

Impact Assessment of the Effects of Trade on Employment: The Role of Mixed Method Approaches

Paul Shaffer

Trent University, Canada

Email: paulshaffer@trentu.ca

Abstract

Following a preliminary discussion of conceptual and definitional questions, this paper examines eight potential contributions of ‘qualitative’ and mixed method approaches in impact assessment related to: lived experiences and locally meaningful measures of impact; weighing the relative importance of impact measures; self-reports of impact; counterfactual thought experiments (CTEs); unpacking mechanisms; combining outcomes and mechanisms; integrating correlates and ‘reasons’ and informing model specification. It concludes with a discussion of how research results may feed into the policy process.

Acknowledgements

This paper was commissioned by the Employment Policy Department of the International Labour Organisation, as part of their Strengthening the Impact on Employment of Sector and Trade Policies (STRENGTHEN) project. It will also appear as a STRENGTHEN Working Paper. I would like to thank David Cheong for his detailed comments which significantly improved the paper and Mutsa Danzvara for her excellent research assistance. All errors are my own.



Table of Contents

1. Introduction.....	1
2. Conceptual and Definitional Issues	2
2.1 The ‘Qualitative’/’Quantitative’ Distinction & ‘Qualitative’ Methods	2
2.2 ‘Participatory’ Research Approaches.....	4
2.3 Employment and Decent Work: Components and Interrelationships.....	6
2.4 Methodological Choices in Impact Assessment.....	8
3. ‘Qualitative’ and Mixed Method Studies in Impact Assessment.....	11
3.1 Lived Experiences and Locally Meaningful Measures of Impact.....	11
3.2 Locally Meaningful Weights of Impact Measures	14
3.3 Self-Reports of Impact	16
3.4 Counterfactual Thought Experiments (CTEs).....	19
3.5 Unpacking Mechanisms	23
3.6 Combining Outcomes and Mechanisms.....	25
3.7 Integrating Correlates and ‘Reasons’	27
3.8 Informing Model Specification	29
4. Key Findings and Recommendations for Policy Making.....	32
4.1 Key Findings: Value-Added of ‘Qualitative’ and Mixed Method Approaches	32
4.2 Feeding Results into Policy Processes	33
References.....	36
Appendix A: Summary Table Of Studies	44

1. Introduction

Assessment of the impact of trade on employment tend to be dominated by modelling exercises. In his review of methods of impact assessment of trade on employment undertaken under the ILO project on ‘Strengthening the Employment Impact of Sectoral and Trade Policies’, Gibson (2011) discusses a range of such approaches from partial equilibrium methods, factor-substitution models and input-output frameworks to Social Accounting Matrices (SAMs) and Computable General Equilibrium (CGE) models along with reduced-form econometric models.¹ Historically, policy deliberations, and academic debates, about the effects of trade have relied heavily on information from one or more of these sources.

Over the past number of years, however, increasing attention has focused on the contribution of so-called ‘qualitative’ and mixed method studies in impact assessment. There has been a growing number of such studies along with a growing appreciation of the insights which they may provide.² Such studies have been applied to a wide range of issues including, to a limited extent, impact assessment of trade.

The purpose of this paper is to document a number of contributions which such studies can make in the general context of trade and employment. Three caveats should be stated at the outset. First, given the limited number of high quality empirical studies addressing trade and employment, some of the empirical cases discussed in section 3 draw on other literatures. Nevertheless, the methodological approaches are generally transferable. Second, in a number of the studies examined, the ‘employment’ aspect is broadened to include other dimensions of decent work. In fact, as discussed in sections 3.1 and 3.2, including and weighting such dimensions is a major contribution of ‘qualitative’ and mixed method approaches. Finally, the discussion of mixed methods is less applicable to modelling exercises which are at a very high level of generality, such as CGEs. Typically, the models discussed are reduced-form econometric estimators and not the simulation exercises reviewed in Gibson (2011).

The format is as follows section 2 addresses conceptual and definitional issues with respect to the ‘qualitative/quantitative’ distinction and ‘qualitative’ methods (section 2.1), ‘participatory’ research approaches (section 2.2), dimensions of employment and decent work (section 2.3), and determinants of methodological choice in impact assessment (section 2.4). Section 3 presents eight potential contributions of ‘qualitative’ and mixed method approaches with examples from the empirical literature related to: lived experiences and locally meaningful measures of impact (section 3.1); weighing the relative importance of impact measures (section 3.2); self-reports of impact (section 3.3); counterfactual thought experiments (CTEs) (section 3.4); unpacking mechanisms (section 3.5); combining outcomes and mechanisms (section 3.6); integrating

¹ See also, Bourguignon and da Silva, Eds (2003) for a survey of modeling approaches in impact assessment.

² See Rao and Woolcock (2003), White (2008) and Shaffer (2013).

correlates and ‘reasons’ (section 3.7) and informing model specification (section 3.8). Section 4 concludes by summarizing the key areas where ‘qualitative’ and mixed method studies can add value to employment impact assessment (section 4.1) and discussing the potential for feeding results into the policy process (section 4.2). A summary table in Appendix A situates all of the studies reviewed in section 3 in terms of the criteria of methodological choice presented in section 2.4.

2. Conceptual and Definitional Issues

2.1 The ‘Qualitative’/‘Quantitative’ Distinction & ‘Qualitative’ Methods

The terms ‘qualitative’ and ‘quantitative’ often appear, at first glance, to have relatively clear meaning.³ The former is often associated with ‘words’, the latter with ‘numbers’. In the social sciences, the related distinction may be between disciplines such as social anthropology on the one hand, and economics or statistics on the other. It may be somewhat surprising, then, that it is actually quite difficult to pin down a working definition of the ‘qualitative/quantitative’ distinction that effectively distinguishes between approaches found in actual mixed method designs. In fact, the ‘qualitative/quantitative’ distinction has been interpreted in quite different ways over the years and has been the subject of controversy (Bryman 1984).

A concrete example of this definitional ambiguity is provided through analysis of a typology of difference between ‘qualitative’ and ‘quantitative’ approaches proposed by Ravi Kanbur (2003). It is based on the following five elements: i) type of information on population (non-numerical or numerical); ii) type of population coverage (specific to general); iii) type of population involvement (active to passive); iv) type of inference methodology (inductive to deductive); v) type of disciplinary framework (broad social sciences to neo-classical economics). Arguably, however, the above distinctions are hard to sustain in light of empirical counterexamples (Shaffer 2005, Kanbur and Shaffer 2007).

More specifically, the numerical/non-numerical distinction poses problems because it is possible to numerically transform almost any type of information by counting, scaling, ranking, and so forth. The second distinction, between specific and general population coverage, is more *incidental* than *essential* to the ‘qualitative/quantitative’ divide in that almost any research technique, ‘qualitative’ or ‘quantitative’, may be conducted in few or many sites. Likewise, population coverage is not intrinsic to research traditions but depends on: the purpose of the research, in particular, whether results are required to be representative of a broader population; the basis of claims to external validity, in particular whether statistical inference is relied upon to estimate standard errors and practical considerations related to cost and standardisation. A final distinction between neo-classical economics and other disciplines understates important ‘quantitative’

³ This discussion is based on Shaffer (2013: 6-8).

traditions within the social sciences such as rational choice political scientists, sociologists schooled in the Lazarsfeld tradition of surveying and model-building and “cliometric” historians (Abbott 2001).

In light of these difficulties in giving precise meaning to the ‘qualitative’/’quantitative’ distinction, the approach here will be to largely eschew these terms in section 3, and refer to the actual methods in question. More specifically, the following techniques have figured in the broader literature on ‘qualitative’ and mixed method approaches to impact assessment, many of which appear in the case studies discussed in section 3:⁴

- i. *Participant Observation/(Ethnographies)*: Extended immersion and participation in the social life of communities, involving multiple data collection methods including many discussed below, (this approach is often regarded as the foundation of social/cultural anthropology);
- ii. *Focus Group Discussions*: Guided discussion of key themes led by a facilitator, usually involving 5-15 participants;
- iii. *Semi-structured Interviews*: Guided interviews with open-ended questions allowing space for emergent themes, but structured around key issues.
- iv. *Self-Reports of Impact*: Fixed response or open ended questions in household surveys, focus group discussions and semi-structured interviews eliciting perceptions of the magnitude impact;
- v. *Counterfactual Thought Experiments*: Fixed response or open ended subjunctive conditional (if/then) questions in household surveys, focus group discussions and semi-structured interviews about some hypothesised link in the causal chain generating impacts;
- vi. *Causal Maps*: Diagrams depicting locally relevant aspects of the causal system generating impact with emphasis on the processes or mechanisms in question;
- vii. *Ranking and Scoring Techniques*: Visual tools used to rank or score the relative importance of different impact measures and/or of causal effects on different population groups or outcomes. One variant, *wealth or well-being ranking*, is used to rank households or individuals in terms of local conceptions of these terms.
- viii. *Life Histories*: Analysis of long-term trajectories of change, including significant events and activities, which have occurred over the life course of individuals or communities. This techniques usually involves focus group discussion and/or semi-structured interviews and may be combined with visual mapping tools such as timelines.

Items ii, ii, vi, and vii figure centrally in so-called ‘participatory’ approaches to impact assessment to which we now turn.

⁴ See Bernard (2002) generally, and Roche (1999), Mayoux and Chambers (2005) and Catley et al. (2014), for application to impact assessment.

2.2 'Participatory' Research Approaches

'Participatory' research approaches are often associated with 'qualitative' and mixed method research. The terms 'participatory' or 'participation', however, have been used in quite different ways to denote very different things. As an example, consider the following typology of forms of 'participation' in development programs or projects (Pretty, 1994: 41):

- i. *Passive Participation*: People participate in a project after being told what is going to happen.
- ii. *Participation in Information Giving*: People answering questions to extractive researchers.
- iii. *Participation by Consultation*: People are consulted though experts define problems and solutions which may be modified in light of people's responses.
- iv. *Participation for Material Incentives*: People participate by providing resources (e.g. labour) in return for material reward, e.g. on-farm research driven by experts.
- v. *Functional Participation*: People form groups to meet predetermined objectives of a project.
- vi. *Interactive Participation*: People participate in joint analysis, action plans, control local decisions and have a stake in maintaining local practices.
- vii. *Self-Mobilization*: People take initiatives independently to effect either incremental or transformative change.

In the context of impact assessment, the term 'participation' is usually associated with techniques of Participatory Rural Appraisal (PRA), including those discussed in the section 2.1. In terms of the above typology, PRA approaches should most closely approximate 'interactive participation', although as discussed below, they often amount to 'participation in information-giving'. There are some clear methodological advantages of participatory approaches as illustrated in the empirical studies in section 3. There are also, however, a number of methodological limitations of these approaches which should be made explicit. First, a brief overview of this tradition will provide some context.

Participatory Rural Appraisal arose in the early 1990s, drawing on a number of sources including Rapid Rural Appraisal (RRA), applied social anthropology, farming systems research, agro-ecology and participatory action research (Chambers 1994). There are particularly close links with RRA, which rejected so-called rural development tourism, whereby foreign researchers descend on easily accessible communities for very short periods, and large fixed-response questionnaire surveys that characterized much applied research in the Global South. Instead, greater importance was afforded local knowledge primarily as a means to facilitate better data quality. While PRA shares with RRA this attempt to validate local knowledge, it adds additional elements which allegedly comprise the 'participatory' import of the approach. Specifically, processes of

information generation, analysis and use are supposed to be driven and owned by local populations with a view to catalyze social change or ‘empowerment’.

PRA has been the subject of a number of wide-ranging critiques from different perspectives (Mosse, 1994; Cooke and Kothari, 2001; Kapoor, 2002). Here, the focus is on methodological shortcomings drawing on the ‘self-criticism’ of a prominent proponent, Robert Chambers. More specifically, Chambers (2003) has flagged the following issues which may undermine data quality as well as the external validity, or ‘representativeness’, of results:

- i. poor quality facilitation of dialogical processes, whereby facilitators ‘dominate, disempower and distort’;
- ii. agenda framing, whereby research processes and results are unduly influenced by the analytical perspectives of facilitators;
- iii. unrepresentative participation, such that the views and interests of a dominant group in a community, or of dominant individuals in focus groups, are over represented and marginalized groups left out;
- iv. unrepresentative site selection, due to purposive sampling of atypical communities.

Such shortcomings can lead to analytical distortions and inaccuracies which are compounded by: large amounts of data; data which are not comparable or difficult to compare; analysts with strong preconceptions and mental templates; pressure and incentives for early and striking policy messages. Specifically, the following types of biases and distortions may arise:

- i. incomplete coverage of the data;
- ii. misleading or inconsistent classification of information;
- iii. selective searching for data which fit the analyst’s preconceptions;
- iv. successive simplification and editing, excluding qualifications and exceptions;
- v. unsubstantiated assertions, in particular omission of qualifications about data;
- vi. oversimplification of complex realities;
- vii. over-attention to striking quotations;
- viii. overgeneralization.

This discussion of the potential shortcomings of participatory approaches is meant to serve as a cautionary note in the context of the review of the significant value-added of PRA techniques in impact assessment in section 3. It also bears on questions of ‘ownership’ and ‘policy buy-in ’ associated with consultative processes. There are two core issues here. The first concerns community ‘ownership’ of research results and policy recommendations generated using participatory techniques. Clearly, this will depend on the nature of ‘participation’ in these exercises, and whether results reflect local perspectives and not those of the facilitator, for

example. Otherwise stated, it depends on whether or not the above mentioned methodological critiques apply to particular PRA exercises.

The second issue concerns feedback of research results into the policy process through consultative processes, often involving Civil Society Organisation (CSOs) and actors. The research in question may include impact assessment studies, other monitoring and evaluation exercises and so forth. At times, such feedback mechanisms are labelled as ‘participatory’ in so far as they involve consultative processes and may be informed by the results of PRA research. This issue is addressed in section 4.2, though it may be noted here that there is cause for some pessimism about the role of ‘participation’ in this context.

2.3 Employment and Decent Work: Components and Interrelationships

A core question in any impact assessment exercise is ‘impact with respect to what?’. A potential contribution of ‘qualitative’ and mixed method approaches is to incorporate locally meaningful dimensions of the measures of impact into the analysis. By locally meaningful, we are referring to dimensions of well-being, among other things, which are relevant to, and valued by, populations who stand to be affected by the policies under review, such as trade reform.

In the context of the present study of the employment effects of trade, a useful starting point may be to revisit what constitutes ‘employment’, that is to see what dimensions ‘employment’ should take on for it be locally meaningful and how these dimensions would relate to ‘decent work.’ The idea of decent work was forcefully articulated in the *1999 Report of the ILO Director* who provided the following definition: ‘opportunities for women and men to obtain decent and productive work in conditions of freedom, equity, security and human dignity’. The definition is broad and open to different interpretation.

For example, Anker et al. (2002) have identified six relevant dimensions of decent work highlighted in this definition namely:

1. Opportunities to find work, defined broadly to include self-employment, unpaid family work and wage employment;
2. Conditions of freedom, which rule out forced or bonded labour, as well as forms of child labour;
3. Productive work, or work which ensures acceptable and sustainable standards of living;
4. Equity, or fair and equitable treatment in the workplace and opportunities for employment;
5. Security, which entails provision of social protection in the event of contingencies, along with pensions and safe working conditions.
6. Dignity, which includes respectful treatment, participation in decision-making and consultations, and opportunities for collective action.

They further subdivide these dimensions into ten categories, which are used for indicator construction: 1) employment opportunities; ii) unacceptable work; iii) adequate earnings and productive work; iv) decent hours; v) stability and security of work; vi) balancing work and family life; vii) fair treatment in employment; viii) safe work; ix) social protection; x) social dialogue and workplace relations. Others, such as Dhai (2003), have suggested an overlapping, yet distinct, set of categories which includes: i) employment opportunities; ii) remunerative employment; iii) conditions of work; iv) social security; vi) forced and child labour; vii) discrimination at work; viii) freedom of association; ix) collective bargaining; x) economic democracy and xi) participation.

It is clear that decent work is a far broader concept than that of employment. The focus of this paper is the latter concept, specifically the quantity and quality of employment. In terms of Anker's, six dimensions of decent work, we are concerned here primarily with work opportunities (#1), productive work (#2), equitable treatment in the work place (#4) and safe working conditions (#5).

The multidimensional nature of employment raises issues of inclusion/exclusion and weighting when addressing the core question 'impact with respect to what'. Specifically, which of the various dimensions of employment should be included and what weight should be accorded those dimensions. These issues are of less significance if the various components are closely correlated so that levels and change in the variables are the same. In such cases, the various components of employment will serve as good 'proxies' for one another, and will constitute good indicators of employment *tout court*, which is important for monitoring and evaluation purposes, for example.

The empirical record, however, suggests significant trade-offs between different components of the quantity and quality of employment. Such issues play out quite acutely in the context of trade policy reform, as evidenced in the findings of many of the empirical studies in section 3. Some of the trade-offs include: increasing work opportunities at the expense of fair and equitable treatment, as in cases where female workers are exposed to sexual harassment in Export Processing Zones (EPZ); conflicts between safe working conditions and work opportunities, as exemplified by dangerous conditions in factories in export industries, and instances where increased work opportunities impose heavy time burdens on female workers to the detriment of caring for family members and other such tasks. As discussed as in section 3.2, there are a range of methods which have been used to attempt to determine the relative importance of competing measures of impact, such as employment.

2.4 Methodological Choices in Impact Assessment

As discussed in the Introduction, there are a wide range of methodological tools and approaches which may be used in impact assessment, including the impact assessment of trade on employment. A core argument of this paper is that the actual choice of methods should be driven by the research questions and the types of information required by users of research results, including policy makers, the research community, civil society organisations and others.

The issues discussed below are not meant to be exhaustive but relate closely to the types of research questions and results found in the empirical studies presented in section 3. We distinguish between general considerations, issues related to the impact measure, or ‘left hand side’ in the language of econometrics, and issues concerning the entire causal system linking causal variables and impact measures.

General Considerations

Extent of Existing Knowledge: A first principle of methodological choice is that the less known about the object of inquiry, the greater the case to begin with open-ended data collection techniques which allow for flexible questioning and emergent issues. This point becomes increasingly relevant as the complexity of the empirical phenomena in question increases. Many of the so-called dialogical techniques involving detailed discussion with research participants, such as focus-group discussions and semi-structure interviews, are examples. As discussed further in section 3, information generating from such sources can subsequently be transformed into fixed response household survey questionnaire modules, such as questions on reasons for impact, or may be used for econometric model specification and hypothesis testing. It should be emphasised however, that this point does not relegate such ‘qualitative’ open-ended techniques to the context of ‘discovery’ as opposed to the context of ‘justification’, sometimes reserved for ‘quantitative’ approaches.⁵ Verification/validation is often an express objective of dialogical techniques and rarely straightforward in statistical analyses of numerical data which are also equally part of a process of ‘discovery’.

Perceptual Information: A second issue concerns the importance of eliciting people’s perceptions. In the context of impact assessment, perceptual knowledge includes, *inter alia*, normative considerations, such as the specification of locally meaningful measures of impact and their weights, along with people’s understandings of the magnitude of impact, the mechanisms generating impact, and behavioural responses in hypothetical scenarios (counterfactual thought experiments). It should be emphasised that in our definition, perceptual information comprises

⁵ This is an old distinction in the philosophy of science (Popper, 1959; Suppe, 1974), variants of which are sometimes applied to distinguish ‘qualitative’ and ‘quantitative’ inquiry in the social sciences (Starr, 2004).

primarily unobservable data.⁶ In terms of employment, we are not referring to perceptions of days or hours worked asked in labour market surveys, for example, which are only a second-best estimate of data which are observable in principle. The perceptions in question concern such issues as trade-offs between dimensions of employment and lived experience of the processes generating impact.

External Validity: A final general consideration concerns whether or not research results are required to be representative of broader population groups, and the basis of the extrapolation exercise from sample to population. For certain policy purposes, statistically representative data are required in the sense that standard errors are calculated for the statistics estimated, which requires some form of probabilistic sampling. In other cases, representative information may not be required, or the basis of the extrapolation exercise may rely on empirical generalisation rather than statistical information. Here, a judgement is rendered on the ‘typicality’ of research findings across broader population groups, though the basis of this judgement is always open to debate (Hammersley, 1992). An interesting hybrid has been to use probabilistic sampling for ‘subjective’ modules in household surveys, with questions about the magnitude of, or reasons for, impact, which allows for standard error estimates of the statistics calculated.

The Impact Measure

Numerical or ‘Narrative’ Information: A first question concerns whether a numerical estimate of impact is required, rather than a narrative account. For many policy purposes, numerical information is important because it carries more weight in policy circles and provides a firmer basis on which to make resource allocation decisions. In other cases, however, narrative information may be more appropriate as discussed above in the context of perceptual information. Such studies add value in their own right, and ‘bring alive’ numerical information on impact measures.

Precision: A second consideration involves the degree of precision of the impact measure required. By precision, we are referring to the scale in which the impact variable is measured and not statistical precision, or the required confidence interval of an estimate. For some purposes, finer precision is required which necessitates use of a cardinal (or ratio) scale, of hours or days of employment, for example. In other cases, ordinal categories may be sufficient, which allow for a relative ranking of impact in terms such as ‘more hours of employment, fewer hours, no difference.’

⁶ The distinction between observables and non-observables is one way of differentiating broad traditions of inquiry in the social sciences, along with the ‘qualitative/quantitative’ divide (Kanbur and Shaffer, 2007; Shaffer, 2013).

The Causal System

Outcomes or Processes: When considering the overall causal system, a first question concerns whether to emphasise causal outcomes (effects) or the processes (mechanisms) generating them (causal mechanisms are defined in section 3.3 below). The distinction between outcomes and processes has some cutting power in the context of the ‘quantitative/qualitative’ divide, as discussed in the broader literature.⁷ As discussed above, for many policy purposes, it is quite important to estimate the magnitude of causal effects (with relatively high precision) and accordingly, answer the ‘how much’ question. In other cases, much more attention is devoted to understanding the underlying causal mechanisms at work, addressing the ‘how’ and ‘why’ questions.

‘Thick’ or ‘Thin’ Description⁸: A related concern is about the degree of detail of the analysis, and in particular, the account of causal processes or mechanisms. All causal analyses include some depiction of processes but differ with respect to the empirical grounding and ‘richness’ of their representations. Analysis which prioritise causal outcomes/effects, such as many of the modelling exercises discussed in Gibson (2011), often present quite a ‘thin’ account of mechanisms, drawing on tenets of economic theory or the predilections of the analyst (as discussed in sections 3 and 4). By contrast, studies that focus on causal mechanisms, typically relying heavily on narrative information and dialogical processes, often present a much ‘thicker’ and more empirically grounded account of processes.

Causal Weighting (and Precision): A final point concerning the importance attached to assigning causal weights (including zero weights) to the many factors potentially affecting causal outcomes and the required precision of the weights. The strength of econometric modelling, in theory, is to provide such weights (after conditioning on all other relevant causal variables). Dialogical information generating processes do not do this, but may provide information on people’s perceptions of the relative importance of causal variables through ranking exercises. Typically the precision of the ensuing weights are lower, expressed in broad ordinal categories such as ‘very important, important and negligible’. An interesting hybrid approach, discussed in section 3.7, involved incorporating dialogically generated information on ‘reasons’ for impact in econometric models to estimate their causal weights.

⁷ See Lipton (1992), Bardhan and Ray (2006) and Shaffer (2013).

⁸ In the broader literature, ‘thick’ description is used in a somewhat different way to denote interpretative or hermeneutic explanation in applied anthropology (Geertz, 1983).

3. ‘Qualitative’ and Mixed Method Studies in Impact Assessment

The sections which follow outline ways that so-called ‘qualitative’ and mixed methods approaches can contribute to impact assessment of policies and programs. Empirical studies are presented which illustrate the methodological issues at hand. Where possible, the selected empirical cases will address trade and employment. In other cases, the methodological issues discussed remain relevant, as they generalise to impact assessment of trade on employment.

The review is not an exhaustive account of the ways that ‘qualitative’ or mixed method approaches have been used or fruitfully combined. It does not address subject areas outside impact assessment where ‘qualitative’ or mixed method approaches have been used, such as wage and price theory, innovation and industrial organisation, willingness to pay and environmental economics and so forth in economics (Starr 2014). Further, there is no discussion of some of the ‘generic’ ways to fruitfully combine research approaches such as using ‘qualitative’ methods to generate hypotheses for subsequent (econometric) testing, to improve household survey design (Chung 2000) and so on. Finally, there are other issues within the impact assessment literature where ‘qualitative’ and mixed methods have been used which are less relevant for the present purposes such as comparison group construction in program evaluation (Rao and Ibáñez 2005), which differs from methodologies typically applied in policy-level assessment.

3.1 Lived Experiences and Locally Meaningful Measures of Impact

The most general use of ‘qualitative’ approaches in the context of impact assessment is to provide information on lived experiences associated with some event, such as a natural disaster, policy change, such as trade reform, or related activity. There have been quite a number of studies which have attempted to document the lived experiences of communities in the aftermath of aspects of trade reform.⁹ Typically, studies of this sort provide narrative information generated by methods such as participant observation, focus group discussions and semi-structured interviews.¹⁰

The main contribution of accounts of ‘lived experience’ is to provide a ‘thicker’ description of the mechanisms generating impact, along with ‘locally meaningful’ measures or ‘metrics’ of impact. As mechanisms are discussed in greater detail in subsequent sections, the focus here is on impact measures.

⁹ A shortlist of recent studies includes: Hamilton and Fischer (2003); Kristiansen (2004); Hancock (2006); Hunt et al. (2007); Blin (2008); Chang (2008); McCarty (2008); Silva (2008); Campbell et al. (2010); Vásquez-León, (2010); Aistara (2011); Hellin et al. (2012); Oueslati-Porter C. (2013); Ansel et al. (2015).

¹⁰ These approaches figure centrally in the discussion of ‘qualitative methods’ in Gibson’s (2011: 113-114) companion piece.

Within such disciplines as cultural anthropology, the importance of understanding locally meaningful categories, or in this case impact measures, is often emphasised. The core reason is to avoid imposing categories which may not be locally relevant. Otherwise, serious conceptual and analytical biases can arise: ‘we interpret all societies in the categories of our own’ (Taylor 1985: 2). The corollary in the context of impact assessment is using measures which imperfectly correspond to relevant objects of value, or dimensions of well-being, of local populations.

There are a few preliminary points to clarify about the limitations of these types of studies, drawing on the discussion in section 2.4:

- *Precision*: In most cases, narrative information is presented about aspects of life which have changed. At times, such changes are represented on an ordinal scale, using categories such as ‘better’, ‘much better’, and so on. Typically, the information collection does not allow for estimates of impact on a cardinal scale, nor for estimation of confidence intervals.
- *Causal Inference*: The assumption underlying these studies is that a policy change, or some such event, is causally relevant to changes in lived experiences. There is no attempt, however, to causally weight the policy change in question relative to other potential causal factors. Typically, the claim of ‘causal relevance’ is based on the account of mechanisms generating the causal effect in question (see section 3.3).
- *External Validity*: Many of these studies are based on a limited number of observations. In addition, in many (though not all) cases, sampling is not done probabilistically and accordingly, standard errors cannot be calculated. As a result, most studies of this sort do not claim to be ‘representative’ of broader population groups, at least in a statistical sense.

A few examples will serve to illustrate certain of the insights provided by these studies.

McCarty (2008) examined the impact of the North American Free Trade Agreement (NAFTA) on the welfare of children and women in rural Mexico. Her analysis drew largely on semi-structured interviews and focus group discussions with 65 women in the states of Guanajuato and Hidalgo over a 9 month period in 2006-2007. The analytical focus of the study was on the effects of NAFTA-induced changes to emigration patterns for communities of source in Mexico. Respondents highlighted the loss of livelihoods for many small-scale rural producers who were unable to compete with agricultural imports from the US and who became dispossessed from their land after the privatization of so-called *ejido* holdings. As a consequence, emigration soared but differed from historical patterns in two ways: i) emigration to the US replaced emigration to urban centers within Mexico; ii) longer-term or permanent emigration became more widespread due to increasingly stringent border enforcement, replacing decades-old patterns of cyclical migration with home visits on a regular basis. The effects on communities of source have been far-reaching. In some cases, men abandon their families and start anew in the United States, leaving their families in Mexico in difficult circumstances due to the absence of remittances. According to the

women interviewed, their situation is compounded by the lack of viable employment options to supplement meagre, or non-existent, remittance flows. Generally, the increase in the scale of migration among family members has led to the loss of emotional support, childcare and community participation, and heightened fear and insecurity of women and children. These sentiments were widely felt even in the presence of remittance income which raises important questions about trade-offs with respect to impact measures.

Two additional studies, by Blim (2006) and Hancock (2006), have chronicled the experience of female workers in export processing zones (EPZs). In both cases, important questions were raised about trade-offs between income gains and other dimensions of well-being. The first study combined econometric analysis of determinants of labour supply in EPZs in Mauritius with detailed semi-structured interviews about working and living conditions. A core issue to emerge was the trade-offs between the imperatives of generating income to meet present and future needs, on the one hand, and devoting sufficient time to care for children and relatives, on the other. Many women felt that ‘they could not give their children the care and attention they needed’ (Blim, 2006: 250).

A different type of trade-off was found in Hancock’s (2006) study of female factory workers in Sri Lanka which combined information from a fixed-response questionnaire from focus group discussions and semi-structured interviews. Certain benefits associated with increased income were reported including better problem-solving capabilities and heightened independence/autonomy. The costs, however, included increased health hazards (dust, poisonous gases and chemicals), ailments (pain, illness, depression) and the pronounced social stigma (and sexual harassment) associated with factory work. According to one respondent: ‘Society has a very bad perception of women factory workers. They look down on us and consider us as immoral people. We are not accepted and people disrespect us’ (Hancock 2006: 235).

A final example concerns the evaluation of an (involuntary) resettlement scheme following a highway development project in Sri Lanka (De Silva and Gunetilleke’s (2008)).¹¹ The methodology included both a household survey, with fixed response and open-ended questions, as well as focus group discussions. The standard indicators used in the evaluation included measures of payment allowances, replacement of agricultural land, physical quality of housing and access to basic utilities. In general, these indicators presented a quite favorable assessment of the resettlement process. On the other hand, focus groups discussions of participants presented a different picture:

Shared ownership of lands among families, the informal social networks where housework such child care is often shared, and open access to assets within the extended family, are characteristics of these villages which the [Southern Transport

¹¹ This discussion is based on Shaffer et al. (2008: 142).

Development Project]STDP has caused to be suddenly severed ... Despite making resettlement decisions that allowed them to maintain their social networks, the whole process of relocation and the change it stimulates has an impact on social well being. A major articulated loss is the loss of the traditional/ancestral village and the lifestyle that goes with it.

Such studies accentuate the fact that the ‘metric matters’ as it may drive the overall assessment of the benefits and costs of policy changes. In all of the case studies, trade-offs were reported which related closely to those identified in the discussion of employment and decent work in section 2.3. Specifically, at times, the benefits of increased work opportunities came at the expense of productive work, equitable treatment in the work place and safe working conditions. These findings direct attention to techniques for weighing the relative importance of different measures of impact.

3.2 Locally Meaningful Weights of Impact Measures¹²

‘Locally meaningful’ weighting schemes can be based on people's perceptions of the relative importance of different dimensions of the impact measure. The literature on the use of mixed method approaches in poverty analysis provides a number of good examples of how to derive ‘locally meaningful’ weighting schemes. Such analyses employ either indirect approaches, which attempt to retrieve weights from correlates of poverty, or direct approaches, which simply ask people to supply the weights in question. These approaches are not specific to poverty and apply, in principle, to any impact measures including dimensions of employment and decent work.

The Indirect Approach

An example of the indirect approach involves attempts to ‘back-out’ weights through econometric analysis. One study of 37 villages in rural districts of Kenya, Malawi, Tanzania and Uganda drew on a dataset which combined wealth ranking results and household survey data (Kebede 2009). As mentioned in section 2.2, in wealth ranking, community members are asked to rank households (or persons) in terms of their own conceptions of wealth or well-being. An ordered logit model was estimated with wealth rank (poor, middle, rich) regressed on household characteristics including income, assets, land, number of adults and housing. The author argues that the resulting coefficients represent an approximation of the social value accorded to such resources as determinants of wealth ranking categories, though the value of less visible resources will be biased downwards.

The approach attempts to integrate statistical analysis of conditional relationships between variables with locally meaningful definitions of poverty as reflected in wealth ranking results. The

¹² This section is based on Shaffer (2013: 53-55).

main assumption is that such variables actually entered into the considerations of those engaged in the wealth ranking exercise. To assess the validity of this assumption, it would have been interesting to compare such results with the reasons given for the categorisation of persons into well-being rankings categories. Nevertheless, it represents an innovative attempt to infer locally relevant weights from observable characteristics of well-being ranking groups.

The Direct Approach

In the direct approach to eliciting local weights, people are simply asked to rank dimensions of poverty or well-being in terms of their relative importance. The direct approach may involve either sequential or simultaneous ranking. One example of the former involves the construction of a composite poverty index, the 'Human Vulnerability Index (HVI)', for the Maldives drawing on data from Vulnerability and Poverty Assessments (VPAs) carried out in 1997/8 and 2004 (de Kruijk and Rutten 2007). The VPAs were nationally representative surveys which covered all 200 inhabited islands along with the capital city, Male. These surveys asked respondents to rank twelve dimensions of well-being in terms of their perceived priority (with the highest priority assigned the value of 1). These 12 dimensions included indicators of: income poverty, electricity, transport, communication, education, health, drinking water, consumer goods, housing, environment, food security and employment. Rankings were averaged separately for men and women, though gender-disaggregated results ended up being identical, and relative weights calculated for use in the HVI.

A second example of sequential ranking is an approach developed by The *Wellbeing Research in Developing Countries (WED)* project at the University of Bath, UK. This project produced the Quality of Life (QoL) questionnaire, which attempted to integrate standardized instruments, such as those in the World Health Organizations' QoL surveys, with open-ended, individualized questionnaires, such as the *Global Person Generated Index* (Camfield et al. 2009, Gough and McGregor (eds.) 2007). As an example, the development of the WEDQoL questionnaire in Thailand, (Woodcock et al. 2009), began with a preliminary phase in which respondents from rural and urban communities were asked 'Describe a time when you felt very happy, giving reasons'. The results were codified into 51 items in the WEDQoL questionnaire, such as 'having sufficient food every day, water, friends, good relationships, public transport etc.' Respondents then rated these items with respect to their perceived necessity for well-being, on a scale of 0–2, and to their satisfaction with the levels of these items, on a scale of 0–3. The mean or median of such scores provided an indication of the relative importance and attainment of dimensions of poverty/wellbeing respectively, while disaggregation allowed for sub-population specific assessment. Additional analysis was performed on the data to assess internal validity.

A drawback of sequential ranking as a means of eliciting weights is that the individual scoring does not necessarily take into account the relative importance of all items jointly considered. Accordingly, individual weights may be biased in any number of ways. One method to address this 'sequential bias' was used in a 2008 study on the relative importance of different literary practices in Mozambique (Esposito et al. 2012). Focus group discussions were initially held with

a view to elicit a short-list of five literary practices which were most highly valued, namely: i) signing one's name; ii) performing simple calculations; iii) dealing with official documents; iv) using mobile phones and v) helping children with homework. In a subsequent survey, data on household characteristics were first collected. Next, respondents were asked to allocate fifty beans among the five pre-identified literary practices according to their relative importance (this is an example of a 'participatory' ranking exercise discussed in section 2.2). Subsequently, analysis was performed to assess the statistical significance of differences among response categories and between populations groups. In terms of the ranking exercise, respondents were forced to simultaneously evaluate the importance of all five practices, thus addressing sequential bias. The drawback, as the authors acknowledge, is that the procedure is unable to detect overall differences in valuation levels across persons because each were assigned a fixed number of beans.

In all of these examples, mixed method approaches have been used to incorporate locally meaningful weights in determining the relative importance of dimensions of poverty or well-being. All of the approaches have shortcomings, but nevertheless, offer a range of options from which to choose.

3.3 Self-Reports of Impact

Another direction in the use of 'qualitative' and mixed method approaches is to elicit local perceptions of the magnitude of impact. Typically, such self-reports (or group reports) are generated through fixed response or open ended questions in household surveys, focus group discussions and semi-structured interviews. Meaningful answers to the question, 'what impact has policy/program x had on impact measure y ', require a mental exercise predicated on an understanding of the mechanisms linking x and y .¹³ More broadly, causal inference in self-reports depends upon identifying the causal mechanisms generating causal effects: 'To assert that A's are causes of B's is to assert that there is a typical causal mechanism through which events of type A lead to events of type B' (Little 1998: 202).

One preliminary point concerns what is meant exactly by 'mechanisms'. In the social sciences, the term has been used to refer to a variety of things including: intervening variables linking causes and effects, mid-level theories which provide information about particular elements of higher-level theories and unobserved entities which have causal effects (Hedström and Swedberg (eds.) 1998; Mahoney 2001). In the context of impact assessment, mechanisms typically refer to the causal processes generating observed outcomes. Processes comprise the causal variables, the links or pathways between them, i.e. the causal 'tree', as well as an explanation of why they are linked. Mechanisms, in this sense, focus on the reasons for observed outcomes. They shed light on the

¹³ This discussion is based on Shaffer (2011, 2012, 2013).

‘how’ and ‘why’ questions surrounding causal effect and form the basis of valid self-reports of the magnitude of impact.

As with the analysis of ‘lived experiences of impact’ discussed in section 3.1, there are a number of limitations of self-report-type data. Biases in either direction may be introduced due to the underlying cognitive processes at work (Gilovich and Griffin 2002) and survey design, due, *inter alia*, to the ordering and phrasing of questions (Sudman et al. 1996). In addition, validity tends to diminish:

- i. the larger the causal distance between policy/program activities and impact measures;
- ii. the greater the number of intervening variables affecting impacts;
- iii. the more complex the pattern of interaction among variables;
- iv. the greater the *changes* in causal variables, and *changes* in their interactions, over the time period of the evaluation and;
- v. the finer the scale in which the outcome/impact variable is measured.

Still, it does not seem unreasonable to ask, for example, about the impact of a water supply project on the time spent collecting water, in broad ordinal categories, where few other intervening variables are driving outcomes and/or where the relative importance of other variables is known to be limited (White 2009: 7). The validity of results is an empirical question, which is partially addressed in the second case study, below.

A first example of self-report data on impact is Houssa and Verpoorten’s (2015) study of the welfare effects of a self-imposed ban on shrimp exports in Benin to pre-empt sanctions by the EU for violations of phyto-sanitary standards. The ban was imposed in 2003 and remained for around 20 months until 2005. The study sought to determine the immediate and mid-term impact of the ban on income along with the coping strategies employed in response. The core source was a fixed response household survey administered in 2009 to a sample of 540 households comprising fisherman and fishmongers (usually the spouses of the fishermen). The questionnaire contained a detailed income and consumption module, questions on household characteristics, self-reports on the magnitude of, and reasons for, impact along with questions on coping strategies employed in the event of negative impact.

With respect to the self-report findings, 85% of respondents indicated that the short term impact of the ban (in 2003) was either ‘strongly’ or ‘rather’ negative. A similar proportion of respondents (82%) considered the medium-term impact (in 2009) to be ‘strongly’ or ‘rather’ negative. These findings are consistent with theoretical results of a simple open-economy supply/demand framework where losses are generated primarily through the loss of the fishermen’s producer surplus in the high-quality shrimp market. They are also consistent with empirical results of semi-structured interviews with traders who reported that the expansion of exports of high quality shrimp to neighbouring countries did not compensate for the loss in access to the EU market. In addition, evidence that shrimp producers did not resume production for export supports the finding

of medium-term negative effect. In terms of the above discussion, key factors enhancing validity included the overriding causal importance of the export ban on the livelihoods of fisherman, the absence of major changes in other causal variables and the fact the broad ordinal categories were used to measure impact, as opposed to measures with higher levels of precision. While it is impossible to assess with certainty the validity of the estimates of impact, these other sources of information suggest that the sign is correct.

A second example is Shaffer's (2012a, 2013) impact assessment of the national Hunger Eradication and Poverty Reduction (HEPR) in Vietnam. The HEPR program comprised a number of targeted projects and policies on health care, education, and social support for the poor. The methodology combined information from a 'Qualitative' Survey (QS) with results of a Propensity Score Matching (PSM) exercise. In the QS, respondents were asked a preliminary, open-ended series of questions on the mechanisms generating impact prior to a follow-up question on the magnitude of impact. Sampling for the QS was done probabilistically which allowed for calculation of standard errors for the statistics estimated.

PSM is a statistical technique which matches household based on their probability of program participation (the propensity score), estimated using logit regressions. The HEPR study assessed the sensitivity of a prior World Bank PSM analysis to the choice of comparison group by presenting results for the nearest one, three and five matched non-beneficiaries¹⁴. In addition, standard errors were calculated and confidence intervals presented for the impact estimates. The data source for the PSM was the *Vietnam Household Living Standard Survey 2002* (VHLSS), a multi-topic nationally representative survey, which contained a module on participation in specific HEPR projects.

In the HEPR impact assessment, the two approaches were combined to assess the credit component of the program. Credit is comprised of small loans disbursed by the Vietnam Bank for the Poor (VBP) for investment in income-generating activities, such as livestock rearing and agriculture. In the QS credit module, an open-ended question was posed and phrased as follows : "What did you do with the loan that you received from the Vietnam Bank for the Poor". As with all the open-ended questions, a list of positive and negative probes was drafted to assist enumerators in probing the mechanisms in question. A followed-up question asked respondents to rank project impact into four broad categories: significantly positive; insignificant; none or significantly negative. The question was phrased as: "Taking into account everything you have told me, what has been the impact of the loan on your household income?"

Around half of respondents ranked the impact of credit on household income to be significantly positive, with around one-third of respondents maintaining it was insignificant. This finding presents a more favourable assessment of program impact than the results of the PSM exercise,

¹⁴ Results were published in the JDR (2004) and appear in Cuong (n.d.).

which found a statistically insignificant effect of credit on household expenditure per capita. There are a number of potential reasons which explain these differences. First, the time frame for the PSM was the twelve month period prior to the VHLSS, whereas the time period used in the HEPR study was between 1999 and 2002. Second, 'impact over time due to credit' is not the same thing as 'impact relative to the situation of the comparison group'. For example, a credit program with significant impact upon income over time may show no or little impact in situations of substitution bias (Heckman 2000), whereby comparison group members benefit from substitute credit. Third, the impact measures used in the two approaches were different, namely consumption per capita for the PSM and population percentages falling into ordinal impact categories based on household income in the self-report analysis. Despite these possible explanations, caution is urged with respect to the self-report findings given their divergence with the PSM results.

The above examples, show the possibilities and limitations of self-reports. In certain situations, they may provide sensible estimates of the magnitude of impact in broad, ordinal categories. The main limitation concerns potential biases which may undermine validity and the difficulties in knowing the size or sign of potential biases.

3.4 Counterfactual Thought Experiments (CTEs)

An additional tool of impact assessment, closely related to self-reports, is a Counterfactual Thought Experiment (CTE). In this approach, subjunctive conditional (if/then) questions are posed about some hypothesised link in the causal chain generating impacts. As with self-reports, such information is generated through fixed response or open ended questions in household surveys, focus group discussions and semi-structured interviews. The same considerations as above apply about the conditions under which CTEs are likely to be more successful, along with the caveats about internal validity.

In the examples below, the focus is on hypothesised behavioural responses to policy or program interventions. Otherwise stated, the emphasis is on one or more specific mechanism generating impact. A distinction is made between CTEs about future (*ex ante*) and past (*ex post*) events.

Ex Ante Analysis

An example of *ex ante* analysis is provided in Shaffer's (2008) mixed method analysis of the response of rural producers in Vietnam to reform measures. One part of this broader study focused on the likely response of sugarcane growers and livestock producers to relative price changes following Vietnam's accession to the World Trade Organisation (WTO). A core objective was to inform policy discussions about populations who were less likely to successfully adjust to such price shocks. As with the HEPR impact assessment discussed above, a so-called 'qualitative' survey was conducted over the period January-February 2007, involving approximately 500 households in twenty communes in Than Hoa province, sampled probabilistically.

The survey asked respondents if sugarcane prices fell by 25%, whether or not they or any household members would “makes changes in the way their household earns income.” The twenty five percent figure was an estimate of the effect of the reduction of the tariff on sugar imports after WTO accession. Around two-thirds of respondents reported that they would make changes (‘adjusters’). Interestingly, there were very small differences in responses between poor and non-poor households, which were not statistically significant at the 95% level.

A follow-up question was posed to ‘non-adjusters’. They were asked why they wouldn’t “switch to other types of activities such as cultivating different crops, raising livestock, etc.” A pre-populated list of reasons were included in the survey, based on results of prior focus group discussions and semi-structured interviews. Results are produced in Table 1 below. Of the many interesting results, it is of policy-relevance to note that lack of access to credit was cited by around 32% of poor compared to only 12% of non-poor respondents.

	2006		2006		Total
	Poor	Non-Poor	Kinh	Non-Kinh	
Sugarcane is still profitable	41.13 [8.66]	55.41 [5.44]	53.20 [5.31]	45.72 [9.42]	50.72 [4.73]
Expect sugarcane price to rise in the future	39.87 [8.76]	41.35 [5.25]	39.36 [4.94]	45.20 [9.27]	41.30 [4.52]
Land unsuitable for other crops	29.79 [7.76]	36.59 [5.51]	36.14 [5.08]	30.56 [9.10]	34.28 [4.54]
Under contract to Sugar Factory	35.34 [8.42]	32.15 [4.94]	34.18 [4.86]	30.30 [8.24]	32.89 [4.26]
Lack of information about other opportunities	0.00 [0.00]	2.83 [1.36]	2.28 [1.25]	1.30 [1.31]	1.95 [0.94]
Lack of labor for other activities	4.97 [2.91]	9.24 [2.55]	10.15 [2.67]	3.34 [2.43]	7.89 [1.98]
Lack of credit	31.99 [9.25]	11.84 [3.68]	16.07 [4.26]	23.47 [8.42]	18.53 [4.01]
In first or second year of the sugarcane crop cycle	17.83 [7.54]	13.84 [3.09]	13.50 [2.96]	19.91 [7.48]	15.63 [3.21]
Already invested heavily in sugarcane	8.10 [4.15]	12.20 [3.25]	12.25 [3.27]	8.11 [4.06]	10.88 [2.58]
Government advice to continue sugarcane	22.47 [7.67]	14.40 [3.79]	14.50 [3.59]	21.22 [7.68]	16.74 [3.52]
Government directive on land use	33.08 [8.69]	45.49 [5.48]	43.07 [5.22]	38.13 [9.23]	41.43 [4.65]
Government negotiation with factory to support prices	2.26 [2.26]	3.03 [1.54]	3.51 [1.76]	1.30 [1.31]	2.78 [1.26]
Other	3.21 [2.43]	11.67 [3.70]	8.37 [2.30]	10.38 [6.66]	9.04 [2.71]

Standard errors in brackets

The core contribution of these CTEs, for policy purposes, is to provide a sense of the magnitude of likely problems associated with the failure to adjust to trade policy reform and importantly, the reasons for such lack of behavioral response. Certain of these reasons cited above have direct policy relevance with respect to facilitating and mitigation measures for improving the welfare impact of trade policies, such as credit for the poor. Further, the fact that a probabilistic sampled survey was conducted allowed for estimates of standard errors, underpinning the external validity of results (within the universe of sugarcane producers in Than Hoa province).

Ex post Analysis

A second example is provided in the impact assessment of the national Hunger Eradication and Poverty Reduction (HEPR) in Vietnam (Shaffer 2012a, 2013), discussed in the last section. In this study, CTEs were combined with propensity score matching analyses to assess the following two project components: the Health Fee Reduction or Exemption program, which assessed utilisation of health services, and the Tuition and School Maintenance Fee Exemption or Reduction, which assessed primary and secondary enrolment. The CTE posed subjunctive conditional (if-then) questions about what respondents would have done in the absence of the program. Unlike the micro-credit assessment discussed above, the impact measures used in the PSM and CTEs were very similar, allowing for a more direct comparison of results.

The Health Fee Exemption or Reduction entailed providing poor households or communes free or subsidised health care. The impact measure, health care utilisation, was defined as the percentage of persons who used health care facilities over the past 12 months. The thought experiment asked respondents whether or not they still would have sought medical attention when they were ill if they had not received the health fee exemption or reduction. Around 92% of respondents maintained that they would have still sought medical care when they were ill even if they had not benefited from the health fee exemption or reduction. Taking into account sampling error, and omitting the 'don't knows' from Table 2, up to 95% of respondents said they would have sought medical care in the absence of the program.

**Table 2 Use of Medical Care in the Absence of HEPR
(Population Proportions, Standard Errors in Parentheses)^a**

	1	2	Total
	Yes	No	
Total Vietnam	91.8 (0.40)	7.3 (0.11)	100

^aData do not sum to 100 because "Don't Knows" have been removed
Data source: HEPR Impact Assessment Qualitative Survey, 2003-4

Table 3 presents results of the propensity score matching analysis which provides a basis for comparison. Data suggest that the program has not had a statistically significant impact on utilisation of healthcare services (excluding traditional healers). These results are very similar to, and provide an explanation for, those of the propensity score matching. The insignificant impact of this program on health care utilisation rates, relative to non-participants, may simply be due to the fact that most people would pay for health services in the absence of the program.

Table 3 Propensity Score Matching: Impact of Health Fee Exemption/Reduction on Utilisation of Healthcare

	Mean Difference	Standard Error^a	95% Conf. Interval
Nearest Match	-0.09	0.06	-0.23 - 0.02
Nearest Three Matches	-0.08	0.05	-0.20 - 0.01
Nearest Five Matches	-0.08	0.05	-0.18 - 0.01

^aStandard errors were bootstrapped with 100 replications

Data source: Vietnam Household Living Standards Survey, 2002

The second relevant program component, the Tuition and School Maintenance Fee Exemption or Reduction, provides students in poor households, as well as certain other eligible groups, exemptions or reductions in the amount they must pay for tuition and the maintenance of schools. In the CTE, around 12% of respondents claimed that they would not have enrolled their children in primary or secondary school in the absence of the program. The PSM results presented in Table 4 show, a modest, but statistically significant, impact of the program on school attendance for all three matched comparisons whose results are presented below. The impact range is between 3% and 15% depending on the comparison group used and taking into account sampling error.

Table 4 Propensity Score Matching: Impact of Tuition and School Maintenance Fee Exemption/Reduction on School Attendance

	Mean Difference	Standard Error	95% Conf. Interval
Nearest Match	0.09	0.03	0.03 - 0.14
Nearest Three Matches	0.09	0.03	0.06 - 0.15
Nearest Five Matches	0.08	0.02	0.04 - 0.13

^aStandard errors were bootstrapped with 100 replications

Data source: Vietnam Household Living Standards Survey, 2002

The above examples have demonstrated the use of counterfactual thought experiments on the magnitude and reasons for impact in *ex ante* and *ex post* impact studies. The Vietnam studies also

examined the question of validity by comparing findings of PSM and CTE analyses. In the HEPR study, the results were quite similar, increasing confidence in the validity of findings.

3.5 Unpacking Mechanisms

Mechanisms were defined in section 3.3 as the causal process generating outcomes which include causal variables, the links or pathways between them, i.e. the causal ‘tree’, as well as an explanation of why they are linked. The account of mechanisms in theoretical and econometric modelling of the impact of trade tends to be at a high level of abstraction and rather ‘thin’. For example, modelling employments effects within a production function framework presents a highly stylised depiction of production processes, where the critical issue of supply response turns on debates about elasticities of substitution between capital and labour (Gibson 2011). Similarly, in econometric exercises of aspects of trade policy, long list of correlates are produced with limited explanation of how correlates may be doing their causal work.

A major contribution of dialogical techniques such as focus group discussions and semi-structured interviews, in this regard, is to provide a ‘thicker’ depiction of the causal mechanisms in question. In the context of trade, the mechanisms can be quite varied given the complexity of the underlying causal system. The examples presented below present information on a range of mechanisms relating to factory workers, small farmers and small scale producers. While such analysis may also be undertaken using visual tools such as causal maps, the main sources of information in the studies below are household surveys, semi-structured interviews and focus group discussions.

The first study by Hunt et al. (2007) explored the employment and welfare impacts of job losses in the clothing sector of Algeria, Morocco and Tunisia due to contraction following increased foreign competition. The research focused on the welfare impact upon displaced workers and on occupational shifts of workers following displacement. The research results are based on a structured questionnaire administered to around 100 respondents in the three countries. The information content is not as rich as in semi-structured interviews, yet the study is relevant given its focus on the employment effects of trade.

According to the authors, displaced workers in Tunisia were least affected by trade liberalisation, as the majority of the women working in the clothing industries were not the sole breadwinners of their respective households. The impact was most severe in Morocco where female head of households were disproportionately affected and least able to secure employment in other sectors due to low levels of education and social barriers. The welfare impact effect in Algeria was apparently mitigated by the fact that few respondents were heads of households, and greater employment opportunities were available in the Algiers region where the survey was conducted.

A richer account of mechanisms is presented in Hellin et al.'s (2012) study of the impact pathways of trade liberalisation on smallholder maize farmers in Mexico. They identified a number of pathways linking macro and micro level processes, and then probed in detail farmer's responses to agricultural policies in the aftermath of the North American Free Trade Agreement (NAFTA). The empirical portion of the study was conducted on four communities in the state of Chiapas in 2006 and 2007 using participant observation, semi-structured interviews and focus group discussion.

The authors uncovered a diverse pattern of responses by small farmers. Some were able to benefit from government support programs by intensifying maize production, through use of improved maize varieties, fertilisers and pesticides. Lack of access to affordable credit, however, was cited as a factor constraining further intensification. Others farmers were able to supplement agricultural income with off-farm employment, in particular migration for short term work in coastal cities or the US. Many others have simply exited from agriculture though this process of urbanisation and the shift from agriculture to services and manufacturing cannot be attributed solely to trade liberalisation. Generally, smallholders have not been able to expand their land holdings or herd size, nor to diversify their cropping pattern by switching to fruits and vegetables. Costs barriers were cited as major constraints impeding such expansion and diversification. This study did not attempt to quantify the overall impact, nor to estimate the relative importance of the mechanisms identified, but simply to provide a better understanding of them.

A final study by Kristiansen (2004) examined the impact of economic liberalization, including trade liberalization, on small scale garment entrepreneurs in the town of Tanga, in northwestern Tanzania. It sought to shed light on competing theoretical propositions about the effects of liberalization on cluster economies. In theory, such environments could be conducive to competitive strengthening in the face of trade liberalization through dynamic learning effects, lower transaction costs, joint action and so on. On the other hand, imperfect information about new technologies, suppliers and customers, for example, and consequent market failure, could have the opposite effect.

Fieldwork on the informal garment industry was conducted in 1999 and 2000, involving semi-structured interviews with twelve licensed tailors, along with members of the entrepreneurs' extended social network. According to respondents, economic liberalization initially stimulated small scale tailoring through improved access to imported machinery and textiles, and the devaluation of the Shilling which made imports of clothing less competitive. Prices and profit margins steadily declined over time, however, due to the influx of both newer and inexpensive garments from Asia and second hand clothing from overseas, the increasing number of local tailors and decreasing local demand. The main constraints on the establishment of dynamic learning processes were the limited flows of business related information and transfers of technological know-how due to asymmetric information within the industry. Overall, the author maintained that

due to asymmetric information within the garment industry, small scale enterprises adopted imitative as opposed to innovative strategies and consequently, did not realise productivity gains. In summary, the detailed study of response mechanisms of petty entrepreneurs provided evidence for the second aforementioned proposition in support of imperfect information and market failure.

The core contribution of the above studies has been to pry open the ‘black box’ of behavioural responses to trade policy changes with respect to workers, small farmers and small scale garment entrepreneurs. Otherwise stated, they have served to unpack some of the mechanisms generating impacts.

3.6 Combining Outcomes and Mechanisms

In addition to unpacking mechanisms, a significant contribution of mixed method analyses has been to better explain outcomes of econometric models, or similar analyses, on the basis of an understanding of the mechanisms generating them. In the broader literature, such approaches are likely the most widely used application of mixed method designs with the greatest value-added (Bardhan and Ray 2006, Shaffer et al. 2008). The studies below provide examples of combining outcomes and mechanisms with specific reference to inequality in Brazil, child labour in Ethiopia and secondary enrolment in Turkey.

In the first example, Weinhold et al. (2013) examined the impact of soybean production for export in the Brazilian Amazon on poverty, inequality and income. They employed a mixed method design which combined econometric results drawing on census data, with ethnographic research conducted in the Santarém region of Pará state over an eight month period. A core finding of the econometric analysis was that soy cultivation had the effect of increasing income, reducing poverty and increasing inequality, though the latter was only marginally statistically significant. These findings stood in contrast to strong local political opposition to soy cultivation and heated conflicts often culminating in violence.

The role of the ethnographic results was to shed light on this puzzle. The expansion of cultivation in the Amazon was driven by newly arrived migrants from southern Brazil with greater financial resources at their disposal. In addition, there were pronounced ethnic and cultural differences between the new arrivals, who were of Northern European descent, and the existing population, who were of Indigenous, African and Southern European ancestry. Cultural differences with respect to housing styles and size, behavioural traits such as perceptions of sociability and work ethic, contributed to hostilities between the groups which was subsequently fuelled by media, religious groups and NGOs. According to the authors, the political and social conflicts around soy cultivation were driven in large part by ethnicity, cultural difference and historical rivalry, variables which were not directly captured in the econometric models.

A second example¹⁵ is Woldehanna et al.'s (2005, 2008) study of child labour in Ethiopia. A household survey was implemented in 2002 and subsequent econometric work undertaken to estimate correlates of child schooling and labour. This analysis was followed-up by semi-structured interviews conducted in 2005 with a view to provide a richer understanding of the econometric results. A number of interesting findings emerged.

First, econometric results suggested, as expected, that the likelihood of children's participation in activities other than full-time schooling declines with paternal education. On the other hand, and surprisingly, the econometric analysis found the opposite result for maternal education. The semi-structured interviews provided an explanation. Women with higher levels of education are more likely to work outside their homes, which increases the domestic work burden of older children who assume responsibility for childcare and other tasks.

Second, regression results did not find a statistically significant effect of landholding size on the probability of children's schooling or participation in the labour market. Information from the semi-structured interviews suggested a number of explanations. First, children in households with more land are frequently working on the farm as would be expected given the high opportunity costs, or income foregone, of not working and imperfect labour or credit markets which precludes hiring sufficient labour. On the other hand, children from households with less land are also frequently working, given the imperative of generating income in poorer families living at the margin. As a result, the 'opportunity cost effect' among households with large land size is offset by a 'wealth or poverty effect' among those with smaller landholdings, which render insignificant the relationship between landholding and schooling.

Third, in the econometric analysis, an attempt was made to estimate the effect of caregiver social capital, measured in terms of variables such as the sense of trust, the number of community organisations that provide social support and so on. One interesting finding was the positive and statistically significant relationship between the number of social support organisations and the probability of child labour. This finding was counterintuitive, in that social support was expected to increase child schooling by diminishing the importance of the wealth or poverty effect, which compels children to work. A potential explanation from the semi-structured interviews was that social support often takes the form of food-for-work schemes in which children participate alongside their parents. Accordingly, the association between social capital and child labour was due to the nature of the social support provided.

A final example is an impact assessment of the Social Risk Mitigation Program conditional cash transfer (CCT) scheme study undertaken by the International Food Policy Research Institute (IFPRI) in Turkey (Adato 2008).¹⁶ The program provided cash payments conditional on school

¹⁵ This discussion is based on Shaffer (2013: 80-81).

¹⁶ This discussion is based on Shaffer (2013: 96).

enrolment for boys and girls along with vaccinations and regular check-ups for children. Ethnographic work in 6 localities was combined with a quasi-experimental analysis drawing on household survey data. The particular technique used, regression discontinuity design, compares outcomes among households who fell just above, and just below, the eligibility threshold for program participation.

The quasi-experimental analysis found that the program raised secondary school enrolment for girls by around 10%, a statistically significant effect. Nevertheless, secondary enrolment rates remained low for program participants in rural areas, at less than 40%. The key contribution of the ethnographic work was to explain some of the reasons why. For boys, doubts were expressed about the value of education in the context of high unemployment and a society where honour is bestowed on those working on the land. For girls, the potential employment or wage effect of additional schooling was not highly valued given the overriding importance of traditional female roles as mothers and wives. Further, concerns were raised about threats to family honour and reputation associated with girls schooling. According to one father in a village in the province of Van: ‘the girls have only their honour as a valuable thing in the village and it is my duty to prevent any bad words about that... No one sends their daughters to school anyway. Why should I send mine? They will look at them in a bad way’ (Adato 2008: 231). As above, the core contribution of the ethnographic work was to provide an account of the mechanisms generating the somewhat disappointing results about program impact.

All of the above examples represent ways that an ethnographically informed understanding of mechanisms has led to a better understanding of outcomes of econometric models.

3.7 Integrating Correlates and ‘Reasons’

An extension of analyses combining outcomes and mechanisms are studies integrating correlates and reasons. By correlates, we are referring to variables in econometric models which are statistically significant such as those discussed in the previous section. Information on ‘reasons’, on the other hand, is closely related to the understanding of mechanisms generating change as discussed in section 3.3. In the studies reviewed below, attempts have been made to include variables representing ‘reasons’ in econometric models with a view to determine their statistical significance and relative weights.

The first example of the integration of reasons and correlates is from the poverty literature. It involves the so-called Stages of Progress (SoP) approach which has been applied to over 35000 households in India, Kenya, Uganda, Peru and North Carolina (Krishna 2010a, 2010b). In the SoP methodology, locally meaningful understandings of poverty, poverty lines and transitions between poverty categories, are elicited on the basis of focus group discussions and semi-structured interviews. Subsequently focus group participants are asked to list reasons for escape from, and descents into, poverty of particular households.

Such information allows for the compilation of lists of reasons for escapes and entries based on the percentage of households to identify them. As it happens, poor health and health-related expenses were found to be the main reasons for descents into poverty across all studies conducted in the Global South (Krishna 2010a: 79). The next stage in the SoP analysis involves modelling 'reasons', with a view to supplement information on mechanisms with information on causal weights and statistical significance. In Uganda and Peru, logistic (logit) regression models were estimated of the likelihood of falling into, or escaping, poverty (Krishna et al. 2006a; Krishna et al. 2006b). In the case of Uganda certain of the 'reasons' found to be statistically significant correlates of descents into poverty included: ill health, health care expenses, death of an income earner, crop disease. Correlates of improvements in well-being included: diversification of livelihood, private sector employment and land improvement. The relative importance of variables, conditional on all others, is inferred by comparing the size of logit coefficients or odds ratios.

A second example, is from Houssa and Verpoorten's (2015) study of the welfare effects of a self-imposed ban on shrimp exports in Benin implemented in 2003, discussed in section 3.3. As mentioned, the core information source was a fixed response household survey administered in 2009 to a sample of 540 households comprising fisherman and fishmongers (usually the spouses of the fishermen). A module in the household survey on coping strategies posed the following question: 'If the impact [of the ban] was negative, how did the household react to cope with it?' A range of strategies were identified include: asset sales, reduced consumption, accessing credit, switching economic activities and so forth.

The effectiveness of switching economic activities was probed econometrically relying on self-reported data in the household survey on the impact of the ban and on changes in economic activity following the ban. A probit model was estimated which included the following 'right-hand side' variables representing coping strategies: diversification into fishing or fish trading; diversification out of the fishery sector; migration for work. The impact measure, or the 'left-hand side' variable, was assigned a value of 1 in the case of 'very negative impact' and '0' otherwise. In the different model specifications, the only statistically significant (negative) correlate of very negative impact was diversification out of fishing (i.e., it lessened the probability of a very negative impact). This finding was consistent with the expectations of the authors, given that diversification within the fishery sector would only increase competition for a limited fishery stock coupled with the fact that migrants faced uncertain prospects for securing productive employment.

Both of these examples demonstrate a greater level of integration of mechanisms and outcomes than those discussed in section 3.6. The inclusion of variables representing mechanisms in econometric models allows for tests of the statistical significance and assessment of the relative importance of the mechanisms in question.

3.8 Informing Model Specification¹⁷

In the previous sections, results of econometric models were presented uncritically, in the context of the discussion of mixed method approaches. Emphasis was placed on comparing results of models and thought experiments (section 3.4), using narrative information to interpret model results (section 3.6) and incorporating ‘reasons’ in econometric models (section 3.7). As discussed in Gibson’s (2011: 71, 103-112) above-mentioned piece, however, econometric modelling faces a number of significant challenges, including those related to reverse causation, or simultaneity, and omitted variables. The problems may be particularly severe in the case of employment effects of trade given the complexity of the underlying causal system characterised by a large number of imperfectly identified variables interacting in a variety of ways.

A final contribution, then of ‘qualitative’ and mixed method studies is to show how they have aided in econometric model specification. By ‘specification’, we refer to the selection of variables, and depiction of their relationship. The examples below are not drawn from the trade literature but are relevant, nonetheless, to econometric analyses of aspects of trade reform. They deal first, with the search for ‘instruments’ to address the problem of endogeneity due to omitted variable bias, and next, with the selection of variables and identification of their interrelationships.

Searching for ‘instruments’

A first example illustrates the so-called ‘participatory econometrics’ approach pursued by Vijayendra Rao of the World Bank (Rao 2002, 2003). In their study of sex workers in Calcutta, Rao et al. (2003) sought to estimate the revenue loss, or compensating differential, associated with condom use. The main econometric problem is that unobserved characteristics of sex workers, which are correlated *both* with condom use and prices, can bias results. For example, if sex workers with more desirable, but unobserved attributes, are better able to require condom use of clients *and* to command high prices, then there will be a downward bias in the value of the differential.

The key contribution was to use dialogical techniques to search for an instrumental variable, or instrument, to deal with the econometric problem. When attempting to estimate the causal effect of x on y , an instrument is a third variable which affects y only through its effect on x . In this case, it must affect the price of sex acts only through its effect on condom use, and not be correlated with unobserved variables which also affect prices. Through semi-structured interviews, the research uncovered just such an instrument. The All India Institute of Public Health and Hygiene has initiated an HIV/AIDs awareness program throughout the area which was implemented in a seemingly random manner. Further, participation in the program appeared to be effective at promoting condom use. Accordingly, participation in this program was used as an instrument to estimate the relationship between condom use and price.

¹⁷ This section is based on Shaffer (2013: 88-90).

Selecting variables and uncovering relationships

A second example is de Weerd's study of poverty transitions in Kagera, Tanzania, which was undertaken as part of the World Bank's *Moving out of Poverty* study (de Weerd 2010). The study drew on the Kagera health and Development Survey (KHDS) which collected panel data in 1994 and 2004 along with focus group discussions and life histories. Econometric analysis was performed on the data with a view to predict 2004 asset values on the basis of 1993 household characteristics. A comparison of model predictions with actual 2004 data revealed significant discrepancies. In particular, only around half of those whose asset values were predicted to increase actually did so.

The narrative information suggested a number of factors as explanations of deviations from predictions of the model. Concerning 'unexpected losers', one explanation concerned intervening events between waves of the panel such as agricultural shocks, mortality, illness and widowhood or death. A second reason for the discrepancies had to do with variables not included in the survey such as alcoholism, bad marital relations and lack of exposure to outside information. With respect to 'surprise winners,' missing variables in the survey, such as exposure to outside ideas and networks, were important as was the incorrect specification of the causal structure of the model, in particular, the interaction between remoteness and initial conditions.

The life histories and focus groups suggested reasons why the interaction between initial conditions and remoteness, and not only their individual effects, was important by contrasting the situation in remote and non-remote villages. In the latter, initial conditions proved less important due to opportunities associated with trade, such as the availability of employment as casual labourers, the emergence of business relationships with outside traders and the influx of money and access to new ideas and networks outside the village. These positive effects were absent in remote villages which compounded the effects of poor initial conditions. In light of these findings, a re-specified model was estimated including an interaction term of remoteness and initial conditions which proved to be statistically significant.

A final example is Quisumbing's (2011) study of the dynamics of poverty in Bangladesh, based on econometric analysis of panel data. The methodology included an initial stage of focus group discussions to refine research questions and identify variables for inclusion in the household survey. A second stage combined panel data with life histories. The contribution of mixed method approaches to variable and model specification is aptly described by Agnes who conducted the econometric work:

Nesting a quantitative analysis of poverty dynamics within a fully-integrated qualitative and quantitative study has also yielded insights that might not have been possible with one approach alone. The focus group discussions, conducted prior to the fielding of the quantitative survey, brought out specific issues that were addressed through the design

or adaptation of specialized questionnaire modules, such as those focusing on shocks. While the shocks module was similar to those administered in other countries, its adaptation to the Bangladesh context—particularly the disaggregation of illness shocks into income losses and medical expenses—was reinforced by the focus group discussions. The life histories work identified the severe deleterious effects of combined dowry and illness expenses as an important factor that put households on a downward life trajectory. This led to the re-specification of the shocks variables to include these combined shocks, which have been found to reduce the probability of moving out of poverty (Quisumbing 2011: 54).

In all of these examples, narrative information from focus groups discussions and semi-structured interviews led to better specification of causal variables and the structure of their interrelationships.

4. Key Findings and Recommendations for Policy Making

4.1 Key Findings: Value-Added of ‘Qualitative’ and Mixed Method Approaches

Section 3 reviewed a number of contributions of ‘qualitative’ and mixed method approaches to impact assessment and illustrated them with examples of recent empirical studies. These studies may be regrouped into the following four core categories based on the value-added which they have brought to impact assessment: i) incorporating locally meaningful impact measures and weights; ii) providing estimates on the magnitude of impact; iii) unpacking and integrating mechanisms; iv) informing model specification.

Incorporating Locally Meaningful Impact Measures and Weights

Assessment of the impact of policy or programs depends fundamentally on the impact measure, or metric, at hand. As shown in section 3.1, a core contribution of the studies has been to incorporate locally meaningful measures of impact to ensure that the assessment metric is locally relevant. The importance of this point was illustrated by the empirical finding that levels or changes in impact measures are not necessarily closely correlated. To recall the discussion in section 2.3, if these measures are not closely correlated, then they will not serve as good proxies for one another. The empirical studies reviewed in section 3.1 uncovered systematic trade-offs between dimensions of employment in that, at times, increased work opportunities came at the expense of productive work, equitable treatment in the work place and safe working conditions. Studies of rural families in Mexico following NAFTA along with female workers in Export Processing Zones (EPZs) in Mauritius and Sri Lanka arrived at this conclusion.

In light of the varied dimensions of impact, a second area of value-added was the use of ‘qualitative’ and mixed method analyses to derive locally meaningful weights in section 3.2. Drawing on the literature on poverty, examples were presented of indirect approaches, which attempt to retrieve weights from correlates of poverty, or direct approaches, which simply ask people to supply the weights in question. While the studies reviewed were poverty-related the methodologies apply to any impact measures including employment.

Providing Estimates on the Magnitude of Impact

A first approach, self-reports, was discussed in section 3.3. Self-reports add potential value in that they base estimates of impact on people’s perceptions, though caution has been urged with respect to the validity of results. Typically, the validity of responses requires a good understanding of the mechanisms generating impact and is enhanced when the causal system is simpler and the impact measure defined in broad ordinal categories. When these conditions are met, as in Houssa and Verpoorten’s (2015) impact assessment of the ban on shrimp exports in Benin, there is more reason to be confident in findings, or at least the sign of results.

The second approach, counterfactual thought experiments (CTEs), was discussed in section 3.4. As above the core value-added is to base estimates of impact on people's perceptions. In this approach subjunctive conditional (if/then) questions are posed, often about behavioural or firm-level responses in hypothetical scenarios. The examples presented include *ex ante* analysis of the likely effects of tariff reduction on sugar cane farmers in Vietnam and *ex post* analysis of the impact of a national poverty reduction program in Vietnam. Importantly, results of the CTE in Vietnam, were very similar to results of propensity score matching techniques, which add credibility to the results.

Unpacking and Integrating Mechanisms

It is likely that the most important contribution of 'qualitative' and mixed method approaches to impact assessment is in unpacking the 'black box' of causal mechanisms and systematically integrating mechanisms into the analysis. Mechanisms refer to the causal processes generating observed outcomes and comprise the causal variables, the links or pathways between them, i.e. the causal 'tree', as well as an explanation of why they are linked. The account of mechanisms in both theoretical and econometric modelling of trade is highly stylised, abstracted and 'thin.'

The empirical studies illustrated this emphasis on mechanisms in a number of different ways. In section 3.5, a range of mechanisms were detailed relating to factory workers, small farmers and small scale producers. Section 3.6 illustrated how an understanding of mechanisms could serve to better explain and interpret of the outcomes of econometric models with specific reference to inequality in Brazil, child labour in Ethiopia and secondary enrolment rates in Turkey. Section 3.7 took this analysis one step further by illustrating how information on mechanisms or 'reasons' could be included in econometric modelling drawing on the examples of poverty dynamics in Uganda and coping strategies in the face of the shrimp export ban in Benin.

Informing Model Specification

The final area of value added in 'qualitative' and mixed method studies concerns the specification of econometric models. Narrative information from focus groups and semi-structured interview have allowed for the identification of 'instruments' to tackle the problem of endogeneity due to omitted variable bias, as evidenced by Rao's (2002, 2003) study of sex workers in Calcutta. Likewise, they have facilitated the selection of variables and the proper depiction of their relationships, as interaction terms, in poverty studies in Tanzania and Bangladesh.

4.2 Feeding Results into Policy Processes

There are a number of reasons why 'qualitative' and mixed method studies may be particularly well suited for policy processes, including trade policy formulation, treaty negotiation and

monitoring/evaluation. Prior to examining these issues, however, there are two important caveats which should be noted.

First, data supply does not create data demand. There are many factors which determine whether or not research results will enter into the policy process. A short list includes: the nature of the decision regime (based on routine, incremental and fundamental decisions types), the nature of the policy context (based on government demand for data and public sector capacity) along with contingent factors relating to crises, personal ties, communication strategies, and so on (Carden 2007). The discussion below, then, about the use of ‘qualitative’ and mixed methods in the policy process presupposes a favourable policy and decision-making environment where data demand exists.

A similar point applies to ‘participatory’ or consultative processes intended to inform policy formulation or monitoring/evaluation (Shaffer 2012b). Their effectiveness depends on the extent to which they are embedded in a favourable political, legal and social context. A number of donor-driven initiatives to involve Civil Society Organizations CSOs in the policy process, such as the World Bank’s Poverty Reduction Strategy Paper (PRSP) process, tend to fall short in this regard. They have had minimal effect on policy (McGee et al. 2002) and stand in marked contrast to successful home grown examples elsewhere. For example, the relative success in Rajasthan, India of the *Workers and Farmers Power Association* (MKSS), who have championed public audits of village-level accounts may be attributed to: committed action of a closely knit and motivated group of activists; connections to journalists, lawyers and social action groups; passage of Right to Information legislation by the Rajasthan State Assembly in 2000 following the election of a sympathetic party in 1998; passage of amendments to the Local Government Act in 2000 which enshrined public hearings at the local level in law (Jenkins and Goetz 1999). As above, ‘participatory’ consultations involving CSOs are unlikely to lead to policy feedback in the absence of a favourable political and legal environment.

With these caveats in mind, there are at least three reasons why ‘qualitative’ and mixed method approaches may be particularly well-suited to policy processes, namely:

First, incorporating narrative information on lived experiences of impact, either in terms of mechanisms or effects, conveys information on the human costs or benefits of impact in a way which numerical data cannot. As such, it can ‘bring alive’ and ‘renders real’ data in ways which resonate with policy makers and CSOs engaged in the policy process. Given the importance, however, of numerical and statistical information in policymaking circles, mixed method designs are likely to be optimal for policy effect.

Second, the modelling literature on trade, from partial equilibrium to CGE models, is far from user friendly for those without an advanced degree in economics. The literature on ‘qualitative’ and mixed methods tends to be much more accessible and accordingly, open to a wider range of actors

engaged in policy work. In fact, often the research results informing CSOs' submissions in consultative process is based on PRA-type research, which, in some cases, they have conducted.

Finally, certain types of information in 'qualitative' and mixed method studies have quite direct policy relevance such as information on behavioural responses to policy change, and in particular, constraints on effective action. Information of this type is directly relevant to the design of facilitating and mitigating strategies in the context of trade policy reform. One specific example was lack of credit as a reason cited by around a third of poor sugar cane farms for their perceived inability to adjust to a future price shock associated with reduction of the tariff on sugar imports (section 3.4, table 1).

References

- Abbott, A. (2001). *Chaos of Disciplines*. Chicago: The University of Chicago Press.
- Adato, M. (2008). 'Combining Survey and Ethnographic Methods to Improve Evaluation of Conditional Cash Transfer Programs', *International Journal of Multiple Research Approaches*, 2(2): 222–236.
- Aistara, G. (2011). 'Seeds of kin, Kin of Seeds: The Commodification of Organic Seeds and Social Relations in Costa Rica and Latvia', *Ethnography*, 12(4): 490-517.
- Anker, R., Chernyshev, P., Mehran, F., Ritter, J. (2002). 'Measuring Decent Work with Statistical Indicators', *Policy Integration Department, Statistical Development and Analysis Group Working Paper No. 2*, Geneva, International Labour Office.
- Ansell, N., Tsoeu, S. and Hajdu, F. (2015). 'Women's Changing Domestic Responsibilities in Neoliberal Africa: a Relational Time-Space Analysis of Lesotho's Garment Industry', *Gender, Place and Culture*, 22(3): 363-382.
- Bardhan, P. and Ray, I. (2006). 'Methodological Approaches to the Question of the Commons', *Economic Development and Cultural Change*, 54(3): 655–676.
- Blin, M. (2008). 'Export-Oriented Policies, Women's Work Burden and Human Development in Mauritius', *Journal of Southern African Studies*, 34(2): 239-253.
- Bourguignon, F. and da Silva, L. (eds.) (2003). *The Impact of Economic Policies on Poverty and Income Distribution. Evaluation Techniques and Tools*. New York: World Bank and Oxford University Press.
- Bryman, A. (1984). 'The Debate about Quantitative and Qualitative Research: A Question of Method or Epistemology?', *The British Journal of Sociology*, 35(1): 75–92.
- Camfield, L., Crivello, G. and Woodhead, M. (2009). 'Well-being Research in Developing Countries: Reviewing the Role of Qualitative Methods', *Social Indicators Research*, 90(1): 5–31.
- Campbell, W., Hernandez, I. Ceremuga, J. and Farmer, H. (2010). 'Globalization and Free Trade Agreements: A Profile of a Nicaraguan Factory Community', *Journal of Community Practice* 18(4): 440-457.

- Carden, F. (2000). *Knowledge to Policy*. Thousand Oaks: Sage Publications and Ottawa: International Development Research Centre
- Catley, A., Burns, J. , Abebe, D. and Suji, O. (2014). *Participatory Impact Assessment*. Friedman School of Nutrition Science and Policy, Tufts University.
- Chambers, R. (1994). ‘Participatory Rural Appraisal (PRA): Analysis of Experience’, *World Development*, 22(9): 1253-1268.
- Chambers, R. (2003). ‘Qualitative Approaches: Self-Criticism and What can be Gained from Quantitative Approaches’, in R. Kanbur (ed), *Q-Squared: Qualitative and Quantitative Methods of Poverty Appraisal*. Delhi: Permanent Black, 28–34.
- Chang L. (2008). *Factory Girls: From Village to City in a Changing China*. New York City: Spiegel and Grau.
- Chung, K. (2000). ‘Qualitative Data Collection Techniques’, in M. Grosh and P. Glewwe (eds), *Designing Household Survey Questionnaire for Developing Countries. Volume Two*. Washington, D.C.: The World Bank, 337-363.
- Cooke, B. and U. Kothari (2001). ‘The Case for Participation as Tyranny’, in B. Cooke and U. Kothari (eds.), *Participation: The New Tyranny?*. New York and London: Zed Press.
- Cuong, N., nd. *Assessing the Impact of Poverty Reduction Programs in Vietnam*, Mimeo, Hanoi.
- de Kruijk, H. and Rutten, M. (2007). ‘Weighting Dimensions of Poverty based on People's Priorities: Constructing a Composite Poverty Index for the Maldives’, *Institute for International and Development Economics (IIDE) Discussion Paper 200708-01*, Rotterdam, the Netherlands, August.
- de Silva, N. and Gunetilleke, N. (2008). ‘On Trying to be Q-Squared: Merging Methods for a Technical Minded Client’, *International Journal of Multiple Research Approaches*, 2(2): 252–265.
- de Weerdt, J. (2010). ‘Moving out of Poverty in Tanzania: Evidence from Kagera’, *Journal of Development Studies*, 46 (2): 331–349.
- Dhai, G. (2003). ‘Decent Work: Concept and Indicators’, *International Labour Review*, 142(2): 113-145.

- Esposito, L, Kebede, B. and Maddox, B. (2012). 'The Value of Literary Practices', *Q-Squared Working Paper No. 59*, Department of International Development Studies, Trent University, Canada.
- Francis, E. and Hoddinott, J. (1993). 'Migration and Differentiation in Western Kenya: A Tale of Two Sub-Locations', *Journal of Development Studies*, 30(1): 115–145.
- Geertz, C. (1983). 'Thick Description. Towards an Interpretive Theory of Culture', Reprinted in M. Martin and L. McIntyre (eds), *Readings in the Philosophy of Social Science*. Cambridge, Mass.: MIT Press, 213-231).
- Gibson, B. (2011). 'Assessing the Impact of Trade on Employment: Methods of Analysis', M. Jansen, R. Peters and J. M. Salazar Xirinachs (eds.), *Trade and Employment: From Myths to Facts*. Geneva: ILO-EU Publications.
- Gilovich, T. and Griffin, D. (2002). 'Introduction – Heuristics and Biases: Then and Now', in T. Gilovich, D. Griffin and D. Kahneman (eds), *Heuristics and Biases. The Psychology of Intuitive Judgement*. Cambridge: Cambridge University Press, 1–18.
- Godfrey, M. (2003). 'Employment Dimensions of Decent Work: Trade-offs and Complementarities', *International Institute for Labour Studies Discussion Paper DP/148/2003*, Geneva, International Labour Organisation.
- Gough, I. and McGregor, J. (eds) (2007). *Wellbeing in Developing Countries. From Theory to Research*. Cambridge: Cambridge University Press.
- Hamilton, S. and Fischer, E. (2003). 'Non-Traditional Agricultural Exports in Highland Guatemala', *Latin American Research Review*, 38(3): 82-110.
- Hammersley, M. (1992). *What's Wrong with Ethnography? Methodological Explorations*. London and New York: Routledge.
- Hancock, P. (2006). 'Women, Work and Empowerment: A Portrait of Women Workers in Two of Sri Lanka's Export Processing Zones', *Norwegian Journal of Geography*, 60(3): 227-239.
- Heckman, J. (2000). 'Comment. Impact Evaluation: Concepts and Methods', in O. Feinstein and R. Picciotto (eds), *Evaluation and Poverty Reduction*. Washington, D.C.: The World Bank, 83–84.

- Hedström, P. and Swedberg, R. (eds) (1998). *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
- Hellin, J., Groenewald, S. and Keleman, A. (2012). 'Impact Pathways of Trade Liberalization on Rural Livelihoods: A Case Study of Smallholder Maize Farmers in Mexico', *Journal of Development Studies*, 1(1): 59-82.
- Houssa, R. and Verpoorten, M. (2015). The Unintended Consequence of an Export Ban: Evidence from Benin's Shrimp Sector', *World Development*, 67: 138-150.
- Hunt, D., Lahlou, M., Musette, S. and Chourou, B. (2007). 'The Employment Impacts of Trade Liberalisation and of Increased Competition in Export Markets: The North African Clothing Sector', *The Journal of North African Studies* 12(4): 453-479.
- Jenkins, R. and Goetz, A.M. (1999). 'Accounts and Accountability: Theoretical Implications of the Right-to-Information Movement in India,' *Third World Quarterly* 20 (3): 603-22.
- Joint Donor Report (JDR) (2004). *Vietnam Development Report 2004. Poverty*. Asian Development Bank, the Australian Agency for International Development, Department for International Development (UK), the German Agency for Development Co-operation, Japan International Cooperation Agency, Save the Children (UK), United Nations Development Programme and the World Bank, Hanoi.
- Kanbur, R. (2003). 'Q-squared? A commentary on qualitative and quantitative poverty appraisal', in R. Kanbur (ed), *Q-squared: qualitative and quantitative poverty appraisal*. Delhi: Permanent Black, 1-21.
- Kanbur, R. and Shaffer, P. (2007). 'Epistemology, Normative Theory and Poverty Analysis: Implications for Q-Squared in Practice', *World Development*, 35(2): 183-196.
- Kapoor, I. (2002). 'The Devil's in the Theory: a Critical Assessment of Robert Chambers' Work on Participatory Development', *Third World Quarterly*, 23(1): 101-117.
- Kebede, B. (2009). 'Community Wealth Ranking and Household Surveys: An Integrative Approach', *Journal of Development Studies*, 45(10): 1731-1746.
- Krishna, A. (2010a). 'Who Became Poor, Who Escaped Poverty and Why? Developing and Using a Retrospective methodology in Five Countries', *Journal of Policy Analysis and Management*, 29(2): 351-372.

- Krishna, A. (2010b). *One Illness Away. Why People Become Poor and How They Escape Poverty*. Oxford: Oxford University Press.
- Krishna, A., et al. (2006a). 'Escaping Poverty and Becoming Poor in 36 Villages of Central and Western Uganda', *Journal of Development Studies*, 42(2): 346–370.
- Krishna, A., et al. (2006b). 'Fixing the Hole in the Bucket: Household Poverty Dynamics in the Peruvian Andes', *Development and Change*, 37 (5): 997–1021.
- Kristiansen, S. (2004). 'Liberalisation, Information and Innovation: Cottage Garment Industries in Tanzania', *The European Journal of Development Research*, 16(2): 375-395.
- Lipton, M. (1992). 'Economics and Anthropology: Grounding Models in Relationships', *World Development*, 20(10): 1541-1546.
- Little, D. (1998). *Microfoundations, Method and Causation*. New Brunswick, U.S.A.: Transaction Publishers.
- Mahoney, J. (2001). 'Beyond Correlational Analysis: Recent Innovations in Theory and Method', *Sociological Forum*, 16(3): 575–593.
- Mayoux, L. and Chambers, R. (2005). 'Reversing the Paradigm: Quantification, Participatory Methods and Pro-Poor Impact Assessment', *Journal of International Development*, 17(2): 271-298.
- McCarty, D. (2008). 'The Impact of the North American Free Trade Agreement (NAFTA) on Rural Children and Families in Mexico: Transnational Policy and Practice Implications', *Journal of Public Child Welfare*, 1(4): 105-123.
- McGee, R. with Levene, J. and Hughes, A. (2002). *Assessing Participation in Poverty Reduction Strategy Papers: A Desk-based Synthesis of Experience in Sub-Saharan Africa*. IDS Research Report 52. Brighton: Institute of Development Studies at the University of Sussex, February.
- Mosse, D. (1994). 'Authority, Gender and Knowledge: Theoretical Reflections on the Practice of Participatory Rural Appraisal', *Development and Change*, 25(3): 497-526.
- Oueslati-Porter C. (2013). 'Infatih and (in)dependence: Bizerte Women's Economic Strategies Three Decades into Tunisian Neoliberal Policy', *The Journal of North African Studies* 18(1): 141-158.

- Popper, K. (1959). *The Logic of Scientific Discovery*. London: Hutchinson.
- Pretty, J. (1994). 'Alternative Systems of Inquiry for a Sustainable Agriculture', *IDS Bulletin*, 25(2): 37-48.
- Quisumbing, A. (2011). 'Poverty Transitions, Shocks and Consumption in Rural Bangladesh, 1996–97 to 2006–07', in B. Baulch (ed), *Why Poverty Persists. Poverty Dynamics in Asia and Africa*. Cheltenham: Edward Elgar Publishing, 29-64.
- Rao, V. (2002). 'Experiments in "Participatory Econometrics"', *Economic and Political Weekly*, May 18: 1887–1891.
- Rao, V. (2003). 'Experiments with "Participatory Econometrics" in India: Can Conversation take the Con out of Econometrics?', in R. Kanbur (ed), *Q-squared: qualitative and quantitative poverty appraisal*. Delhi: Permanent Black, 103-113.
- Rao, V. et al. (2003). 'Sex Workers and the Cost of Safe Sex: the Compensating Differential for Condom Use among Calcutta Prostitutes', *Journal of Development Economics*, 71 (2): 585–603.
- Rao, V. and Woolcock, M. (2003). 'Integrating Qualitative and Quantitative Approaches in Program Evaluation', in F. Bourguignon and L. da Silva (eds), *The Impact of Economic Policies on Poverty and Income Distribution. Evaluation Techniques and Tools*. New York.: World Bank and Oxford University Press, 165–190.
- Rao, V. and Ibáñez, M. (2005). 'The Social Impact of Social Funds in Jamaica: A "Participatory Econometric" Analysis of Targeting, Collective Action, and Participation in Community-Driven Development', *Journal of Development Studies*, 41(5): 788–838.
- Roche, C. (1999). *Impact Assessment for Development Agencies*. Oxford: Oxfam Publications
- Shaffer, P. (2005). 'Assumptions Matter: Reflections on the Kanbur Typology', *Focaal - European Journal of Anthropology*, 45(1): 18–32.
- Shaffer, P. (2008). 'Behavioral Responses to Changes Accompanying Reform: A Q-Squared Analysis from Than Hoa Province, Vietnam', *Q-Squared Working Paper No. 53*, Centre for International Studies, University of Toronto.
- Shaffer, P. (2011). 'Against Excessive Rhetoric in Impact Assessment: Overstating the Case for Randomised Controlled Experiments', *Journal of Development Studies*, 47(11): 1619–1635.

- Shaffer, P. (2012a). 'Beneath the "Methods Debate" in Impact Assessment: Baring Assumptions of a Mixed Methods Impact Assessment in Vietnam', *Journal of Development Effectiveness*, 4(1): 134–150.
- Shaffer, P. (2012b). 'Demand-Side Challenges to Poverty Monitoring and Assessment Systems: Illustrations from Tanzania', *Development Policy Review*, 30(1): 29-48.
- Shaffer, P. (2013). *Q-Squared. Combining Qualitative and Quantitative Approaches in Poverty Analysis*. Oxford: Oxford University Press.
- Shaffer, P. et al. (eds) (2008). 'Introduction to Q-Squared in Policy: the Use of Qualitative and Quantitative Methods of Poverty Analysis in Decision-Making', *International Journal of Multiple Research Approaches*, 2(2): 134–144.
- Silva, J. (2008). 'A Multilevel Analysis of Agricultural Trade and Socioeconomic Inequality in Rural Mozambique', *The Professional Geographer* 60(2): 174-189.
- Starr, M. (2014). 'Qualitative and Mixed-Methods Research in Economics: Surprising Growth, Promising Future', *Journal of Economic Surveys*, 28(2): 238-264.
- Sudman, S., Bradburn, N. and Schwartz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers.
- Suppes, F. (ed) (1974). *The Structure of Scientific Theories*. Urbana: The University of Illinois Press.
- Taylor, C. (1985). 'Interpretation and the Science of Man', in C. Taylor (ed), *Philosophy and the Human Sciences*. Cambridge: Cambridge University Press (original work published 1971), 15–57.
- Vásquez-León, M. (2010). 'Free Markets and Fair Trade, Collective Livelihood Struggles, and the Cooperative Model : Two Case Studies from Paraguay', *Latin American Perspectives* 37(6): 53-73.
- White, H. (2008). 'Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation', *IDS Bulletin*, 39(1): 98–109.

- White, H. (2009). 'Some Reflections on Current Debates in Impact Evaluation', *International Initiative for Impact Evaluation Working Paper No. 1*, International Initiative for Impact Evaluation, New Delhi, India.
- Woldehanna, T et al. (2005). 'Gender Labour, Gender Inequality and Rural/Urban Disparities: How Can Ethiopia's National Development Strategies be Revised to Address Negative Spillover Impacts on Child Education and Wellbeing?', *Young Lives Working Paper No. 20*, Young Lives, Save the Children UK, London.
- Woldehanna, T., Jones, N. and Tefera, B. (2008). 'The Invisibility of Children's Paid and Unpaid Work. Implication for Ethiopia's National Poverty Reduction Policy', *Childhood*, 15(2): 177–201.
- Woodcock, A. et al. (2009). 'Validation of the WEDQoL-Goals-Thailand Measure: Culture-Specific Individualised Quality of Life', *Social Indicators Research*, 94(1): 135–171.

Appendix A: Summary Table Of Studies

The summary table presented below situates the studies reviewed in section 3 in terms of the criteria of methodological choice developed in section 2.4. To recall, the methodological criteria include the following elements:

General Considerations:

Background Knowledge: Do the studies employ open-ended and flexible data collection technique to provide information about the impact measure or the causal system?

Perceptual Information: Do the studies provide information on perceptions/understandings of locally meaningful measures of impact and their weights, the magnitude of impact, the mechanisms generating impact, behavioural responses in hypothetical scenarios and so on?

External Validity: Do the studies claim to be ‘representative’ of broader population groups?

The Impact Measure

Narrative or Numerical: Is the impact measure presented in numerical terms or in words?

Ordinal or Cardinal: If numerical, is the precision of the impact measure higher, as depicted in cardinal terms, or lower, as represented on an ordinal scale, for example?

The Causal System:

Process-based: Is analytical attention directed to causal processes/mechanisms, as opposed, or in addition to, outcomes?

Thick Description: Do the studies provide a rich depiction of causal mechanisms drawing on information from dialogical processes?

Causal Weights: Is there an attempt to assign weights to the various causal variables?

	General Considerations			Impact Measure		Causal System		
	Background Knowledge	Perceptual Information	External Validity	Narrative or Numerical	Ordinal or Cardinal	Process-based	Thick Description	Causal Weights
1. Lived Experience & Locally Meaningful Measures of Impact								
McCarty (2008): Impact of NAFTA, Mexico	✓	✓	✗	NAR	✗	✓	✓	✗
Blim (2006): Export Processing Zones, Mauritius	✓	✓	✗	NAR	✗	✓	✓	✗
Hancock (2006): Female Factory Workers, Sri Lanka	✓	✓	✗	BOTH	BOTH	✓	✓	✗
De Silva et al. (2008): Project Impact Assessment, Sri Lanka	✓	✓	✗	BOTH	BOTH	✓	✓	✗
2. Locally Meaningful Weights of Impact Measures								
Kebede (2009): Indirect Weighting Approach, sub-Saharan Africa	✓	✓	✗	BOTH	BOTH	✗	✗	✗
De Kruijk and Rutten (2007): Human Vulnerability Index, Maldives	✗	✓	✓	NUM	BOTH	✗	✗	✗
Woodcock et al. (2009): Quality of Life Questionnaire, Thailand	✓	✓	✓	NUM	ORD	✗	✗	✗
Esposito et al. (2012): Sequential Ranking, Mozambique	✓	✓	✓	BOTH	BOTH	✗	✗	✗
3. Self-Reports of Impact								
Houssa and Verpoorten (2015): Survey of Fishermen, Benin	✗	✓	✓	NUM	ORD	✗	✗	✗
Shaffer (2012a, 2013): Program Impact Assessment, Vietnam	✗	✓	✓	NUM	ORD	✗	✗	✗
4. Counter-Factual Thought Experiments								
Shaffer (2008): Study of Sugarcane Growers, Vietnam	✓	✓	✓	NUM	ORD	✓	✗	✗
Shaffer (2012a, 2013): Program Impact Assessment, Vietnam	✓	✓	✓	NUM	ORD	✓	✗	✗

	General Considerations			Impact Measure		Causal System		
	Background Knowledge	Perceptual Information	External Validity	Narrative or Numerical	Ordinal or Cardinal	Process-based	Thick Description	Causal Weights
5. Unpacking Mechanisms								
Hunt et al. (2007): Study of Textile Workers, North Africa	✗	✓	✗	NUM	CAR	✓	✗	✗
Hellin et al. (2012): Study of Maize Farmers, Mexico	✓	✓	✗	NAR	✗	✓	✓	✗
Kristiansen (2004): Study of Garment Entrepreneurs, Tanzania	✓	✓	✗	NAR	✗	✓	✓	✗
6. Combining Outcomes and Mechanisms								
Weinhold et al. (2013): Study of Soybean Producers, Brazil	✓	✓	✓	BOTH	CAR	✓	✓	✓
Woldehanna et al. (2005, 2008): Study of Child Labour, Ethiopia	✓	✓	✓	BOTH	CAR	✓	✓	✓
Adato (2008): Program Impact Assessment, Turkey	✓	✓	✓	BOTH	CAR	✓	✓	✗
7. Integrating Correlates and Reasons								
Krishna (2010a, 2010b): Poverty Dynamics Studies, Global	✓	✓	✓	BOTH	BOTH	✓	✓	✓
Houssa and Verpoorten (2015): Survey of Fishermen, Benin	✗	✓	✓	NUM	ORD	✓	✓	✓
8. Informing Model Specification								
Rao et al. (2003): Searching for 'Instruments', India	✓	✓	✓	BOTH	CAR	✓	✓	✓
de Weerd 2010: Modelling Poverty Dynamics, Tanzania	✓	✓	✓	BOTH	ORD	✓	✓	✓
Quisumbing's (2011): Modelling Poverty Dynamics, Bangladesh	✓	✓	✓	BOTH	ORD	✓	✓	✓