

# Q-Squared in Impact Assessment: A Review\*

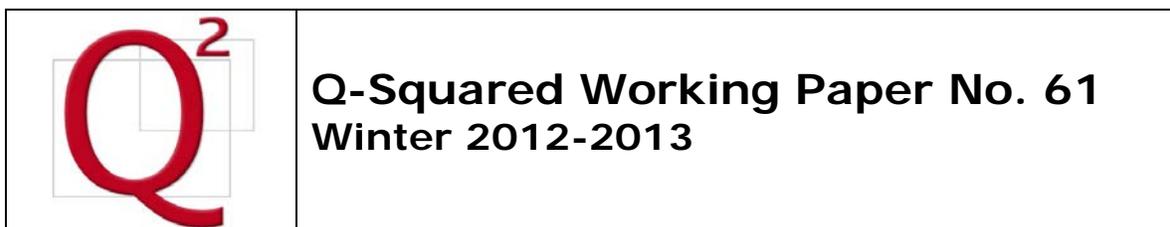
**Paul Shaffer**  
Trent University, Canada  
Email: [paulshaffer@trentu.ca](mailto:paulshaffer@trentu.ca)

## Abstract

*This paper reviews the contribution of mixed methods to impact assessment. Empirical examples are presented which demonstrate the value-added of  $Q^2$  approaches through the combined analysis of causal effects (results) and mechanisms. Other examples include the use of narrative information in the construction of comparison groups, in the case of unmeasured or unobservable variables, the use of thought experiments to create counterfactual scenarios and the use of narrative information to investigate if the assessment of program performance is sensitive to the definition of benefits.  $Q^2$  approaches have provided a richer causal analysis of program effects by integrating different models of causal inference and by bolstering analyses of counterfactual dependence.*

**Key words:** impact assessment; methods; mixed method research; methodological pluralism; poverty

\*This paper is based on Chapter 7 of Shaffer, P. (2013 (forthcoming). *Q-Squared. Combining Qualitative and Quantitative Approaches in Poverty Analysis.* (Oxford: Oxford University Press).



## 1. Introduction

This paper presents a selective review of empirical examples of the use of mixed method, or Q-Squared, approaches in impact assessment. The objective is to present an empirical case for methodological and causal pluralism in impact assessment on grounds that such approaches add value for understanding and explaining impact.

Following an opening section which defines impact assessment and reviews methods (section 2), this paper is structured around four core contributions that Q<sup>2</sup> analyses have made to assessing the impact of development programs and projects. Specifically, such analyses have facilitated integrated analysis of mechanisms and results (section 3), and comparison group construction (section 4), allowed for alternative ways of constructing counterfactuals such as thought experiments (section 5) and directed attention to the question of the valuation of different program benefits by participants (section 6). A final section concludes.

The issues addressed in this paper were chosen to illustrate the contributions of Q<sup>2</sup> to impact assessment. The focus is on impact assessment of programs or projects not policies. There are other approaches to impact evaluation, such as partial and general equilibrium analyses which address multi-sector or economy-wide impacts of policies, (Bourguignon and da Silva (eds), 2003), which are not addressed. Also, there is limited consideration of the role of mixed methods in program monitoring and evaluation (Rallis and Rossman, 2003), which differs from impact assessment in a number of ways.<sup>1</sup>

## 2. Definitions and methods

There is no consensus definition of impact assessment and it consequently ends up referring to different things (White 2009a).<sup>2</sup> For the present purposes, impact assessment is defined in terms of two core characteristics:

- i. it is concerned with outcomes or impacts and not program inputs, activities or outputs;
- ii. it attempts to 'attribute' outcomes or impacts to specific programs, and does not simply track changes in them and as such, uses techniques to address the so-called 'attribution' problem.

The input-impact terminology draws on the concept of the impact chain presented in Figure 7.1 below (Roche 1999). The impact chain depicts the causal relationships in programs or projects from inputs through activities and outputs, culminating in outcomes or impacts. Inputs refer to the financial, physical

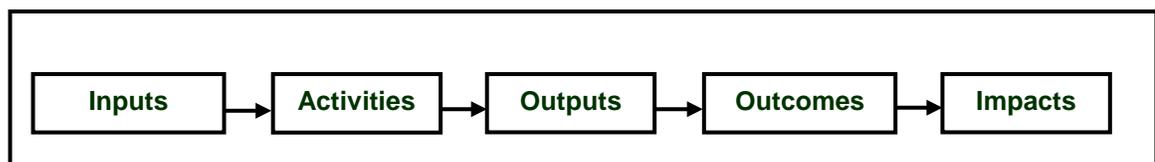
---

<sup>1</sup> Other surveys with additional examples include Bamberger et al. (2010), Rao and Woolcock (2003) and White (2008).

<sup>2</sup> This discussion is based on Shaffer (2011).

and human resource requirements for the implementation of project activities. Activities are the tasks undertaken to realize program objectives. Outputs are the goods or services resulting directly from the successful implementation of activities. Outcomes and impacts reflect longer-term project effects usually on some dimension of well-being. While these distinctions tend to blur in practice, the key point is that impact assessment tends to focus on the right hand side of figure 1, whereas standard project monitoring and evaluation on the left-hand side.

**Figure 1 The Impact Chain**



The second defining characteristic of impact assessment, addressing the attribution problem, stems from the fact that most outcome and impact indicators are affected by many variables which have nothing to do with a particular development project or program. One of the features of impact assessment is its attempt to control for the effects of these other ‘confounding variables’ when assessing the impact of a project or program. Otherwise stated, impact assessment attempts to make a *causal link* between the project or program under investigation and the outcome or impact variables of interest. This causal claim is the second distinguishing characteristic of impact assessment.

Of the different ways to establish a causal link, this paper focuses on those which are based on counterfactual dependence and mechanism-based conceptions of causation and models of causal inference, (Shaffer 2011, 2013). In the context of impact assessment, counterfactual analyses involve either experimental or quasi-experimental approaches. For the former, program participation is randomly assigned and consequently, population characteristics of treatment and controls groups are identical in expectation over large enough numbers (Duflo et al., 2008). The main complicating factors arise when assumptions of the underlying model are violated. Examples include incomplete compliance with assignment, externalities or spill-overs between treatment and control groups, non-random attrition among treatment and controls and so forth. There is a large literature on how to address these applied problems which usually involves using econometric models to adjust for them in a number of ways (Duflo et al., 2008).

There is a much wider variety of quasi-experimental approaches to impact assessment which differ in the statistical techniques used to construct comparison groups. Some of the better known approaches include regression discontinuity designs, where individuals just above and just below project eligibility thresholds are compared, pipeline approaches where eligible persons in line for project participation are compared with actual project participants, and matching procedures whereby individuals in comparison and treatment groups are matched on the basis of similar observable characteristics (Ravallion, 2008). One widely used technique, propensity score matching, uses logistic regressions to estimate the probability of participation, or the propensity score, which serves as the basis for the subsequent matching, though other non-parametric matching estimators are also available (Abadie and Imbens, 2009). In both experiments and quasi-experiments, the estimate of project impact, or the average treatment effect on the treated, is the difference in value of outcome indicators between treatment and control groups.

Applied mechanism-based approaches attempt to forge causal links between program activities or outputs and development outcomes or impacts, drawing on theory or the results of dialogic inquiry. Examples of such approaches include the Context-Mechanism-Outcome (CMO) model of Realistic Evaluation (Pawson and Tilly, 1997; Pawson, 2002), applied instances of theory-based evaluation (Weiss, 2000, White 2009b), some types of participatory assessment (Mayoux and Chambers, 2005) applied ethnographic evaluations (Adato, 2008) and others. Most of the examples presented in this chapter rely heavily on dialogical methods, such as focus group discussions and semi-structured interviews, to establish the causal links in question, though participant observation is also used.

### **3. Combining results and mechanisms**

A core contribution of Q<sup>2</sup> approaches in impact assessment has been to combine analyses of program outcomes and impacts, or 'results', with an examination of the underlying mechanisms generating them. There are many good examples in the literature of the combined analysis of results and mechanisms using a range of analytical techniques. Two such studies were undertaken by the International Food Policy Research Institute (IFPRI) in Nicaragua and Turkey to assess the impact of conditional cash transfer (CCT) schemes (Adato, 2008). Both of these impact evaluations integrated experimental or quasi-experimental designs with ethnographic studies to provide a combined account of results and mechanisms.

The Nicaraguan CCT program, the *Red de Protección Social* (RPS) provided cash and in-kind benefits to participants on condition that they attend health and nutrition workshops and their children participate in

growth monitoring and vaccination programs. Children also received nutritional supplements, including iron, and anti-parasite treatments.

The experimental component of the impact assessment drew on the fact that program placement was random. Specifically, 21 of 42 poor<sup>3</sup> administrative areas (*comarcas*) in the northern part of the Central Region were randomly selected for program participation. Household surveys, with a range of modules on social outcomes indicators, were administered in 2000, prior to the start of the program, and again in 2002. Accordingly, program impact was measured as the ‘double-difference’ between program participants and non-participants between 2000 and 2002. The ethnographic component involved village stays of 4 to 5 months and entailed multiple household visits by field researchers who conducted semi-structured interviews and engaged in participant observation.

One striking finding from the experimental analysis was the absence of any program impact on anaemia despite a sharp rise in the percentage of children receiving iron supplements. The double difference analysis found a slightly negative, but not statistically significant, program impact despite the increase, from 24% to 80%, of the percentage of children receiving iron supplements. The ethnographic study provided a potential explanation. Despite respondent claims in semi-structured interviews that they were providing iron supplements, only 3 of 60 case study households were observed to be doing so<sup>4</sup>. The narrative information attributed this apparent reluctance to the bad taste of iron and adverse side effects including vomiting and diarrhoea (Adato 2008: 229). In this case, information from the ethnographies provided a potential explanation for lack of program impact uncovered through the experimental analysis.

The second IFPRI impact evaluation was conducted on Turkey’s Social Risk Mitigation Program which became fully operational in 2004. The program provided cash payments conditional on school enrolment for boys and girls along with vaccinations and regular check-ups for children. As above, 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. For example, secondary enrolment rates for girls were below 40%.

---

<sup>3</sup> These 42 *comarcas* were selected on the basis of a set of poverty indicators including access to potable water, latrines and illiteracy rates (Mallucio and Flores, 2005: 4–5).

<sup>4</sup> This example illustrates the use of observation, not simply dialogue, in ethnographic inquiry and highlights differences between applied social anthropology and PRA techniques discussed in section 2.1.

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.

A final example of combining mechanisms and results involves a mixed methods impact assessment of the Transport Sector Support Program (PAST) rural infrastructure project in Nicaragua (Broegaard et al., 2011). In one area covered by this study, *Las Segovias*, narrative information from focus group discussions and semi-structured interviews was combined with results of a double difference quasi-experimental design involving propensity score matching. The data source for the latter included a baseline survey carried out in 2001 along with a resurvey undertaken in the context of the impact assessment.

One important finding from the double difference analysis was that the program had a positive and statistically significant effect on employment in both agriculture and construction. It was estimated that the project increased the number of hours worked per week by between 9.5 and 12.3 hours (Rand, 2010). The narrative information corroborated this finding and identified a number of the underlying reasons.

In the construction sector, there was a learning effect from the PAST projects in that other municipalities increasingly adopted similar labour-intensive methods. Further, past project participants gained employment due, in part, to their construction experience with PAST. With respect to agriculture, the employment boost was attributed to changes in relative prices between inputs and outputs and the attendant modifications to incentives. More specifically, employment was generated through the following processes: i) prices received by farmers rose due to more frequent, timelier and less expensive contacts with buyers; ii) revenue increased through reduction in post-harvest losses of higher value, perishable products; iii) land area under cultivation rose, increasing demand for on-farm employment. The narrative results provided evidence on the mechanisms generating results and provided empirical support for the validity of the quasi-experimental findings.

#### **4. Identifying comparison groups**

As discussed in section 2, applied examples of counterfactual dependence are predicated on the use of a control or comparison group of program non-participants. A core issue for quasi-experimental approaches, however, concerns unobservable population characteristics which may be affecting program participation and outcomes. If the distribution of such characteristics differs between treatment and comparison groups, then impact results may be biased either positively or negatively. A core argument in favour of randomisation is that it minimises this ‘selection bias’ in that characteristics of treatment and comparison groups are equal in expectation over large enough numbers (Duflo et al., 2008).

One contribution of  $Q^2$  analysis has been to draw on a wider range of sources of information to address the problem of selection bias. A hypothetical example is provided in Martin Ravallion’s (2001) fictional account of a chance encounter between ‘Ms. Analyst’ and ‘Ms. Sensible Sociologist’. The former is on a quest to find ways to address selection bias in her quasi-experimental impact assessment of an education project. A chance encounter with Ms. Sensible Sociologist reveals information on the determinants of program participation, namely that is heavily influenced by the school board in which one happens to live. Since all school boards receive the same allocation, a poor household living in a better-off school board has a higher chance of participation. Accordingly, a variable representing budget allocation to school boards could be used to estimate a model of participation, such as propensity score matching, which subsequently can be used as an instrumental variable in a model of schooling. In this case, narrative information on the determinants of program participation uncovered an observable variable which could be used in subsequent modelling.

In addition to uncovering observable variables, attempts have been made to incorporate either unmeasured or unobserved information for use in the construction of comparison groups. An example is provided in Rao and Ibáñez’s (2005) impact assessment of the Jamaica Social Investment Fund (JSIF), which financed a range of small scale projects proposed by local communities. In the impact assessment, detailed dialogic inquiry, including focus group discussions and semi-structured interviews, preceded the administration of a household survey and subsequent propensity score matching analysis.

Five communities were chosen for the impact assessment. The first stage in comparison group construction was to pair these communities with five other similar ones. This pairing exercise first took into account poverty scores used by the JSIF drawing on census data. Focus group discussions were then held to narrow down the pool of potential pairs, incorporating unmeasured variables such as geography,

occupational structure, number of churches, youth groups and so forth. Field visits were also conducted to improve the match, by taking into account ‘unobservables’ such as political culture and social structure.

The second stage consisted of data collection for the propensity score matching. Here, a number of contextual variables were included in the household survey on the basis of the previously collected narrative information such as Rastafarian affiliation and the availability of social networks. Overall, the authors maintained that use of these additional sources of information contributed to minimising the problem of selection bias (Rao and Ibáñez, 2005: 81). Unfortunately, it was not possible, within the design of this study, to test if the ‘improved’ comparison group mattered for impact results by examining, for example, if results were sensitive to the choice of comparison groups.

## **5. Conducting counterfactual thought experiments**

In addition to providing information on unobservable or unmeasured variables,  $Q^2$  approaches have contributed to comparison group construction through the use of thought experiments or mental simulation exercises. An example was provided in the 2003–2004 impact assessment of the national Hunger Eradication and Poverty Reduction (HEPR) in Vietnam<sup>5</sup>. The HEPR program comprised a number of targeted projects as well as policies on health care, education, and social support for the poor. The impact assessment was innovative in that it attempted to combine two ways of constructing a comparison groups with a view to determine if they would generate similar results or, if not, to spur reflection as to why.

The first approach involved propensity score matching. The primary objective of this exercise was to conduct sensitivity analysis on the results of a similar, prior analysis conducted by the World Bank in 2003.<sup>6</sup> Specifically, it assessed the sensitivity of results of the World Bank analysis to the choice of comparison group by presenting results for the nearest one, three and five matched non-beneficiaries. In addition, standard errors were calculated and confidence intervals presented for the impact estimates. The data source for the analysis 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.

---

<sup>5</sup> This section is based on Shaffer (2012).

<sup>6</sup> Results were published in the JDR (2004) and appear in Cuong (n.d.).

The second approach used in the HEPR study relied on a thought experiment, rather than intersubjectively observable information, to arrive at an appropriate counterfactual. As discussed in section 5.3, intersubjective observability is quite integral to the Holland-Rubin framework which defines the problem of causal inference as a problem of observation. Subjunctive conditional (if-then) questions were posed about what respondents would have done in the absence of the program.

In order to get meaningful answers to questions of this type, a mental simulation exercise is required which faces potential biases relating to human judgement (Elster, 1987; Gilovich and Griffin, 2002), survey design (Sudman et al., 1996), and the nature of dialogic processes (Chambers, 2003). Further, meaningful responses become increasingly difficult the greater the causal distance between program activities and the outcome/impact variable, the greater the number of intervening variables affecting outcomes/ impacts, the more complex the pattern of interaction among variables and the finer the scale in which the outcome/impact variable is measured (e.g. cardinal vs. ordinal). Nevertheless, it is worthwhile to examine whether such perceptual information differs systematically from intersubjective observables in the HEPR study. In the HEPR impact assessment, the two approaches were used for 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 Health Fee Exemption or Reduction entailed providing poor households or communes free or subsidised health care through: i) the distribution of health insurance cards or poor household certificates which entitle the holders to free or subsidised care and ii) the direct provision of free services in certain healthcare facilities or through mobile health units. The outcome variable, health care utilisation, was defined as the percentage of persons who used health care facilities over the past 12 months.

Table 1 presents results of the propensity score matching. Data suggest that the program has not had a statistically significant impact on utilisation of healthcare services (excluding traditional healers), which paralleled the findings of the World Bank study. Healthcare utilisation appears, in fact, to be lower among project participants than among their matched comparator for all three comparison groups but none of these results are statistically significant.

**Table 1 Propensity Score Matching: Impact of Health Fee Exemption/Reduction on Utilisation of Healthcare**

	<b>Mean Difference</b>	<b>Standard Error<sup>a</sup></b>	<b>[95% Conf. Interval]</b>
<b>Nearest Match</b>	-0.089	0.060	-0.226 - 0.020
<b>Nearest Three Matches</b>	-0.078	0.049	-0.196 - 0.006
<b>Nearest Five Matches</b>	-0.077	0.048	-0.176 - 0.009

<sup>a</sup>Standard errors were bootstrapped with 100 replications  
 Data source: Vietnam Household Living Standards Survey, 2002

The self-report exercise attempted to assess the impact on utilisation of health care services by asking 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. Table 2 presents results of this exercise. The vast majority 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. 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 2 Self-Reported Assessment of Use of Medical Care in HEPR Absence (Population Proportions, Standard Errors in Parentheses)<sup>a</sup>**

	<b>1 Yes</b>	<b>2 No</b>	<b>Total</b>
<b>Total Vietnam</b>	91.81 (0.40)	7.28 (0.11)	100

<sup>a</sup>Data do not sum to 100 because "Don't Knows" have been removed  
 Data source: HEPR Impact Assessment Qualitative Survey, 2003-4

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. The outcome variable, school attendance, was defined as the percentage of children aged 6–17 who attended school over the past 12 months.

Table 3 reveals a modest, but statistically significant, impact of the program on school attendance for all three matched comparisons, which paralleled the findings of the World Bank Study. The impact range is between 3% and 15% depending on the comparison group used and taking into account sampling error.

**Table 3 Propensity Score Matching: Impact of Tuition and School Maintenance Fee Exemption/Reduction on School Attendance**

	<b>Mean Difference</b>	<b>Standard Error</b>	<b>[95% Conf. Interval]</b>
<b>Nearest Match</b>	0.089	0.028	0.032 - 0.135
<b>Nearest Three Matches</b>	0.092	0.026	0.060 - 0.152
<b>Nearest Five Matches</b>	0.076	0.023	0.037 - 0.128

<sup>a</sup>Standard errors were bootstrapped with 100 replications  
Data source: Vietnam Household Living Standards Survey, 2002

Table 4 presents results of the self-report exercise. Around 12% of respondents claimed that they would not have enrolled their children in primary or secondary school in the absence of the program. This figure is within the range of program impact found in the propensity score matching exercise which examined *actual* differences in attendance between program participants and non-participants.

**Table 4 Self-Reported Assessment of Primary or Secondary Enrolment in HEPR Absence (Population Proportions, Standard Errors in Parentheses)**

	<b>1 Yes</b>	<b>2 No</b>	<b>3 DK<sup>a</sup></b>	<b>Total</b>
<b>Total Vietnam</b>	87.64 (1.01)	11.56 (0.32)	0.80 (0.69)	100

<sup>a</sup> Don't Know  
Data source: HEPR Impact Assessment Qualitative Survey, 2003-4

In summary, the HEPR impact assessment combined intersubjectively observable data and thought experiments to determine what would have happened in the absence of the program. These techniques generated similar results for the education and health project components of HEPR and as such, serve to enhance the overall validity of results. Further, the findings provide preliminary support for the validity of thought experiments as a means of constructing a counterfactual situation though additional research is required to establish their validity across a range of situations.

## 6. Defining benefits

A final contribution of Q<sup>2</sup> analysis to impact assessment concerns the metric which should be used to gauge program success or failure. How should project benefits be defined and by whom?

Q<sup>2</sup> analyses have contributed to addressing this question in two ways. First, in the most general sense, a wide range of information sources can inform a broad understanding of the full scope of intended and unintended program effects and the differentiated impact upon population groups. With respect to the former issue, White (2011) has drawn attention to unintended effects of rural road projects often missed in impact assessments. Drawing on a range of secondary sources, his list of potentially adverse consequences includes increases in traffic accidents, air and noise pollution and opening up an area to military and police control. In terms of affected populations, White argues strongly for disaggregated results based on categories of social differentiation such as caste, gender, ethnicity and so forth, and the nature of livelihood strategies. In Laos, for example, rural roads appear to have increased rural inequality by disproportionately benefitting those able to take advantage of greater market access, including men relative to women.

A second contribution of Q<sup>2</sup> has been to provide a sense of the relative importance of different evaluative metrics for project participants. An example is de Silva and Gunetilleke's (2008) evaluation of resettlement schemes implemented under the Southern Transport Development Project (STDP) in Sri Lanka. This highway development project led to the displacement of around 1400 households who were subsequently resettled. The methodology included a household survey, with fixed response and open-ended questions, as well as focus group discussions. While the study focussed on monitoring, rather than impact assessment per se, it, nevertheless, addressed the relevant issue that 'metrics matter'.

In general terms, household survey data suggested a significant improvement in the quality of new housing provided by the project, as measured by household size, access to toilets, water and energy. High levels of satisfaction with these amenities were reported. On the other hand, focus group discussions recorded considerable unease with the new living environment due to the loss of a quiet rural environment, coolness, shade, access to fruit and space for garbage disposal and family burials. Interestingly, the very poor and landless did not share such concerns in that they had limited space previously. According to one respondent: 'in our previous place, we had no place even to spit' (de Silva and Gunetilleke 2008: 258).

A second issue concerned the loss and replacement of land devoted to paddy cultivation. A working assumption among project staff was that such land is not highly valued because paddy cultivation generates very low returns and necessitates arduous effort. The low valuation of paddy land was reflected in the low compensation paid by the project and the fact that households did not replace paddy lands lost to land acquisition in their new surroundings.

A different picture of the local value assigned paddy emerged from the focus group discussions. There was widespread unhappiness about the loss of paddy land. Three factors figured prominently in the discussion. First, there was dissatisfaction with the financial burden of purchasing paddy which had previously been produced on the farm. Second, there was sadness at the ending of an important aspect of social life, sharing the harvest amongst kin. Finally, there was a pervasive sense of loss at having been dispossessed from land held for generations in one's family. Overall, the assessment of the resettlement experience was much less favourable according to results of focus group discussions than to household survey data. According to the authors (de Silva and Gunetilleke 2008: 260).:

*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 STDP has caused to be suddenly severed ... A major articulated loss is the loss of the traditional/ancestral village and the lifestyle that goes with it.*

## **7. Conclusion**

Impact assessment is about establishing a causal link between program activities or outputs and development outcomes or impacts. This paper has presented an empirical case for methodological and causal pluralism in impact assessment on grounds that Q<sup>2</sup> designs add value of understanding and explaining impact.

More specifically, a number of cases were presented of attempts to combine results and mechanisms. The value added of this design, to explain the reasons underlying observed results, represents one of the key benefits of Q<sup>2</sup> in impact evaluation. Two other categories of approaches were discussed which presented innovative ways of constructing a comparison group or counterfactual scenario, drawing on narrative information, in the case of unmeasured or unobservable variables, and on thought experiments. While these examples leave room for optimism about the role of such approaches, more research is required to establish their validity. A final category of examples reviewed Q<sup>2</sup> approaches which investigated if the assessment of program performance is sensitive to the definition of benefits and concluded that the metric does indeed 'matter'.

Overall, the examples reviewed suggest that Q<sup>2</sup> approaches have provided a richer causal analysis of program benefits by integrating different models of causal inference and by bolstering analyses of counterfactual dependence. As such, they have contributed to enriching our understanding of the core questions of impact assessment namely, ‘who benefits, how much, and why?’ Such studies support the empirical case for methodological and causal pluralism.

## References

- Abadie, A. and Imbens, G. (2009). 'Matching on the Estimated Propensity Score', *NBER Working Paper 15301*. Cambridge, M.A: National Bureau of Economic Research.
- 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.
- Bamberger, M., Rao, V., and Woolcock, M. (2010). 'Using Mixed Methods in Monitoring and Evaluation. Experiences from International Development', *Policy Research Working Paper 5245*, World Bank, Washington, D.C.
- Bourguignon, F. and da Silva, L. (eds) (2003). *The Impact of Economic Policies on Poverty and Income Distribution. Evaluation Techniques and Tools*. Washington, D.C.: World Bank and Oxford University Press.
- Broegaard, E. , Freeman, T. and Schwensen, C. (2011). 'Experience from a Phased Mixed-Methods Approach to Impact Evaluation of Danida Support to Rural Transport Infrastructure in Nicaragua', *Journal of Development Effectiveness*, 3(1): 9–27.
- 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.
- Cuong, N. (n.d.). 'Assessing the Impact of Poverty Reduction Programs in Vietnam', Mimeo, Hanoi.
- 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.
- Duflo, E., Glennerster, R. and Kremer, M. (2008). 'Using Randomization in Development Economics Research: A Toolkit', in T.P. Schultz and J. Strauss (eds), *Handbook of Development Economics. Volume 4*. Amsterdam: North Holland, 3895–3962.
- Elster, J. (1987). *Sour Grapes. Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.
- 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.
- 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.
- Maluccio, J. and Flores, R. (2005). 'Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Protección Social', *Research Report 141*, International Food Policy Research Institute, Washington, D.C.

- 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.
- Pawson, R. (2002). 'Evidence-based Policy: The Promise of "Realistic" Synthesis', *Evaluation*, 8(3): 340–358.
- Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage.
- Rallis, S. and Rossman, G. (2003). 'Mixed Methods in Evaluation Contexts: A Pragmatic Framework,' in A. Tashakkori and C. Teddlle (eds), *Handbook of Mixed Methods in Social & Behavioral Research*. Sage: Thousand Oaks, California, 491–512.
- Rand, J. (2010). 'Evaluating the Employment Generating Impact of Rural Roads in Nicaragua', *Journal of Development Effectiveness*, 3(1): 28–43.
- 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.
- 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*. Washington, D.C.: World Bank and Oxford University Press, 165–190.
- Ravallion, M. (2001). 'The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation', *The World Bank Economic Review*, 15(1): 115–140.
- Ravallion, M. (2008). 'Evaluating Anti-Poverty Programs', in T.P. Schultz and J. Strauss (eds), *Handbook of Development Economics. Volume 4*. Amsterdam: North Holland, 3787–3846.
- Roche, C. (1999). *Impact Assessment for Development Agencies*. Oxford, U.K.: Oxfam Publications.
- 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. (2012). '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. (2013 (forthcoming)). *Q-Squared. Combining Qualitative and Quantitative Approaches in Poverty Analysis*. Oxford: Oxford University Press.
- Sudman, S., Bradburn, N. and Schwartz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers.
- Weiss, C. (2000). 'Theory-based Evaluation : Theories of Change for Poverty Reduction Strategies', in O. Feinstein and R. Picciotto (eds), *Evaluation and Poverty Reduction*. Washington, D.C.: The World Bank, 83–84.

- White, H. (2008). 'Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation', *IDS Bulletin*, 39(1): 98–109.
- White, H. (2009a). '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.
- White, H. (2009b). 'Theory-based Impact Evaluation: Principles and Practice', *International Initiative for Impact Evaluation Working Paper No. 3*, International Initiative for Impact Evaluation, New Delhi, India.
- White, H. (2011). 'Achieving High-Quality Impact Evaluation through Mixed Methods: the Case of Infrastructure', *Journal of Development Effectiveness*, 3(1): 131–144.