Beneath the 'Methods Debate' in Impact Assessment: Baring Assumptions of a Mixed Methods Impact Assessment in Vietnam*

Paul Shaffer
Dept. of International Development Studies, Trent University
Email: paulshaffer@trentu.ca

Abstract

The past decade has seen renewed interest in the use of mixed method approaches across the social sciences and in the field of impact assessment. This body of work has focused on questions of method with insufficient attention devoted to foundational issues. The objective of the present article is to bare the foundational concepts which guided a mixed method impact assessment of the Hunger Eradication and Poverty Reduction (HEPR) program in Vietnam. Specifically, it shows how the applied methods used in the HEPR study rested on foundational differences concerning: conceptions of causation and models of causal inferences (probabilities vs. mechanisms); analytical focus (outcomes vs. processes) and external validity (empirical generalization vs. statistical inference); and constituents of 'objective' knowledge (inte
rsubjective observables vs. perceptual data).

* A version of this paper is forthcoming in the *Journal of Development Effectiveness*.

Q-Squared Working Paper No. 56
Winter 2012

1. Introduction and Context

The past decade has seen renewed interest in the use of mixed method approaches across the social sciences. This trend has been reflected, *inter alia*, in the publication of a *Handbook of Mixed Methods in Social and Behavioral Research* (Taskahkkori and Teddle, eds., 2003) and the emergence of journals dedicated to this theme such as the *Journal of Mixed Method Research* and the *International Journal of Multiple Research Approaches,* etc. The integration of qualitative and quantitative methods has been particularly prominent in a number of areas of applied research, such as poverty analysis, which has seen a burgeoning number of contributions.[1]

This same trend is visible in the literature on evaluation and impact assessment, though there is a long history of interdisciplinarity in both fields (Roche 1999). A number of journals promoting mixed methods have recently emerged including the *Journal of Multidisciplinary Evaluation* along with the present journal. The recent Network of Networks for Impact Evaluation (NONIE) Guidance Paper (Leeuw and Vaessen 2009) makes a strong case for the use of mixed methods in impact assessment as do a number of important recent contributions in the field (e.g. White 2008, White and Bamberger 2008).

Much of the work on mixed methods in impact assessment has highlighted questions of method. Inquiry has focused on optimal strategies of mixing and/or empirical results of mixed method studies.[2] Such analysis is useful in promoting the case for mixed method research and in informing better practice. Nevertheless, insufficient attention has focused on foundational concepts on which different approaches to impact assessment rest. This point applies equally to broader debates within impact assessment about say, the relative merits and demerits of randomised control trials, which have focussed almost exclusively on issues of internal and external validity (e.g. see contributions in Banerjee, 2007).

The objective of the present article is to bare the foundational concepts which guided a mixed method impact assessment of the Hunger Eradication and Poverty Reduction (HEPR) program in Vietnam. Specifically, it shows how the applied methods used in the HEPR study rested on foundational differences concerning: conceptions of causation and models of causal inferences (probabilities vs. mechanisms); analytical focus (outcomes vs. processes) and external validity (empirical generalization vs. statistical inference); and constituents of 'objective' knowledge (intersubjective observables vs. perceptual data). Arguably, many of applied debates about the merits or demerits of approaches to impact assessment are really debates about underlying issues such as these.

The present article makes three contributions to the literature: first, it outlines foundational differences between approaches to impact assessment and shows how they matter for practice; second, it presents the

methodology of an innovative mixed method impact assessment, the aforementioned HEPR impact assessment in Vietnam; third, it presents select empirical results which bear on questions of the internal validity of mixed method approaches.

The format of the paper is as follows. Section 2 outlines the approaches to impact assessment which informed the HEPR impact assessment and presents the methodology of the study. Section 3 reviews the conceptions of causation and models of causal inference which underlie the said approaches to impact assessment and discusses the conflicting empirical results which they generated. Section 4 addresses the different analytical foci, and ways of establishing external validity, in approaches to impact assessment and shows how the HEPR impact assessment attempted to integrate them. Section 5 discusses questions of 'objective' knowledge and shows how different sources of information were used in the context of 'counterfactual causality'. Section 6 concludes.

There is one point which should be made explicit at the outset. The core objective of the present article is to illustrate how philosophical and theoretical assumptions 'matter' for practice, on the basis of an actually existing impact study undertaken in Vietnam. As with all real world impact assessments, the HEPR study had a number of limitations due to the methodology selected as well as time and financial constraints. These limitations are discussed at different points below. There is no attempt, however, to redo the analysis of 2003-4 which would respond to a different core objective than that of this article.

2. The Hunger Eradication and Poverty Reduction (HEPR) Impact Assessment

The national Hunger Eradication and Poverty Reduction (HEPR) program was launched by the government of the Socialist Republic of Vietnam in 1998. The program comprised a number of targeted projects as well as policies on health care, education, and social support for the poor. HEPR was renewed in 2001 (for the period up to 2005) and merged at that time with the Employment Creation Program. The merged program combined 12 targeted projects plus support policies on social safety nets.

The impact assessment of the HEPR took place in 2003-2004. It combined a range of approaches to impact assessment along with a number of methods of inquiry. Of particular note, in this regard, was the use of a 'Qualitative Survey' which aimed to provide information which was not already available in Vietnam. The following discussion outlines the approaches to impact assessment which informed the HEPR impact assessment before reviewing the methodology of the study.

Approaches to Impact Assessment in the HEPR Study

The term 'impact assessment' is often used in different ways and consequently ends up referring to different things (White 2009). In the HEPR study, the term was defined in terms of two core characteristics. First, it is concerned with well-being outcomes or impacts and not project activities or outputs. Second, it is concerned with attributing outcomes or impacts to project interventions and not simply tracking changes in them (i.e. it addresses the 'attribution problem'). The critical point is the impact assessment must make a causal claim linking program activities or outputs to outcomes or impacts.

There are many ways to attempt to make the causal links in question, all of which have strengths and limitations (Shaffer 2011). Furthermore, practical concerns, related to time and resource constraints, along with considerations of policy-relevance, often are key factors driving the choice of approach. In the present case, such considerations led to the three broad approaches to impact assessment which informed the methodology design of the HEPR study, namely: i) Process Tracing; ii) Self-Reports; iii) Quasi-Experiment.

The first, following the terminology in George and Bennett (2004), was *process tracing*. In this approach, causal links are established between program outputs and well-being outcomes or impacts on the basis of information from, *inter alia*, interviews, focus groups, causal maps, etc. PRA-type tools are often used to establish these causal links. The objective of this approach is to understand the processes generating, or failing to generate impact. It addresses the following question: *how and why has the project succeeded or failed to have an impact on well-being?* Answers to these sorts of questions are particularly relevant for program redesign.

The second approach involved *self-reports*. After first establishing causal links between project outputs and outcomes/impacts, ranking exercises may be conducted, or questions posed, to assess the perceived well-being impact of project components (often expressed on an ordinal scale). It addresses the following question: *how much impact has a project had over time according to people's perceptions*. It should be noted that there are well-known potential biases associated with self-report data, which raises caution about interpretation of results (see Section 5).

Third, a *quasi-experimental* component was integrated. In this approach, a comparison group is created statistically which is supposed to be identical to project participants, the 'treatment group', in all respects except for project participation. Well-being outcomes or impacts are compared in treatment and comparison groups and the difference is assumed to be the impact of the project. This approach aims to provide an estimate of the magnitude of impact. It addresses the following question: *how much impact has the project had relative to what would have happened if the treatment group had not participated in the program?*

Methodology of the HEPR Study

The three aforementioned approaches to impact assessment were integrated in three main components of the HEPR impact assessment: i) Literature Review; ii) Qualitative Survey (QS) and iii) Propensity Score Matching.

The *literature review* was a straight-forward application of the *process tracing* approach. The primary objective was to provide a detailed understanding of the processes, or transmission mechanisms, linking outputs to outcomes or impacts. Attention focused on the reasons why project components have succeeded or failed to generate a well-being impact. Information was compiled and synthesised from a range of sources including PRAs, project evaluations, research studies, etc. The primary use of this information was to guide questionnaire design for the 'Qualitative' Survey and to provide background information about select processes driving outcomes. There were two key limitations of the data contained in the literature reviews. First, the 'representativeness' of the findings was unclear. Second, these data generally, did not provide an estimate of the magnitude of impact.

The key contribution of the Qualitative Survey (QS) was to combine *process tracing* exercises and *self-reports*. Its two primary objectives were to present a 'representative' account of peoples' perceptions of the magnitude of project impact along with factors limiting project impact. In the QS, respondents were asked a preliminary, open-ended series of questions on the processes generating impact (*process tracing*). A follow-up question asked respondents to rank the well-being impact of the program in light of their responses to the open-ended ones (*self-reports*). The latter question was prefaced by the former because one is unlikely to get meaningful responses to a lead question such as 'how much impact did the project have on your lives.' Next, a question was administered on the reasons for negative, no or insignificant impact, using pre-coded reasons drawn, *inter alia*, from the results of the literature review (*process tracing*). A secondary contribution of the Qualitative Survey was to provide a mental simulation-equivalent of the information provided by quasi-experimental approaches (see Section 5). It asked respondents, for select program components, what they would have done in the absence of the program.

The claim of 'representativeness' of data from the Qualitative Survey was based on the fact that sampling was done probabilistically. Accordingly, standard errors could be calculated for the statistics generated (usually population proportions within different response categories). The sampling strategy involved a four stage process including selection of provinces, communes, villages and households.[3] Total sample size was 3700 households. Data collection was completed in early 2004.

It must be acknowledged that the structure of the Qualitative Survey, which placed emphasis on ensuring comparability of responses in different sites, and generating statistically significant numerical information, limited its flexibility as an information gathering tool. Accordingly, the open-ended probing was used to 'set up' subsequent fixed-response questions and not used in the write-up of the report. Time and financial constraints precluded analysis of these narrative data, along with concerns about ensuring the comparability of results. As in many real world impact studies, there was a tradeoff between comprehensiveness, on the one hand, and producing timely, comparable, nationally representative, on-budget results, on the other.

The final component, the *propensity score matching*, is an example of a *quasi-experimental* approach. 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.[4] 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. A logistic (logit) regression was estimated to calculate the value of the propensity score, which formed the basis of the matching between comparison and treatment groups. Matching was done within sub-populations grouped by poverty status (according to commune authorities), followed by ethnicity, urban/rural strata and region. 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.

There are at least four limitations of the PSM exercise which should be noted.[5] First, using single cross-sectional data to capture the causal effect of a program rests on the assumption that there is a similar distribution of relevant population characteristics between treatment and control groups. A problem arises if this assumption is violated as say, in the case of credit, where there is an element of self-selection and where 'unobservables' such as entrepreneurial ability may be driving results. Second, the match of observable characteristics over the entire range of the propensity score was close, but not perfect, as shown in the kernel density estimates in Appendix. Third, the PSM did not conduct sub-group analysis, which could have contributed to explaining average treatment effects on the treated. Fourth, the bootstrapping exercise undertaken to calculate standard errors used 100 replications which may be on the low side and may generate invalid standard errors for nearest neighbour estimates.[6]

In terms of the program subcomponents which figured in the impact assessment, it was necessary to draw from the twelve HEPR projects. Selection criteria included: i) financial importance; ii) policy relevance; iii) strength of the link between a program output and a well-being outcome. On the basis of such criteria, the following program components were selected for inclusion in HEPR study: i) Credit Project (Vietnam Bank

for the Poor); ii) Health Care Support Policy (Health Fee Exemption or Reduction); iii) Education Support Policy for the Poor (Tuition or School Maintenance Fee Exemption or Reduction); iv) Extension Services; v) Resettlement and Development of New Economic Zones; vi) Sedentarisation for Ethnic Minority Groups. The present article focuses on the first three HEPR components. It should be noted that the impact study did not attempt to estimate the combined effect of the program, but only the impact of its individual components.

It is important to emphasise that the HEPR study was a collaborative effort involving the Ministry of Labour, Invalids and Social Assistance (MOLISA), the General Statistical Office (GSO), the Institute of Economics, the United Nations Development Programme (UNDP) Support Project (VIE/02/001) and the author. MOLISA staff made significant contributions to the formulation of the methodology and the design/pretest of the questionnaire for the 'Qualitative' Survey. The General Statistical Office (GSO) played a very significant role in designing, pretesting, revising, administering and entering data from the QHS questionnaire. The Institute of Economics, in particular Dr. Nguyen Thang, were responsible for developing the QS sampling strategy, analysis of data from the Qualitative Survey, conducting sensitivity analysis on the propensity score matching exercise and undertaking literature reviews. The UNDP Support Project organised the training sessions for enumerators and backstopped the entire process. The role of the author included formulation of the methodology for the impact assessment, questionnaire design/pretest/revision, training of enumerators, review/oversight of the process as well as drafting the final consolidated report.

3. Causal Concepts and Causal Inference: Combining Approaches[7]

As discussed above, the distinguishing characteristic of impact assessment is its attempt to make a causal link between project activities or outputs and outcome or impact measures. The HEPR impact assessment attempted to combine two different conceptual approaches to causation and two models of causal inference. The intent was to bolster the internal validity of results, if both approaches ended up telling a broadly consistent story, or, if not, to spur reflection on why they may differ.

As discussed in the preceding section, two (of three) approaches to impact assessment used in the HEPR study were the propensity score matching and process tracing exercises. The following discussion will begin by outlining the conceptions of causation and models of causal inference which underlie these approaches to impact assessment and proceed to review empirical results.

The literature on causation is vast. In Schaffer's (2008) survey piece, a wide range of causal theories are reviewed whose foundations include: counterfactual dependence, nomological subsumption, statistical

correlation, agential manipulability, contiguous change, physical processes and property transference. One way of distinguishing between these approaches relies on the contrast between probabilities and processes: 'the nomological, statistical, counterfactual and agential accounts [of causation] … understand connection in terms of probability: causing is making more likely. The change, energy, process and transferring accounts [of causation] converge in treating connection in terms of process [mechanism]: causing is physical producing' (Schaffer, 2008).

Propensity score matching is probabilistic and based on a counterfactual conception of causation. It model of causal inference is probabilistic in that causation is inferred if there are statistically significant differences in outcomes between comparison and treatment groups. It is a counterfactual conception of causation in that the causal claim that project *a* causes outcome *e*, depends on the counterfactual claim about what would have happened to *e* in the absence of *a (*Menzies 2009*).* As phrased in a recent discussion of the 'archetypal evaluation problem' in the context of quasi-experiments*:* 'an "impact evaluation" assesses a program's performance in attaining well-defined objectives against an explicit counterfactual, such as the absence of the program' (Ravallion 2008).

Process tracing is 'physical producing' and relies on a mechanism-based model of causal inference. It is physical producing in that it aims to identify mechanisms through which project impact is generated. Otherwise stated, causal inference depends upon identifying the causal mechanisms generating causal effects. According to Little (1998: 202): '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.'

It should be noted that there is debate as to the precise definition of causal mechanism (Hedström and Swedberg 1998, Pickel 2004).[8] In the context of impact assessment, three aspects of mechanisms are relevant, namely: the causal variables; the links or pathways between them, i.e. the causal 'tree'; as well as an explanation of why they are linked. Mechanism, in this sense, focuses on the reasons for observed outcomes. In the HEPR impact assessment, mechanisms were also used to elicit information on the magnitude of impact (as discussed below).

In the HEPR impact assessment, propensity score matching and process tracing techniques were combined to assess the credit component of the program. Credit is comprised of loans disbursed by the Vietnam Bank for the Poor (VBP). The VBP provides small loans for investment in income-generating activities, such as livestock rearing and agriculture. Interest rates are highly subsidized at rates which generally, do not exceed one percent per year. Loan duration varies from one to five years.

Table 1 presents results of the propensity score matching. Data suggest that the program has not had a statistically significant impact on household expenditure per capita. Average expenditure per capita appears, in fact, to be lower among credit recipients than among their matched comparator for all three comparison groups but none of these results are statistically significant. These results parallel those of the World Bank study. This finding may be due to the fact that the VHLSS questionnaire asked about receipt of credit *anytime* within the last twelve months. For very recent recipients, this is clearly not sufficient time to generate a return.

Table 1 Propensity Score Matching: Impact of Credit on Per Capita Household Expenditure (000 VND)

	Mean Difference	Standard Error[a]	[95% Conf. Interval]
Nearest Match	-14.44	132.93	-469.9 - 183.6
Nearest Three Matches	-11.96	107.37	-341.2 - 149.3
Nearest Five Matches	-47.72	95.73	-254.7 - 96.1

[a]Standard errors were bootstrapped with 100 replications
Data source: Vietnam Household Living Standards Survey, 2002

The credit module in the Qualitative Survey used the *process tracing* + *self-report* format discussed in the preceding section. 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". The objective was to probe the transmission mechanisms or pathways from receipt of credit to changes in living standard. As with all the open-ended questions, a list of positive and negative probes was drafted to assist enumerators in probing pathways leading to positive or negative impact. 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?"

Table 2 presents results of the follow up question. 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. Much smaller proportions felt impact was significantly negative or non-existent.

Table 2 Self-Reported Impact of the Loan on Household Income (Population Proportions, Standard Errors in Parentheses)

	1 Sig. Pos.	2 None	3 Insig.	4 Sig. Neg.	5 DK	6 2+3+4	Total
Total Vietnam	52.57	8.25	33.20	5.95	0.03	47.40	100
	(2.08)	(0.27)	(3.56)	(1.24)	(0.03)		

Data source: HEPR Impact Assessment Qualitative Survey, 2003-4

This finding presents a more positive assessment of program impact than the results of the PSM exercise. There are a number of potential reasons which may explain these differences. First, as mentioned above, the time frame for the PSM was the twelve month period prior to the VHLSS, whereas the present time period 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 good credit program with significant impact upon income over time may show little or no impact in a PSM if a suitably matched comparison group also has access to credit from other sources. Both approaches legitimately measure 'impact' but relative to different scenarios. Third, the outcome indicator in the PSM, household expenditure per capita, is different from that in the process tracing, household income. Fourth, it is possible that the open-ended questions were not sufficiently probed, leading to a positive response bias.

To conclude, two different conceptions of causation and models of causal inference were used in the HEPR impact assessment with a view to determine if they generated broadly consistent results. The propensity score matching exercise, which relies on probabilistic causal inference and counterfactual conceptions of causation, did not find significant impacts of credit on household expenditure per capita. The process tracing exercise, which is a 'physically producing'/mechanism-based account of causation, followed by a self-report on the magnitude of impact, came to a much more positive conclusion, with around half of respondents claiming a significantly positive effect on household income. While these conflicting results did not serve the end of bolstering the validity of study results, they did spur dialogue on why different approaches to impact

assessment may generate different results. Further, they highlight the potential role of foundational differences related to causation as drivers of conflicting results.

4. Outcomes, Processes and External Validity: Integrating Approaches

The distinction between probabilities and (physically producing) mechanisms discussed in the preceding section relates closely to a distinction between outcomes and processes in the literature on disciplinary differences in the social sciences and methodological differences between quantitative and qualitative approaches. The HEPR impact assessment attempted to combine information on outcomes and processes with a view to explain observed outcomes, and to provide a basis for determining the 'representativeness' of information on processes. The following discussion will begin by further unpacking the process/outcome distinction and proceed to review empirical results.

A classic statement of the distinction between outcomes and processes is from Michael Lipton (1992):

> Economics is mainly about outcomes… [not] about processes. Economists, of course, have models of perfect competition, or bargaining to reach a Nash equilibrium, or surplus extraction and use by the dominant class. But economist's tests show only whether a modelled process is consistent with the measured outcomes … Only seldom does the economist empirically explore the processes themselves.

Likewise, Bardham and Ray (2006) rely on this distinction to contrast economic and anthropological analyses of social phenomena: 'while anthropologists are thus better at telling us how a variable mattered to the outcome, economists are often better at measuring how much it mattered.'

In the impact assessment context, the affinity between outcomes or effects and, propensity score matching is made explicit in the canonical model which underlies experimental and quasi-experimental approaches to impact assessment. This is known as the Holland-Rubin framework in reference to seminal papers by its authors (Rubin 1974, Holland 1986). According to Holland (1986: 945): 'Others are interested in understanding the details of causal mechanisms. The emphasis here will be on *measuring the effects of causes* because this seems to be a place where statistics, which is concerned with measurement, has contributions to make.'

With respect to processes, in the literature on qualitative and quantitative integration, there are many good examples of the use of detailed small *n* studies, using a variety of methods, to explain outcomes in the context

of impact assessment, (Adato 2008), and more broadly, (Mosse 2006). A limitation of such small *n* studies, however, concerns external validity, or the extrapolation of study results to different populations in time and space. Usually, empirical generalisation of this sort requires establishing the typicality of results over a broader population (Hammersley 1997). Establishing typicality is extremely difficult, however, and, unlike statistical inference, there are no hard set rules about how to do it.

It is important not to overstate the linkages between the outcome/process distinction and either the divide between disciplines or approaches to impact assessment. There are good examples of empirically informed econometric models of social processes including Rao et. al. (2003) and de Weert (2006). In addition, in the context of impact assessment, it is possible to create comparison or control groups for project subcomponents to determine to what extent these intermediate links (i.e., processes) are driving program impact. Nevertheless, there is a mapping between the outcome/process distinction and the distinction between probabilistic and mechanism-based approaches in impact assessment.

As discussed above, the HEPR impact assessment drew on an extensive literature review of small *n*, 'qualitative', studies undertaken in Vietnam on project impact. It opted not to conduct additional similar studies because the existing body of literature was reasonably comprehensive. Instead, a decision was taken to include a number of pre-coded responses in the questionnaire on the reasons for lack of impact for those respondents who fell in the 'no, insignificant or significantly negative impact' categories. These pre-coded responses were drawn directly from those found in the literature review. Given that sampling was done probabilistically, standard errors could be calculated for population proportions within the different response categories, and claims of statistical 'representativeness' could be made.

The credit component of the HEPR study combined information on outcomes and processes. For those respondents who answered that the program did not have a significantly positive impact on household income, a follow up question inquired as to the reasons why. The pre-coded options included: 1) Loan was too Small; 2) Loan was too Big; 3) Duration of Loan was too Short; 4) Lack of Support Services such as Training, Extension, Veterinary Services and Financial Management; 5) Forced to Invest in Only Livestock or Agriculture; 6) Lack of Knowledge about Investment Opportunities; 7) Other.

Table 3 (below) presents the results of this question for all regions in the study. The two dominant reasons for the lack of a significantly positive result at the national level were: small loan size and short loan duration. There is variation between and within regions. In the South, lack of support services figured very prominently as it did in the North West (Sonla) and the Central Highlands. Lack of knowledge about investment opportunities was a major response item only in the SouthEast and did *not* appear to disproportionately affect

ethnic populations[9]. These latter findings provide important qualification to the earlier *small n* results, which did not provide a ready basis for determining the external validity of results, and at times implied much greater generality of results than warranted.

In summary, the HEPR impact assessment attempted to combine data on outcomes, i.e. the magnitude of impact, with information on processes, i.e. the reasons for lack of impact. Given that sampling was done probabilistically, standard errors were calculated for population proportions within different response categories, allowing for judgements about the 'representativeness' of results. Accordingly, it provided the added dimension of breadth, or external validity, to the existing body of small *n* studies of the program, while retaining, in skeletal form, the information on processes provided by the latter.

Table 3 Reasons why the Loan had No, Insignificant or a Significantly Negative Impact (Population Proportions, Standard Errors in Parentheses)

	1 Loan was too small	2 Loan was too big	3 Duration was too short	4 Lack of Support services	5 Forced to invest in livestock or agriculture	6 Lack of knowledge on investment opportunities	7 Other	8 DK[a]	Total
Location	29.23	0.00	29.82	12.24	4.26	8.24	16.20	0	
North	(2.06)	(0.00)	(5.00)	(4.20)	(0.56)	(0.31)	(2.62)	(0)	100
Red River Delta	24.72	0	23.12	11.36	0	13.61	27.19	0	100
	(1.86)	(0)	(2.44)	(3.23)	(0)	(7.81)	(6.74)	(0)	
North East	31.08	0	32.49	11.81	5.48	6.56	12.58	0	100
	(1.33)	(0)	(4.70)	(4.26)	(1.24)	(3.55)	(4.08)	(0)	
North West	17.18	0	13.55	27.07	4.41	11.29	26.48	0	100
	(0.33)	(0)	(0.17)	(3.00)	(0.49)	(3.22)	(6.21)	(0)	
Sonla	17.18	0	13.55	27.07	4.41	11.29	26.48	0	100
	(0.33)	(0)	(0.17)	(3.00)	(0.49)	(3.22)	(6.21)	(0)	
Centre	31.25	0.03	26.79	11.24	1.97	13.32	15.39	0	100
	(0.13)	(0.03)	(1.69)	(0.31)	(1.01)	(0.53)	(3.64)	(0)	
North Central Coast	38.21	0.00	28.39	11.34	0.00	7.39	14.67	0	100
	(0.65)	(0.00)	(0.12)	(3.41)	(0.00)	(0.27)	(2.38)	(0)	
South Central Coast	26.18	0.05	27.15	9.19	3.48	18.61	15.35	0	100
	(2.85)	(0.06)	(3.35)	(0.09)	(1.30)	(0.91)	(6.45)	(0)	
Central Highland	34.24	0	18.52	23.04	0.70	5.09	18.40	0	100
	(6.43)	(0)	(6.37)	(0.50)	(0.99)	(2.41)	(9.90)	(0)	
Kontum	17.60	0	21.08	19.16	2.33	16.82	23.01	0	100
	(4.47)	(0)	(7.31)	(0.12)	(2.25)	(0.57)	(9.07)	(0)	
South	22.93	2.28	14.86	27.34	2.15	12.95	17.50	0	100
	(0.49)	(0.45)	(5.78)	(0.74)	(0.24)	(2.48)	(3.85)	(0)	
South East	21.38	4.75	11.63	26.63	1.22	21.22	13.17	0	100
	(0.97)	(0.12)	(1.37)	(1.50)	(0.14)	(0.75)	(2.11)	(0)	
HCMC	40.49	0	18.25	9.48	0	9.68	21.69	0.41	100
	(1.29)	(0)	(0.37)	(1.85)	(0)	(1.07)	(3.02)	(0.58)	
Mekong River Delta	23.85	0.81	16.77	27.76	2.70	8.05	20.06	0	100
	(0.93)	(0.03)	(7.52)	(1.87)	(0.54)	(1.26)	(8.55)	(0)	
Travinh	18.86	5.01	7.77	26.09	3.12	14.10	25.06	0	100
	(3.26)	(0.89)	(2.46)	(5.77)	(2.77)	(0.01)	(2.90)	(0)	
Ethnicity									
Kinh & Hoa	29.47	0.39	26.21	13.83	2.00	11.98	16.12	0	100
	(0.53)	(0.18)	(3.63)	(0.53)	(0.24)	(0.86)	(3.41)	(0)	
Minorities	28.16	0.23	27.10	14.60	5.78	8.17	15.96	0	100
	(2.75)	(0.01)	(5.39)	(8.38)	(1.47)	(1.49)	(2.73)	(0)	
Total Vietnam	29.10	0.34	26.46	14.05	3.06	10.91	16.07	0	100
	(1.19)	(0.12)	(4.10)	(2.00)	(0.14)	(0.11)	(3.21)	(0)	

[a] Don't Know

Data source: HEPR Impact Assessment Qualitative Survey, 2003-4

5. Intersubjective Observables and Perceptual Data: Inferring Counterfactual Causality[10]

A final example of integration across theoretical and methodological boundaries in the HEPR impact occurred in the context of 'counterfactual causality'. Specifically, intersubjectively observable data and perceptual information were combined to determine what would have happened in the absence of the program. As with the combination of different concepts of causation and models of causal inference in Section 3, the intent was to bolster the validity of results if both approaches ended up telling a broadly consistent story, or, if not, to spur reflection on why not. The following discussion begins by elaborating on the distinction between intersubjective observables and perceptual data and proceeds to review empirical results.

Historically, intersubjectively observable data have occupied a privileged place in the philosophical tradition of Empiricism.[11] Allegedly, such data are 'objective' in the sense of ensuring the subject-invariance of properties of objects.[12] Rom Harré (1985) paraphrases this requirement: "many qualities [of objects] vary with the state of the subject, the perceiver, while for scientific purposes we should choose those qualities which are subject invariant". Karl Popper (1959) provided a classic statement on the centrality of intersubjective observability as the foundation of objective knowledge, in his discussion of 'basic statements': … a basic statement must also satisfy a material requirement … this event must be an "*observable"* event; that is to say, basic statements must be testable, inter-subjectively, by "observation"…

Intersubjective observability is quite integral to the Holland-Rubin framework discussed in the previous section. They define the fundamental problem of causal inference as a problem of observation. Specifically, the same person, household, village, etc. cannot partake in both treatment and control/comparison groups simultaneously and as such, differences in outcomes between treatment and controls cannot be observed. According to Holland (1986: 947): 'It is impossible to *observe* the value of Yt (u) and Yc (u) on the same unit, therefore, it is impossible to observe the effect of t on u. The emphasis is on the word *observe*.'

One alternative to intersubjective observables involves perceptual or self-report data, generated in dialogue. In the context of counterfactual causation, the challenge for dialogical techniques is to get meaningful answers to the subjunctive conditional (if-then) type question of what would have happened in the absence of the program. It is necessary to conduct 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). Meaningful responses to an if-then question 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, propensity score matching, based on intersubjectively observable data, was conducted alongside subjunctive conditional questioning for two project components: the Health Fee Reduction or Exemption program, which assessed utilisation of health services; the Tuition and School Maintenance Fee Exemption or Reduction which assessed enrolment in primary and secondary school. Results for each are reviewed in turn below.

The Health Fee Exemption or Reduction was part of HEPR's Health Care Support policy. It entailed providing poor households or communes free or subsidised health care through: 1) the distribution of health insurance cards or poor household certificates which entitle the holders to free or subsided care; 2) 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 4 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 4 Propensity Score Matching: Impact of Health Free Exemption/Reduction on Utilisation of Healthcare

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

[a]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 5 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 PSM. 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 5 Self-Reported Assessment of Use of Medical Care in HEPR Absence (Population Proportions, Standard Errors in Parentheses)[a]

	1 Yes	2 No	Total
Total Vietnam	91.81 (0.40)	7.28 (0.11)	100

[a]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 is one component of HEPR's Education Support policy. The program 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 6 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 percent depending on the comparison group used and taking into account sampling error.

Table 6 Propensity Score Matching: Impact of Tuition and School Maintenance Fee Exemption/Reduction on School Attendance

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

[a]Standard errors were bootstrapped with 100 replications
Data source: Vietnam Household Living Standards Survey, 2002

Table 7 presents results of the self-report exercise. Around twelve percent 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 7 Self –Reported Assessment of Primary or Secondary Enrolment in HEPR Absence (Population Proportions, Standard Errors in Parentheses)

	1 Yes	2 No	3 DK[a]	Total
Total Vietnam	87.64 (1.01)	11.56 (0.32)	0.80 (0.69)	100

[a] Don't Know
Data source: HEPR Impact Assessment Qualitative Survey, 2003-4

In summary, the HEPR impact assessment combined intersubjectively observable data and perceptual information to determine what would have happened in the absence of the program. These techniques were used to assess the Health Fee Reduction or Exemption program on utilisation of health services and the Tuition and School Maintenance Fee Exemption or Reduction on enrolment in primary and secondary school. For both these project components the results of both exercises arrived at very similar results. In such instances, there is a stronger case to assert the internal validity of results.

6. Conclusion

The preceding analysis aimed to redress a shortcoming in the literature on mixed method impact assessment. Excessive attention has focused on questions of method to the detriment of foundational issues. The objective of the article has been to show how such issues mattered for the impact assessment of the Hunger Eradication and Poverty Reduction (HEPR) program in Vietnam.

Specifically, it was argued that the approaches to impact assessment in the HEPR study rested on foundational differences with respect to: i) conceptions of causation and models of causal inferences (probabilities vs. mechanisms); ii) analytical focus (outcomes vs. processes) and external validity (empirical generalization vs. statistical inference) and iii) constituents of 'objective' knowledge (intersubjective observables vs. perceptual data).

The first point was illustrated in the discussion of the credit module which combined a propensity score matching exercise, which relies on probabilistic causal inference and counterfactual conceptions of causation, with a process tracing exercise, which is a 'physically producing'/mechanism-based account of causation, followed by a self-report on the magnitude of impact. These two exercises produced quite different results which did not serve the end of bolstering internal validity.

The second point was again illustrated by the credit module in which data on the magnitude of impact (outcomes) were combined with information on the perceived reasons for lack of program impact. Probabilistic sampling provided the basis for claims of external validity in that standard errors could be calculated for population proportions within different response categories (unlike existing small *n* studies). According, outcomes were combined with processes in statistically 'representative' fashion.

The modules on the Health Fee Reduction or Exemption program and the Tuition and School Maintenance Fee Exemption or Reduction were used to illustrate the final point. Intersubjectively observable data and perceptual information were combined to determine what would have happened in the absence of the program. For both these project components the results of both exercises arrived at very similar results, strengthening the case for the internal validity of results.

The focus on foundations in the context of impact assessment is important for at least two reasons. First, conflicting empirical results in mixed method designs may be due to underlying foundational differences. For example, as discussed in the case of the credit program, there is a conceptual difference between 'impact over time due to credit' and 'impact relative to the situation of the comparison group.' The root of the difference concerns conceptions of causation and models of causal inference.

Second, and more importantly, impact assessment should not be depicted primarily as a technical exercise, preoccupied with method. Choice of conception of causation and model of causal inference, determination of what should count as 'objective' knowledge or 'hard' evidence, as well as what should be the object of inquiry, are not, at root, technical questions. A more informed debate will ensue if our assumptions are bared and put on the table.

Notes

[1] See, for example, Addison, Hulme and Kanbur, eds. (2009), Hulme and Toye, eds. (2007), Kanbur (2003), Kanbur and Shaffer (2007a), Shaffer et. al. (2008).
[2] See, for example, contributions in Shaffer et. al. (2008), Broegaard et. al. (2011) and White (2011).
[3] In the first two stages, selection was conducted using Probabilities Proportional to Size (PPS) techniques while in the third and fourth stages selection was random. The sample was stratified by region (in the first stage) and agro-ecological zone (in the second stage). Full details of the sampling strategy are available from the author.
[4] Results were published in the *Vietnam Development Report 2004* and appear in Cuong (nd).
[5] I thank two anonymous referees for noting these issues.
[6] See Abadie and Imbens (2008). I thank an anonymous referee for drawing my attention to this cite.
[7] Parts of this section draw on Shaffer (2011).
8 Leeuw and Vaessen (2009) present a good discussion of causal mechanisms in the context of impact assessment.
[9] Ethnicity was determined in terms of characteristics of the household head.
[10] Parts of this section draw on Shaffer (2011).
[11] These issues are discussed at greater length in Shaffer (2002) and Kanbur and Shaffer (2007b).
[12] Chalmers (1999, Chs. 1 and 2) provides a good critique of this claim and statement of the fallibility of observation statements.

References

Abadie, A., and Imbens, G., 2008 On the Failure of the Bootstrap for Matching Estimators. *Econometrica*, 76(6), 1537-1557.

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.

Addison T., Hulme D., and Kanbur, R., eds. 2009. *Poverty Dynamics: Interdisciplinary Perspectives.* Oxford: Oxford University Press.

Banerjee, A., 2007. *Making Aid Work.* Cambridge, Mass.: MIT Press.

Bardhan, P. and Ray, I., 2006. Methodological approaches to the question of the commons. *Economic Development and Cultural Change*, 54, 655 – 676.

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.

Cartwright, N., 2007a. *Hunting Causes and Using Them. Approaches in Philosophy and Economics.* Cambridge: Cambridge University Press.

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., nd. *Assessing the Impact of Poverty Reduction Programs in Vietnam*, Mimeo, Hanoi.

de Weerdt, J., 2010. Moving Out of Poverty in Tanzania's Kagera Region. *Journal of Development Studies*, 46 (2), 331-349.

Elster, J., 1987. *Sour Grapes. Studies in the Subversion of Rationality.* Cambridge: Cambridge University Press.

George, A. and Bennett, A., 2004 *Case Studies and Theory Development in the Social Sciences.* Cambridge Mass.: MIT 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.

Hammersley, M., 1992. *What's Wrong with Ethnography? Methodological Explorations.* London and New York: Routledge.

Harré, R., 1985. *The Philosophies of Science.* Second Edition. Oxford: Oxford University Press.

Holland, P., 1986. Statistics and Causal Inference. *Journal of the American Statistical Association*, 81 (396), 945-960.

Hedström, P., and Swedberg, R., eds., 1998. *Social Mechanisms: An Analytical Approach to Social Theory.* Cambridge: Cambridge University Press.

Hulme, D., and Toye, J., 2006. The case for cross-disciplinary social science research on poverty, inequality and well-being, *Journal of Development Studies* 42(7): 1085-1107.

Kanbur, R., ed. 2003. *Q-Squared: Qualitative and Quantitative Poverty Appraisal.* Delhi: Permanent Black.

Kanbur, R., and Shaffer, P., eds. 2007a. Special Issue: Experiences of Combining Qualitative and Quantitative Approaches in Poverty Analysis. *World Development* 35 (2), 183-353.

Kanbur, R., and Shaffer, P., 2007b. Epistemology, Normative Theory and Poverty Analysis: Implications for Q-Squared in Practice. *World Development* 35(2), 183-196.

Leeuw, F., and Vaessen, J., 2009. *Impact Evaluations and Development. NONIE Guidance on Impact Evaluation.* Network or Networks on Impact Evaluation (NONIE), Washington, D.C.

Little, D., 1991. *Varieties of Social Explanation.* Boulder: Westview Press.

Little, D., 1998. *Microfoundations, Method and Causation.* New Brunswick, U.S.A.: Transaction Publishers.

Lipton, M., 1992. Economics and Anthropology: Grounding Models in Relationships. *World Development,* 20 (10): 1541-1546.

Menzies, P., 2009. Counterfactual Theories of Causation. *In*: E. Zalta, ed. *The Stanford Encyclopedia of Philosophy*. URL = http://plato.stanford.edu/archives/fall2009/entries/causation-counterfactual.

Mosse, D., 2006. Collective Action, Common Property and Social Capital in South India: An Anthropological Commentary. *Economic Development and Cultural Change* 54(3): 695-724.

Pickel, A., 2004. Systems and Mechanisms. A Symposium on Mario Bunge's Philosophy of Social Science. *Philosophy of Social Sciences*, 34 (2), 169-181.

Popper, Karl., 1959. *The Logic of Scientific Discovery.* London: Hutchinson.

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.* New York City: World Bank Publications and Oxford University Press, 165-190.

Rao, V., Gupta, I., Lokshin, M. and Jana, S., 2003. Sex Workers and the Cost of Safe Sex: the Compensating Differential for Condom Use among Calcutta Prostitutes. *Journal of Development Economics* 71: 585-603.

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.

Rubin, D.B., 1974. Estimating Causal Effects on Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66, 688-701.

Schaffer, J., 2008. The Metaphysics of Causation. In: E. Zalta, ed. *The Stanford Encyclopedia of Philosophy.* URL = http://plato.stanford.edu/archives/fall2008/entries/causation-metaphysics/.

Shaffer, P., 2002. Poverty Naturalized: Implications for Gender. *Feminist Economics* 8(4), 55-75.

Shaffer, P., 2011. Against Excessive Rhetoric in Impact Assessment: Overstating the Case for Randomised Controlled Experiments. *Journal of Development Studies,* 47(8), 1-17.

Shaffer, P., Kanbur, R., Thang, N., Bortei-Doku Aryeetey, E., 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.

Sudman, S., Bradburn, N. and Schwartz, N. 1996. *Thinking About Answers*. San Francisco: Jossey-Bass Publishers.

Tashakkori, A., and Teddle, C., eds., 2003. Handbook of Mixed Methods in Social & Behavioral Research. Sage: Thousand Oaks, California.

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*, New Delhi, India.

White, H., 2011 Achieving High-Quality Impact Evaluation Design through Mixed Methods: the case of Infrastructure. *Journal of Development Effectiveness*, 3 (1), 131-144.

White, H., and Bamberger M., eds. 2008. Impact Evaluation in Official Development Agencies. *IDS Bulletin,* 39 (1), 1-11.

Appendix - Kernel Density Estimates of Propensity Scores

Education

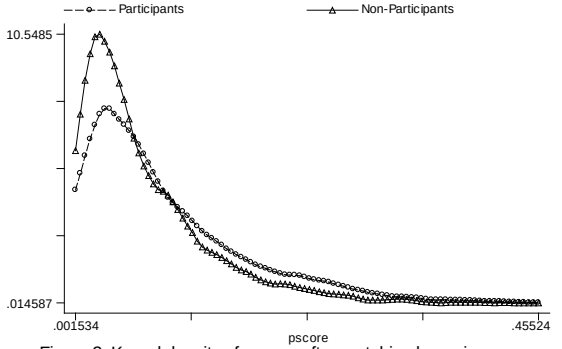


Figure 2: Kernel density of scores after matching by regions

Credit

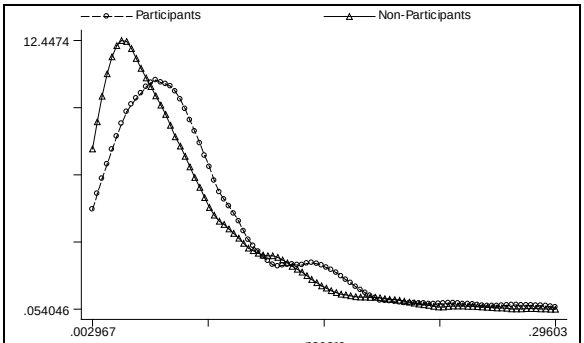
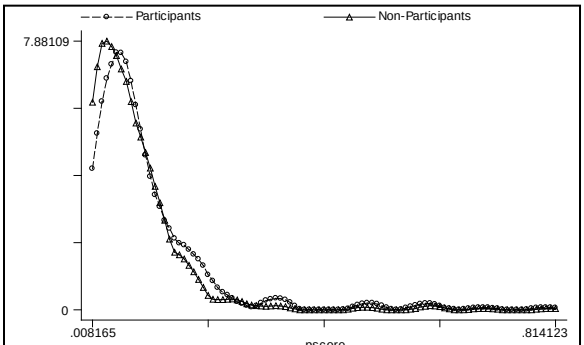


Figure 2: Kernel density of scores after matching by regions

Health



Kernel density of scores after matching by regions