

Exploring issues of rigour and utility in Fairtrade impact assessment

VALERIE NELSON and ADRIENNE MARTIN

This article explores the evolution of Fairtrade impact assessment, which reflects the wider context of international development evaluation practice and debates. Appropriate designs and methods in evaluation are hotly contested, ultimately reflecting different development philosophies and values. Earlier Fairtrade impact studies were primarily case studies involving qualitative methods. As Fairtrade has grown and scrutiny from different stakeholders has increased, there has been increased demand for more rigour and criticism of studies that do not include a 'credible' counterfactual. More recently, there have been increasing numbers of impact evaluation studies using mixed designs as well as mixed methods. But challenges remain as to how to balance utility and rigour in Fairtrade impact assessment, because there are trade-offs in terms of skill and resource requirements and in relation to ethical issues. Yet all sustainability standards are being asked to both demonstrate impact and to inform impact. Achieving utility not only at higher levels of organizations in Fairtrade, but also for producers at the local level is a significant challenge, when 'credibility' in impact assessment is judged in some quarters as being the same as using counterfactual logics. In many cases the construction of a counterfactual is very difficult if not impossible. In this paper we seek to provide some practical suggestions for improving both rigour and utility.

Keywords: Fairtrade, impact assessment, sustainability standards, evaluation, fair trade

FAIRTRADE HAS GROWN RAPIDLY in recent decades and this has led to increased scrutiny of its impact by researchers, donors, and the press. This paper explores how Fairtrade impact assessment has evolved. Firstly we provide a brief overview of the evolution of monitoring and evaluation in international development more generally and then we chart the development of impact evaluation and learning in Fairtrade. We identify some of the specific challenges posed for impact assessment for this particular standard system and discuss the ways in which scholars and practitioners are seeking to improve rigour and/or utility – both of which have been lacking in the past to varying degrees in many studies. There are difficulties in achieving both improved rigour and utility simultaneously, because there are often trade-offs involved. However, we seek to show some practical ways of doing this. This article draws on secondary literature and on the authors' own extensive experience of impact assessment generally and specifically in Fairtrade, and sustainability standards.

Valerie Nelson (v.j.nelson@gre.ac.uk) is Principal Researcher and Adrienne Martin is Head of Programme Development at the Natural Resources Institute, University of Greenwich, UK.

© Practical Action Publishing, 2014, www.practicalactionpublishing.org
<http://dx.doi.org/10.3362/2046-1887.2014.003> ISSN: 2046-1879 (print) 2046-1887 (online)

The evolution of impact evaluation in the wider international development context

Fairtrade impact assessment reflects overall practice in evaluation in international development. Between the 1950s and 1970s there was little study of development impact and effectiveness: 'the (assumed) need for aid was seen as a sufficient basis for providing it' (Riddell, 2009 cited in Ramalingam, 2011: 1). Early impact assessments were *ex ante* economic, social, environmental impact assessments conducted *before* a project to scope the potential impacts of an intervention to inform funding decisions. *Ex post* impact assessments emerged in time, with a time lag of several years after project end (Roche, 1999). During the 1970s and 1980s project planning tools emerged, such as logical frameworks. Logical frameworks present the main elements of a project, the links between them, and notes risks and assumptions, with widespread uptake in the 1990s, driven by donors, focusing on results and delivery. This approach possibly focused attention on project deliverables, potentially diverting attention from social change processes (Edward and Tallontire, 2009).

During the 1980s and 1990s participatory approaches emerged in development, including in impact assessment (for example see Guijt et al., 1998; Estrella et al., 2000). Interpretations of participation have varied from approaches which support appreciative enquiry and a learning process guided by participants' own interests and decisions to approaches which rely only on consultation of beneficiaries and which tend to conflate a participatory *process* with specific participatory rural appraisal (PRA) techniques. While there was a flowering of participatory approaches and ideas, many multilateral agencies stopped short of adopting 'extended' participation and instead allowed only 'limited participation' in impact assessment (Roche, 1999: 19). The values of participatory development infer that a 'pluralist, evolutionary and iterative' approach to evaluation is important. They can be rigorous and include both qualitative and quantitative techniques and, crucially, they give greater voice to those affected by a project and much more weight to their experiences when compared with conventional methods (Chambers, 2009). The ideas and values of participatory development have had a major impact on international development, including in evaluation. However, the gains achieved in the 1990s and 2000s by proponents are also under pressure from the new focus on evidence, certain interpretations of rigour, and impact evaluation which we explain below.

During the 1990s impact assessment became more systematic, with a greater focus on outcomes and consequences of a project. There were increased efforts to define and measure impact and lots of activities and debate, but still a lack of progress in understanding impacts and no overarching evidence to provide a clear steer to policy-makers on what works (Riddell, 2007; Ramalingam, 2011). Donor projects were, for the first time, assessed against their logframe outputs and outcomes. The increased pressure on NGOs to demonstrate results and impact led to exaggerations of achievements by some and overblown criticisms by opponents (Roche, 1999). A lack of professional norms and standards in the sector, a growing demand for high profile and press coverage to raise funds, and poor institutional learning

systems and weak accountability mechanisms led to a widening gap between agency rhetoric and the realities of what had been achieved. This in turn contributed to growing scepticism of the value of aid and exposed NGOs to public criticism and the odd polemical attack (Roche, 1999). Demand for greater accountability and learning in donor-led interventions was also growing in this period, with some innovations such as outcome-oriented approaches. Outcome Mapping, developed by the International Development Research Centre (Earl et al., 2001), for example, tracks changes in knowledge, attitudes, and practices, rather than more traditional impact indicators such as income and productivity in recognition of complexity and to promote learning.

In the 1990s the scientific realism school of evaluation developed. Pawson and Tilley (1997) of the scientific realist school argued that experimental and quasi-experimental methods never reach expectations and they critique 'the epistemological assumptions about causation and their lack of fit with the nature of social programs' (Pawson and Tilley, 1997: 30). Ramalingam (2011: 1) concurs that initially ambitious studies have often had to be 'scaled back, narrowed in scope and made "more realistic", in the face of the complex realities of development efforts. As a result, the arguments for and against the effectiveness of aid policies and practices remained patchy, partial and inconclusive'. By focusing on the mechanisms by which an intervention seeks to effect change (articulated in a theory of change) a more realistic understanding of impact can be achieved (Pawson and Tilley, 1997). Recently, Eyben (2013) has cautioned against the use of theories of change in a mechanistic way and merely replicating linear logical framework assumptions of how change happens. With theories of change and theory-based evaluation currently in vogue, it is important to remember that these are just tools and they need to be used in a process of learning which leads to rethinking assumptions and making changes to strategies, where evaluations show existing assumptions do not hold and current strategies are not working.

In the 2000s there has been a significant rise of 'an evidence agenda' among donors and development agencies (Garbarino and Holland, 2009) with growing demands for 'rigorous' evidence on the impact of development interventions to inform policy-makers and to justify aid budgets. In the next section we further explore this focus on rigour in monitoring and evaluation among aid agencies and some evaluation specialists.

The new focus on rigour in impact studies

Increased attention is being paid to improving rigour in impact assessment and to understanding 'what works' to increase aid impact. While few would argue with the overall goal, there are issues in terms of how this is achieved. There are also important divergences in how rigour is understood.

Among many researchers, evaluators, and aid agencies rigour is most associated with experimental and quasi-experimental approaches involving statistical surveys of controlled comparisons between treatment and control groups. Value for money

assessments in planning stages and formal impact evaluation are increasingly required by some donors. Such approaches restrict impact assessment to assessing the magnitude of change which is specifically attributable to a programme or intervention, rather than broader approaches which define impact as learning about change processes, involving before and after comparisons, but not necessarily employing counterfactuals, i.e. what might have happened without an intervention.

3iE is an organization supported by several donors and its own rapid growth trajectory and (evolving) interpretation of rigour reflects the way in which rigour has come to the fore in international development evaluation. For example, 3iE sees only studies with credible counterfactuals and with a design based on the underlying programme theory to learn what works, why, and at what cost as being up to standard (3iE website). This does not only mean experimental and quasi-experimental designs, but these are now central to what is considered high quality evidence by donors, 3iE, and many evaluation specialists.

But some researchers and practitioners have highlighted that randomized control trials (RCTs) and quasi-experimental designs are not appropriate for some types of interventions, are costly, and have flaws. For example, while they might give strong evidence of whether an intervention had an impact in a particular place, they are weak on issues of replication elsewhere. In this more sophisticated understanding of rigour, rigour is not determined solely by the use of a particular method, but rather the appropriateness of the 'fit' between the nature of the problem being assessed and the particular methods (singular or in combination) used in response to the problem, and the time, political, financial, ethical, and logistical constraints (Patton, 2008; Woolcock et al., 2010). Interventions vary in their characteristics, with more complex interventions necessitating different evaluation designs compared with more simple examples (e.g. vaccination programmes).

While some argue that credible evidence can *only* be generated by experimental and quasi-experimental designs (see for example Blackman and Rivera, 2010), others proffer a broader set of designs which may be appropriate for different situations and may have *equivalent* robustness and credibility. In their DFID review, Stern et al. (2012) provide an overview of different evaluation designs and the logics upon which they rely, as well as the mix of methods which can be employed. Thus, while many would assume that a counterfactual is the basis of any evaluation (the rules of causal inference to support causal claims stemming from a comparison between carefully selected treatment and control groups), there are other understandings of causation that can be used. Theory-based evaluation draws on generative causation as its rules for causal inference. In a holistic way it interrogates the mechanism being employed: 'How does it work? Which elements are important?' and 'How can it be improved?' (Befani, 2012; Yin, 2014). It is of course possible to mix not just methods but designs so that the strengths of one can complement the flaws of the other and vice versa.

In terms of mixing methods there has also been recent and growing convergence between qualitative and quantitative approaches to data generation. Examples include the quantification of stakeholder perceptions, the use of software to code qualitative data to make it more manageable for analysis, and participatory

generation of numbers. Holland (2013), for example, provides multiple examples of participatory statistics that aggregate data gathered from individuals using PRA tools in group settings. The traditional distinction could therefore be recast as 'data collected from structured, closed-ended questions and non-structured, open-ended, modes of enquiry' (Woolcock et al., 2010: 3). While many evaluations in the past have focused on the specific inputs and outputs of a project or programme, theory-based evaluation considers the whole chain. There are also methods for analysing the other plausible interventions or contextual factors which have created change, as well as the initiative being evaluated, such as contribution analysis. However, less structured techniques are emerging which instead essentially ask, 'What changes have happened here?', and only secondarily ask, 'what factors caused these changes?' They thus seek to assess the impact of an initiative in a more participatory, less structured way and may therefore deliver a less biased assessment. At the same time the direct and detailed causal mechanisms may become less clear to evaluators and perhaps less easy to improve.

The expectation of many donors of rigour in studies which they commission, involving 'credible' counterfactuals, theory-based evaluation or mixed designs, and of mixed methods has substantial resource and skill implications. These approaches are costly and require sophisticated skills both in advanced statistics and quantitative techniques, but also in qualitative methods and in combining all of these together to answer evaluation questions. The level of investment may not always be appropriate in the context of non-governmental organizations – certainly for smaller ones. For first generation studies aimed at informing policy there may be more justification (Stern et al., 2012), but this approach is both hard to resource and hard to justify for smaller organizations.

Balancing accountability with participation is also a critical issue here and adds to the challenges for those commissioning and undertaking studies. The methods involved in rigorous studies of the kind proposed by 3iE and many donors tend toward the more extractive end of the spectrum and can distance interviewees from the research process itself. Ethical issues arise in experimental methods (who gets treatment and who does not is a particularly loaded issue where participants are already disadvantaged). Quasi-experimental methods are also less likely to support ongoing learning, participation, and flexibility. These techniques take up the time of participants, without them being involved in decision-making or seeing any immediate benefits, and resources may be diverted away from support for participants' own processes of learning. Balancing accountability and learning objectives for participants is thus tricky. Proponents of rigorous experimental and quasi-experimental approaches would argue that in the longer-term much larger numbers of people will benefit from policies which are more informed by evidence. Some policy questions merit such an intensive approach, but only for 'first generation' studies – i.e. once a base level of evidence has been gathered it should not be necessary to keep repeating the studies (Stern et al., 2012). Much therefore depends upon the overall priorities for the study.

However, the generation of evidence is only one part of policy-making, which is itself a messy, non-linear process. The existence of rigorous evidence does not

guarantee that it shapes final decisions on policy processes – which should be publicly debated and driven by values, but which are also influenced by many other factors and interest groups. The framing of which type of policy option on which to gather evidence is also an important issue: there are dominant policy narratives at work and these frame the nature of the evidence collected. We see this reflected perhaps in the much greater scrutiny of Fairtrade compared with other sustainability standards and market mechanisms to date, for example, because of its high visibility and marketing as an alternative form of trade. A critical current is emerging from some NGOs, practitioners and academics who are challenging a results and evidence agenda which they think neglects the power dimensions involved and which presents evidence generation and use as a neutral, technical exercise alone. In fact, evidence can be misrepresented, is often inconclusive, and can be used in a partial manner to suit interests. According to Eyben (2013) official aid agencies tend to focus on measuring effectiveness in a way that assumes problems are bounded and simple, with their emphasis on linear cause–effect logical planning. Power relations, complexity, surprises, and unexpected impacts and the partiality of knowledge are neglected (Eyben, 2013). However, while this may be true in the past, it is our experience that there are also changes occurring within donors and in the development debate. For example, there is recognition within DFID of a wide set of evaluation approaches (see Stern et al., 2012). In our experience this is also influencing their terms of reference for evaluations of market-based interventions, which take account of challenges to complexity and the need for mixed designs and methods.

Complexity and its implications are currently in the spotlight in development debates with implications for evaluation, as well as aid in general (see for example Ramalingam, 2013). Complex systems are a collection of parts, but collectively they have a range of dimensions; the parts share a physical or symbolic environment or space, and action by any part can affect the whole (Ramalingam, 2008). This means that spillover effects are highly likely and creating a counterfactual as the sole basis for evaluation design and measurement becomes problematic, because the comparison between treatment and control group can be invalidated. It is rarely possible (or desirable) to exert control over treatment and control groups in private sector-led interventions and longitudinal studies are particularly challenging. Thus, different approaches may be needed, such as theory-based evaluation, but also changes in the way monitoring and evaluation is commissioned and positioned vis-à-vis the entity being evaluated.

The development of Fairtrade impact assessment

This broader backdrop of monitoring and evaluation, the evidence agenda, and tensions between accountability and learning approaches and objectives provides the context for sustainability standards, such as Fairtrade, as they have firstly recognized and secondly sought to assess their impact. The evolution of Fairtrade impact assessment reflects the wider picture in international development, although

Fairtrade, as a market-based mechanism, has perhaps lagged behind practice in international NGOs. In this section we explore how Fairtrade impact assessment has evolved and identify some of the specific challenges posed.

Early studies of fair trade in the late 1990s were fairly exploratory and covered cases which could be termed fair or ethical trade (see NRET, 1999; Roberts et al., 1999). A number of in-depth, qualitative studies on fair trade followed in the early 2000s: for example Ronchi (2002a) conducted an impact study in Costa Rica and a participatory monitoring and evaluation (M&E) exercise with a certified Ghanaian cocoa cooperative funded by the alternative trade organization called Twin (Ronchi, 2002b). The first comparative studies – those that compared outcomes and impacts of conventional and fair or ethical trading chains in Peruvian brazil nuts and Ecuadorian cocoa (Nelson et al., 2002) – did not involve counterfactuals in a statistical sense, but provided comparisons that were qualitative in nature and helped to tackle some research bias issues, and identified stakeholder groups being neglected by the fair trade scheme being studied.

During the mid- to late 2000s, a series of rich case studies was undertaken (see the meta reviews of Nelson and Pound, 2009 and the ITC, 2011 review of sustainability standards impact on producers) including some very in-depth studies (see for example, Jaffee, 2007), but few involved the kinds of counterfactual logics described above. Ruben et al. (2008) is one key exception. By 2009 there had been a proliferation of studies. Nelson and Pound (2009) found a number of gaps in the evidence base including a dearth of Asian and African studies, few non-coffee studies, the lack of attention to gender and environmental impacts, and noted the mainly qualitative nature of the studies undertaken. To our knowledge there were no theory-based evaluations.

In 2010 Blackman and Rivera reviewed the evidence on sustainability standards in agricultural commodities and tourism to establish whether these standards improve the social and environmental performance of farms and firms. They identified *ex post* empirical farm level studies and classified them according to whether they employed methods likely to generate credible results. Their definition of credible studies was based on the inclusion of a counterfactual; that is, studies with an experimental or quasi-experimental design (e.g. the latter involving matching of certified producers and non-certified producers and using advanced statistical techniques to address potential selection bias, such as propensity score matching; see Ruben et al., 2008). The latter approach requires large numbers of observations and therefore can be expensive, and as a result it means the coverage of a wide-ranging system such as Fairtrade is very limited, with findings drawn from a small number of organizations. It is also the case that quasi-experimental studies which are not nested within a mixed methods and theory-based approach cannot explain very well how and why outcomes and impacts have been achieved. Many detailed case studies are dismissed as irrelevant when they do not involve counterfactuals in this way, but this seems to ignore both the light they can shed on causal mechanisms and, when used systematically in comparative analysis, the fact that they can allow researchers to generalize (although not to universalize) beyond one particular situation.

Case studies may be the only option in some situations where the construction of a counterfactual is just not possible. In one study (B. Pound, unpublished internal report) for the Fairtrade Foundation, the researchers found that all of the farmers in a particular industry fell within a certain producer organization which holds Fairtrade certification (Belize sugar). Therefore, there is no counterfactual at all.

It is increasingly the case that in some industries and countries all producer organizations are now certified to one sustainability standard or another. If they are not yet certified it is because they sell on a different market or are newly established; that is, they exhibit characteristics which mean they are not a valid comparison. In some way they are not a good match for the organizations which are certified. This situation is only likely to increase as certification expands in different sectors. In this context it is necessary to move to the next best level of counterfactual – namely the non-certified farmers in similar zones and of similar characteristics who do not sell through an organization, but who sell directly to intermediary traders. It is possible that organizations change their certification status during the study and this means the comparisons are undermined, with some taking up new certifications, others stacking up multiple certifications, and others dropping them when they are deemed no longer useful. The comparisons between the certified and non-certified producers may still have some value, but only where a detailed qualitative analysis using theory-based evaluation is also conducted to explain how and why changes are occurring. This has resource implications. While it may be desirable to have high levels of rigour, it is not always practical or achievable given the nature of and patterns of distribution of certified groups in Fairtrade.

The development of theories of change by researchers and standard systems is now helping to provide some structure to studies operating in such high levels of complexity. It is significant that Fairtrade International has now developed its own theory of change and indicators, which will improve its own monitoring of results, can provide greater clarity for researchers, and will enable more standardization across cases to build up a stronger picture overall – even where studies are conducted by different researchers or standard representatives. Complexity in the Fairtrade system stems from the broad reaching nature of the standards, as well as the variation in the other inputs which vary over time and are less standardized (e.g. producer support, networking, and growing markets). In other words Fairtrade has multiple impact pathways. The inputs are also fairly volatile in market-based interventions – the Fairtrade Minimum Price is only active when market prices fall below it and at other times may provide price uplifts for individual farmers, but benefits to individuals also depend upon the operating costs and relative efficiency of the producer organization in returning benefits to members. The Fairtrade Premium is generated according to what a producer organization can sell – and this will vary according to market demand for Fairtrade-certified products and the willingness of buyers to buy from the specific producer organization. The particular training provided by the producer support services of Fairtrade may also vary from place to place. Critically, the additional support provided by external NGOs, alternative trade organizations, and donors varies from place to place. The impact trajectory of Fairtrade will also vary over time – there may be significant early gains prior

to certification in the preparation phase, for example, and while continuous improvement is intended, there may be significant jumps or setbacks for different contextual reasons. The uptake of Fairtrade co-produces outcomes in interaction with the local context (Neilson and Pritchard, 2009; see also Nelson and Martin, 2012 on the factors shaping the impact of Fairtrade). For example, individual farmers' views of their producer organization can be shaped by the history of cooperatives in that particular country, as well as their current performance.

Studies which try to manage all of this complexity are highly challenging for researchers, but also for those commissioning the studies (donors and Fairtrade organizations), because the findings may not be as clear as they would like. Many rigorous studies are unable to establish statistically significant differences between those certified with sustainability standards and those not certified at individual household level. Many impact evaluations in this field are showing mixed results, although this does not mean that no impacts are being achieved. Further, there is a risk of generalizing from too few cases across a whole standard system. Given the breadth of the Fairtrade system, which currently spans 70 countries, building up evidence will take time and should not consume disproportionate amounts of resources in comparison with those invested by consumers, companies, donors, and others in the Fairtrade system. A recent impact evaluation of Fairtrade and Rainforest Alliance is both multi-year and multi-enterprise, but the findings are still restricted to two commodities and four countries (see Nelson and Martin, 2013a). The cost and challenges of employing mixed designs should not be under-estimated. Policy-makers and Fairtrade organizations frequently want clearer, less nuanced findings than can be realistically delivered by such studies – especially when given an ambitious scope at the outset.

Participatory statistics provide one way in which participatory approaches and impact evaluation can converge (see Holland, 2013). Community level data can be generated to assess livelihood outcomes, as well as process issues, and this data can be more accurate than small numbers of responses from community leaders or individual households. Thus accuracy is a key aspect of this debate. However, it is also the case that participatory statistics and qualitative data need to be representative, otherwise inaccuracies can occur and received wisdom about what works may go unchallenged. For example, a reliance on village case studies rather than nationally representative surveys led many social scientists to inaccurately interpret the impacts of the Green Revolution in Bangladesh, according to Orr (2013).

Improving rigour in Fairtrade impact assessment

The choice of design for an impact study should be about aligning evaluation questions with the tools and methods which are available and the specific features or attributes of the intervention being studied (Stern et al., 2012). Misalignment between methods, questions, and attributes can mean that the evaluation cannot actually answer the questions posed by the study. The ethos or development philosophy of the organization commissioning or undertaking the study is also

relevant. Many organizations, including Fairtrade ones, face difficult choices about which objectives (learning or accountability) to prioritize, as these two objectives do not always sit easily together and resources are limited. Some NGOs have taken a public stance to focus only on learning-based evaluation as a matter of principle. Solidaridad, a Dutch international NGO, has opted for ‘improving not proving’ (Solidaridad Network, Annual Report, 2012), although this does not render them immune from external criticism of a lack of rigour and therefore credibility in their evaluation findings.

When selecting an approach to impact assessment, the focus has been on mixing methods, but mixing of designs is also possible. All evaluations draw upon an idea of what causes change (they have a specific understanding of causality, even if this is not explicitly articulated). Statistical surveys rely on controlled comparisons and quasi-experimental and experimental methods with carefully selected treatment and control groups. In this type of study the counterfactual is the ‘without treatment’ group. However, it is not always easy to sustain these groups when private sector companies are involved as they may decide to drop or take up new certifications or decide not to continue participating in the study. Spillover effects are also common, with practices taken up by one group copied by others in the same industry. These resource-intensive studies can help us to understand the extent of impact in a particular location, but are weak at explaining whether the approach might work elsewhere. Experimental and quasi-experimental designs require large numbers of observations. This is not often the situation when studying Fairtrade, or other sustainability standards. At the organizational level there are limited numbers of certified and non-certified organizations that could be matched and Fairtrade works through the producer organization. At the individual member level higher numbers may be found but only in some instances and a focus only on the individual members fails to capture the role of the producer organization in shaping impact and the diversity among producer organizations.

Theory-based evaluation is an alternative in such situations and such approaches are based on generative causation; that is, the ‘mechanisms’ employed as a whole in a case are identified and detailed analysis is carried out to understand how these have generated effects and how much other explanations are responsible (Nelson and Martin, 2011). In a recent study conducted for Fairtrade we explored the impact of Fairtrade on cocoa in Peru (Laroche et al., forthcoming). In such an approach the focus is on generating data along the theory of change, focusing on the transitions (e.g. have the inputs led to outputs and so on) to build a rigorous ‘within case’ analysis. The influence of context grows as one moves along the theory of change, which means that it is usually more feasible to provide rigorous attribution only on the earlier phases of the theory of change (i.e. inputs and outputs) and to provide validation of plausible outcomes and impacts at the latter stages – where both broader contextual factors and alternative interventions contribute. Theory-based evaluation is strong on explanation, but weaker on estimating quantities or the extent of impact (Stern et al., 2012).

It is possible to combine designs which are complementary to strengthen the overall study (Stern et al., 2012; Yin, 2014). For example, in a DFID-funded study

of the poverty impact of sustainability standards (Nelson and Martin, 2013a) we covered multiple cases at country-industry level and compared these in a systematic way. We included multiple producer organizations in each country and compared them systematically also. In each country we included non-participating producer organizations and companies and collected data at both certified and non-certified entities to allow for comparisons (quasi-experimental study). This represents a nested and mixed design, with diverse methods also used, and so the findings are as rigorous as they can be in light of the complexity inherent within sustainability standards and in the different contexts of study. One approach which may be promising for Fairtrade and which has not yet been tried is qualitative comparative analysis (QCA). QCA is useful for identifying which causes and conditions are necessary and sufficient to achieve certain outcomes. This type of analysis could be used with sets of case studies, where the number of case studies makes it difficult for a researcher to manage them comfortably, and as a way of teasing out the aspects of an intervention and the contextual conditions for success (Ragin, 1987) – something which has eluded many Fairtrade case study researchers to date.

There are a number of practical steps to improving rigour which could be considered. Firstly, the development of a research protocol is important. If more than one case is being covered in the impact study it is important to employ standardized questions to support cross-comparative analysis. The theory of change already provides a level of standardization, but it is important to ensure that the evaluation questions and research propositions are clearly articulated (Yin, 2014). A research protocol can then be developed to guide each country team. While this sounds prescriptive it need not be so if there is sufficient time and effort undertaken to engage all those involved in conducting the study in the design, although it is unlikely to support a participatory learning process driven by farmers and workers. A protocol can support shared understanding within the research team, clarity on how Fairtrade works, and communication of the study objectives. As well as the theory of change and priority evaluation questions, it should specify the relevant indicators to be used (including scope for participatory indicator development), the data collection strategies and process, the analytical strategies, the writing up process (responsibilities and structure of report), and feedback processes and opportunities.

Secondly, stakeholders, especially producers and workers, can be engaged in analysing whether the intended theory of change of Fairtrade has indeed led to intended outcomes and impacts. However, this can be time consuming and fairly challenging and requires careful explanation and adequate time allocation. An alternative is to ask in a more open way what has generated change, with the contribution of Fairtrade explored secondarily. As Fairtrade works along the value chain it is important that impact studies engage with and consider the full value chain in understanding whether intended inputs and outputs have led to outcomes and impacts. While this has been relatively neglected or under-resourced in many studies, greater attention is being paid with methodologies being developed to quantify perceptions of change in terms of the 'fairness' of the value chain relationships (see Unilever et al., 2012; Twin, 2012) and given many of the difficulties in obtaining commercially sensitive information from value chain actors. The use of

theory of change and diverse stakeholder participation in assessing change and its different causes using new techniques such as contribution analysis (Mayne, 2008) is a potentially powerful approach and one that may increase rigour. Because the Fairtrade theory of change is complex it is important to ensure that impact studies focus on priority evaluation questions, particular impact pathways or specific themes, otherwise the study is likely to become too ambitious and may find it difficult to answer the questions adequately.

Thirdly, new methods are emerging which could be applied in the Fairtrade field. Process tracing methods, for example, are 'tools to study causal mechanisms in a single case research design' (Beach and Pedersen, 2013: 2; Oxfam, 2013). Once a rigorous 'within case' analysis is built up for different cases these can be systematically compared. Typologies can be used to guide the selection of cases and support wider conclusions to be drawn than unconnected case studies, although not universal answers (Yin, 2014; George and Bennett, 2005). For sustainability standards such as Fairtrade, it is possible to develop a typology as we are currently doing in a coffee impact study for Fairtrade International of the types of producer organizations in Fairtrade coffee and selecting from these. QCA has not been used as yet in Fairtrade or sustainability standard impact studies, but offers the potential to review a larger number of cases and to establish the necessary and sufficient conditions for the achievement of specific outcomes (Ragin, 1987). This may be a useful approach because Fairtrade is such a wide-ranging system, exhibiting inherent complexities, and it is important that advances are made in understanding not whether it works or not, but in a more nuanced way we need to know how well it works under different conditions and how it can be improved.

Assessing the relative contribution of plausible rival interventions or contextual factors should be part of the discussion and analysis at each point in the theory of change to establish their relative significance in bringing about impact (Stern et al., 2012). We have done this in an impact evaluation with key informants, but it could be done more explicitly with multiple stakeholders in a workshop (Nelson and Martin, 2013a). Contribution analysis (Mayne, 2008) is a participatory technique which can be used to assess the relative contribution of different interventions to change processes. In relation to Fairtrade, this technique is particularly important because other NGO, donor, and private sector initiatives commonly invest in producer organizations as well as, and often because of, Fairtrade. While many impact studies have noted this (e.g. Nelson and Smith, 2010a), it is also the case that few have attempted to engage stakeholders in quantifying their perceptions of relative contribution and more could be done in this regard.

Local stakeholder workshops are useful, not just for feedback, but as part of the process of data gathering, to draw in diverse stakeholder perspectives including farmers and workers. In an ongoing Fairtrade coffee study, we have used force field analysis to understand how producer organization board managers and leaders see their organizations changing over recent years and the negative and positive forces shaping this trajectory of change.

At an organizational level, where producer groups have already developed clear plans and strategies of their own and collect key data for monitoring, it is easier to

build upon this with the impact study. Outcome mapping could also contribute to producer organization learning and contribute to mixed method evaluations. While it is important not to conflate participatory rural appraisal (PRA) methods with a participatory *process*, it is the case that some methods are more understandable to many rural farmers where there are high levels of illiteracy. For example, participatory gross margin analysis, also being trialled in an ongoing Fairtrade coffee impact study, is a technique for estimating the costs and benefits of a specific enterprise with a group of carefully selected farmers, in a way that is of value to them.

Development of poverty ladders and rural typologies is an area where more innovation is also urgently needed. This is because while Fairtrade often brings positive benefits, impact evaluations are indicating that poverty impacts – moving significant numbers of disadvantaged groups up the poverty ladder – are less likely (Nelson and Martin, 2013a). Therefore, a focus on the wider picture and the structural challenges for Fairtrade, producer organizations, and wider stakeholders in a locality is important. Overlapping and complementary interventions will be needed to tackle poverty far beyond Fairtrade (Vorley et al., 2013). An analysis of

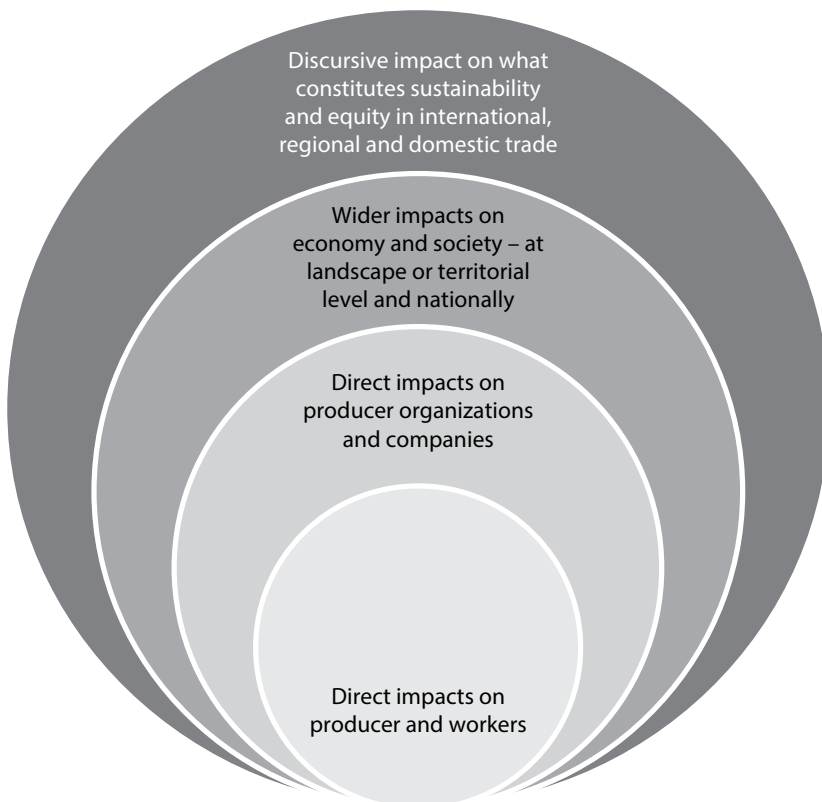


Figure 1 Different types or scales of Fairtrade impact

rural differentiation is important and impact studies can be enriched by seeking to understand this bigger picture.

While specific Fairtrade impact studies may need to focus on specific themes or priority questions, it is important that the different scales at which Fairtrade operates are taken into account in the wider debate on its effectiveness. Too often Fairtrade has been valued based on its impact for certified producers and workers alone. Yet there are different scales at which Fairtrade operates (see Figure 1). As well as the impacts at individual and organizational level, there may be wider impacts in the local economy. Spillover effects can occur in the local community or beyond (e.g. others may benefit from the improvement of a school or crop collection centres). Demonstration and learning effects are also common: For example, new agricultural practices introduced by Fairtrade and sustainability standards can spread among neighbouring farmers. Cases have been found where Fairtrade has raised prices offered by non-certified buyers of cocoa in order to compete with the certified buyers (Nelson and Galvez, 2002; Laroche et al., 2014). Thus Fairtrade can be having an impact, but a simple comparison between certified and non-certified groups would not show significant statistical differences as a result. This is why theory of change analysis and mixed methods are needed – to contextualize and explain such processes and effects (positive or negative).

Increased attention is being paid to landscape level issues in relation to sourcing decisions by companies and in monitoring and evaluation. Agribusinesses are being urged to consider ‘reducing risks by adopting landscape approaches to sustainable sourcing’ (Kissinger et al., 2013). This is because to protect and enhance ecosystem services requires consideration of scale effects (Tallontire et al., 2012). Assessing change at an individual farm level does not necessarily capture the cumulative effects at a wider scale: reduced pesticide use may be environmentally beneficial at the individual farm level and produce health benefits for farmers less exposed to harmful agrochemicals, but a greater carbon footprint may be the overall result; as many farmers reduce their pesticide use this can lead to more land being used and more yields lost to pests (Tallontire et al., 2012). To fully understand environmental impacts would thus require full life-cycle analysis.

There is an aspect of Fairtrade impact which tends not to be given much attention in impact studies, namely its discursive impact. In other words, Fairtrade has an influence on (and is influenced by) the discourse around agricultural trade and sustainability. As the pioneer, Fairtrade deserves credit for the later development of other sustainability standards and initiatives in agricultural food chains. Whatever the current challenges or future preferred strategies, Fairtrade and other sustainability standards have made an important contribution to putting sustainability on the agenda by building consumer, corporate, and development agency understanding and interest in sustainability issues in agri-food chains (SustainAbility, 2011). However, it is also worth noting that some sustainability issues are not currently covered by Fairtrade as part of its core approach, such as greenhouse gas emissions.

Improving utility in Fairtrade impact assessment

So far the discussion has focused on ways of improving rigour in Fairtrade impact assessment, but for moral reasons an equally important issue is that of learning and utility and for whom? The increased focus on rigour can leave utility at best neglected and at worst undermined – particularly for farmers and workers, as well as for higher level stakeholders. Different impact studies have differing objectives and therefore not all need the same levels of rigour. As a first step it is important for those considering investing in impact work to understand what is appropriate for them, given their resources and considering who will benefit and who will bear the cost of the evaluation.

It is absolutely essential to engage stakeholders in impact assessment and not to sacrifice this objective in the pursuit of rigour. But how can this be achieved given the trade-offs involved? Firstly, it is important to focus on the overall evaluation process rather than the specific methods being employed.

For Fairtrade as a wide-ranging stakeholder movement there is the opportunity to engage with stakeholders in terms of developing their overall monitoring and evaluation framework. Fairtrade International has conducted various workshops with regional producer networks, for example, in developing the theory of change and in planning indicators. Fairtrade has the most participatory governance structure of any of the voluntary sustainability standards, but it is not clear how far Southern stakeholders have been engaged in decision-making on impact assessment and the commissioning process for studies (which is when many decisions are made about the type of study which can be undertaken). As a global sustainability standard and as a member of ISEAL, Fairtrade International is developing global indicators to allow for comparative analysis linked to their theory of change, but there would also be opportunities in specific impact case studies to allow for more participatory indicator development to be undertaken. Where stakeholder engagement can be facilitated prior to commissioning this is to be encouraged. Capacity strengthening among Southern evaluation specialists in this field is also needed to support ongoing learning processes and investment is needed in producer organizations' own capacity to monitor and learn about their performance and impact.

Once a study is under way, stakeholders can be invited to directly debate the theory of change or, in a more open, less structured way, they can be asked to identify change and its causes. A compromise is to ask about organizational change over the past x years and the forces which have shaped change (positive and negative during that period).

Participation can be increased even in quasi-experimental studies which are mixed with qualitative methods by engaging key stakeholders, including producers and workers, in the research process. This is particularly important in situations of complexity, as is the case with Fairtrade impact assessments which have to take account of multiple impact pathways, diverse partnerships, plausible alternative interventions, and highly variable contextual conditions. By holding a stakeholder meeting, key informant interviews, and using PRA techniques this can support engagement by farmers, workers, and broader stakeholders. The purpose of the

study can be fully explained, time given to gathering diverse stakeholder perspectives, and a feedback process established. There is less scope though for supporting learning by farmers and workers as it is not a process driven by their own learning and there are constraints on what can be changed and adapted in terms of the evaluation design if rigour is to be sustained.

Within the Fairtrade system there are possibilities for participatory methods to be used, such as value chain dialogues or participatory video, to support producer and worker communication along the value chain and with Fairtrade International, even within a mixed design or method evaluation, but only if resources are available. The more attention given to farmers' and workers' voices the better. While some studies adopt a more participatory approach, the voices of farmers and workers are still filtered through researchers and presented in reports. There is scope to use and adapt participatory video approaches to overcome literacy barriers at the local level and to cross distances in terms of communicating issues. How far such approaches can effect change depends upon the willingness of Fairtrade actors to act upon the findings.

For a longer-term strategy and where Fairtrade itself is funding impact studies there is scope to build up relationships in a particular sub-region, landscape, or territory. Fairtrade could support a learning alliance to emerge, involving stakeholders at this more localized level in which a producer organization or estate is located, to support a process of learning linked to action. Such a learning alliance should involve stakeholders in the horizontal landscape, but also in the vertical value chain and in the governance structures of Fairtrade International. The incentives for participation of stakeholders in a learning alliance cannot be assumed, however, and much would rest on the ability of Fairtrade to act upon key findings and insights.

Conclusions

In this paper we have explained how impact assessment has developed in the field of Fairtrade covering studies commissioned by Fairtrade organizations and those by donors. Greater attention has been paid to impact assessment in Fairtrade, reflecting the picture in international development more widely. But the increased pressure for rigour in impact assessment, interpreted as experimental and quasi-experimental methods, presents some challenges for Fairtrade in the light of its inherent complexity and its wide scope of implementation, and given the importance of achieving utility. Utility is about ensuring that Fairtrade impact studies are useful to key stakeholders, with findings communicated to and acted upon by the wider organization, and also, critically, involving farmers and workers. There is scope to improve rigour, and not only through the use of counterfactual logics, as explained in this paper, but the challenge is to try to improve both rigour and utility. There are trade-offs between accountability and learning objectives in evaluation studies which are hard to avoid, but in this paper we have sought to provide some suggestions for how to achieve this. We have explained that rigour is about matching evaluation questions to impact design in the light of the specific

attributes of the programme being studied and organizational ethos. It should not be conflated only with experimental and quasi-experimental designs. Other types of causality and theory-based evaluation can be employed, or combined. Mixed methods are also desirable, but we note the resource implications. Any resources invested in evaluation need to be proportionate to the scale of investment in the intervention (in this case Fairtrade). As Fairtrade is an approach seeking to achieve sustainable trade it is also important that it does not overly rely on donor funds. It is appropriate that donor funds be channelled into establishing better institutional learning systems and accountability mechanisms to help ensure that claims and achievements are consistent and credible, and support improvement and rethinking of strategies and assumptions where necessary.

While rigour can be improved it is also important that researchers do not over-claim from small numbers of case studies, and more systematic comparative analysis between cases is needed. Not all studies can or should be ‘first-generation’ type studies seeking to inform policy with robust evidence (which is difficult enough anyway). It is appropriate that all sustainability standard systems, including Fairtrade, also pay attention to the realities of fieldwork and to organizational ethos – if sustainability includes producer empowerment then it is important that the utility of an impact study is given full consideration. This includes ensuring that Fairtrade organizations take up the findings as far as possible and find ways that producers and workers can have their voices heard more directly than in the past. Facilitation of learning alliances at the sub-regional level would be one way of doing this.

Acknowledgement

This article draws upon various research studies, including a major study funded by the UK Department for International Development on the poverty impact of sustainability standards.

References

- Beach, D. and Pedersen, R.B. (2013) *Process Tracing Methods: Foundations and Guidelines*, Ann Arbor, MI: The University of Michigan Press.
- Befani, B. (2012) *Models of Causality and Causal Inference* [online] Department for International Development (DfID) <http://betterevaluation.org/resources/guide/causality_and_causal_inference> [accessed 1 February 2014].
- Blackman, A. and Rivera, J. (2010) ‘The evidence base for environmental and socio-economic impacts of “sustainable” certification’, Discussion Paper, Washington, DC: Resources for the Future.
- Chambers, R. (2009) ‘So that the poor count more: using participatory methods for impact evaluation’, *Journal of Development Effectiveness* 1(3): 243–6.
- Earl, S., Carden, F. and Smutylo, T. (2001) *Outcome Mapping: Building Learning and Reflection into Development Programs*, Ottawa, Canada: International Development Research Centre.

- Edward, P. and Tallontire, A.M. (2009) 'Business and development: towards re-politicisation', *Journal of International Development* 21(6): 819–33.
- Estrella, M. (ed.) with Blauert, J., Campilan, D., Gaventa, J. Gonsalves, J., Guijt, I., Johnson, D. and Ricafort, R. (2000) *Issues and Experiences in Participatory Monitoring and Evaluation*, Rugby, UK: Intermediate Technology Publications.
- Eyben, R. (2013) *Uncovering the Politics of 'Evidence' and 'Results': a Framing Paper for Development Practitioners* [pdf] bigpushforward.net <<http://bigpushforward.net/wp-content/uploads/2011/01/Uncovering-the-Politics-of-Evidence-and-Results-by-Rosalind-Eyben.pdf>> [accessed 18 December 2013].
- Garbarino, S. and Holland, J. (2009) *Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results* [pdf] Governance and Social Development Resource Centre, Discussion Paper, University of Birmingham, UK <<http://epapers.bham.ac.uk/646/1/eirs4.pdf>> [accessed 18 December 2013].
- George, A.L. and Bennett, A. (2005) *Case Studies and Theory Development in the Social Sciences*, Cambridge, MA: MIT Press.
- Guijt, I., Arevalo, M. and Saladores, K. (1998) *Participatory Monitoring and Evaluation: Tracking Change Together*, PLA Notes Issue 31, pp. 28–36, London: IIED.
- Holland, J. (ed.) (2013) *Who Counts? The Power of Participatory Statistics*, Rugby, UK: Practical Action Publishing.
- International Trade Centre (ITC) (2011) *The Impacts of Private Standards on Producers in Developing Countries*, Literature Review Series on the Impacts of Private Standards; Part II (Doc. No. MAR-11-201.E), Geneva: ITC.
- Jaffee, D. (2007) *Brewing Justice: Fair Trade Coffee, Sustainability, and Survival*, Berkeley, CA: University of California Press.
- Kissinger, G., Brasser, A. and Gross, L. (2013) *Reducing Risk: Landscape Approaches to Sustainable Sourcing*, Scoping Study [pdf] Washington, DC: Landscapes for People, Food and Nature Initiative <http://landscapes.ecoagriculture.org/documents/files/reducing_risk_synthesis_report.pdf> [accessed 19 December 2013].
- Laroche, K., Jimenez, R. and Nelson, V. (2014) *Assessing the Impact of Fairtrade in Peruvian Cocoa*, study commissioned by Fairtrade International, Chatham, UK: NRI.
- Mayne, J. (2008) *Contribution Analysis: An Approach to Exploring Cause and Effect* [pdf] Institutional Learning and Change Brief No. 7 <www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf> [accessed 1 February 2014].
- Murray, D., Reynolds, L. and Taylor, P. (2003) *One Cup at a Time: Poverty Alleviation and Fairtrade in Latin America*, Fort Collins, CO: Fairtrade Research Group, Colorado State University.
- Neilson, J. and Pritchard, B. (2009) *Value Chain Struggles: Institutions and Governance in the Plantation Districts of South India*, Chichester, UK: Wiley-Blackwell.
- Nelson, V. and Galvez, M. (2000) *Assessing the Social Impact of Ethical and Conventional Cocoa in Ecuador*, NRI report, Chatham, UK: University of Greenwich.
- Nelson, V. and Martin, A. (2011) 'Impact evaluation of social and environmental voluntary standard systems (SEVSS): using theories of change', NRI Working Paper [pdf], Chatham, UK: University of Greenwich <www.nri.org/projects/tradestandards/docs/final_dfid_paper_on_using_theories_of_change_in_ie_of_standards.pdf> [accessed 5 February 2014].

Nelson, V. and Martin, A. (2012) 'The impact of Fairtrade: evidence, shaping factors, and future pathways', *Food Chain* 2(1): 42–63 <<http://dx.doi.org/10.3362/2046-1887.2012.005>>.

Nelson, V. and Martin, A. (2013a) *Final Technical Report: Assessing the Poverty Impact of Sustainability Standards* [pdf] Chatham, UK: Natural Resources Institute <www.nri.org/images/documents/project_websites/AssessingPovertyImpacts/AssessingThePovertyImpactOfSustainabilityStandards.pdf> [accessed 1 February 2014].

Nelson, V. and Martin, A. (2013b) *The Strategic Use of Case Studies by Standard Systems*, NRI Working Paper, Chatham: UK, Natural Resources Institute.

Nelson, V. and Pound, B. (2009) *The Last Ten Years: A Comprehensive Assessment of the Impact of Fairtrade*, NRI Report, Chatham, UK: Natural Resources Institute.

Nelson, V. and Smith, S. (2011) *Fairtrade Cotton: Assessing Impact in Mali, Senegal, Cameroon and India: Main Report* [pdf], Chatham, UK: NRI <www.nri.org/projects/fairtradecotton/docs/Impact%20of%20Fairtrade%20Cotton%20-%20main%20report%20-%20final%20Apr2012.pdf> [accessed 5 February 2014].

Nelson, V., Tallontire, A. and Collinson, C. (2002) 'Assessing the potential of ethical trade schemes for forest dependent people: comparative experiences from Peru and Ecuador', *International Forestry Review* 4: 99–110 <<http://dx.doi.org/10.1505/IFOR.4.2.99.17440>>.

Natural Resources and Ethical Trade Programme (NRET) (1999) 'Ethical trade and rural livelihoods', in *Sustainable Rural Livelihoods: What Difference Can We Make?*, London: DFID.

Orr, A. (2013) 'Why were so many social scientists wrong about the Green Revolution? Learning from Bangladesh', *The Journal of Development Studies* 48(11): 1565–86 <<http://dx.doi.org/10.1080/00220388.2012.663905>>.

Oxfam (2013) *Process Tracing: Draft Protocol* [online] <<http://policy-practice.oxfam.org.uk/blog/2013/02/~media/C396B507E01C47AB880D7EEF9ECCD171.ashx>>.

Patton, M.Q. (2008) *Utilisation-Focused Evaluation*, 4th edn, London: Sage.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*, London: Sage Publications.

Ragin, C.C. (1987) *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*, Berkeley, CA: University of California Press.

Ramalingam, B. (2008) 'Evaluation and the science of complexity'. Presented at the Evaluating Complexity Conference, NORAD, 29–30 May.

Ramalingam, B. (2011) *Learning How to Learn: Eight Lessons for Impact Assessment that Make a Difference*, Background Note, London: ODI.

Ramalingam, B. (2013) *Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World*, Oxford, UK: Oxford University Press.

Riddell, R. (2007) *Does Foreign Aid Really Work?* Oxford, UK: Oxford University Press.

Roberts, S., Robins, N. and Abbot, J. (1999) *Who Benefits? A Social Assessment of Environmentally-Driven Trade*, London: IIED.

Roche, C. (1999) *Impact Assessment for Development Agencies: Learning to Value Change*, Oxford, UK: Oxfam Publications.

Ronchi, L. (2002a) *The Impact of Fairtrade on Producers and their Organisations: A Case Study with COOCAFE in Costa Rica*, PRUS working paper, No. 11 [pdf], Poverty Research Unit, University of Sussex <www.sussex.ac.uk/Units/PRU/wps/wp11.pdf> [accessed 5 February 2014].

Ronchi, L. (2002b) *Monitoring Impact of Fairtrade Initiatives: A Case Study of Kuapa Kokoo and the Day Chocolate Company* [pdf] London, UK: TWIN <http://portals.wi.wur.nl/files/docs/ppme/TwinMEKuapaandDayA_5version.pdf> [accessed 5 February 2014].

Ruben, P., Fort, R. and Zuniga, G. (2008) *Final Report. Fairtrade Programme Evaluation: Impact Assessment of Fairtrade Programs for Coffee and Bananas in Peru, Costa Rica and Ghana*, study assignment by Solidaridad Coordinated by the Centre for International Development Issues (CIDIN), The Netherlands: Radboud University Nijmegen.

Solidaridad (2012) *The Solidaridad Network 2012 Annual Report* [pdf] Netherlands: Solidaridad <<http://solidaridadnetwork.org/sites/solidaridadnetwork.org/files/The%20Solidaridad%20network%202012%20annual%20report.pdf>> [accessed 5 February 2014].

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befani, B. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations* [pdf], Working Paper 38, April 2012, report of a study commissioned by DFID, London: DFID <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf> [accessed 5 February 2014].

SustainAbility (2011) 'Signed, sealed, delivered: behind certifications and beyond labels', London and Washington, DC: SustainAbility <www.sustainability.com/library/signed-sealed-delivered-1#.UrLWGImYbIU> [accessed 19 December 2013].

Tallontire, A., Nelson, V., Dixon, J. and Benton, T. (2012) *A Review of the Literature and Knowledge of Standards and Certification Systems in Agricultural Production and Farming*, NRI Working Paper no 2, Chatham, UK: Natural Resources Institute.

Twin (2012) *Making International Supply Chains Work for Smallholder Farmers: A Comparative Study of Six Fair Trade Value Chains* [pdf] London: Fairtrade Foundation <www.fairtrade.org.uk/includes/documents/cm_docs/2012/M/Making_international_supply_chains_work_for_smallholder%20farmers.pdf> [accessed 18 December 2013].

Unilever, Oxfam, IIED (2012) *Measuring Fairness in Supply Chain Relationships: Methodology Guide* [pdf] London: IIED <<http://pubs.iied.org/pdfs/16042IIED.pdf>> [accessed 18 December 2013].

Vorley, B., Cotula, L. and Chan, M.K. (2013) *Tipping the Balance: Policies to Shape Agricultural Investments and Markets in Favour of Small-Scale Farmers*, research report, December 2013, London: IIED; Oxford: Oxfam.

Woolcock, M., Rao, V. and Bamberger, M. (2010) 'Using mixed methods in monitoring and impact evaluation', *Policy Research Working Paper* 5242 [online], World Bank Poverty and Inequality Team <<http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-5245>> [accessed 17 December 2013].

Yin, R.K. (2014) *Case Study Research: Design and Methods*, 5th edn, Thousand Oaks, CA: Sage Publications.

Website

3iE website <www.3ieimpact.org/> [accessed 10 December 2013].