the local bean karwi. Due to the changing product preferences in the local school food programme, they had to specialise in baked goods, based on inputs bought on the local market.

### Dynamics

With significant co-investment in cash from each member, they invested the grant in equipment for the bakery. However, they failed to get the contracts with the Municipality of Sucre after 2010, and managed to get only minor contracts with other institutional buyers. In 2011 and 2012 they managed to win additional contracts in another municipality, Monteagudo, and, in 2013, with the regional hospital in Sucre.

As a result they stopped sourcing from members but continue as a micro-enterprise, with the objective to generate employment for the (female) members of the group. In 2012 they managed to get a subsidy from the Employment programme (PAE) to pay part of the labour costs. All profits are now reserved for future investments and not distributed anymore at the end of the month.

### Grant influence

ASAFOP invested the grant in an industrial oven and food mixer to provide to the school feeding programme. However, they did not get the contract and used the equipment to sell directly to consumers, and later to institutional buyers. The FONDOECAS grant has been a key factor in adapting their processing activities to changing product requirements. Without it, they would probably have had to stop operations. However, from an organisation creating market access for agricultural products, they converted into a micro-enterprise. The grant did not help to improve market access for member products.

They consider their dependency to only one buyer as the negative factor that affected the business plan, though they consider the government procurement a positive factor, together with the good quality equipment and the role of members and board. Their tension containment capacities were already quite high in 2010, after having received the grant and having participated in the school meal programme, which helped them to find solutions when the government procurement market was constrained.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: Yes
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

## ASAPROF

### Context

ASAPROF is a relatively large organisation that exports beans and sesame, sourced from members and non-members through a contract-farming arrangement. They have a steady number of 300 to 350 farmers that contract with them, of which less than half are formally members. ASAPROF provides them with seeds and inputs as a pre-harvest loan. Many members see ASAPROF as one of the potential buyers of their product. Many farmers have contracts with two or more firms, one of them ASAPROF, and prioritise their sales to the firm that pays the better price, independent of whether this firm has pre-financed the production or not. ASAPROF has good access to trade capital from various financial institutions: FOPREPO, BID, OIKOCREDIT. They have a professional staff of seven persons.

To stimulate the production of sesame (e.g., in a rotation with soya) and stimulate good agricultural practices (reduction of toxic agrochemicals, e.g., to get access to the Japanese market), ASAPROF participates in an extension programme with funding of the International Development Bank in coordination with the sesame export chamber (CABEXSE).

### Dynamics

ASOPROF faces problems of disloyal behaviour due to the limited commitment and identification of the members with their organisation. Due to a problem of side-selling and credit default, the organisation has introduced stricter loan conditions, including the requirement of formal guarantees such as machinery or infrastructure.

Exports have been growing steadily between 2010 and 2012, but meanwhile their relations with members have deteriorated. Increasingly, the organisation sources from non-members.

ASOPROF did not access credit from the BDP. They consider this a result of political decisions of the government. Politics is heavily polarised in the Eastern Lowlands between organisations that supported the MAS and those that supported the opposition to the Morales government. ASAPROF explicitly prioritised activities to improve their relations with the government, as they consider public support essential to the expansion of their operations. They also are considering becoming more active in CIOEC. They see CIOEC's role principally to assist in managing the difficult relations with the government, which is necessary for example when exporting to countries such as Cuba and Venezuela, very interesting markets for ASAPROF.

Also, they see a role in CIOEC's Leadership School for improving the commitment of members, as they want to promote a more pro-active participation of the board in decision making. The dominant role of contracted staff in communications with the farmers results in many supplying farmers (even those who are members) viewing the organisation as a private firm.

### Grant influence

ASOPROF did not apply for a FONDOECA grant.

### Outcome summary

- Increased access to markets for members: NR
- Increased organisational capacities: NR
- Increased access to loans: NR
- Increased access to grants: NR
- Increased income to pay organisational expenses: NR

# ASOCOM

## Context

ASOCOM is an association that emerged when the community seized a stone mine owned by the former, expelled president Sanchez de Losada. It is an enterprise of 72 members that in 2006 choose the legal identity of an association. They do not really source from members, but work as a collective enterprise with central negotiation of contract and work assignments to mining teams. Most of the stones are sold to pave the urban roads in La Paz.

## Dynamics

ASOCOM decided to convert to the legal form of a cooperative in 2011, changing its name to COCACOM. This was more appropriate to the form and function of the organisation, and made it possible to join the organisation of mining cooperatives FEDECOMIN, which could better serve their interests than CIOEC. They adapted their internal regulations concerning the distribution of profits, which is the main difference between the legal forms of association and cooperative. The turn-over is growing steadily, without many changes in the type of buyers.

## Grant influence

In 2008 they received a FONDOECAS grant for a compressor to facilitate stone mining. They calculated a service fee to be charged when using the compressor to recover operating costs. However, the equipment proved to be too heavy to handled easily. It needed transport to be moved from one location to another. The group ceased to use it. The increased sales of the group is unrelated to the grant-supported business plan. Nevertheless, the ASOCOM board acknowledges the support of FONDOECAS, as the compressor is until now their only collectively owned asset. The sales effect of the grant can be considered as being zero, even though the total sales of ASOCOM/COCACOM has increased substantially during this period. The grant had, likewise, no effect on the tension containment capacities.

## Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No



## ASPASA

### Context

ASPASA is a relatively new organisation that started as a dairy association, in 2006, providing technical assistance, but increasingly works with quinoa production and marketing. Dairy is still an important activity of the association. The 37 members started with collective marketing of quinoa because as dairy farmers they paid for dairy inputs and equipment with quinoa, instead of cash. They sold the quinoa through ANAPQUI, becoming one of their regional sourcing centres. This caused an increase in quality requirements, which first created some problems but later was accepted by the members. All activities are performed by unpaid board members. They have several members with professional skills. The president, for example, has a university degree, specialised in quinoa production and skilled in the formulation of projects. They diversified their activities to stimulate quinoa production among members, next to their dairies. For the dairy producers, quinoa is a secondary activity. The average yearly income from quinoa (average production 10 ha) is similar to their monthly income from dairy production. In 2011 they did not have organic certification, even though production took place without chemical inputs.

### Dynamics

ASPASA grew to a membership of 52 families in 2012, many of them quinoa producers without dairy production. They are close to being certified as organic quinoa producers, which would need a stricter planning of production and can increase the price of their product. Twenty-five members are in the final stages of being certified by BIO LATINA for organic exports. For the local market, they are working with the government to be recognised as ecological producers, following a new legal regulation.

Through the higher margin on certified organic quinoa, ASPASA plans to introduce a system of levies to pay for some staff and other organisational costs, now paid by the board members themselves.

They increased the volume of quinoa sold from 1,000 bushels in 2010 to 5,000 bushels in 2012, partly sourced from non-members. Most is channelled through ANAPQUI but, using a loan from FONDOECAS to increase their working capital, they also started to supply to other buyers who offered higher prices for first-quality quinoa grains. ASPASA could serve this more demanding market and sold the smaller grains (second grade) to other buyers. They opened a website to interest buyers through a web-portal developed by CIOEC.

They managed to get a long-term loan from BDP to buy six tractors. The loan is individual but the group did the work to channel the support. They also managed to get micro-irrigation equipment from the innovation platform Fundación Altiplano, which is rented out as a service to members.

### Grant influence

ASPASA never received a FONDOECAS grant. However, in 2012, a FONDOECAS loan permitted increased sourcing of quinoa and cash payment to members. This undoubtedly increased the price of the quinoa received from this alternative buyer. However, logically, there is no sales effect due to the grant, as no grant was provided by FONDOECAS but a loan.

### Outcome summary

- Increased access to markets for members: NR
- Increased organisational capacities: NR
- Increased access to loans: NR
- Increased access to grants: NR
- Increased income to pay organisational expenses: NR

# CECAOT

## Context

CECAOT is a federation of 13 local cooperatives that produce quinoa for export and processes quinoa products for the local market. One of the oldest OECAs in Bolivia, founded in 1974, and initially supported by the Belgian development cooperation, IAF and BID (Healy, 2001). CECAOT started to export quinoa in 1984 to the EU. CECAOT can be considered as the pioneer of quinoa marketing at a time that quinoa was still considered an inferior 'backward' grain by urban consumers .

CECAOT has an industrial processing plant and owns some machinery, rented out as a service to farmers. CECAOT used the grant to repair the optic sensor of the selection machine necessary to meet uniform export quality of quinoa grains. However, the optic sensor only worked for one year after being repaired. As a result, the removal of black-pointed-quinoa, not allowed in exports, continued manually by workers in the plant.

## Dynamics

In 2010, CECAOT lost a traditional client in the EU and, without this forward sales contract, could not access a loan for trade capital. As result they could not buy much quinoa. Due to the rising quinoa prices, turnover remained stable, but volumes declined. This caused a decline in membership, which recovered in 2012. CECAOT suffered from competition of other buyers who paid farmers in cash. In 2012 they managed to get a loan which enabled them to pay cash at the moment of sourcing from members, which caused a sharp increase in turn-over from 1,800 bushels in 2011 to 7,000 bushels in 2012.

The sale of processed quinoa is considered to be a promising complementary activity. In 2012 they managed to gain the contract to supply to the local school meal programme in the municipality of Colcha 'K' (Uyuni). This caused an increase in the turnover of processed products.

## Grant influence

CECAOT used their first FONDOECAS grant, in 2009, to repair an optical quality control unit in their plant, to limit labour costs in the plant. The maximum amount available from FONDOECAS (US\$10,000) motivated them to repair the equipment instead of buying a completely new machine (US\$40,000). However, the equipment broke down again in 2010, partly due to improper handling. The optic sensor has not been repaired anymore due to the high costs. Instead, CECAOT considered buying a completely new optical sensor, which they did not do however, partly due to the crisis and resulting internal organisational problems in 2011 which resulted from the failure to get a pre-harvest sales contract.

CECAOT benefited from a second grant, in 2012, intended to strengthen their quinoa processing activities and diversifying the range of processed products offered. International buyers indicated demand for quinoa meal. The implementation of the grant-funded business was still ongoing in 2013, and had therefore not yet influenced the group sales.

## Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## CELCCAR

### Context

CELCCAR, founded in 1965, specialises in organic coffee, generally marketed under the Fair Trade label. The Fair Trade premium is divided between CELCCAR and the member cooperatives, and is generally used as co-funding for support projects rather than distributed. CELCCAR has a membership of 11 cooperatives, three of which joined in 2011. Only four of these cooperatives are specialised in coffee. Other cooperatives produce stevia, citrus fruits and the natural colouring *achiote*. CELCCAR's prime collective marketing activity is the export of coffee for their member cooperatives. The cooperatives pay farmers 40% of the price in cash at the moment of delivery, and complete the payment five months later, when the processing and export process has been completed by CELCCAR. CELCCAR managed a program to expand production with new coffee shrubs, two hectares per member. This programme was supported by the Swedish NGO SCC, who paid the salaries of five technical staff. Board members receive a daily allowance of US\$10 when they work for the organisation. CELCCAR also managed an Internet café and rented out shops to cover part of the organisational costs of the organisation.

### Dynamics

Three cooperative are specialised in citrus fruits. CELCCAR used the FONDOECAS grant to invest in juice processing equipment, installed in one of these cooperatives. With SCC, they invested in a processing unit for the natural sweetener stevia. Export of coffee by CELCCAR has been growing steadily between 2010-2012, even though coffee prices declined between 2011 and 2012.

### Grant influence

CELCCAR channelled the FONDOECAS grant to one of its member cooperatives. They experimented with fruit processing on a pilot scale. They mention internal organisational problems and lack of complementary equipment as the major factors that negatively affected the business plan. The capacity of the equipment was considered by the 2013 interviewees to be too low to seriously create market access. The juice produced was made traditionally, without using the machinery bought with the grant, had only been used for internal consumption and sales in the Internet café. The yearly sales effect of this pilot experience, estimated in US\$952, has been very small in relation to the number of members. An expansion of production capacity is needed to obtain real access to the market. The maximum amount of support provided by FONDOECAS, approximately US\$10,000, could well be the reason for this under-scaled investment.

Because they managed the business without making use of the grant investment and because they already had a high tension containment capacity (and declined between 2011 and 2013), we consider that the grant cannot be considered a contributory factor to an increase in organisational capacities.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

# CEMUR

## Context

CEMUR is an association of women's groups organised around capacity building and business development. It manages collective production units to sell products to their members. It has around 150 women as members. CEMUR develops many activities to stimulate employment and income for the women, but does not procure agricultural products from their members.

CEMUR has a large patrimony and a diversity of social and economic activities with women groups (Clubes de Madres). As part of CEMUR, they manage intensive husbandry, poultry and pig farming, they have 16,000 hens for egg production, manage a small slaughterhouse and sell animal feed. All these activities have the objective to generate employment for women and income for CEMUR.

In addition to animal production, they organise activities to stimulate economic activities by women in their own farm or house. This did however never result in a production that was collectively marketed. The only link between CEMUR and the farming system of members is for the production of the animal feed, where CEMUR tries to buy preferentially cereal from members. However, due to quality and logistics, only 20% of the cereals needed could be sourced from members.

CEMUR also provides many social services, such as legal assistance, capacity building and education. These activities are supported by NGOs that use CEMUR as their outreach structure to rural women.

## Dynamics

CEMUR worked for many years more as a development NGO than as a farmer-led organisation. In decision-making, the director had a dominant role and the board members had little influence. In 2012, the organisation started a process to change its internal governance system and make it more member-driven. One of the intentions was to make a distinction in economic and social activities in the internal bookkeeping and financial management. They discussed the possibility of passing the production-related patrimony to a new legal entity, with participation only of the women's groups engaged in economic activities. However, during the period 2010-2013 this process had not yet been completed.

## Grant influence

The growth in turn-over was generated by several production lines (broiler hens, pigs, animal feed). The FONDOECAS grant supported the development of a semi-industrial production line in meat processing (*embutidos*) for which sales were low. One reason for this was the incomplete infrastructure, which did not comply with the food-safety requirements. Registration is needed to make it possible to sell to supermarkets and institutional buyers, and this leaves them with an abundant supply of very cheap meat products for the informal market. Production stayed far below capacity and the products were sold exclusively to the women in the member groups. Four persons worked part-time in the activity, while they had planned to have 20 persons involved. The 2013 interviewees indicated internal organisational problems and price competition in the consumer markets as the negative factors and the quality of the equipment as the positive factor that influenced the development of the business plan

## Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes

## CEPLACH

### Context

CEPLACH is a small women's association of dairy processors, founded in 2001. It specialises in the production of yoghurt and cheese, whenever there is a sales opportunity for the products. Its main objective is the generation of part-time employment and complementary cash income for the female worker-members. CEPLACH also creates market access for a minority of the members who supply milk, selling in the city of Oruro and on regional markets. In the past they were one of the first farmer associations in Bolivia to supply the school meal programme. When delivering on a daily basis to this school food programme (in 2002, 2003, and 2008), the various teams within the association took turns, each team delivering products for one week.

### Dynamics

The women work in the processing of yoghurt and cheese, in addition to their housekeeping. The earnings (approximately US\$70/week) are modest compared to the time needed to work in the processing. This situation demotivated some of the members. In 2010 they had 40 members and in 2013 they were a group only 20, with ten more women hoping to join the group.

One of the benefits of being a member of CEPLACH is the access to micro-credit and other support from development institutions. For example, the members gained access to the micro-finance institution CRECER because the infrastructure of CEPLACH could serve as collateral.

In 2011-12, sales decreased partly because of a relocation of their main selling point, and partly because they suffered increasing costs due to higher prices paid for raw milk by LACTEOBOL, which affected the profit margin when selling on the (low-price) informal market.

### Grant influence

CEPLACH used the FONDOECAS grant to buy a site to install their activities instead of renting it. They also invested in a new product based on whey, as a by-product in cheese production (*Chicolac*).

Their sales declined as a result of the relocation of their plant, and was a direct result of the grant-supported business plan. The negative sales growth can, therefore be attributed to the grant.

Nevertheless, they are positive about the grant, as the fact that they did not have to pay rent increased the profit margin on their products: more income with fewer sales. They point to competition in the market and lack of complementary equipment as factors that negatively influenced their business, and considered the support of NGOs, the quality of their product and the role of the board as positive factors.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: Yes
- Increased access to grants: Yes
- Increased income to pay organisational expenses: Yes

# CIAPEC

## Context

CIAPEC is a relatively young organisation (2003) and has a direct membership of persons who formerly belonged to other cooperatives but left as a result of disagreement about the price-determination system. CIAPEC started with significant member contribution and took several decisions to organize itself differently from other cooperatives, e.g., those related to profit distribution and capitalisation. They employ three permanent staff and 20 persons in their processing unit. They export organic coffee, principally to Germany. In 2010, they paid 30% at the moment of purchase from members and complement the 70% after the export sales have been finished at the price level of the transaction and have been paid by the German buyer. They rent a small plant in El Alto to process coffee before export, and were preparing to build one themselves. With support of USAID, they also managed a rejuvenating programme for specialty coffee in 2009-2010, and in 2011 they started with a project supported by PAR to invest in natural control of coffee pests.

In 2008, they changed the pre-processing system to provide the plant with dried coffee. Before, they mixed de-pulped fermented coffee from different farmers at the drying centre before transporting it to El Alto for final processing. Now, each individual farmer is responsible for the quality of his or her coffee, which is paid once the whole processing is finished and the quality of the coffee provided by the farmer is determined by the taster (catador). These quality points depend on altitude, soil type, plant variety and post-harvest handling (Kawai, 2011).

## Dynamics

The export volume and turn-over of CIAPEC grew steadily between 2010 and 2012. They managed to construct and open their own processing plant in El Alto in 2011. Just like most other cooperatives, they suffered problems in 2010 due to the rise in world coffee prices. They had contracted an export volume with a predetermined price some months pre-harvest, before actual procurement from the farmers. They had problems purchasing enough coffee because members side-sold the coffee that they produced to private buyers that offered a higher price and cash payment. Because of this, almost half of the members left the cooperative.

CIAPEC increase the price paid to farmers up to 50% of the price at the time of delivery. To withstand competition from intermediaries who pay cash, they provide access to credit for their members, based on a loan obtained from FINCAFE, the financial institution of coffee cooperatives. As an additional service to members, CIAPEC facilitates using equipment of the cooperative for private use. Access to training on coffee-growing practices is another service provided to raise the commitment of farmers to their organisation..

## Grant influence

CIAPEC wanted to develop a production line for roasted coffee for the national market in La Paz, including expectations for export. It started to experiment with roasting and packaging but the production capacity was lower than expected and they experienced technical problems with the equipment after only one year of operation. They consider the equipment not suitable for processing on an industrial scale. CIAPEC planned to buy new equipment using their own resources. The lower coffee price in 2013 was considered an opportunity to upscale the processing business. At the moment, the yearly sales of processed coffee is estimated to be US\$2,856. This sales volume is insignificant in relation to their overall sales volume.

Nevertheless, the coffee roasting pilots were considered by the interviewed board members to have served as a learning experience about the technical and administrative needs incurred in the domestic consumer market of coffee. However, the capacity, access to finance and patrimony of CIAPEC is such that they could have developed the activity themselves with their own resourc-

es, had they considered it a promising business opportunity (Prudencio, 2010). They also did not mention any positive factor for the development of the business plan; they mention technical and organisational issues as major factors that negatively affected the business opportunity.

### **Outcome summary**

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## COAINE

### Context

COAINE manages several coffee processing facilities for pulping and removal of mucilage and parchment. COAINE was the first Bolivian organisation to export organic coffee, and it sells most of the coffee under the Fair Trade label to the Netherlands. COAINE employs four permanent staff and approximately 30 persons who work in the drying and parchment centres. They have access to various credit lines (including a large loan from the BDP) that permit 40-50% advance payments to farmers. The business plan submitted to FONDOECAS consisted of roasting equipment to access the national market with a finished product. Most of their coffee is exported. COAINE wanted to grow in the national market to improve the visibility and recognition of COAINE as a cooperative with a quality product.

Most of the coffee farmers dry their coffee themselves and sell directly to COAINE's plant in El Alto. One member organisation uses a centralised system for de-pulping and drying coffee berries, which generates a more uniform quality of coffee beans (Kawai, 2011).

### Dynamics

Due to the steep increase in international coffee prices between 2008 and 2011, roasted coffee processing relied on second-quality coffee not suitable for export to the Fair Trade market. COAINE used to contract the exports with a predetermined price some months before shipment. With this contract, the organisation could access credit for trade capital to pay the farmers. However, due to price hikes in the months before the actual shipment, they had problems purchasing enough coffee because members side-sold the coffee that they produced to private buyers that offered a higher price. In response, they had to complete shipments by buying coffee at a higher price than they could sell it, both from members and non-members. This was aggravated by a local incident some months later. The person responsible for the finances of one of the Colonias was robbed and killed at his home and the money to pay the coffee producers in this Colonia disappeared. The colonial committee had already paid the producers 40% in total amount in advance, so the amount of 60% was ready to pay in cash. This theft affected the internal organisation of COAINE, reducing trust in the collective marketing process.

With the FONDOECAS grant, COAINE wanted to establish a production line for roasted coffee for the national market (Cafe COAINE), a service that previously had to be externally contracted to a private roaster. In 2009, it started to process roasted coffee on a pilot scale and worked with the Health authorities to get their food safety certification, which would allow them to access the market. By 2013, COAINE had not yet managed to obtain the certificate and coffee roasting was limited to the supply for internal consumption, and sales at sporadic festive events (ferias). The equipment is sometimes provided as a free service to members to roast coffee for individual consumption.

### Grant influence

The grant proposal of COAINE was rejected twice in the FONDOECAS grant system. At the third attempt, in 2009, COAINE improved the proposal with the help of a technician from CIOEC headquarters. They had preferential access to this skilled support because a member of COAINE was treasurer in the national board of CIOEC at that time.

The equipment bought with the grant was far too small for the use that COAINE projected. COAINE considers the service provided to members as positive. However, additional access to markets has not been created, nor has COAINE visibility in the market been enhanced by the grant. The average yearly turnover of processed coffee was only US\$1,393 for COAINE, an insignificant amount when compared to the size of the total turnover and size of membership. The interviewees mentioned competition in the market and the role of board and technical staff as key factors that negatively influenced the development of the business plan.



## **Outcome summary**

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

# COMART

## Context

COMART is a second-tier artisan organisation (1,200 members in 32 affiliate organisations), founded in 1997. They originally targeted the Fair Trade export market. However, in practice they depend on the national tourist market. Exports require volume of uniform quality, which is a challenge when working with small artisan associations as members.

COMART has four shops to sell handicrafts. To increase the capacity of their showroom, they invested in furniture for one of their shops, and remodelled all four shops. COMART has a coordinator and employs vendors in their shops. After experimenting some years with having board members as salespersons in the shop, they reverted to professional vendors, although this resulted in additional operational costs. Part of these staff costs are covered by support from the Belgian development organisation SOS FAIM.

## Dynamics

COMART suffered a reduction in members due to changing social and economic conditions. In 2010 they had 42 affiliated grassroots organisations and in 2012 this was reduced to 32. After 2009, their total sales started to decline, increased competition being the major cause of this decline. The private handicraft shops in La Paz have improved the quality of their products. Private shops are now partly sourcing from COMART's member organisations, paying them cash at the moment of delivery. Due to working capital constraints COMART must use a system of consignment, in which the product is paid to the artisan only after having been sold, with a percentage retained to pay for COMART's intermediation. To improve sales in the shops, they started a strategy of making commercial alliances with other artisan groups, allowing them shelf space in the COMART shops.

## Grant influence

COMART received two grants. In 2009 they invested in the equipment for a new shop, and in 2011 they used a second grant to improve working conditions in 20 member organisations. In 2012, the newly furnished shop had to close down, due to a reduction in sales and an increase in competition of other shops entering the market of quality weavings.

The decline in total sales, an average of US\$14,286 per year, cannot reasonably be attributed to the two FONDOECAS grants. Without the grant investments, this decline would probably also have taken place. Nevertheless, we consider that the grant to COMART did not result in a positive sales effect.

The investments with the 2011 grant did improve the relations with their member organisations and helped to dissipate some tensions within the membership around the collective marketing. However, one of the reasons for success was the fact that they could divide the 2011 grant over all members instead of targeting a subgroup or more focussed business opportunity, which may indicate that the grant was only a short-term solution to one of the (non-core) agency dilemmas in collective marketing.

The 2013 interviewees considered the role of the board and staff, and the support of NGOs to be positive factors that helped in the effective implementation of the grant-supported business plan.

## Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: Yes
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## COPROQUINACC-T

### Context

COPROQUINACC-T is the smallest regional organisation that is part of ANAPQUI. It started in 1998, but only in 2006 did they manage to obtain legal status, with 60 members. Its activities are principally the bulking of quinoa that is transferred to ANAPQUI for processing.

### Dynamics

The organisation is growing in membership and registered 130 members in 2013. The volume of quinoa bulked for ANAPQUI increased from 3,000 bushels in 2008 to 7,000 in 2012.

In 2007, COPROQUINACC started a project to build a processing unit for which they used the FONDOECAS grant. However, they needed to change the location of the plant due to shortage of electricity and water in their current locality. In 2012, they asked for a second grant to finish the project, which was however rejected by FONDOECAS technical committee.

### Grant influence

COPROQUINACC-T used the grant to prepare for a relocation of the processing plant. They invested in additional equipment to streamline the processing process in the new location, which however were not operational yet in 2013. The 2013 interviews blame the (former) board and members for neglect in resolving these issues and implementing the project. There is no relation between the large increase in the volume of sales and the grant-supported business plan.

COPROQUINACC-T arranged their legal status to gain access to the FONDOECAS grant. The project was formulated with the support of a professional of the NGO Buena Vida. The equipment bought with grant is the only patrimony owned by the group. The process for getting the FONDOECAS grant may have helped to organise the group, even though the business plan did not prosper, but rules and regulations about collective marketing activities have not changed much in response to the grant, as they are principally discussed and defined in ANAPQUI.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## FENCA

### Context

FENCA, Federación Nacional de Cooperativas Arroceras de Bolivia), the oldest organisation in the sample (formed in 1964), is a federation of sixty rice cooperatives in the lowland area of Santa Cruz, with an estimated membership of 1,200 members, 500 of whom are considered to be active members (2011). FENCA's members are the relatively small and medium producers in the Eastern Lowlands. FENCA manages a rice mill and sells white rice on the national market. FENCA realises its logistic activities as a first-tier organisation directly with the individual producer, as their member organisations do not count on working capital and facilities for processing. Next to this processing service, the rice farmers are given a registration card which they can use to access a preferential credit line from the governmental Banco de Desarrollo Productivo (BDP).

For many years, most of their organisational expenses could be paid with the levy per bushel of rice transported from the centre of production, Ichilo province, to the rest of Bolivia. However, due to political pressures of competing farmer unions (the ones that supported the Morales government), the operations of the road blocks had been disrupted for several years, and, since 2010, the authorisation to raise these taxes had not been renewed. In addition to their services in processing, FENCA managed a programme on seed improvement and technical assistance provided by the Centro Internacional de Agricultura Tropical (CIAT) through external funding from the Fondo Latinoamericano de Arroz de Riego (FLAR). In the absence of sufficient group income, the required co-funding for this programme has been paid with public funds from the regional government.

### Dynamics

In 2011-2012, the rice cooperative FENCA lost its role in collective marketing as a result of the emergence of parallel groups initiated by the government-managed EMAPA. In 2013, it intended to re-establish itself as a representative sector organisation, to regain recognition by the government as a representative body of rice farmers. Overall, their capacities to manage collective marketing decreased between 2010 and 2013: many internal regulations needed to be re-worked. It is not sure if FENCA will be able to recover its market share with its activities in processing and collective marketing, due to the continued presence of government-induced initiatives by EMAPA (Cordoba, 2014; Cordoba and Jansen, 2013). EMAPA works with a large sector of rice producers, including most of FENCA's members, and provided similar services to its constituents, such as input credits and preferential prices. FENCA will continue as an organisation. They prioritise their role as a representative body of rice producers to advocate for enabling policies and support programmes. Some of their traditional members moved away from rice to produce other crops such as sugarcane, and a large portion of rice farmers has settled in the new agricultural frontier. They decided to turn to a direct membership organisation, (re-) affiliating individual rice producers. In the absence of collective marketing activities, they have sought other ways to generate income to pay their recurrent organisational costs. In 2013, they used existing funds (part of their patrimony), income from consultancy services provided by FENCA staff, and they started to get a cash income out of their re-affiliation process (US\$1 per hectare, approximately US\$20-30 per member).

### Grant influence

FENCA has a strong political network in the region, including active participation in the powerful association of commercial farmers in the Eastern Lowlands, CAO. This makes CIOEC less important as their representative body, and as broker for support. They were active in the re-founding process of CIOEC in 2002-2002, but currently only participate in the Leadership School to prepare high-potential members for leadership positions.

FENCA has not applied for a FONDOECAS grant.

Their mill is only working at low capacity, to provide rice for special markets, such as the procurement for the police in 2011. The mill needs a significant investment to reopen, and FENCA needs trade capital to manage it. The amount available with a FONDOECAS grant is far too low to be of interest to FENCA. In 2013, FENCA was negotiating a loan of US\$500,000 with the BDP to do so.

### **Outcome summary**

- Increased access to markets for members: NR
- Increased organisational capacities: NR
- Increased access to loans: NR
- Increased access to grants: NR
- Increased income to pay organisational expenses: NR

## INCA PALLAY

### Context

INCA PALLAY is a direct-membership organisation that manages various production units within two geographically separated cultural regions, east and north of the city of Sucre. INCA PALLAY coordinates the production plans with the representatives of the production units but buys directly from the individual artisans, paying in cash at the moment of purchase.

It is specialised in weaving art targeted to tourists and high-end markets. It has shops in Sucre and La Paz, and a museum-shop near the tourist market of Tarabuco. To diversify the offer to clients, they have partnerships with other economic farmer organisations such as COMART and ARAO.

Five persons work full-time in INCA PALLAY, plus a Belgian volunteer. INCA PALLAY has support from various NGOs in the salary costs of the coordinator (SNV, SOS FAIM).

### Dynamics

INCA PALLAY showed a steady but slow increase in sales, with fluctuations principally related to political unrest that affected the flow of tourists to Bolivia. They changed from working with a professional vendor to working with one of the women in the board to directly communicate with potential buyers in the shop, and considered this to result positively in sales, and positive also for raising the commitment of members to their organisation. Several other buyers compete with INCA PALLAY in procuring the highest quality weavings from the women in the area of production.

Sales in Sucre are growing, but stagnating in La Paz, where competition with private handicraft shops is tough due to the better quality products that these are selling. It is exploring possibilities to access the Fair Trade market in Europe, supported in this initiative by a Belgian volunteer. However, this has not resulted in significant sales.

Membership is stable, though some women leave and others enter the association. Due to a constraint on the market for the weavings, they decided not to accept new members in 2013.

### Grant influence

INCA PALLAY invested the grant in a production unit, a weaving and dyeing centre, in one of the production areas (Paredón). The investment gave the group a place to work and keep their material. The members can more easily operate and organise themselves without interference of the village authorities. The effects of the grant are social and political rather than economic. Nevertheless, the number of weavers in this production centre has been declining due to more lucrative other income sources (road construction, dairy).

The investments were made in 2007. Compared with the sales volume in 2006 reported in the grant proposal, the sales in 2008-2012 only increased with a modest US\$681/year. It is clear that there is no sales effect due to the better conditions in this peculiar production unit, as weaving is generally an in-house activity. Though it is unlikely that there had been an influence of this investment in INCA PALLAY's core tension containment capacity, it may well have improved for the concerned unit of production. Internal organisational issues were considered to have negatively affected the development of the business plan.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: Yes
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## OMCSA

### Context

OMCSA started in 1994 as wheat-seed producers, with an impulse from the development NGO ACLO. Wheat seed was a high-value niche product that received significant governmental support (through USAID PL-480) during the 1980s and 1990s. The high altitudes of the Andean Valleys were conducive to the multiplication of virus-free seeds to be used in the extensive agriculture in the Lowlands of Santa Cruz. Technical possibilities for reproduction changed, however, and the market collapsed around the year 2000. In the same year, the NGO ACLO also stopped their support activities in the area. OMCSA tried to find alternatives products for collective marketing, such as other quality seeds, including certified potato seeds. In 2007 OMCSA wanted to restart the wheat seed production, but the state enterprise EMAPA imposed a direct contractual arrangement with the farmers, not through the existing associations of farmers.

### Dynamics

OMCSA's main role between 2010 and 2012 was to continue looking for business opportunities and projects to create markets and employment in their local area. OMCSA had only one staff member, who worked as coordinator and developer of projects. He was hired because of his previous experience in working in another economic farmer organisation in the area, AOCEMM. In addition to the two projects funded by FONDOECAS, OMCSA also worked with the French NGO Vétérinaires Sans Frontières to develop and expand their processing activities. The idea for a business plan submitted to FONDOECAS in 2009 emerged from a participatory planning process with the municipal authorities to define the local economic development plan. The business plan had been presented that year, 2006, for municipal funding but was not considered as a priority. Subsequently, the project had been submitted the same proposal to FONDOECAS in 2009. With a second FONDOECAS grant, received in 2011, OMCSA invested in a processing unit to provide wheat popcorn and sesame bars to the school food programme. However, the business was not yet operational in 2013. In 2013, the coordinator left and OMCSA activities became entirely managed by board members. The board decided to rent the bakery out to a private baker. In 2013, OMCSA functioned as a social organisation without collective marketing activities, although having some income from the renting out of the productive infrastructure (bakery, silos) that was originally intended to be operated as collective marketing activity. June 2013, the interviewed board members mentioned the role of the coordinator as a positive factor in the implementation of the business plan and acknowledged the role of the (former) board and the role of the members as factors that hindered the development of the business, along with adverse market conditions. In the 2011 interview, the board already admitted that the plan had been formulated without a proper market analysis.

### Grant influence

The bakery business was planned to involve 61 members. These members all signed the proposal, a standard eligibility requirement of FONDOECAS. The business never prospered as expected, and, in 2012, its operations involved only a group of 35 active members. Initially, when they made the proposal, they also wanted to create employment opportunities for members. But when implementing the bakery they decided to work with skilled labour from Tarija instead. In 2013, they rented all infrastructure to an external baker, who pays a monthly rent (US\$100) that serves to cover the organisational expenses of the board.

The sales volume of wheat through the bakery was rather insignificant in relation to the number of farmers that they represent. Based on the data provided, and even attributing the sales of processed products completely to the FONDOECAS grant, the estimated average yearly sales effect would be only US\$ 1,193.

The bakery activities were operational in 2011, with high expectation for expansion, which is reflected in a relatively high tension containment capacity at the 2011 measurement. In 2013,

most of the tensions were considered to be irrelevant by the interviewed board members and local researcher, which resulted in a low TCC-score. The externalisation of the activities may also be considered as an effective solution to resolve some of the tensions in collective marketing, and the 2013 of zero points might thus be an underestimate. In any case, we consider OMCSA as an organisation that has a far lower tension containment capacity in 2013 than in 2011.

### **Outcome summary**

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: Yes



## ORLIPA

### Context

ORLIPA (Líderes Productivos Agropecuarios Pampajasi) is an association of livestock herders initiated in 2004 by ten farmers, who had been trained as veterinary promoters, and provide services to the rest of the members. They were formalised as an association in 2007, and in 2008 they grew to 54 members. They manage several services to support and improve cattle herding by their members. As a collective marketing activity, ORLIPA manages a slaughterhouse, a cereal thresher, a milling machine, and a solar drier. Members pay a fee when they use these services. Members use or sell their own products after having them processed. As a group they sell processed products whenever they can get a contract with an institution. In these cases, they use a system of delayed payment. They are located close to the Peruvian border, which means that currency rates influence cattle and meat prices as well as the demand for their services.

### Dynamics

With the support of FONDOECAS in 2009, they started to experiment with the production of various meat products (sausages, etc.). In 2010, 42 of the members were actively involved in starting up these activities. ORLIPA has no staff; most of the work was done by the president of the association. After two years of pilot production, without having a collective market for their goods, they started to deliver products to the school food program of the municipality of Humanata in 2012. In addition to their own products, they were required to deliver other food items (quinoa, biscuits, yoghurt, etc.), which they had to buy elsewhere. However, the contract was changed just before they started to deliver, and excluded the product that ORLIPA processed using equipment bought with FONDOECAS grant. In 2013 the contract was discontinued by the municipality. Moreover, the price received was not attractive, and the members had to wait too long before ORLIPA could pay for their supplies. With a loan from FONDOECAS they managed to resolve the latter. In spite of this experience, at the time of the interview in September 2013, they had not managed to arrange a new contract. They continued providing services to members through the slaughterhouse, thresher and motorised mill, and were working with the authorities on food safety regulations to become a certified slaughterhouse. In 2012, they changed leadership and the newly elected president was later also elected as secretary-general of the communal organisation. By 2013, he had not yet dedicated himself to the re-launching of the collective commercial activities, and the business done with the school feeding programme had not yet been properly evaluated by the members. It is clear that the revenue was less than had been expected by the members. The members pressed for distribution of the margin gained in the business, paying only for ORLIPA's operational costs but leaving nothing for reinvestment. In the 2013 interview, the president and board member who operated the slaughterhouse mentioned the internal organisational issues as negative factors, together with the operational costs of the equipment.

### Grant influence

ORLIPA invested in meat processing equipment to sell boiled dried meat (*charque*) combined with beans (*haba*) or maize in pre-cooked meals, and to enter the market of the school meal programme. In 2012 they managed to get the contract, complying with the quality requirements. The activities were further facilitated by a FONDOECAS loan to resolve payment delays in the contract with the local government. However, at the moment of actual contracting, the product that they intended to sell was removed from the specifications. Most products that remained on the list were grocery products. Therefore, the increased group sales in that year cannot be attributed to the grant-supported business plan. The impact on their organisational capacities is ambiguous. ORLIPA formalised its organisation in 2006-2007, and explicitly mentioned that this was done in view of the opportunity of a FONDOECAS grant. They developed the technical proposal with the help of the NGO CUNA and supported by CIOEC-La Paz. The grant worked as a mecha-

nism to define internal organisational issues around collective marketing. At the time of the first interview, they had just started to implement the business plan, with some pilot products to test the equipment, and low sales of processed products. In the second interview, ORLIPA had gained experience with the implementation of the business. However, in 2013 they did not continue the processing activities, and several tensions were therefore considered to be less relevant. The tension containment score in 2013 was far lower than in 2010 and we consider that the intended outcome of organisational strengthening did not materialise.

### **Outcome summary**

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## SOPROQUI

### Context

SOPROQUI is one of the nine regional organisations that are members of the national quinoa organisation, ANAPQUI. These regional associations procure quinoa for processing and export by ANAPQUI, which manages an industrial processing plant to clean the quinoa seeds of saponin, a substance which results in a bitter taste. The members of SOPROQUI are those quinoa producers in the area that are certified as organic producers (BOLICERT), or those that are in transition to becoming certified. ANAPQUI has one person that works with SOPROQUI to organise logistics and provide technical assistance, largely funded through the quinoa programme in FAUTAPO, supported by the Dutch Embassy. Through this programme, members can also access some support in production (organic fertiliser, small implements, etc.). Farmers tend to sell only part of the quinoa at harvest time, selling bit-by-bit during the rest of the year, whenever they have need cash.

In addition to their core business of procuring quinoa for ANAPQUI, SOPROQUI managed a shop where members can get basic supplies of food, with a credit facility to pay with quinoa at the time of harvest. In 2008, they started to pilot quinoa processing for quinoa popcorn (*pipocas, estrusados*) and quinoa soup, with a view to accessing the local school meal programme.

### Dynamics

Between 2007 and 2010, they grew from 150 to 260 members. Volumes of quinoa declined in 2009-2010 due to adverse climatic conditions, but prices of quinoa in the world market began to rise to unprecedented heights. Between 2010 and 2012, the price of quinoa in the international market continued to rise. Many traders compete with SOPROQUI to source quinoa in their area of influence. Because of the high processing costs, working capital became a constraint. However, through ANAPQUI the quinoa organisation managed to get access to a loan from the BDP. They can pay farmers cash in hand when they sell their quinoa to SOPROQUI/ANAPQUI.

### Grant influence

SOPROQUI benefited from a FONDOECAS grant in 2008. It wanted to invest in processing and packaging equipment to supply processed quinoa products (quinoa popcorn, quinoa soup) to the market, including the school meal programmes. They projected a turnover of US\$20,000. However, the equipment was never properly delivered and installed. The FONDOECAS grant was one of the motivations to start with this new business activity, next to their core business, but the project never took off. The current board members do consider quinoa processing still as an interesting business opportunity but indicate that other machinery and skilled personnel is needed to start doing so. The increased sales effect is considered to be zero.

### Outcome summary

- Increased access to markets for members: No
- Increased organisational capacities: No
- Increased access to loans: No
- Increased access to grants: No
- Increased income to pay organisational expenses: No

## Annex 2 Changes in the sample

COMPONENT NAME	DEPARTMENT	SECTOR	COMPONENT	EXPLANATION	
Not included in 2011 baseline					
1	AMUPAAC	Potosí	Quinoa	Beneficiary	Logistic problems made interview impossible
2	APROMICAL	Potosí	Honey	Beneficiary	Logistics made interview impossible
3	ARCCA	Potosí	Camelids	Beneficiary	Regional organisation. Logistics made interview impossible
4	ASAGROP	Potosí	Camelids	Beneficiary	Logistics made interview impossible
5	ASPATI	Potosí	Fruits	Beneficiary	Logistics made interview impossible
6	ISLA COJATA	La Paz	Dairy	Beneficiary	Logistics made interview impossible
7	AMPASAN	Tarija	Cereals	Comparison	Group no longer existed
8	APROGRAY	Santa Cruz	Cereals	Comparison	Logistics made interview impossible
9	ASACHA	Tarija	Agriculture	Comparison	Group no longer existed
10	ASMAGOR	Chuquisaca	Amaranth, tarwi	Comparison	Logistics made interview impossible
11	ATHIPA	Chuquisaca	Fruits	Comparison	Group no longer existed
12	AZCCA - MUSAP	Potosí	Camelids	Comparison	Logistics made interview impossible
13	C-ITAYOBAY	Santa Cruz	Cereals	Comparison	Logistics made interview impossible
14	CETHA SOCAMANI	Oruro	Agriculture	Comparison	Vocational training institute without collective marketing activities
15	CORACA IRUPANA	La Paz	Coffee, honey, amaranth	Comparison	Refused to give information due to internal problems at the time of baseline interview
16	PICOTANI	Oruro	Fishery	Comparison	On the verge of disappearing; interviews could not be held. Members have other main activities (musicians).
17	REGANTES TACAGUA	Oruro	Agriculture	Comparison	Refused to give information due to internal problems at the time of baseline interview
Excluded in 2013 analysis					
18	APEMAK	Chuquisaca	Cereals	Beneficiary	The organisation seized to function as a collective marketing group in 2012. All assets had been passed to the villages syndicate. The board could not be contacted for a 2013 interview.
19	APLEPO ARASAYA	La Paz	Dairy	Beneficiary	The organisation broke down before the baseline measurement, after equipment bought with the grant was stolen. They continued to be a collection point for dairy but without collective marketing activities.
20	APRODESA	Potosí	Agriculture, dairy	Beneficiary	A second-tier organisation without direct collective marketing activities, supporting member organisations that do so. The baseline had not been conducted with the member beneficiary of the grant.
21	APROLAC	La Paz	Dairy	Beneficiary	The baseline measurement of capacities and sales volumes referred to the situation some years before the grant. In fact, the investments acquired with the grant have never been operational and used.
22	AMPAAIG	Tarija	Agriculture	Comparison	Group had not yet developed any collective marketing activities before or after the baseline.
23	APROMAJÍ	Chuquisaca	Peppers, peanuts	Comparison	The organisation had seized to function at the baseline after a one-time export intent through a semi-public support institution. The to-be-exported product had never been paid to the member suppliers due to alleged quality and ownership issues. They have not managed to restart any collective business yet.
24	CATCOPA	Cochabamba	Cereals	Comparison	The organisation COMUVA facilitated the use of processing equipment to women from the village organisations in a unit called CATCOPA-COMUVA. These women are not engaged in collective marketing but sell individually. COMUVA attracts donor support for social services to their wider membership. They had no collective sales and a huge undefined membership. Data-collection on sales and internal organisation in was incomplete in both measurements.



## Annex 4 Time-series data sheet

ITEM	DESCRIPTION	2008	2009	2010	2011	2012	ANY IMPACT DUE TO FONDOECAS FUNDED INVESTMENTS?		
							NO	LITTLE	MUCH
<b>1. ESTIMATED INCOME FROM SALES (Bs)</b>									
	unprocessed products								
	processed products								
	inputs for production								
	others								
<b>2. EVOLUTION OF NUMBER OF MEMBERS (ACTIVE)</b>									
<b>3. DETAILS ON THE ACTIVITY INTENDED TO BE ENLARGED ACCORDING TO THE BUSINESS PLAN</b>									
	value of sales of the products								
	value of products sourced from members								
	number of producers from which products have been sourced								
<b>4. PATRIMONY</b>									
	value of infrastructure owned								
	value of equipment								
	value of working capital used for sourcing								
<b>5. RECURRENT COSTS</b>									
	office costs								
	paid staff costs								
	organisational costs								
	communication and travel costs								
		<b>CURRENT ESTIMATE</b>				<b>CHANGE IN LAST THREE YEARS</b>			
		<b>(indicate currency Bs or US\$)</b>				<b>RELEVANT (YES/NO)</b>			
							<b>DECLINE</b>	<b>STABLE</b>	<b>INCREASED</b>

# Annex 5 Questionnaire included in the AR-LAT household survey (part J and K)

Complemento de Formulario F.1.	Corresponde a : /	F.3. HOGAR (Complemento de F.1)	
<b>J PERCEPCION DEL HOGAR EN SU CONTEXTO</b> Las áreas de datos en adelante deben ser llenadas por respuesta del Entrevistado.			
<b>COMMUNAL</b>	No responde No sabe. EN TOTAL DESECCION NO EN SI DE ACURSION MUNICI	No responde No sabe. EN TOTAL DESECCION NO EN SI DE ACURSION MUNICI	
1. La comunidad esta bastante unida en comparacion a otras comunidades en la zona.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
2. Existen conflictos entre familias que afectan al buen funcionamiento de la comunidad?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
3. Cuando se necesita mejorar el camino o mejorar la escuela, u otros, todos en la comunidad participan y ayudan?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
4. Cuando un profesor/profesora no llegue a clases por varios dias seguidos, la comunidad actua y suele resolverlo.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
5. Cuando una casa se destruye (trueno o quema) la comunidad apoyara en reconstruirla?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
6. Cuando alguien en la comunidad se enferma y no tiene como pagar los gastos, la comunidad apoya con dinero para conseguir la medicina o el transporte al hospital?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
<b>ORGANIZACION FUNCIONAL</b>			
1. La asociacion/OECA es un instrumento importante para el desarrollo economico de la zona?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
2. La asociacion/OECA esta conformada por un grupo de familias que tiene caracteristicas diferentes al promedio de la comunidad?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
3. Los Socios de la asociacion/OECA tienen mas tierras o capital de trabajo que el promedio en la comunidad?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
4. Los Socios de la asociacion/OECA tienen casas más cómodas que el promedio?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
5. Los Socios de la asociacion/OECA tienen mas influencia politica que el promedio?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
6. La asociacion/OECA les ayuda más que el sindicato o el ayllu?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
7. ¿Algun miembro de su hogar es miembro de la Asociación? (OECA u otra)	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
8. ¿Nadie de su hogar es miembro?, pero alguien tiene interés para hacerse parte de la asociación	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
9. ....	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
10. La asociacion/OECA es un medio que le sirve para acceder a apoyos (ayuda)	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
11. La asociacion/OECA es un medio que le sirve para acceder a mercados	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
12. La asociacion/OECA es un medio que le sirve para la vida social en la comunidad	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
13. La asociacion/OECA es un medio que le sirve para mejorar su produccion	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
14. Conoce al menos dos nombres de dirigentes de la asociacion/OECA?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
15. Los tecnicos que trabajan en la asociacion/OECA tienen parentesco con los dirigentes?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
16. La asociacion/OECA está entre las organizaciones campesinas que mejor funciona en la zona	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
17. Recomendaría a sus vecinos integrarse a la asociacion/OECA?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
<b>SENTIMIENTO INDIVIDUAL</b>			
18. A la hora de comercializar mis productos le vende a quien ofrece mejor precio	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
19. Esta orgulloso de ser productor agrícola	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
20. La situación económica en el país esta mejorando	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
21. En el futuro, sus hijos tendrían oportunidad de hacer dinero en agricultura	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
22. Le gusta experimentar con nuevos cultivos en sus parcelas o tierras	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
23. Le gusta vender sus productos junto con otros en forma organizada	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
24. En general, los compradores de su producción son confiables	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
25. Cuando produce una mayor calidad de producto, le pagan un mayor precio	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
26. En general, las personas a quien compra sus insumos son confiables	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
27. La ley le protege cuando otros tratan de engañar	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
28. Su familia es más pobre que el año pasado	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
29. Han pasado hambre más veces este año que el año pasado?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
30. Sus hijos van a la escuela más días al año que el año pasado?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
31. Han logrado ahorrar más dinero para pagar los gastos médicos este año que el año pasado?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
32. Tienen más diversidad de productos o cultivos que el año pasado?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	





## Summary

The development of value chains has emerged as an important area of donor interventions for poverty reduction in developing countries. Chain performance can be enhanced by policies and projects that support farmer organisations to, for example, increase the scale of operations, improve service provision to producers, develop capacities to comply with (buyer-driven) quality requirements or address the process of value creation and value distribution. Donors such as the World Bank and a large number of development NGOs propose to increase the development support to them in order to strengthen their capacities for responding to market demands. This development support must be able to prove its effectiveness. Impact evaluation is a method used to generate this information. This thesis is about the design of impact evaluations and how research methods can be combined to obtain credible evidence on effectiveness. The study covers various interventions that support smallholders' market access, such as innovation grants, certification schemes, the supply of micro-irrigation technology, and investment subsidies to economic farmer groups. In all these interventions, the impact on smallholder market access was mediated or moderated by farmer organisations.

Chapter 1 distinguishes between two contrasting approaches to impact evaluation design, caricaturised as 'randomistas' and 'realistas'. The 'randomistas' and 'realistas' differ in the kind of evidence that they prioritise in impact evaluation, and the criteria used to judge the rigour or credibility of conclusions. The 'randomistas' focus on the measurement of effects of an intervention and analyse the differences between beneficiaries and non-beneficiaries of the support to verify if the support is a relevant causal factor (Does it work?). 'Realistas', in contrast, highlight the differential effects that an intervention may have in different contexts, and focus especially on the exploration and explanation of causal configurations that define effectiveness (Why does it work, and under what conditions?). This relates to differences in the way that causal relations are being analysed in both archetypical approaches to impact evaluation, different conceptualisations of causality in complex social systems, and differences in the preferred way of expressing this complexity in causal models and detecting patterns in data sets.

Chapter 2 present the results of a systematic review on the effectiveness of innovation grants to smallholders. It maps the studies according to the type of grant system and reflects on causal assumptions in the respective impact pathways: voucher systems, business development matching grants and farmer-driven innovation support funds. The rationale behind these impact pathways is not contested by the studies but the supportive evidence from impact studies is fairly small. This is partly because of methodological limitations to measure key outcomes, especially those of human and social capital.

Chapter 3 argues that there are limits to the accuracy of net-effect estimates on outcomes that are outside the 'span of direct influence'. Baseline data of an impact evaluation in Ivory Coast is used to show that the sample size needed to measure the expected effects on income and yields would need to be higher than feasible for most real-world impact evaluations in certification. We therefore propose to focus more on intermediate outcomes in knowledge on and implementation of good agricultural practices where net-effect estimates are more likely to be measured with sufficient statistical power. Information on this level of intermediate outcomes is also more useful for adjusting the training activities, generally the most important activity in these certification programmes.

Chapter 4 points to three interrelated challenges in impact evaluation of value chain development support. These are the measurement of outcome patterns, attribution of effects in open systems and the generalisation of findings from the particular research context. The main point we stress in this chapter is the role of theory in impact evaluation design and data analysis, and the potential of theory-based evaluation and realist case-studies for presenting evidence in a format that facilitates learning for replicating or scaling of development interventions.

Chapter 5 shows that it is useful for the dynamics of interdisciplinary research design to start with a provisional core methodology to address a specific research question, and identify the main threats to validity. After this reflection, complementary research methods are added, resulting in a creative, interrelated mix of quantitative and qualitative methods. The process is illustrated with the examples of an impact evaluation of micro-irrigation technology supply in Nepal, Zambia and Ethiopia, and an impact evaluation of a Bolivian grant fund FONDOECAS (*Fondo para el Desarrollo de Organizaciones Económicas Campesinas*) that supports investments in processing and collective marketing.

To assess the effectiveness of FONDOECAS grants it was necessary to develop a new tool to measure organisational social capital in collective marketing groups, called Tension Containment Capacity (TCC). The measure is based on the information from semi-structured interviews about the presence and effectiveness of rules and regulations in a group, which are needed to overcome the inherent tendency to break down due to opportunistic behaviour and free-riding. Chapter 6 documents the field test of this measure and shows that the measure is suited for cross-sectional research. The thesis argues that the TCC provides a framework for the analysis of organisational strengthening and an agenda for further research. While this interview process and conceptualisation of the construct Tension Containment Capacity is firmly embedded in qualitative research, the conversion of the information from the interview into a summary sheet, and subsequently into a quantitative proxy- by implication indicator of organisational social capital makes the link with quantitative research, be it configurational comparative or regressional-analytic research. The TCC-score makes it possible to use the level of organisational social capital as an independent variable in regressional-analytic models.

'Randomistas' use regressional analytic methods to detect or verify causality, whereas 'realistas' prefer configurational comparative methods. These differences imply different logics of causal analysis. The regressional analytics label a factor a cause when it covaries with the outcome: 'the more/less of the cause, the more/less of the outcome'. Configurational comparativists, in contrast, label a factor a cause by implication: 'if the cause is absent/present, then the outcome is absent/present'. Qualitative Comparative Analysis (QCA) is a relatively new research approach that uses the method of implication. QCA uses Boolean algebra to identify the most concise (parsimonious) recipe of conditions that are consistently related to success or failure. In Chapter 7, it is demonstrated, using the software applications fsQCA and Kirq, that QCA helped to detect relevant causal patterns in a small data set of 26 observations. Because some of these predictors were single conditions, logistic regression was used to triangulate these findings and increase the validity of the predictors. QCA also identified a peculiar causal configuration: the grants to the best-endowed organisations were consistently unsuccessful, a pattern that was supported by the case study interview reports, in which all these organisations

explained the failure of the grant-supported business plan with reference to the small amount of the grant and resulting under-scaled investments. This combined use of distinct methods of causal analysis in a real-world impact evaluation is quite new in the evaluation literature.

Chapter 8 reports on the FONDOECAS impact evaluation. The initiators of the grant fund intended to develop an effective grant allocation system that could be replicated within other agricultural development programmes. The impact evaluation identified three interlinked assumptions of impact in the intervention logic of FONDOECAS: relevance of the support to collective marketing groups, effectiveness of the grants to strengthen these, and efficiency of the grant allocation system. Each of these assumptions in the intervention logic had a specific research design. A household survey of 1,945 households in 40 municipalities showed that there was majority support for economic farmer organisations among the rural population. Households that participated in economic farmer organisations considered them more supportive than the village organisations. There are prospects for growth, considering the willingness of two-thirds of the farmers to market their products collectively. This survey confirmed that the support of FONDOECAS was relevant. We used the feasibility assessments of the technical committee concerning grant proposals from 150 organisations to assess the efficiency of FONDOECAS' grant allocation system. Though we found a positive correlation between feasibility scores and progress in implementation of the investments, there was no correlation with progress on organisation, production and marketing related issues. This suggested that the efficiency of the technical committee to target grants to the most feasible business plans was low.

We used comparative case studies to assess the effectiveness of the grants in strengthening these organisations. Interviews with board members took place in 2011 and 2013, and time-series data on sales, membership and organisational expenses was collected for the period 2008-2012. Overall, the organisations had become commercially stronger during this period, especially due to higher agricultural prices (coffee, quinoa) and preferential procurement by public nutrition programmes (honey). However, the contribution of FONDOECAS to this growth was modest. Only five out of 29 organisations created market access for members with the grant-supported business plan. FONDOECAS did help to build organisational capacities in only a third of organisations, mainly the capacities needed for quality assurance. FONDOECAS contributed to the capacity to pay organisational costs in almost half of the supported groups. There was no evidence of an improved access to formal credit as a result of the grants. The effectiveness of the grants in strengthening economic farmer organisations was less than initially expected. These findings have wider relevance. Grant funds should differentiate between organisations that source and those that do not source their raw materials from members. Market access for members is only possible with sourcing organisations, and grants to larger organisations are more likely to be successful. Instead, organisations that source their raw material from spot markets are more likely to invest the grant in activities that raise additional group income to pay for organisational expenses. When organisations are already well-endowed, with relatively strong sales, large membership and high patrimony, small grants are less important and are likely to result in under-scaled investments in secondary activities. Access to trade finance may be a more effective strategy to strengthen these organisations.

Based on the experiences described briefly above, Chapter 9 highlights seven principles for creative, credible and appropriate impact evaluation design that help to create synergy be-

tween the ‘randomista’ and ‘realista’ approaches to impact evaluation:

### **1. Anticipate generalisation**

In order to increase the relevance of the findings, the research design should include methods that help to answer the question ‘What works for whom under what conditions?’ Currently the balance in commissioned impact research is too much on the What works? question, and the learning potential of impact evaluations would be enhanced by more attention to the questions For whom? and Under what conditions?, in order to identify causal mechanisms that explain effectiveness.

### **2. Map the intervention logic**

Mapping the intervention logic is a good way to differentiate components in an intervention, in order to focus on those causal process that seem most critical for the effectiveness of this intervention logic and most informative for future intervention design.

### **3. Use theory-based evaluation**

A theory-based impact evaluation collects data to verify whether the theorised causal processes indeed took place and reflects on the role of important influencing factors and necessary context conditions. Not all assumptions in the intervention logic can be a focus of impact research. A decision will have to be made about the key causal assumption on which the research will focus.

### **4. Explore the existing literature**

Impact evaluation needs to feed and use the wider body of knowledge. Past studies on similar interventions and causal processes help to reflect on the key assumptions in the intervention logic and learn about methods and indicators to verify them.

### **5. Define the span of direct influence**

The attribution of effects to a support intervention is only possible within a span of direct influence, and, of course, only when outcomes can be properly measured or observed. Each impact pathway of an intervention will have a different boundary of this span of direct influence. Moreover, some research methods are better able to capture some of these ‘borderline’ outcomes than others.

### **6. Apply the threats to validity check**

Inter-disciplinary research and mixed-methods designs are enhanced by a process in which the main threats to validity related with the core method are identified, and other methods are added to reduce them. To do so, researchers with different methodological traditions come together, which stimulates pro-active and creative thinking, and learning about new methods that are complementary to those in their own preferred tool-kit.

### **7. Combine different logics to detect causal patterns**

The regressional-analytic school uses ‘covariation’ to detect causal patterns in data; the configurational comparative methods use the logic of ‘implication’. The combination of both logics in one and the same research helps to detect patterns in data sets and may increase the validity of causal inferences.

## Samenvatting

Het ontwikkelen van waardeketens is een steeds belangrijke strategie geworden in de ontwikkelings samenwerking. Door het versterken van boerenorganisaties kan de keten efficiënter worden, bijvoorbeeld doordat er schaalvoordelen ontstaan, boeren beter bereikt kunnen worden met de noodzakelijk inputs en dienstverlening, of door te zorgen dat de productie aan de steeds strengere kwaliteitseisen van klanten gaat voldoen. Ook kunnen boerenorganisaties er voor zorgen dat er waarde wordt toegevoegd aan landbouwproducten door deze te verwerken of door een betere prijs te onderhandelen. Donoren, zoals bijvoorbeeld de Wereldbank, en een groot aantal ontwikkelingsorganisaties stellen daarom voor om de ontwikkelingshulp aan boerengroepen te vergroten. Maar deze vorm van ontwikkelingshulp moet dan wel in staat zijn om aan te tonen dat ze effectief is. Impactevaluatie is een methode om informatie te verzamelen om daar iets over te kunnen zeggen. Deze thesis gaat over het ontwerpen van impactevaluaties, en over de manier waarop onderzoeksmethoden gecombineerd kunnen worden zodat ze geloofwaardige informatie over effectiviteit opleveren. Deze studie gaat over verschillende project-interventies waarin geprobeerd wordt om de toegang van kleine boeren tot de markt te vergroten: innovatiesubsidies voor boeren, certificeringsprogramma's, het aanbieden van goedkope micro irrigatie-technologie, en het geven van geld aan boerenorganisaties voor investeringen in productverwerking. In al deze interventies was het de bedoeling om de toegang van de boeren tot de markt te verbeteren, en speelden boerengroepen daarin een belangrijke rol.

Hoofdstuk 1 maakt een onderscheid in twee tegenover elkaar staande benaderingen van impact evaluatie, die ik als karikaturen neerzet als 'randomistas' en 'realistas'. Deze 'randomistas' en 'realistas' verschillen in de soort bewijsvoering die ze het belangrijkste vinden in impact evaluatie, en in de criteria die ze gebruiken om de geloofwaardig daarvan te beoordelen. De 'randomistas' richten zich vooral op het meten van effecten van een interventie door groepen die steun ontvangen te vergelijken met groepen die dat niet krijgen. Zo kan gekeken worden of die hulp inderdaad een relevante rol heeft gespeeld in de veranderingen (Werkt het?). 'Realistas' daarentegen benadrukken de contextafhankelijkheid van de effecten, en richten zich vooral op het zoeken van de combinaties van factoren die bepalen of de ondersteuning effectief is (Waarom werkt het voor wie en onder welke condities?). Bij impact evaluaties gaat het er om de complexe werkelijkheid meetbaar te maken. De beide archetypische evaluatiebenaderingen verschillen in wat ze als causaal verband beschouwen, in hoe ze sociale processen modeleren, en in de manier om patronen te ontdekken in de onderzoeksgegevens.

Hoofdstuk 2 presenteert de resultaten van een systematische literatuurstudie naar de effectiviteit van innovatiesubsidies die direct verstrekt worden aan boeren. Het maakt een overzicht van de studies naargelang het type subsidiefonds, en reflecteert over de vooronderstellingen in de logica achter de manier waarop ze denken effect te hebben: vouchersystemen, cofinancieringssystemen van business plannen en door boeren bestuurde innovatiefondsen. De logica achter deze resultaatketens wordt door de studies niet in twijfel getrokken, maar het ondersteunende bewijs uit de impact studies is behoorlijk mager. Dit komt gedeeltelijk door methodologische beperkingen die er kleven aan het meten van sommige resultaten, vooral wanneer die betrekking hebben op de capaciteiten van mensen en organisaties.

In hoofdstuk 3 benadrukken we dat er een grens is aan het precies kunnen meten van netto-effecten, waar de directe invloed van een interventie ophoudt. Met gegevens uit een referentiestudie in Ivoorkust tonen we aan dat de steekproefgrootte, die nodig zou zijn om de verwachte effecten op inkomens en gewasopbrengsten te meten, hoger is dan haalbaar in de realiteit van impact evaluaties in certificering. We stellen daarom voor om meer te focussen op het meten van effecten op minder ver gelegen resultaatengebieden, met name op het kennisniveau van boeren en op hun toepassing van betere landbouwmethodes. Op die gebieden is het waarschijnlijker dat netto-effecten nog wel gemeten kunnen worden met voldoende statistische sterkte. Bovendien is Informatie over deze tussenliggende resultaten vaak ook nuttiger, bijvoorbeeld voor het bijsturen van de trainingsprogramma's, voor boeren meestal de belangrijkste activiteit van certificeringsprogramma's.

Hoofdstuk 4 gaat over drie met elkaar gerelateerde problemen bij impact evaluaties van ketenontwikkelingsprojecten: het meetbaar maken van resultaten, het toeschrijven van resultaten aan een project-interventie terwijl er veel andere factoren en partijen bij betrokken zijn, en de vraag tot waar het nog mogelijk is om de onderzoeksuitkomsten te generaliseren omdat ze sterk context-afhankelijk kunnen zijn. De kern van het betoog is dat 'theorie' heel belangrijk is zowel voor het ontwerpen van impact evaluaties, maar ook bij het analyseren van de data. We propageren het gebruik van theory-based evaluation en een op het kritisch realisme gebaseerde manier van case studies opschrijven, zodat het makkelijker wordt om van impact evaluaties iets te leren, en te beoordelen of een bepaalde project-interventie ook elders, of op een grotere schaal, uitgerold kan worden.

In Hoofdstuk 5 toont ik aan dat het nuttig is om bij het ontwerpen van een interdisciplinair onderzoek eerst één kernmethodologie als uitgangspunt te nemen om een bepaalde onderzoeksvraag te beantwoorden, en daarvan de mogelijke validiteitsproblemen (threats to validity) op een rijtje te zetten. Op basis van deze reflectie worden er dan complementaire onderzoeksmethodes toegevoegd, wat er dan meestal op uitdraait dat er een creatieve, aan elkaar gerelateerde mix van kwantitatief en kwalitatief onderzoek ontstaat. Dit ontwerpproces illustreer ik met een impact evaluatie van een project rond micro-irrigatie technologie in Nepal, Zambia en Ethiopia, en een impact evaluatie van een Boliviaans subsidiefonds, FONDOECAS (Fondo para el Desarrollo de Organizaciones Económicas Campesinas), voor investeringen in verwerking en gezamenlijke vermarkting door boerenorganisaties

Om de effectiviteit van de door FONDOECAS verleende subsidies te kunnen bepalen was het nodig om een nieuw instrument te ontwikkelen dat het organisatorische vermogen van dit soort groepen kan meten, dat ik Tension Containment Capacity (TCC) heb genoemd. Het instrument is gebaseerd op semigestructureerd interviews waarin wordt gevraagd naar de aanwezigheid en effectiviteit van afspraken en regels op de gebieden waarbij dit soort groepen neigen om uit elkaar te vallen door opportunistisch gedrag en free-riding. Hoofdstuk 6 beschrijft hoe dit instrument in het echt is uitgeteerd en toont aan dat het inderdaad bruikbaar is in vergelijkend onderzoek. In het laatste hoofdstuk stel ik dat TCC een conceptueel raamwerk biedt voor de analyse van organisatieversterking en aangrijpingspunten geeft voor onderzoeksagenda. Hoewel het interviewproces en de conceptualisering van Tension Containment Capacity sterk geworteld is in het kwalitatieve onderzoek, maakt het samenvatten van de informatie en het berekenen van een TCC-score het juist geschikt voor kwantitatief

onderzoek. De TCC-score maakt het bijvoorbeeld mogelijk om de sterkte van organisaties te gebruiken als een van de factoren in een regressie-analytisch model, om zo te kijken of dit de verschillen in effectiviteit van project-interventies misschien deels kan verklaren .

‘Randomistas’ gebruiken vooral regressie-analytische methoden om causale verbanden te ontdekken of te verifiëren, terwijl de ‘realistas’ vooral de vergelijkend configurationele methodes prefereren. Dit komt deels door verschillende logica’s van causale analyse. De regressie-analysici beschouwen iets als een oorzaak wanneer deze samenhangt (covarieert) met de uitkomst: ‘hoe meer/minder van de oorzaak, des te meer/minder van de uitkomst’. Andersom, noemen de mensen, die vergelijkend configurationeel onderzoek doen, iets een oorzaak wanneer deze impliciet aanwezig is: ‘als de oorzaak aan/afwezig is, dan is de uitkomst aan/afwezig’. Qualitative Comparative Analysis (QCA) is een relatief nieuwe onderzoeksbenadering die werkt met dit principe van implicatie. QCA gebruikt Booleaanse algebra om de bondigste formule te zoeken voor de combinatie van factoren die gerelateerd zijn aan succes of falen. In hoofdstuk 7 toon ik aan, met gebruikmaking van de software applicaties fsQCA en Kirq, dat QCA inderdaad hielp om bepaalde relevante causale verbanden te vinden in een relatief kleine data set van 26 observaties. Sommige van die verbanden kon ik met logistische regressies verifiëren en daardoor sterker onderbouwen. QCA kon echter ook nog een andere, specifieke combinatie van factoren vinden die het resultaat van de subsidies bleek te bepalen. Het bleek namelijk dat subsidies aan de meest-getalenteerde organisaties, met de meeste omzet, kapitaal en ledental, desondanks bijna altijd onsuccesvol bleken te zijn geweest. Dit verrassende patroon in de data-analyse bleek inderdaad te kloppen met de informatie in de interviewrapporten, waarbij al deze organisaties als één van de belangrijkste redenen voor de mislukking van het business plan aangaven dat het subsidiebedrag voor hen te laag was en dat daardoor de schaal van de investering die ze hadden gedaan te klein was om economisch interessant te worden. Deze combinatie van analysemethodes, met echte gegevens van een reëel bestaande project, is nieuw in de evaluatie literatuur.

Hoofdstuk 8 beschrijft de resultaten van de impact evaluatie van het subsidiefonds FONDOECAS. De initiatiefnemers wilde laten zien dat FONDOECAS een effectief systeem van subsidieverdeling is, dat ook in andere landbouwontwikkelingsprogramma’s toegepast zou kunnen worden. De impact evaluatie identificeerde drie vooronderstellingen in de resultaatketen van FONDOECAS, namelijk dat steun aan dit soort economische boerengroepen relevant is, dat de subsidies er inderdaad in slagen om deze groepen te versterken, en dat het subsidieverdelingssysteem efficiënt is ingericht. Om na te gaan of deze drie vooronderstellingen inderdaad klopten, ontwierp ik voor elk een specifiek onderzoeksontwerp. Met een enquête bij 1.945 huishoudens in 40 gemeentes toon ik aan dat een meerderheid van de rurale bevolking deze economische boerengroepen belangrijk vindt. De huishoudens die erin deelnamen vonden ze zelfs nuttiger dan de traditionele dorpsorganisaties. Er lijken ook groeimogelijkheden, gezien de bereidheid van twee-derde van de boeren die in de enquête aangaven dat ze hun producten best gezamenlijk zouden willen verkopen. Dit onderzoek onderschreef dus de vooronderstelling dat de steun van FONDOECAS relevant was. Om een idee te krijgen van de efficiëntie van het subsidieverleningssysteem gebruikten we de administratieve data van FONDOECAS met de scores van de beoordelingscommissie over de verwachte uitvoerbaarheid van 150 ingediende plannen. Hoewel er een positief verband was tussen de gegeven score en de voortgang van de investeringen, vond ik geen verband met de verwachte voortgang op

het gebied van de interne organisatie rond die investeringen, en ook niet bij de voortgang in productie en vermarkting. Dit suggereert dat de beoordelingscommissie niet erg adequaat was in het toedelen van subsidies aan de organisaties met de best uitvoerbare plannen.

Ik gebruikte vergelijkende casestudies om te onderzoeken of de subsidies effectief waren voor de versterking van de organisaties. In 2011 en 2012 hebben lokale onderzoekers interviews gehouden met de bestuursleden van een veertigtal organisaties, waarbij ook kengetallen verzameld werden over de omzet, ledenaantal en organisatiekosten in de periode 2008-2013. In het algemeen bleken de organisatie commercieel gezien sterk te zijn gegroeid gedurende deze periode, maar vooral vanwege de hogere landbouwprijzen (koffie, quinoa) en het voorkeursbeleid bij aanbesteding van publieke voedingsprogramma's (honing). De bijdrage van FONDOECAS aan deze groei lijkt echter niet zo groot. Alleen bij 5 van de 29 organisaties was er sprake van een positief effect door de subsidie op de marktmogelijkheden van leden. In een derde van de organisaties had FONDOECAS bijgedragen aan organisatieversterking, met name die rond kwaliteitswaarborging. En in bijna de helft van de gevallen hielp de subsidie van FONDOECAS om inkomsten te genereren waarvan de organisatiekosten betaald konden worden. Ik zag weinig tekenen dat een organisatie meer toegang tot bankkrediet had gekregen als gevolg van de subsidie. Al met al was de effectiviteit van de subsidies in het versterken van de economische boerenorganisaties minder dan aanvankelijk verwacht. Deze onderzoeksbevindingen zijn niet alleen relevant voor FONDOECAS, maar zijn ook relevant voor andere subsidiefondsen. Dit soort subsidiefondsen zouden er goed aan doen om een onderscheid te maken tussen boerengroepen die hun producten van de leden kopen, en groepen die dat niet doen. Als het doel is om boeren een betere toegang tot markten te geven dan is dat natuurlijk alleen mogelijk bij organisaties die de producten van leden kopen. En subsidies aan de grotere organisaties hebben daarbij ook meer kans op succes dan aan de kleine. Anderzijds lijken de groepen die hun grondstoffen op de lokale markt kopen wel weer beter in staat om met de subsidie geld te verdienen om hun organisatiekosten te betalen. Bovendien zijn dit soort subsidies minder belangrijk voor organisaties die al in behoorlijk goede doen zijn, en leiden ze veelal tot ondermaatse investeringen in minder belangrijke bedrijfsactiviteiten. Het bieden van een betere toegang tot handelskrediet is waarschijnlijk een meer effectieve manier om dit type organisaties te versterken.

Op basis van de ervaringen opgedaan in de verschillende onderzoeken die hierboven kort beschreven staan, schetst Hoofdstuk 9 zeven principes om tot een creatief, geloofwaardig en passend impact evaluatieontwerp te komen, waarin er synergie kan optreden tussen de twee geschetste impact evaluatie benaderingen van 'randomistas' en 'realistas':

### **1. Anticipeer op het trekken van meer algemene conclusies**

Om de relevantie van de onderzoeksuitkomsten te vergroten, moet een onderzoeksontwerp methodes opnemen die kunnen helpen bij het beantwoorden van de vraag 'Wat werkt voor wie en onder welke condities?' Momenteel ligt de balans bij gefinancierde onderzoeksopdrachten nog teveel op de 'Wat werkt?' vraag, en de mogelijkheden om van impact evaluaties te leren zou worden vergroot wanneer er meer aandacht zou zijn aan de vragen 'Voor wie?' en 'Onder welke condities?', om daarbij causale mechanismes te vinden die kunnen verklaren waarom project-interventies wel of niet slagen.



## **2. Schets de resultaatketen**

Het schetsen van een resultaatketen is een goede manier om verschillende componenten in een interventie te onderscheiden, om daarmee in te zoomen op de processen die het meest kritiek zijn voor de effectiviteit van de interventie en/of het meest informatief voor het ontwikkelen van nieuwe interventies.

## **3. Gebruik theory-based evaluation**

In een op theorie gebaseerde evaluatie wordt er data verzameld om te beoordelen of de causale processen zich inderdaad voltrekken zoals verwacht in de theorie, en er wordt gekeken wat de invloed is van context en andere factoren. Niet alle vooronderstellingen in een resultaatketen kunnen natuurlijk dezelfde aandacht krijgen in een impact onderzoek. Er moet daarom een besluit genomen worden over de meeste cruciale causale vooronderstellingen waar het onderzoek zich op kan gaan richten.

## **4. Lees wat er al over het thema geschreven is**

Impact evaluatie moet de bestaande kennis-pool voeden en gebruiken. Eerdere studies over soortgelijke interventies of vergelijkbare causale processen kunnen helpen bij het reflecteren over vooronderstellingen in de resultaatketen, en ook om te leren welke methodes en indicatoren er al gebruikt zijn om deze kritisch tegen het licht te houden.

## **5. Bepaal waar de directe invloedssfeer van een project-interventie ophoudt**

De toerekening van effecten aan een bepaalde interventie is alleen mogelijk binnen een directe invloedssfeer, en, vanzelfsprekend, alleen wanneer die goed te meten of te observeren zijn. Iedere component van een interventie zal een andere grens hebben waar deze directe invloedssfeer ophoudt. Natuurlijk zijn sommige onderzoeksmethodes beter in staat om uitkomsten op dat grensvlak te vangen dan andere.

## **6. Anticipeer op mogelijke validiteitsproblemen**

Interdisciplinair onderzoek en mixed-methods onderzoek worden gestimuleerd door een proces waarbij eerst de belangrijkste validiteitsproblemen van een kern-onderzoeksmethode worden onderkent, waarna deze met andere methodes zo goed mogelijk kunnen worden afgedekt. Het werkt goed als onderzoekers uit verschillende methodologische tradities bij elkaar komen om dat te doen. Dit stimuleert een proactieve en creatief denkproces, waarbij mensen nieuwe methodes leren die hun bestaande methodologische gereedschapskist aanvullen.

## **7. Combineer verschillende logica's om patronen in data te ontdekken**

De regressie-analytische school gebruikt 'covariantie' om oorzakelijk verbanden in data te ontdekken; de configuratieve vergelijkende methodes gebruiken de logica van 'impliceren'. De combinatie van beide logica's in één en hetzelfde onderzoek helpt om patronen te ontdekken in data sets, en kan de validiteit van de gevonden causale verbanden vergroten.

## Resumen

El desarrollo de cadenas de valor es una estrategia cada vez más importante en la lucha contra la pobreza. Estas cadenas pueden mejorar su eficiencia si se fortalecen las organizaciones de productores que participan en ellas. Así, se pueden lograr economías de escala, una mayor vinculación de pequeños productores con los mercados de insumos y servicios o una producción que cumpla mejor con las normas de calidad y sanidad requeridas por compradores y consumidores. Además, las organizaciones de productores pueden agregar valor a los productos agrícolas mediante el procesamiento de los mismos o a través de la negociación de un mejor precio en el mercado.

Las organizaciones internacionales de desarrollo, como el Banco Mundial y otras, proponen aumentar el apoyo a este sector de productores organizados. Sin embargo, esta forma de apoyo debe demostrar su efectividad. Las evaluaciones de impacto son maneras de recabar la información necesaria para sacar conclusiones sobre su efectividad. La presente tesis trata sobre el diseño de evaluaciones de impacto y sobre las maneras de combinar métodos de investigación para que suministren información creíble. Abarca distintos proyectos de desarrollo donde se trató de mejorar el acceso de los pequeños productores al mercado: subsidios para estimular la innovación, programas de certificación, tecnología de micro-riego, subsidios para procesamiento y comercialización en grupo. En todos estos proyectos, las organizaciones de productores jugaban un papel importante.

En el Capítulo 1 se contraponen dos enfoques de cómo hacer evaluación de impacto, caricaturizados aquí como 'randomistas' y 'realistas'. Estos enfoques difieren en el tipo de pruebas que buscan para respaldar conclusiones sobre impacto y en los criterios que utilizan para juzgar la validez de estas conclusiones. Los 'randomistas' se enfocan más en la medición de los efectos netos de una intervención y usan las diferencias existentes entre los beneficiarios y no beneficiarios del apoyo para determinar si un proyecto realmente tiene efecto (¿funciona?). Los "realistas", en cambio, se enfocan principalmente en la influencia que ejerce el contexto sobre los efectos y sobre todo tratan de detectar y explicar las configuraciones de factores que determinan si el apoyo es efectivo (¿Por qué funciona? ¿Para quiénes? ¿Bajo qué condiciones?). Ambos arquetípicos enfoques de evaluación se diferencian en lo que consideran como relación causal, en cómo modelan los procesos sociales y en la manera de descubrir patrones en las bases de datos.

El Capítulo 2 presenta los resultados de una revisión sistemática de los estudios de efectividad referentes a subsidios de innovación para pequeños productores. Se hace un mapeo de estudios respecto del tipo de fondo y su respectiva lógica para generar impacto: fondos que otorgan vales, los que cofinancian planes de negocios y fondos de innovación cogestionados por productores. La lógica detrás de estos tres tipos de apoyo no está en duda, pero la cantidad de estudios de impacto que presentan pruebas de ellos es bastante limitada. En parte, esto se explica por las limitaciones metodológicas para medir resultados, especialmente los resultados en capital humano y capital social.

El Capítulo 3 argumenta que hay un límite para la estimación precisa de efectos netos en resultados que están fuera del límite de la influencia directa de una intervención. Se usa un

estudio de línea de base en Costa de Marfil para mostrar que en la mayor parte de las evaluaciones de impacto de programas de certificación, el tamaño de la muestra que se necesitaría para medir los efectos esperados con suficiente poder estadístico, en realidad resulta ser más alto que viable. Proponemos dar más atención a los resultados inmediatos e intermedios, al conocimiento sobre buenas prácticas agrícolas y su aplicación, donde es más probable que se puedan captar los efectos con suficiente poder estadístico. La información sobre resultados intermedios es también más útil para ajustar los programas de extensión, generalmente la actividad más importante de estos programas de certificación.

El Capítulo 4 trata sobre tres desafíos, relacionados entre sí, que se presentan en las evaluaciones de impacto. Estos son: cómo medir los resultados, cómo atribuir efectos en sistemas abiertos y cómo generalizar las conclusiones de una investigación en un contexto particular. El punto más importante que subrayamos en este capítulo es la importancia de la teoría para el diseño de evaluaciones de impacto y para el análisis de datos. También se indica el potencial del enfoque de la ‘evaluación basada en la teoría’ y del método de hacer estudios de caso comparativos para presentar las pruebas de efectividad de una forma que facilite el aprendizaje sobre una posible replicación o un aumento de la escala de un proyecto.

El Capítulo 5 indica que en la dinámica de diseño interdisciplinario es útil empezar con la definición de un método principal para abordar la pregunta de investigación e identificar las amenazas a la validez más relevantes, relacionadas con este método. Después de esta reflexión, se añaden métodos de investigación complementarios, lo que resulta en una mezcla creativa de métodos cuantitativos y cualitativos. Se presentan dos ejemplos de diseños de investigaciones de impacto, uno para un proyecto de micro-riego en Nepal, Zambia y Etiopía y otro para la evaluación de impacto del fondo de subsidios FONDOECAS (*Fondo para el Desarrollo de Organizaciones Económicas Campesinas*), el cual apoya inversiones en procesamiento y comercialización colectiva en Bolivia.

Para medir la efectividad de FONDOECAS fue necesario desarrollar un nuevo método para medir el capital social organizativo en las organizaciones económicas campesinas, lo que se ha llamado ‘Capacidad de Contención de Tensiones’ (TCC, por sus siglas en inglés). Esta medida se calcula sobre la base de la información recabada mediante entrevistas semiestructuradas, donde se pregunta sobre la presencia y efectividad de las reglas internas del grupo para superar algunas tendencias a la desintegración que son inherentes a la acción colectiva. El Capítulo 6 describe los resultados de la prueba de campo y muestra que esta medida es apropiada para comparaciones transversales. La tesis sostiene que la TCC ofrece un marco conceptual para el análisis del fortalecimiento organizacional y para la elaboración de una agenda de investigación. Si bien el proceso de recabar información mediante entrevistas y el marco conceptual de la TCC están fuertemente enraizados en la investigación cualitativa; el resumen de las informaciones y el cálculo de un puntaje TCC, hacen que estos instrumentos sean justamente más apropiados para una investigación cuantitativa. El puntaje de la TCC hace posible, por ejemplo, utilizar las fortalezas de las organizaciones como uno de los factores en un modelo analítico de regresión, para ver si de esta manera es posible aclarar las diferencias de la efectividad de los proyectos de intervención.

Los ‘Randomistas’ usan regresiones econométricas para detectar y verificar causalidad, mientras que los ‘realistas’ prefieren métodos de análisis configuracionales comparativos. Esto implica diferentes lógicas de análisis causal. Los ‘randomistas’ tildan un factor como causa, cuando éste tiene covarianza con el efecto: ‘a mayor/menor causa, mayor/menor efecto’. Los que usan métodos configuracionales comparativos, al contrario, tildan un factor como causa por implicación: ‘cuando la causa está presente/ausente, el efecto está presente/ausente’. El Análisis Comparativo Cualitativo (QCA, por sus siglas en inglés) es un método de análisis relativamente nuevo que aplica la lógica de implicación. El QCA aplica algebra Booleana para identificar la fórmula de condiciones, escrita de manera más corta, que está relacionada al éxito o al fracaso. En el Capítulo 7, se demuestra, utilizando las aplicaciones fsQCA y Kirq, que el QCA efectivamente encontró factores causales relevantes en una muestra pequeña de 26 observaciones. Algunos de estos factores, predictores de efectividad, resultaron ser condiciones únicas, por lo cual fue posible confirmarlos mediante regresión logística. Adicionalmente, el QCA identificó una configuración de causas: los subsidios a las organizaciones mejor dotadas – mayor patrimonio, más socios y volumen de venta alto - resultaron consistentemente infructuosos, un patrón que fue comprobado en las entrevistas, en las que estas organizaciones indicaron como causa de su falta de éxito, el reducido monto del subsidio recibido y la escala pequeña de las inversiones realizadas. Esta combinación de métodos de análisis de causalidad con datos reales de una evaluación de impacto es bastante novedosa en la literatura académica.

El Capítulo 8 presenta los resultados de la evaluación de impacto de FONDOECAS. Los iniciadores de este fondo querían desarrollar un sistema de asignación de subsidios que pudiera ser replicado en otros programas de desarrollo agrícola. La evaluación de impacto identificaba tres suposiciones en la lógica de impacto de FONDOECAS: la relevancia del apoyo a las organizaciones económicas campesinas, la efectividad de los subsidios para fortalecerlas y la eficiencia del sistema de asignación. Cada una de estas suposiciones en la lógica de impacto tenía su propio diseño de investigación. Una encuesta a 1.945 hogares rurales mostró que la mayoría apoyaba a este tipo de organizaciones económicas campesinas. Los hogares con algún socio en estas organizaciones consideraban que éstas les brindaban más apoyo que la organización comunal. Existen amplias posibilidades para que este sector pueda crecer, considerando que dos tercios de los productores manifestaron su voluntad de vender de forma colectiva. La encuesta confirmaba la suposición de que el apoyo de FONDOECAS es relevante. Para evaluar la eficiencia del sistema de asignación de subsidios se utilizaron las evaluaciones de viabilidad del comité técnico referentes a los planes de negocios de 150 organizaciones. Aunque se detectó una correlación positiva entre el puntaje que daban a la viabilidad del plan de negocios y el progreso registrado en su implementación, no existía una correlación con el progreso en asuntos organizativos, productivos y de comercialización. Esto sugiere que el comité de asignación de subsidios no resultó eficiente a la hora de juzgar qué planes de negocios serían los más viables.

Se utilizó el método de estudio de casos comparativos para averiguar la efectividad de los subsidios en el fortalecimiento de estas organizaciones. En 2011 y 2013 se llevaron a cabo entrevistas con dirigentes que aportaron datos sobre sus ventas, membresía y gastos organizativos durante el período 2008-2012. En general, las organizaciones se habían fortalecido comercialmente durante este período, especialmente debido a los altos precios agrícolas (café, quinua)

y a las políticas preferenciales de compras estatales para programas nutricionales (miel). Sin embargo, la contribución de FONDOECAS a este crecimiento fue módica. Sólo cinco de las 29 organizaciones realizaron un mayor acceso a mercados para sus socios con los planes de negocios financiados. FONDOECAS logró el fortalecimiento de capacidades organizativas en sólo un tercio de ellas, sobretodo en cuanto a su capacidad de control de calidad. En casi la mitad de las organizaciones beneficiadas se logró mejorar la capacidad de pagar gastos organizativos. No había indicios de que con el subsidio se hubiera mejorado el acceso al crédito formal. Así, la efectividad del fondo en fortalecer a las organizaciones económicas campesinas fue menor que lo esperado al inicio. Estos resultados de la evaluación tienen una relevancia más amplia. Los fondos de subsidios deberían diferenciar bien entre organizaciones que compran productos de socios y organizaciones que no lo hacen. Es obvio que para los socios, el acceso al mercado, sólo es posible cuando las organizaciones adquieren sus materias primas de ellos, y hay más chance de lograr este objetivo cuando las organizaciones apoyadas son más grandes. Al contrario, las organizaciones que compran su materia prima en el mercado tienden a ser más exitosas en generar ingresos adicionales para pagar gastos organizativos. Además, cuando las organizaciones ya están bien dotadas, con ventas relativamente altas, membresía grande y un buen patrimonio, los subsidios pequeños son menos importantes y tienden a resultar en inversiones demasiado pequeñas en actividades secundarias. Facilitar el acceso a capital de trabajo podría ser una estrategia más efectiva para fortalecer estas organizaciones.

Sobre la base de las experiencias descritas arriba, el Capítulo 9 resalta siete principios para diseñar creativamente evaluaciones de impacto creíbles y apropiadas, que ayuden a crear sinergia entre los enfoques ‘randomista’ y ‘realista’.

### **1. Prepárese para poder generalizar conclusiones**

Para aumentar la relevancia de los resultados de una evaluación, el diseño metodológico debe incluir métodos que ayuden a responder las preguntas ‘¿Qué funciona?’, ‘¿Para quiénes?’ y ‘¿Bajo qué condiciones?’. Actualmente, los que encargan evaluaciones de impacto están demasiado inclinados a la pregunta ‘¿Qué funciona?’. Además, hay mayores posibilidades de aprender de evaluaciones de impacto cuando éstas dan más atención a las preguntas ‘¿Para quiénes?’ y ‘¿Bajo qué condiciones?’, y así encontrar factores y mecanismos que expliquen la efectividad de un proyecto.

### **2. Haga un mapeo de la lógica de impacto**

El mapeo de la lógica de impacto es una manera útil para diferenciar entre componentes de un proyecto para luego enfocar la evaluación de impacto en los procesos causales más críticos para la efectividad de los proyectos y/o los más informativos para el diseño de nuevos proyectos.

### **3. Utilice una evaluación basada en la teoría**

La evaluación basada en la teoría busca datos para verificar si el proceso efectivamente se desarrolló como se esperaba en teoría y refleja la importancia de factores de influencia y condiciones del contexto que necesariamente deben estar presentes. No se pueden investigar todos los supuestos de una lógica de impacto. Se debe tomar una decisión sobre las suposiciones más importantes sobre las que basarse en la investigación.

#### **4. Examine estudios existentes**

La evaluación de impacto necesita nutrir y nutrirse de un cuerpo de evidencia más amplia. Estudios realizados anteriormente sobre proyectos o procesos similares ayudan a esclarecer las suposiciones en la lógica de impacto y pueden informar sobre métodos e indicadores para verificarlas.

#### **5. Establezca la esfera de influencia directa**

La atribución de efectos a un proyecto de apoyo se puede hacer solamente dentro de su esfera de influencia directa y, evidentemente, sólo cuando estos efectos pueden ser observados o medidos. Cada componente de la lógica de impacto tendrá un límite diferente en esta esfera de influencia directa. Además, algunos métodos de investigación pueden medir mejor que otros los efectos que están en esta zona límite de influencia.

#### **6. Haga un chequeo de las posibles amenazas a la validez**

La investigación interdisciplinaria y la de métodos múltiples están incentivadas por un proceso donde primero se identifican las mayores amenazas a la validez relacionadas con un método principal, después de lo cual se agregan otros métodos para reducir estas amenazas. Investigadores con diferentes enfoques disciplinarios se juntan para hacer este chequeo, lo que estimula una reflexión proactiva y crítica, de la que se aprenden métodos nuevos que son complementarios a los que cada uno normalmente utiliza.

#### **7. Combine las distintas lógicas para detectar patrones en bases de datos**

El enfoque de análisis econométrico utiliza la lógica de 'covarianza' para detectar relaciones causales; el enfoque configuracional comparativo utiliza la lógica de 'implicación'. La combinación de ambas lógicas en un solo diseño de investigación puede ayudar a detectar patrones en bases de datos y puede mejorar la validez de las inferencias.

## About the author

Giel Ton (1963) graduated from Wageningen University as an agricultural economist. Since 2006, he has been working in the Netherlands for the research institute LEI Wageningen UR, as a senior researcher on projects relating to business training, contract farming, export promotion and certification schemes. In the academic field, he has published on methodologies of impact evaluations in agricultural value chains and private-sector development. He favours mixed methods research designs, where in-depth qualitative interviews are combined with survey data collection. He has a special interest in case-based comparative analysis, in which regression-analytic and configurational comparative methods are combined.



At the present time, Giel is coordinating the programme Pioneering Real-time Monitoring and Evaluation of Small and Medium Enterprises ([www.primepartnership.nl](http://www.primepartnership.nl)), as well as a collaborative research programme with farmer organisations in developing countries, Empowering Smallholder Farmers in Markets ([www.esfm.org](http://www.esfm.org)).

Before joining Wageningen UR, Giel worked for Friends of the Earth Netherlands as an advocacy officer on timber certification. Between 1998 and 2004, he worked in Bolivia for the Bolivian platform of economic farmer organisations (CIOEC), as a policy officer on sector policies, business regulation, and international trade. In 1997, he worked for the Dutch workers' union FNV on the Westland Reconstruction Plan. In the period 1989-1996, he worked in Condega, Nicaragua, for the Belgian development organisation COOPIBO on participatory technology development and extension; he set up a credit facility for land acquisitions and a collective storage facility. The Condega project resulted from a crowdfunded eight-months identification study by the TLOD Foundation (Technology, Labour Organisation and Development), an organisation which he co-founded. In 1991, he worked for Wageningen University to develop the interdisciplinary course entitled Development Problems, and, in 1987, he co-organised the course Political Economy of Agriculture in Development Countries.

The research described in this thesis was financially supported by the Netherlands' Ministry of Economic Affairs (KB-11-004 and BO-o-010-129), the Australian Agency for International Development (AusAID ABN 62 921 558838) and the Dutch Interchurch Development Organizations, ICCO.

Financial support from LEI Wageningen UR and Wageningen University for printing this thesis is gratefully acknowledged.

### **Photographs**

Shutterstock (p.1); iStockphoto (p.25, 65); Enrique Uribe (p.51); IDE (p.85); FONDOECAS (p.107, 167, 137, 241); Dreamstime (p.205)

### **Cover art**

Jeanie van der Heiden, 2012, *Detecting Patterns*, Mixed media

### **Layout and design**

Mariëtte Boomgaard, Ocelot Ontwerp

### **Printed by**

Uitgeverij BOXPress || Proefschriftmaken.nl



