GSDRC
Governance and Social Development Resource Centre

Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results

Issues Paper

Sabine Garbarino and Jeremy Holland

March 2009

GSDRC Emerging Issues Research Service

This Issues Paper was commissioned by the UK Department for International Development (DFID) through the Emerging Issues Research Service of the Governance and Social Development Resource Centre (GSDRC). The service addresses emerging priority agendas of concern to the international development community.

For more information see www.gsdrc.org

ISBN 0 7044 2756 7 / 9780704427563

Acknowledgements

We are grateful to colleagues in the Department for International Development (DFID) and amongst the research and practitioner communities for their time and inputs to the development of this paper.

This is a DFID/OPM/GSDRC study, which was commissioned by DFID. The paper was produced by Sabine Garbarino and Jeremy Holland from Oxford Policy Management (OPM). We are grateful to colleagues in DFID for their expert insight and for providing material, resources and comments: Jane Hobson (Gender and Social Exclusion Workstream Leader, Equity and Rights Team), Sue Holloway (Deputy Chief Economist), Penny Innes (Senior Stats Adviser, Aid Effectiveness and Accountability Department), Rachael Beaven (Deputy Head of Profession, Statistics), Tim Robertson (Theme Leader for Governance, Social Development, Humanitarian and Conflict), Sue Kinn (Research Manager, Central Research Department), Helen Wedgewood (Deputy Head, Evaluation Department) and Ellen Wratten (Head of Profession, Social Development). We are specifically thankful to Jillian Popkins (Deputy Head of Profession Social Development) and Adrian Nembhard (Economist, Evaluation Department) for the essential guidance and detailed comments on the draft report. Finally, we would like to thank all DFID staff, researchers and practitioners who participated in the workshop on 21 January 2009 and contributed to the revision of the paper through their comments and discussions.

Contents

List of tables and figures

Abbreviations

3ie	International Initiative for Impact Evaluation
CGD	Center for Global Development
CoIMpact	Consultative impact monitoring of policies
CRC	Citizen Report Card
CSC	Community Score Card
DAC	Development Assistance Committee
DFID	Department for International Development
GEM	Gender Empowerment Measure
GSEA	Gender and Social Exclusion Analysis
HDP	Human Development Project
IFPRI	International Food Policy Research Institute
IMD	Index of Multiple Deprivation
NONIE	Network of Networks on Impact Evaluation
OECD	Organisation for Economic Co-operation and Development
OPM	Oxford Policy Management
PPA	Participatory Poverty Assessment
PSIA	Poverty and Social Impact Analysis
QIM	Qualitative Impact Monitoring
RCT	Randomised Control Trial
SCPS	Social Capital and Poverty Survey
SHGs	Self-Help Groups
UN	United Nations
UNDP	United Nations Development Programme

Glossary of terms relating to impact (common and technical)

Term	Definition
Accountability	Obligations of partners to act according to clear defined responsibilities, roles and performance expectations, often with respect to the prudent use of resources particularly to taxpayers and citizens.
Activity	Actions taken or work performed through which inputs, such as funds, technical assistance and other types of resources are mobilised to produce specific outputs.
Attribution	Refers to that which is to be credited for the observed changes or results achieved. It represents the extent to which observed development effects can be attributed to a specific intervention or to the performance of one or more partners, taking account of other interventions (anticipated or unanticipated) confounding factors, or external shocks.
Base-line Study	An analysis describing the situation prior to a development intervention, against which progress can be assessed or comparisons made.
Beneficiaries	The individuals, groups, or organizations, whether targeted or not, that benefit, directly or indirectly, from the development intervention.
Counterfactual	The situation or condition which hypothetically may prevail for individuals, organisations, or groups were there no development intervention.
Development Intervention	An instrument for partner support aimed to promote development.
Effect	Intended or unintended change due directly or indirectly to an intervention.
Evaluation	Assessment, as systematic and objective as possible, of ongoing or completed aid activities, their design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, developmental efficiency, effectiveness, impact and sustainability.
Goal	The higher-order objective to which a development intervention is intended to contribute.
Impact (as used by Management)	The difference international aid makes to either indicators or people's lives; and having the ability to prove it.
Impact (as used in evaluation)	Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.
Indicator	Quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor.
Inputs	The financial, human, and material resources used for the development intervention.
Monitoring	A continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives

Term	Definition
	and progress in the use of allocated funds.
Outcome	The likely or achieved short-term and medium-term effects of an intervention's outputs.
Outputs	The products, capital goods and services which result from a development intervention; may also include changes resulting from the intervention which are relevant to the achievement of outcomes
Results	Describing what has been achieved following the provision of international aid.
Results Chain	The causal sequence for a development intervention that stipulates the necessary sequence to achieve desired objectives beginning with inputs, moving through activities and outputs, and culminating in outcomes and impacts.

Source*:* The glossary of terms is presented in the DFID Draft Policy Statement on Impact Evaluation and is based on OECD DAC 'Evaluation of Key Terms in Evaluation: Glossary of Key Terms in Evaluation & Results Based Management'—apart from 'impact' (as used by Management) and 'evaluation' which is taken from DFID's new economists guide (Brook at al 2006).

1 Introduction

There has been a renewed interest in impact evaluation[1] in recent years amongst development agencies and donors. Additional attention was drawn to the issue recently by a Center for Global Development (CGD) report calling for more rigorous impact evaluations, where 'rigorous' was taken to mean studies which tackle the selection bias aspect of the attribution problem (CGD, 2006). This argument was not universally well received in the development community; among other reasons there was the mistaken belief that supporters of rigorous impact evaluations were pushing for an approach solely based on randomised control trials (RCTs). While 'randomisers' have appeared to gain the upper hand in a lot of the debates—particularly in the United States—the CGD report in fact recognises a range of approaches and the entity set up as a results of its efforts, 3ie, is moving even more strongly towards mixed methods (White, nd).

The Department for International Development (DFID) in its draft policy statements similarly stresses the opportunities arising from a synthesis of qualitative and qualitative approaches in impact evaluation. Other work underway on 'measuring results' and 'using numbers' recognises the need to find standard indicators which capture non-material impacts and which are sensitive to social difference. This work also stresses the importance of supplementing standard indicators with narrative that can capture those dimensions of poverty that are harder to measure.

This paper contributes to the ongoing debate on 'more and better' impact evaluations by highlighting experience on combining qualitative and quantitative methods for impact evaluation to ensure that we:

1. measure the different impact of donor interventions on different groups of people and
2. measure the different dimensions of poverty, particularly those that are not readily quantified but which poor people themselves identity as important, such as dignity, respect, security and power.

A third framing question was added during the discussions with DFID staff on the use of the research process itself as a way of increasing accountability and empowerment of the poor.

This paper does not intend to provide a detailed account of different approaches to impact evaluation nor an overview of proposed solutions to specific impact evaluation challenges[2]. Instead it defines and reviews the case for combining qualitative and quantitative approaches to impact evaluation. An important principle that emerges in this discussion is that of equity, or what McGee (2003, 135) calls 'equality of difference'. By promoting various forms of mixing we are moving methodological discussion away from a norm in development research in which qualitative research plays 'second fiddle' to conventional empiricist investigation. This means, for example, that contextual studies should not be used simply to confirm or 'window dress' the findings of non-contextual surveys. Instead they should play a more rigorous role of observing and evaluating impacts, even replacing, when appropriate, large-scale and lengthy surveys that can 'overgenerate' information in an untimely fashion for policy audiences.

[1] DFID's draft policy statement adapts White's (2006) definition of impact evaluation as the 'counterfactual analysis of the impact of an intervention on final (human) welfare outcomes'.

[2] These include the definition of the counterfactual, selection bias or lack of a baseline.

The remainder of the paper is structured as follows. Section 2 briefly sets the scene by summarising the policy context. Section 3 clarifies the terminology surrounding qualitative and quantitative approaches, including participatory research. Section 4 reviews options for combining and sequencing qualitative and quantitative methods and data and looks at recent methodological innovations in measuring and analysing qualitative impacts. Section 5 addresses the operational issues to consider when combing methods in impact evaluation. Section 6 briefly concludes.

2 The policy context: Tackling the evaluation gap

2.1 From principles to practice

Embedded in the international initiative for more and better impact evaluation, DFID's work on impact evaluation is a contribution towards a new evaluation policy and strategy for DFID and part of the Results Action Plan. DFID's draft policy statement stresses the urgent need to 'tackle the evaluation gap' and address the shortage of rigorous and policy-relevant impact evaluation studies. The paper presents a set of guiding principles for impact evaluation which are summarised in Box 2.1.

Box 2.1 DFID policy paper: Guiding principles for impact evaluation

Doing more, and higher quality, impact evaluations: given the identified impact evaluation gap, and the potential for impact evaluations to improve the effectiveness and accountability of development policy, the first guiding principle is that DFID will play a role in filling the impact evaluation gap. The rest of the principles cover how DFID should go about helping to fill this gap.

International coordination and cooperation: because impact evaluations are a global public good, DFID should achieve its objective of increasing the number and quality of impact evaluations in an internationally coordinated way. This means participating in international initiatives, doing joint impact evaluations with other donors, but also coordinating on what issues to cover—for example, DFID could take the lead in doing impact evaluations in an area of policy where it has particular expertise.

Relevant: Making sure impact evaluations are done in ways that are relevant to policy. This means asking the right questions, at the right time, but also answering why something worked or did not work, not just whether it worked or not. Combining qualitative and quantitative methods can help achieve this.

Ethical: Impact evaluations need to be done in an ethical way. Having a counterfactual usually means having a comparison group that has not benefited from a programme or policy. This raises ethical issues about denying coverage to some groups for the sake of evaluation. However, resources are often constrained and choices often have to be made. Programmes are not always introduced everywhere all at the same time because of limited capacity. Impact evaluations can exploit these existing constraints to learn from them about the impact of the programme.

Use the appropriate methodology: To ensure that impact evaluations are high quality and therefore useful, it is important that they use the most appropriate methodology to answer the counterfactual. DFID will utilise internationally agreed guidance from 3ie when this becomes available.

Source: DFID 2008.

This renewed commitment to impact evaluation should be seen in the wider context of stressing the importance of evidence-based decision making and for both monitoring and evaluation of results. Managing for results means focusing on outcomes (what is being achieved) as well as outputs (what is being produced) and inputs (how much money is being spent). While this is by no means new, there has been an increasing focus on results both within DFID and internationally. Driven by the international consensus around the Millennium Development Goals and the Paris Declaration for Aid Effectiveness, development agencies want to know 'what works, what doesn't, where and why', and want to identify who benefits from development assistance, including poor, vulnerable and excluded groups (DFID, 2007). Evidence on the effectiveness and impact of projects and programmes can help make DFID

more accountable, to parliament, the UK public, partner country governments and the citizens of developing countries.

In response to this growing consensus there are a number of international initiatives on impact evaluation. The Network of Networks on Impact Evaluation (NONIE), chaired by DFID, brings together representatives from donors (the DAC, the UN and the multilateral banks) and developing countries (the evaluation associations) to share information and develop standards and guidance on impact evaluation. The International Initiative for Impact Evaluation (3ie) is a major new initiative funding new high quality impact evaluations and carrying out design, meta-analysis and review of research proposals.

2.2 Challenges to doing more and better impact evaluations

Why is it case that what gets evaluated is only a fraction of what gets done in the name of development—given that the benefits and the need are so clear? Impact evaluation is resource and data intensive; it is a public good—while the benefits are shared the costs are not. High quality impact evaluations generate learning relevant across the international community, yet they are typically paid for by one agency, country or project.

While there is agreement that not every project or programme can be evaluated, many existing impact evaluations are motivated by a desire to prove the worth of a given intervention, meaning that external funding for evaluations is almost entirely demand driven. This process is likely to favour projects and policies that are expected to have benefits by their advocates. Why would someone commission an impact evaluation if the results are expected to be negative (Ravallion, 2008)? An important counterbalance to this tendency for 'promotional' impact evaluation is created when national government partners/borrowers become strong and intelligent customers of more and better evaluation.

In addition to the challenges to doing more impact evaluations is the challenge for better quality impact evaluations that arises as development agencies have moved upstream from project aid towards sector and general budget support. There is a lot of experience with doing impact evaluations of specific projects and programmes and rigorous methods for evaluating impact are designed for projects. There is far less experience, however, with conducting impact evaluations of these new aid instruments, with no general rules yet established for the measurement of the impact of sector or general budget support (No author, 2007).[3]

This move upstream has been accompanied by a growing interest amongst donors in institutional change, which brings the methodological challenge of attributing complex change to donor interventions. It is relatively easy to attribute—and establish a counterfactual for—changing infant mortality rates to investments in an immunisation campaign. However, evaluating the effectiveness of donor interventions in areas such as good governance, institutional processes or human security is much more challenging and fewer impact evaluations are typically undertaken even as donor funds are increasingly channelled towards these areas (DFID, 2008; Prowse, 2007).

[3] There are studies trying to bridge the gap between the existing methods of impact evaluation (designed for projects) and the much needed instruments to evaluate the impact of the new aid instruments (Gunning 2006).

2.3 Opportunities

While many constraints have to be tackled before the impact evaluation gap can be filled there are various opportunities arising from this agenda. These include the opportunity to operationalise many of the documented gains from combining methods, using powerful qualitative analytical frameworks and measuring non-material dimensions of poverty. In recent years, the use of combined methods and social analytical frameworks in Poverty and Social Impact Analysis (PSIA) has demonstrated this very clearly (see for example Box 2.2). While PSIA is usually *ex ante* analysis and, when conducted *ex post*, rarely employs a counterfactual, the potential for combing methods in impact evaluation has been powerfully demonstrated.

Impact evaluations are now being done in many non-traditional ways. New methodologies are increasing the capacity to measure and analyse different dimensions of poverty—including those not readably amenable to quantification, such as trust, respect, governance, empowerment and security. Impact evaluations can also play a crucial role—as noted in DFID's policy statement guiding principles (Box 2.1)—in revealing the reasons underlying poverty outcomes and trends as well as in disaggregating poverty impacts in ways that ensure better targeting, effectiveness and efficiency (DFID, 2008). Applying mixed methods in impact evaluation can help to identify who benefits from development aid and why, in particular in order to measure the impact on girls and women and on traditionally excluded groups. An impact evaluation that combines qualitative and quantitative methods can generate both a statistically reliable measure of the magnitude of the impact as well as a greater depth of understanding of how and why a programme was or was not effective and how it might be adapted in future to make it more effective (DFID, 2008).

Furthermore, there is now widespread recognition amongst development agencies of the importance of process and participation in all stages of the policy cycle. A qualitative approach can sharpen the focus on the policy process in which impact evaluations are embedded. Recent discussion concerning ways to undertake PSIA has included a focus on a shift from donor-driven to country-owned PSIA (World Bank, 2008). Within this shift many stakeholders have expressed a concern that getting the PSIA process right is as important as improving the quality of the analysis itself (Schnell at al, 2005). Similarly the UNDP Oslo Governance Centre is supporting 'Joint Governance Assessments' that are nationally owned and embedded so that the evaluation of impacts can be more democratic and can lead to reflection, action and change. We can learn lessons from these approaches for the way we undertake impact evaluations with the ultimate aim of improving policy making.

Box 2.2 Combined methods in ex-post PSIA: Abolition of user fees in health units in Uganda

In Uganda, through the careful sequencing of methods and data analysis, a strong picture emerged that abolishing user fees has made health care more accessible to poor people and consequently they have increased their consumption of these services.

The Uganda **Participatory Poverty Assessment** Process (UPPAP) revealed the significance of health shocks on the vulnerability of the poor, particularly in rural areas. Addressing this popular concern, which had become a major election issue, in March 2001, the President of Uganda scrapped user fees for Government health units. This decision was also informed by finding that the user fees failed to raise much revenue.

Time series data on outpatient attendances showed that the combined effects of abolishing fees and increasing the supply of health services were dramatic, with an 84% increase in outpatient attendances between 2000/01 and 2002/03. Immunisation rates among children also increased dramatically (e.g. the proportion of children who received their third DPT immunisation increased from 48% to 84% between 2000/2001 and 2002/03).

To confirm that the pick up in outpatient attendance was progressively weighted towards poorer households, the WHO/MoH conducted **participatory research** with carefully-sampled communities. A **wealth ranking** exercise conducted by villagers revealed that since the abolition of user fees, the poorest quartile had consistently used Government health centres more than any other group: in 2002 the poorest quartile used these facilities at a rate of 0.99 visits per person per year whereas for the wealthiest group the rate was 0.77. There was only limited data available for the period preceding the abolition of fees, but the average utilisation rate for the months of January and February 2001 was 0.52 and 0.42 respectively. Subsequent analysis of the 1999/2000 and the 2002/3 **household surveys** confirmed that poorer income groups had increased their utilisation of services more than richer groups. For hospital services, the rate of increase of consumption for the poorest two quartiles was double that of the richest group.

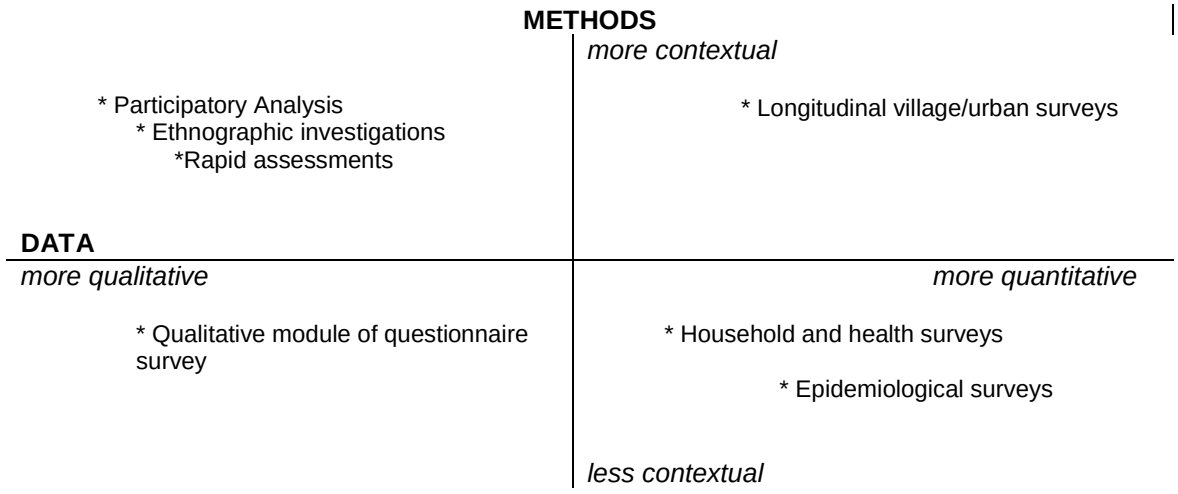
Source: Yates *et al* (2008)

3 Quantitative and qualitative research: Clarifying the terminology

In this section we introduce a precise definition of the terms 'quantitative' and 'qualitative', and explain where participatory research fits into this definition. We also discuss different ways of combining quantitative and qualitative approaches that add value in impact evaluation.

Simply put, the terms 'qualitative' and 'quantitative' should refer to the type of *data* generated in the research process.[4] Quantitative research produces data in the form of numbers while qualitative research tends to produce data that are stated in prose or textual forms. In order to produce different types of data, qualitative and quantitative research tend to employ different *methods*. Using the terminology from Hentschel's (1999) method-data framework (see Figure 3.1) non-contextual methods—applied across the population universe, often a country or region—are designed to achieve breadth in coverage and analysis. Typically, the random sample survey produces quantifiable data that can be statistically analysed with the main aim of measuring, aggregating, modelling and predicting behaviour and relations. Contextual methods in contrast are applied to a specific locality, case or social setting and sacrifice breadth of population coverage and statistical generalisability in order to explore issues in depth (Booth et al, 1998). Contextual research includes ethnographic techniques, such as participant observation, interviews and participatory tools that are often group-based and visual. Using open-ended questions these methods are designed to capture judgements and perceptions and allow complex analyses of often non-quantifiable cause-and-effect processes.

Figure 3.1 The method data framework



Source: Adapted from Hentschel (1999)

In common with qualitative research, *participatory research* tends to employ more contextual methods and elicit more qualitative and interpretive information, but brings an important additional commitment to respect local (emic) knowledge and facilitate local ownership and control of data generation and analysis (Chambers, 1994, 1997). This aspect of ownership and control in participatory research is intended to provide space for local people to establish

[4] This section draws on Holland (2007)

their own analytical framework and to be in a position to challenge 'development from above' (Mukherjee, 1995, 27). In contrast to the individualised observation and discussions in much qualitative investigation, participatory research focuses on public and collective reflection and action.

At its most extractive, participatory research simply uses a suite of participatory methods to improve outsiders' understanding of local context (while adhering to certain ethical principles relating to behaviour, transparency and ownership). At its most political, participatory research is a *process* in which reflection is internalised and promotes raised political consciousness. In this way, population involvement in research shifts from *passive* to *active*.

Participatory methods generate both qualitative and quantitative data. 'Participatory numbers' can be generated and used in context, but have also been taken to scale, most notably through participatory surveys or through aggregation of group-based scoring and ranking activities. Participatory methods can be quick and efficient, producing data in a timely fashion for evidence-based analysis and action. In a significant recent application participatory methods generating quantitative data as part of a project evaluation were utilised to generate robust population estimates (see Box 3.1). Research teams showed that it was possible to generate statistics which would be taken seriously by policy makers from research using participatory methods. One key requirement, however, is to produce results from a sufficiently large sample for national level inference and analysis. This can imply working in a larger number of research sites than is usually the case with participatory research.

Box 3.1 Participatory numbers and going to scale in Malawi

In Malawi in 1999-2002 a research team conducted studies using participatory methods to generate population estimates; estimates of the proportion of people in a population with certain characteristics (e.g. the very food insecure); and estimates of the proportion of people in a population that should be targeted by an intervention. We claim that the statistics generated in these studies are at least as reliable as statistics obtained from traditional methods such as surveys. A key requirement is to produce results from a representative sample, which can be generalised in order to reach conclusions for the population of interest. This implies working in a larger number of sites than is common for most studies that use participatory methods. However, we argue that the Malawi experience shows that it is possible to organise such studies at a reasonable cost.

Other key requirements are for the study design to incorporate statistical principles; and for PRA tools to be adapted to meet the demands of standardisation and comparability of data produced across sites. We argue that this can be done without undermining participatory approaches, behaviours and methods. If research studies using participatory methods follow this approach, the data generated will be suitable for standard statistical analysis. The statistics produced by such studies should be capable of informing policy at national level. However, there are some concerns about empowerment and ethical issues, which present challenges for the future.

The methods developed as well as the lessons learnt in Malawi were used again in 2006 in Uganda to design and implement a qualitative study of the dimensions of poverty that was linked by design to the sampling process of the Uganda National Household Survey (UNHS III). The experience demonstrates that the methodology proposed can be adapted to be part of the operations of National Statistical Systems, but that the challenges of integration of statistical methods and participatory approaches also face institutional constraints

Sources: Barahona and Levy (2005; 2002); Carlos Barahona pers comm.[5]

Participatory numbers can throw up striking insights with policy implications, and do this in a statistical form. The Bangladesh Participatory Poverty Assessment (PPA) (UNDP, 1996), for instance, found that the second and third 'doable' priorities for poor urban women (their first was water) were a private place for washing, and action on dowry. Moser and McIlwaine's (2002) work in nine urban communities in Colombia elicited numerous types of violence, and produced the unexpected finding that 54 per cent of the types of violence identified were economic (related to drugs, insecurity, robbery, gangs etc) as against only 14 per cent political (related to police abuse, war, paramilitaries, assassinations etc) contrary to the common belief that political violence was the bigger problem.

When considering the use of participatory numbers in impact evaluation research, there are a number of methodological and ethical trade-offs that should be considered and which have been well documented (see Box 3.2).

Finally, qualitative research—particularly in Europe—is strongly associated with non-economic social science disciplines, while quantitative research has an economics and natural sciences leaning. While in the US, social sciences are more firmly embedded in the positivist school than the European tradition, economics and the natural sciences have largely failed to embrace qualitative research because of the positivist-empiricist embeddedness of those disciplines.

[5] Carlos Barahona (c.e.barahona@reading.ac.uk) is a Senior Statistician at the Statistical Services Centre, University of Reading.

Box 3.2 Participatory numbers: Opportunities and trade-offs

Participatory methods are by no means restricted to create qualitative data; people map, count, estimate, compare and value using numbers during participatory research. Chambers (2007). Methodological challenges, however, do exist. Some of the issues concern applying statistical principles, others do involve optimising trade-offs (Chambers, 2007), for example:

1. Standardised, closed and commensurable versus open, diverse and empowering: The more standardised the process, the more extractive and less empowering and accommodating of local priorities and realities it is likely to be. The less standardised it is, the harder the outcomes are to analyse.
2. Scale, quality, time, resources, and ethics: the issues here are far from simple. Smaller scale, more time, and more resources can allow for higher quality and better ethics but loss of representativeness; and vice versa.
3. Quality of facilitation versus speed, scale and cost of implementation. In these approaches, the quality of facilitation is critical. To achieve good facilitation requires time and resources devoted to careful selection of facilitators, their training and then their supervision in the field. This may add to costs and slow implementation and limit its scale, even if the outcomes are still highly cost-effective compared with alternatives.
4. Ease and spontaneity of convening groups versus representativeness. Where groups are involved, and as is well known with focus groups generally, those who are most easily convened may be unrepresentative or dominated by one or a few people, or by one sort of person (for example, men in a mixed group of men and women). Care in selection, in judging size of group, and observation and facilitation of process can offset these dangers but takes time and effort and can entail a loss of spontaneity.

Source: Adapted from Chambers (2007)

When these additional dimensions of quantitative and qualitative research are taken into account, our original precise definition of quantitative and qualitative as describing the types of data produced can be broadened to describe a set of five dimensions that collectively characterise the two research traditions (Kanbur, 2003, see Figure 3.2). These broader dimensions are useful reference points when we consider the potential for impact evaluation *frameworks* to employ dynamic social analytical lenses and for impact evaluation *processes* to engage a wider range of stakeholders and increase accountability.

Figure 3.2 Qualitative and quantitative dimensions of poverty appraisal

More 'qualitative' research	More 'quantitative' research
1. Non-numerical information	Numerical information
2. Specific (contextual) population coverage	General (non-contextual) population coverage
3. Active population involvement	Passive population involvement
4. Inductive inference methodology	Deductive inference methodology
5. Broad social sciences disciplinary framework	Neo-classical economics (and natural sciences) disciplinary framework

Source: Adapted from Kanbur (2003, 1)

4 Utilising quantitative and qualitative approaches in impact evaluation

The case for combined methods in poverty assessment—and in evaluations more generally—has been convincingly made. Through its support of combined methods in World Bank Poverty Assessments in the 1990s, through to the more recent use of a mix of methods and analytical frameworks in PSIA, DFID has consistently supported innovation in the use of methods for evaluation and its mainstreaming development research. There is therefore a firm basis on which to adapt and implement these methods in the specific design of impact evaluation.

When considering ways to combine quantitative and qualitative methods and data, it is important to be aware of their comparative advantages and to recognise that 'strong fences make good neighbours' (Appleton and Booth, 2005). In short, while quantitative methods produce data that can be aggregated and analysed to describe and predict relationships, qualitative research can help to probe and explain those relationships and to explain contextual differences in the quality of those relationships. Qualitative research is able to use social analytical frameworks to interpret observed patterns and trends—including analysis of socially differentiated outcomes—and to analyse poverty as a dynamic process rather than a static outcome. One rapidly growing area of qualitative analysis is political economy analysis, which was operationalised by DFID through the Drivers of Change initiative for country system analysis, widely used for analysis of sector and policy reform in PSIA and which is now being applied to 'problem-focussed' analysis, such as recent DFID-funded analysis of growth policy options in Uganda.

Without these analytical insights into the complex 'missing middle' between interventions and impacts, researchers and policy analysts tend to make 'interpretive leaps' of analysis based on what is measured. The danger here is that what is not quantifiable becomes unimportant while 'what is measurable and measured then becomes what is real and what matters' (Chambers, 1995, 8). Conversely, if qualitative research inductively throws up interesting, often surprising and sometimes counterintuitive relationships and patterns, quantitative research is then able to ask 'how much?' and establish how confident we can be in these 'working hypotheses'. This iterative relationship between describing and explaining provides the key to effective combination of methods and data. Good practice from DFID's support to Poverty Reduction Strategies includes the use of combined methods for poverty monitoring in Uganda which was at the cutting edge of poverty-monitoring practice at the time. As well as being focused on final outcomes for people, the monitoring instrument relied upon a) a good *panel* survey of household expenditure, a well designed PPA with linkages built between the two via a well-run Ministry of Finance. This combination permitted the asking of some of the 'why?' and 'how?' questions that distinguish evaluation from mere monitoring, with the results have been used for evaluation purposes by the Government (David Booth, pers comm[6]).

With this recognition that qualitative and quantitative methods and data are often more powerful when combined, at different levels and in different sequences, we can categorise different ways of combining and sequencing. Carvalho and White (1997, 18) usefully describe three ways of combining the best of qualitative and quantitative approaches: (1) *integrating* methodologies for better measurement, (2) *sequencing* information for better

[6] David Booth (d.booth@odi.org.uk) is a Research Fellow at the Overseas Development Institute.

analysis, and (3) *merging* findings for better action. We elaborate briefly on these three areas below and apply them to their potential use in impact evaluation.

4.1 Combining methods in impact evaluation

4.1.1 Integrating methodologies

During the design and fieldwork phases of impact evaluation integrating methods and data involves a number of steps or iterations between qualitative and quantitative research. Single steps within this longer process might include:

1. A survey is used to select a qualitative investigation sample;
2. A survey highlights priority issues to be covered in qualitative research;
3. Qualitative analysis identifies knowledge gaps to be filled by survey (here White (2008) describes how crucial information was not collected before a quantitative baseline survey took place and consequently no information was collected on one key explanatory variable (see Box 4.1);
4. Qualitative analysis enables surveys to predict more accurately which issues (sectoral, cross-sectoral or other) and which options are of importance to local people and what explanations they might give, thus improving the definition of survey modules and questions and the categories of choice of answer available;[7]
5. Qualitative analysis identifies what is highly contextual information and therefore what should *not* be subject to standardising quantitative methods;
6. Qualitative research suggests the importance and means of construction of indicators[8] that usefully complement or replace existing indicators. There has been considerable activity around constructing indicators of qualitative (non-material) development impacts, including governance, empowerment, social capital and social exclusion. These are discussed in more detail in Section 4.2 below and in Annex A.
7. Insights from qualitative and quantitative studies help to define population sub-group sampling frames; and
8. Qualitative analysis (including that of key informants) helps determine 'appropriate stratification' of the quantitative survey and subsequent disaggregation of survey cross-tabulation analysis (e.g. along gender, age, socio-economic, political, socio-cultural or ethnic lines).

4.1.2 Sequencing information

Sequencing information can allow for *examining, explaining, confirming, refuting, and/or enriching* information from one approach with that from the other for better analysis. Means of sequencing might include:

1. A qualitative study generates 'working hypotheses' that can be further examined through quantitative research with specific pre-defined questions;

[7] White (2002) gives the example of the potential for qualitative research on labour exchange options to improve the content of rural labour market surveys in Africa.

[8] A powerful example here is use of qualitative research to explore perceptions of social capital in Tanzania and from these perceptions generate weighted indicators with cardinal scores for econometric testing (Narayan, 1997). A similar process was followed in the construction of indicators for participation in water projects (Isham et al, 1995).

2. A contextual study is conducted as a sub-sample of larger, non-contextual surveys. This allows for comparisons to be made between the results of contextual investigation and those of larger surveys *for the same community*;[9]
3. Qualitative investigation assesses how important the 'the average is at the local level'; i.e. assesses how important the heterogeneity of local conditions is for challenging the worth of abstracted/ standardised findings/ conclusions;
4. Qualitative research *explains* relationships/ trends/ patterns emerging from survey;[10]
5. Qualitative investigation triangulates (verifies or refutes) survey results;[11] and
6. Qualitative research enriches analysis of relationships/ trends/ patterns emerging from survey through new learning.

Box 4.1 You can't carry electricity on boats: rural electrification in Lao PDR

In Lao PDR the electricity utilities, Electricitie du Lao—like most electricity utilities—followed the least cost strategy of extending medium voltage cables along the line of road, running low voltage lines into communities with enough households of sufficient income to afford the household connection fee. Using the data to model which communities were connected, the analyst expected that the three variables (distance from road, distance from provincial headquarters—as the service was not yet fully extended along all major roads—and average community income) would explain most of the variation. Against any expectations, the R2 remained low to various variable specifications and a dummy variable for the three ethnic groups stayed significant. Puzzled by those results the researcher went on a field trip to rural Lao PDR and found that many of the unconnected villages were neither particular distant or poor but simple were on an island and had to be reached by boat. There was no grid connection, however, since it is not economic to run the connection across the river for less than 30 households. A key explanatory variable in the regression of whether a community was connected or not was whether it was on an island or not—but the survey's village questionnaire hadn't collected this piece of information.

Source: White (2008)

4.1.3 Merging findings

During the analysis phase of impact evaluation emphasis shifts to ensuring that data are merged sufficiently for improved analysis and policy influence. Analytical frameworks shape powerfully the use and interpretation of data. Social analytical frameworks have been used increasingly in Poverty and Social Impact Analysis and examples of these analytical

[9] This comparative analysis might address, for example, whether stratification according to the household survey is matched by the results of rankings drawn up by local analysts (Booth et al, 1998, 60).

[10] Croll (2000), for example, uses ethnographic methods to explore the statistical relationship between sex and infant mortality in India and China and develops a thesis on 'daughter discrimination ' that highlights the economic status of women. White (2002) points out that these arguments have subsequently been incorporated into economic models of the household.

[11] Hanmer, Pyatt and White (1997) argue that by establishing the ability of qualitative approaches to collect data normally gathered from more expensive techniques, verification plays two important roles: i) Data users will gain the confidence to use the results from qualitative investigation in advance of receiving results from formal sample surveys; and ii) for some uses it should in time become accepted that the formal survey is not required at all (and here they draw a parallel with the difficulty proponents of sample surveys had in initially having their methods accepted).

approaches that trace the political economy and institutional transmission of policy reforms through to distributional outcomes are discussed in the DFID/World Bank *Tools for Institutional, Political and Social Analysis of Policy Reform* sourcebook (Holland, 2007).

An example of an impact evaluation of self-help groups in Andhra Pradesh (see Box 4.2) shows convincingly how merging the results from the qualitative and the quantitative evaluation yield a powerful set of policy recommendations and how qualitative data helped explain and enrich survey data.

Box 4.2 Self-help groups in Andhra Pradesh

Since the mid-90s, the government of the Indian state of Andhra Pradesh has been encouraging women-only grassroots organisations at the village level called Self-Help Groups (SHGs). By 2007, over 700,000 such groups had been formed, partly facilitated by two externally funded programmes supported by DFID and the World Bank which provided funds and technical training to SHGs. IEG's evaluation of these programmes utilised panel data collected in 2005 and 2007. Responses to the village questionnaire, which listed all the SHGs in the village, confirmed a continued rise in the number of these organisations but the individual-level data showed a drop in participation in SHGs. This apparent discrepancy was readily explained by the qualitative data collected alongside the quantitative survey which revealed the build up of non-functioning SHGs—through a lack of skills, non-payment, factionalism and so on—which nonetheless remained on the books. Had the researchers anticipated this attrition of SHGs—through preceding qualitative research—the survey could have included questions regarding the reasons for dropouts.

SHG dropout had affected the poor most, with participation rates for the upper deciles over twice those of the lower deciles. The qualitative fieldwork pointed to some possible policy responses to this problem, including support to illiterate groups in record-keeping (and adoption of simpler bookkeeping systems suitable for semi-literates), finding alternative payment arrangements for the poorest households (lower payments or not requiring payment on a monthly basis), the need for animal insurance to accompany livestock loans, and defining a different (social protection) model to assist those unable to engage in productive activities.

Further policy implications came from the quantitative analysis of the membership decision. Households with multiple eligible female members were not receiving a higher level of loans and so were not participating. The policy implications were clear: either to revise the goal of 100 percent coverage downwards or attempt to change village-level behaviour so that households with multiple members did receive multiple loans.

Source: White (2008)

4.2 Measuring qualitative impacts

We have demonstrated above the comparative advantage of qualitative research for *analysing* and *explaining* impact. However, one can also measure qualitative impacts such as good governance, social capital, empowerment and social inclusion. Details on recent research on measuring non-monetary or non-material dimensions of poverty are provided in Annex A.

There are often good reasons for considering quantifying qualitative impacts in order to measure change. None of these reasons suggest that qualitative approaches are trying to 'fit in' with the 'dominant' quantitative paradigm. In particular, the use of quantification can be very effective in opening up policy space for discussing non-monetary impacts and linking

this discussion to a broader policy debate that incorporates process issues of governance, empowerment, social inclusion and so on.

However, in order for the measurement of qualitative impacts not to become too reductionist it should always be sequenced with qualitative *analysis*, for example of a sub-sample of the surveyed population. In this way the evaluation and subsequent policy learning can be enriched by qualitative analytical studies.

Quantification involves developing and/or applying indicators or indexes that measure changes in qualitative impacts, including both perception scoring data and observable changes in behaviour. These indicators will allow for measurement and aggregation of non-material and often complex multi-dimensional impacts.

The indicators can be collected in different ways. One efficient method is to develop qualitative module to add to an existing longitudinal survey instrument that is being applied to a relatively large sample of the targeted population and, where required, to a comparator population. An alternative method is to use contextual methods in a random stratified sample of sites to generate and aggregate indicator data on qualitative impacts. These data can be collected and using a mix of individual and group based scoring—often described as a community score card (CSC) (see Box 4.3). It can alternatively be collected through a locally conducted beneficiary or household survey—sometimes referred to as a citizen report card (CRC).

A community score card was designed and implemented in Jamaica, for example, as part of a community based monitoring and evaluation of social policy impacts on police-youth relations in a cross section of communities. The score card included indicators of empowerment, designed to measure the existence of choice, exercise of choice and impact of choice for young people in their interactions with the police (see Table 4.1).

Box 4.3 Community Score Cards

A Community Score Card (CSC) is an interactive monitoring tool used to increase accountability of service providers by soliciting male and female user perceptions on the quality, accessibility and relevance of various public services. Here we distinguish the CSC from the Citizen Report Card (CRC) survey module. Unlike the CRC, the CSC is conducted in a focus group setting with a stratified sample of 6-12 service users. CSC data are usually less reliable than survey-generated data because they are elicited from a relatively small sample of respondents. CSC trustworthiness can be increased if necessary by triangulating the CSC data with equivalent data generated by bigger sample-size survey instrument.

The CSC is described as a 'mixed method' tool because it generates both quantitative and qualitative data and analysis. The quantitative data comprise perception scores of specific qualities of service provision, usually scored on a 4 or 5 point scale. These scores can then be aggregated from all the focus group discussions held and can be compared across groups and over time. The key to a successful CSC session, in contrast with a survey module, is that the scores are not simply elicited as an end in themselves but feed qualitative discussion. The interactive focus group setting of a CSC exercise allows the facilitators to use the scores generated to encourage an in-depth diagnostic discussion by the group. The scoring is used to prompt a discussion of three questions: (a) Defining the problem/issue; (b) Diagnosing the problem; and (c) Identifying solutions. Follow up action might involve service users taking action or engaging with service providers to resolve some of the problems identified during the CSC session. If appropriate, the CSC facilitators can extend their role to facilitating 'interface' sessions between groups of users and service providers in which the results of the CSC sessions are discussed and action agreed.

Source: Holland et al, 2007

A third option is to sequence these two methods to strengthen confidence in the findings and to sequence the group based scoring with qualitative analysis. An ongoing longitudinal survey in the Maldives, for example, is monitoring and evaluating the impact of the World Bank's Integrated Human Development Project (IHDP) (Riemenschneider and Holland, forthcoming). This implements the *same perception scoring questions* through a module of a survey (CRC) and through a group-based CSC. The group based activity generates perception data that can triangulate with the survey data, but more importantly prompts a deeper diagnostic discussion to justify and explain the satisfaction scores that the group has given. In other words, scoring is intrinsically useful to the qualitative exercise because the act of being required to score something that is subjective sharpens the qualitative analysis that follows. Participants have to justify their (relatively precise) scores. This process shows variance in the opinion and the facilitator can call upon those people who have scored differently than the majority and can ask for the reasons why they have done so.

Table 4.1 Empowerment scorecard results in the three research communities, Jamaica

	Community Score (Scale: 1= Very poor; 2= Poor; 3= Fair; 4= Good; 5= Excellent).		
Indicator	Harasson Gardens	Poyuton Terrace	Coolblue Gap
	Violent, poor urban	Stable, poor urban	Poor rural
Original indicators (first round scores from February in brackets)			
Level of trust youth have in the police	1 (2)	4 (5)	4 (1)
Level of respect and courtesy displayed by the police	2 (1)	5 (5)	5 (2)
Level of fairness displayed by police	1 (1)	4 (5)	4 (1)
Level of responsiveness of police	3 (--)	3 (4)	2 (1)
Level of effort made by police to interact with the youth	2 (3)	5 (5)	3 (2)
Additional empowerment indicators			
Level of youth access to information about police activities and services	3	5	1
Level of youth willingness to use police services (e.g. reporting incidents)	4	5	4
Ability of youth to officially complain about inappropriate police behaviour / action	5	5	2
Level of youth willingness to officially complain about inappropriate police behaviour / action	1	4	4
Level of youth hope that police-youth relations can improve	2	5	5

Source: Holland et al (2007)

5 Operationalising combined methods for impact evaluation

With a strong case for qualitative research and combined methods in impact evaluation, DFID staff will be faced with a number of issues when considering when and how to support combined methods approaches. These areas—and a checklist of questions that DFID staff will need to consider—are summarised in Table 5.2. Questions and prompts focus on the identification and design of impact evaluations as well as on the choices of data collection methods and analytical tools. They also include questions and prompts on how best to manage the 'political economy' of impact evaluation.

Identification and design questions cover the importance of developing a theory of change and a clear research strategy, of designing a robust sampling protocol and of defending the cost effectiveness of the study as you designed it. Data collection and analysis questions prompt clear and transparent thinking on how best to combine methods for better impact evaluation. They also urge reflection on what types of methods and data will be used to observe and analyse change, including the possibility of contextual alternatives to questionnaires and quantitative measurements of qualitative impacts.

Indeed, the context for evaluation will be a significant determinant in identifying and combining methods in impact evaluation. The notion of a 'fit for purpose methodology' suggests that there is rarely an ideal type impact evaluation that can be appropriately rolled out in all contexts. This was brought home particularly forcefully in comments from DFID colleagues in country offices describing the challenges they face when conducting research and generating data in fragile states (see Box 5.1). These methodological challenges include the need for a 'back to basics' approach when faced with a 'data free environment' and the difficulties in reaching geographical areas.

In highly insecure contexts, methodologies must be sufficiently flexible to allow research teams to manage the risks they face in the field. In Afghanistan, for example, DFID has been supporting an annual country wide assessment called the National Risk and Vulnerability Survey (NRVA). The purpose of this study is to elicit insights into exposure to -- and management of -- risk amongst different social groups in rural populations. The survey is conducted in a challenging and changing context of conflict and insecurity. The NRVA team took a number of measures to manage the security risks associated with the study including: (i) adjusting the monthly sampling procedure by give field staff the opportunity to assess the local security situation each month. If the enumeration area was not considered secure within the assigned month the team were able to conduct field work in subsequent months, while not missing the entire season; (ii) recruiting localized staff better able to deal with security and to travel during times of insecurity. In the highly insecure south western Zone, for example, the field staff coordinated with village elders to rehearse answers to questions they might be asked if stopped en route. They might say that they were going to ask for engagement of their son with one of the daughters living in this enumeration area and that is why they had women with them, or explain that they were visiting a *Zyarat* (grave of a good person) which is common in Afghanistan; (iii) taking care of the field staff through good behaviour and close communication, including 24-hour phone support; (iv) being flexible about transporting questionnaire forms to and from field sites, under the guidance of regionally recruited supervisors; and (v) being flexible about the recruitment of women to the enumeration teams. In some areas, such as Uruzgan province, where the presence of women would put the team under greater risk, male-only teams were used. In insecure and highly remote areas, such as Nooristan province, they used local male enumerators who

were able to move more freely, speak the local dialect and negotiate local cultural constraints on accessing female respondents (Amanullah Assil, pers comm.).[12]

In addition to these methodological challenges, questions, there is an increasing acceptance—recognised by the DFID Evaluation Department and reflected in the DFID Policy Paper—that demand for impact evaluation will need to come from policy-makers in-country if the results are to successfully feed back into policy making. Often DFID staff face what has been called 'political economy' of impact evaluation. Sometimes, there is no interest in undertaking impact evaluations especially when the results are expected to be bad news. The result can be a biased selection of impact evaluation towards interventions where results are good.

Moreover, there can in many instances be an embedded culture where evidence is not rated very highly, or a situation in which evaluation capacity is extremely low or where you simply do not have data in the first place. Overcoming these problems goes beyond impact evaluation methodology and prompts concerns with capacity building and institutional demand for evidence-based policy making.

Thus the task of strengthening qualitative and combined methods approaches to impact evaluation is far more than a technical challenge. DFID staff will need to consider how best to manage the risks and opportunities associated with the political economy of impact evaluation.

[12] Amanullah Assil (Amanullah.assil@mrrd.gov.af) heads the NRVA National Surveillance System implementation project and is senior advisor to the Ministry of Rural Rehabilitation and Development on food security, poverty and vulnerability issues.

Box 5.1 The Challenge of impact evaluation in fragile states: Thoughts from DRC

With the first democratic elections held in 2006 in DRC, DFID started to design a suite of programmes covering 11 different areas of development. Given the opportunity of having a range of programmes start at similar times as well as the challenge of a poor human rights context, the DFID country office is very aware of the importance of having a results-focus and being able to demonstrate impact. In each of the 11 programme areas, DFID started from the basics of ensuring a clear log frame, vetting M&E plans and requiring baseline studies to be completed within the first year of the programme, when targets would then be revised. Additional funds for monitoring and evaluation have been allocated (often requiring DFID to top-up multilateral partners) and, as appropriate, independent monitoring and evaluation mechanisms established. Two programmes illustrate where monitoring systems are combining methods to generate baseline and longitudinal data.

(1) In the programme area of community reconstruction, a baseline was created by combining randomised control trials (conducted independently by Colombia University) with qualitative work on perceptions of security and wealth ranking (led by the implementing NGOs International Rescue Committee and CARE). This survey work will be repeated in Year 4 of the programme to assess overall programme impact.

(2) In the programme area of media, DFID asked 'What is the media environment and how is this changing?'. Attribution to DFID programmes becoming a second order question. The country office has generated a robust baseline by combining a qualitative, peer reviewed study on perceptions of media freedom with a quantitative survey on ownership of radios, media access and so on. Longitudinal monitoring includes objective assessment of the quality of governance content of key media (internet, radio, press and television), six-monthly qualitative reviews by five panels assessing different aspects of the media as well as periodic national audience surveys.

Monitoring and evaluation is being conducted in a very challenging context. The DRC is essentially a 'data-free' environment, lacking even basics national data on population size and distribution and so there are virtually no national or regional poverty or sector secondary data to draw from. The security and infrastructure context means that survey work is subject to delays and changes, time intensive and expensive: the community recovery baseline required a population mapping exercise to develop the sampling frame. This identified some villages not recognised by local administrators and left a survey team out of contact for weeks at a time walking from place to place. There are also accompanying capacity challenges to build monitoring capacity both in the national bodies and the university 'think tanks' where some of the programme monitoring should eventually by housed. There can also be limited technical capacity within some DFID advisory cadres and multilateral partners (notably within the UN) to take on this type of results work substantively. DFID DRC recognises that—despite the importance of measuring change and results—there are huge time and cost implications of building data bases and methodologies from scratch in this fragile context. Finally there is a significant process challenge in attempting to anchor and communicate the monitoring of sensitive governance indicators—e.g. on media freedoms—with a partner government that is transitioning from conflict to stability.

Source: Jo Abbot, Senior Social Development Adviser/Results Champion, DFID DRC

Improving technical approaches and methods is not very effective if you cannot manage the political economy risks and opportunities through the process of designing, implementing and reporting on the impact evaluation. Your results (as good as rigorous as they may be) will not feed back into policy if Government and other stakeholders do not have ownership of the process and the results. Hence the process of building ownership—and associated capacity—amongst policy makers and bureaucrats early and sustainably into the research process is key to ownership and traction of the impact evaluation, as illustrated by the case of a recent DFID-financed impact evaluation in Ethiopia (Box 5.2).

Box 5.2 Impact evaluation of the Ethiopia social protection programme

A recent experience with impact evaluation of Ethiopia's social protection programme has shown that the process how the impact evaluation is undertaken is key to its success. It became clear that if the results are to be accepted by policy makers there is a need to integrate the impact evaluation in government processes. Long discussions between donors and the GoE took place on many issues including:

(1) How to make use of Government resources

(2) Design on the evaluation

(3) Recognition of Government that analytical capacity was needed (at the end reporting arrangements directly to GoE)

At the end the GoE demanded for resources to be seconded to the Statistics office and the International Food Policy Research Institute (IFPRI) subsequently worked with the Bureau of Statistics on the design of the impact evaluation. This agreement resulted in a complex process which was difficult and time-consuming to manage but the alternative—no government buy-in—the results of the impact evaluation would have had a low chance of being fed back into policy-making.

The careful management of the process around the impact evaluation in addition helped with dealing with the political economy of impact evaluation, i.e. the reluctance of many stakeholders to undertake the evaluation as results are expected to include bad news. In the case of Ethiopia several issues contributed to the GoE's acceptance of the need to evaluate the programme's impact, including:

(1) Political incentives for impact evaluation and around the programme

(2) GoE wanted to demonstrate that they moved away from food-aid towards cash benefits

(3) Support to the economy and growth agenda by having people graduating from the programme, i.e. not depending on cash benefits anymore (participants had also access to other livelihoods services)

Besides the importance of the process as such this experience showed that impact evaluations that take into account political motivations are more likely to be accepted. Involving political economists to work alongside technical impact evaluation specialists may be one starting point.

Source: Tim Robertson pers comm[13]

This process approach to impact has been formalised by GTZ in developing the CoImpact tool (see Box 5.3). CoImpact has been successfully applied in Malawi, Kenya and Jordan. The Government of Malawi used CoIMpact—called QIM (Qualitative Impact Monitoring)—to assess the impacts policy reform, including its policy to remove fertilizer subsidies.

[13] Tim Robertson (t-robertson@dfid.gov.uk) is Theme Leader for Governance, Social Development, Humanitarian and Conflict at DFID.

Box 5.3 Consultative impact monitoring of policies (CoIMpact): Qualitative Impact Monitoring (QIM) in Malawi

CoIMPact is designed to assess and monitor the effectiveness of poverty-focused policies and programs on their target group. It collects data using participatory methods and combines this where appropriate with survey and other data sources. At the same time, CoIMPact focuses on the process of policy dialogue, trying to involve representatives from an array of institutions both governmental and nongovernmental, and ensuring that the results of the exercise find their way into the policy process.

CoIMPact builds on three principles which smooth the transition between research and policy advice:

- First, the instrument is **institutionally anchored** in the national policy review process. In Malawi, for example, CoIMPact is called QIM (Qualitative Impact Monitoring) and is managed by the Ministry of Economic Planning and Development. This ministry formulated the Policy Framework for Poverty Alleviation in 1996 and is now the secretariat for the poverty reduction strategy
- Second, CoIMPact is **embedded in existing monitoring systems** and complements quantitative monitoring efforts. QIM has been integrated in the national Poverty Monitoring System since1998, and was adopted as the tool for participatory monitoring of the Malawi poverty reduction strategy in 2000
- Third**,** capacity **building,** predominantly of government officials, but also some NGO staff and local consultants, is achieved through training, and more importantly, exposure. In the most recent QIM round in Malawi, 71 people were trained and applied the instrument, of which 24 were national-level government officials and 45 were district-level officials and local NGO representatives. Seven people who had also participated in the first QIM round, passed their skills on through their roles as trainers and facilitators.

Source: Holland (2007); see also http://www.gtz.de/

The work of the UNDP Oslo Governance Centre (OGC) includes a strong focus on nationally –owned governance assessments. For UNDP a 'democratic governance assessment' provides a critical accountability mechanism for government and for citizens to engage on governance issues and voice their opinions. Through a Global Programme coordinated by the OGC, UNDP assists developing countries produce disaggregated and non-ranking governance indicators to enable national stakeholders to better monitor performance in democratic governance reforms. The programme aims to provide cutting edge knowledge on methods, country experiences, good practices as well as a network of partners in the field. The OGC is only too aware of the challenges to such a process. Although nationally owned, country-led governance assessments often face common challenges of capacity deficits, a lack of political commitment, competing assessment interests, poor coordination, indicator fatigue and unsustainable resources (Joachim Nahem, pers comm.[14]).

Lessons for impact evaluation processes can also be learnt from PSIA. The donor community has increasingly recognised that getting the PSIA process right is as important as the quality of the analysis (World Bank, 2004; DFID, 2005). Past experience has shown that a PSIA that is conducted with due concern to national policy processes and stakeholder interests can expand the evidence basis for policy. Similarly, understanding issues related to

[14] Joachim Nahem (Joachim.Nahem@undp.org) manages the UNDP Oslo Governance Centre's Global Democratic Governance Assessments.

political economy improves the likelihood that the results of the PSIA will influence decisions and policy implementation. Box 5.4 outlines key principles of good PSIA processes and Holland (2007) provides examples of lessons learnt from past PSIA processes.

Box 5.4 Key principles of good PSIA process

PSIA can be built on an understanding of policies and policy processes. Policy processes are not technical procedures that respond in a neutral fashion to emerging evidence.

PSIA can be embedded in local policy cycles and be a transparent part of the policy process.

The choice of topic for PSIA can be transparent and consultative.

The key actors leading the PSIA process can understand their complementary roles: these are commissioners, practitioners, and facilitators.

The appraisal of PSIA proposals can take good process into account. Where it is evident that PSIAs are likely to be extractive, with weak local engagement and ownership, PSIAs are unlikely to be effective.

Communication and dialogue can be promoted to encourage broadened participation from a wide range of stakeholders. Existing or new sets of relationships that are inclusive and empowering can be further institutionalized through the PSIA process.

Wherever possible, PSIA can build the capacity of local partners, including research practitioners, policy makers, and civil society organizations.

Source: Schnell et al (2005)

Table 5.2 Using qualitative and combined approaches in impact evaluation: A checklist for DFID staff

Questions	Prompts
1. Impact Evaluation identification and design	
1.1. Have you identified what you are trying to evaluate?	Do you have a clear statement of evaluation aims and objectives?
1.2. Have you established a theory of change that can be tested in the evaluation?	Can you articulate the main hypotheses and effect assumptions that will be tested in the impact evaluation? Do you have a clear theoretical framework that can be used to interpret the impact evaluation findings?
1.3. Do you have a clear research strategy?	Can you describe how the overall research strategy is designed to meet the aims of the impact evaluation? Do you need to collect additional data to inform your desk review (e.g. through a limited number of key informant interview)? Can you make a convincing argument for the different features of the research strategy (components, stages, methods, data sources, time frames etc)?
1.4. Do you have a well defended sampling and/or case selection methodology?	Do you have a clear understanding of what can be generalised to a wider population from which your sample is drawn or case selection made? Do you have a clear understanding of the limitations of the design for drawing wider inference?
1.5. Can you demonstrate that the impact evaluation is cost effective?	Have you considered which research instruments will generate the impact evaluation data and analysis most efficiently (time and resources)? Have you established that the evaluation is fit for purpose and hasn't been over designed and over resourced?
2. Data collection and analysis	
2.1. Have you considered how you might combine methods for better impact evaluation?	
Integrating methodologies	Using an (existing) survey to select qualitative investigation sample. Using an (existing) survey to highlight priority issues to be covered in qualitative research. Using qualitative analysis to identify knowledge gaps to be filled by survey. Using qualitative analysis to identify categories for responses in surveys. Using qualitative analysis to identify what is contextual and therefore what should *not* be standardised in surveys. Using qualitative research to construct indicators of qualitative (non-material) impacts.

Questions	Prompts
	Using insights from qualitative and quantitative studies help to define population sub-group sampling frames. Using qualitative analysis (including that of key informants) to determine 'appropriate stratification' of a quantitative survey and subsequent disaggregation of survey cross-tabulation analysis.
Sequencing information	Using a qualitative study to generate 'working hypotheses' that can be further examined through quantitative research with specific pre-defined questions. Conducting a contextual study as a sub-sample of a larger, non-contextual survey. This allows for comparisons to be made between the results of contextual investigation and those of larger surveys for the same community. Using qualitative research to assess how important the 'the average is at the local level'. Using qualitative research to *explain* relationships/ trends/ patterns emerging from a survey. Using qualitative research to triangulate (verifies or refutes) survey results. Using qualitative research to enrich analysis of relationships/ trends/ patterns emerging from survey through new learning.
Merging findings	Using social analytical frameworks that can interpret qualitative and quantitative data
2.2 Have you identified what types of qualitative methods and data you are going to use and generate?	
Observing change	Using a qualitative perception scoring module of a survey. Using measurements of observable behavioural change. Using group based scoring (community score card). If so, have you ensured that community score card activities are linked to qualitative diagnostic discussion?
Analysing change	Using contextual methods (including participatory research methods, ethnography, interviews, focus group discussions, theatre, documentary and video analysis, and observation)
2.3. Have you considered using contextual (including participatory) research as an alternative to a standard survey instrument?	If so, have you considered what the aim of the participatory research is primarily extractive (extracting information to understand the local context) or transformative (empowering local communities through the research process)? Following on from the above, have you considered the costs and trade offs that might be involved choosing participatory research over 'conventional' qualitative or quantitative research?
2.4. Have you considered quantifying and measuring qualitative (non-material) impacts?	If so, have you considered the most appropriate quantification of qualitative impacts: (i) perception scoring; (ii) observable change; (iii) a mix of perception scoring and observable change?

Questions	Prompts
	Is there existing or ongoing secondary survey data that can provide proxy data? If not, can you justify further survey modules in contexts where there may be capacity and resource challenges in national agencies? Have you identified an existing/ongoing survey instrument that can be adapted? Have you considered sequencing this module with in-depth sub-sample qualitative analytical tools?
3. The Political Economy of Impact Evaluation	
3.1. Have you considered the political risks and opportunities to impact evaluation?	How do you deal with policy stakeholders whose interests are threatened by attempts to tackle the evaluation gap? Do you know which stakeholders these are likely to be and what their political motivations are? How do you deal with policy stakeholders that create a selection bias through favoured policies, programmes or projects? Do you have a strategy or entry point for the impact evaluation, e.g. finding supporters of the impact evaluation among national policy makers?
3.2. Have you considered the risks and opportunities faced by impact evaluation *process*?	Do you have a strategy for making the impact evaluation process more transparent, inclusive and accountable? Do you have relevant and sufficient inputs (financial and human resources) to manage the impact evaluation process successfully?
3.3. Have you assessed and managed the likely risks and opportunities to using qualitative research?	Is there a general openness to mixed methods? If not have you considered strategies for introducing qualitative approaches? Is there a champion of qualitative research in the MDA? Have you ensured Bureau of Statistics seal of approval for sampling protocol and research methodology? Have you considered strategies for building policy ownership of the research process and findings early and sustainably?
3.4. Have you considered the capacity building risks and opportunities for qualitative approaches in impact evaluation?	Have you conducted a capacity assessment of local research partners? Do you have a strategy for capacity building, which may include linking 'northern' and 'southern' institutions in technical cooperation? Have you explored the possibility of collaboration with Government stakeholders, e.g. line ministries or the Bureau of Statistics?

Source: Authors' analysis; see also Spencer et al (2004)

6 Concluding comments

In this issues paper we have reviewed the case for promoting and formalising qualitative and combined methods for impact evaluation as part of a broader strategy amongst donors and country partners for tackling the evaluation gap.

The case for qualitative and combined methods is strong. Qualitative methods have an equal footing in impact evaluation and can generate sophisticated, robust and timely data and analysis. Combining qualitative research with quantitative instruments that have greater breadth of coverage and generalisability can result in impact evaluations that make the most of their comparative advantages.

Operational staff need sufficient knowledge to consider the methodological options available to them when identifying and designing impact evaluations. Perhaps even more significantly, they need to consider the political economy risks and opportunities for impact evaluation in any given policy context. Managing these risks and opportunities effectively will mean embedding the impact evaluation in a policy process which is locally owned, inclusive and sustainable.

Bibliography

Alsop R, Bertelsen, M. and J Holland. (2006). *Empowerment in practice: From Analysis to* implementation. The World Bank: Washington D.C.

Appleton, S. and D. Booth. (2001). '*Combining participatory and survey-based approaches to poverty monitoring and analysis'*. Background Paper for Workshop held in Entebbe, Uganda 30 May-1 June 2001. Overseas Development Institute: London.

Appleton, S. and D. Booth. (2005). 'Strong fences make good neighbours: survey and participatory appraisal methods in poverty evaluation and poverty reduction strategy monitoring' in Holland, J. with J. Campbell (eds) (2005). *Methods in development research: Combining qualitative and quantitative approaches.* ITDG Publications: London.

Barahona, C. and S. Levy. (2002). *'How to generate statistics and influence policy using participatory methods in research'*. Statistical Services Centre Working Paper, University of Reading.

Barahona, C. and S. Levy. (2005) '*The best of both worlds: Producing national statistics using participatory methods*'. Working Paper No. 11. Q-Squared. Toronto, Canada.

Blattman, C. (2008) *Impact Evaluation 2.0. Presentation to the Department of International Development*, 14 February 2008.

Booth, D., Holland, J., Hentschel, J., Lanjouw, P. and A. Herbert. (1998) '*Participation and combined methods in African poverty evaluation: Renewing the agenda*'. Department for International Development, Issues, February, pp 127, ISBN 86192 027 X. http://www.dfid.gov.uk/pubs/files/sddafpov.pdf.

Brook, S., Thomson, A. and M. Strode. (2006) *DFID New Economists Guide: Monitoring Chapter*, Oxford Policy Management: Oxford.

Carraro, L., Weinzierl, S and S. Hunt. (2005*). 'A quantitative evaluation of social exclusion in Pakistan'.* Oxford Policy Management: Oxford.

Carvalho, S. and H. White. (1997) *'Combining the quantitative and qualitative approaches to poverty measurement and analysis'*, Technical Paper 366. The World Bank: Washington D.C.

Center for Global Development. (2006) *'When will we ever learn? Improving lives through impact evaluation.*' Report of the Evaluation Gap Working Group. Washington, D.C.

Chambers, R. (1994) 'The origins and practice of participatory rural appraisal' in: *World Development* 22, pp: 959-961.

Chambers, R. (1995) '*Poverty and livelihoods: Whose reality counts?'*, IDS Discussion Paper No. 347, IDS: Brighton.

Chambers, R. (1997) '*Whose reality counts? Putting the first last'*. ITDG: London.

Chambers, R. (2003) 'The best of both worlds', in Kanbur, R. (ed) *Q-Squared: Qualitative and quantitative methods of poverty appraisal*, Permanent Black: Delhi.

Chambers, R. (2007) '*Who counts? The quiet revolution of participation and numbers.'* Working Paper 296. Institute of Development Studies: Brighton.

Chung, K. (1997) '*Identifying good practice for the integration of qualitative and quantitative methods for poverty research'*, Providence, RI, Brown University, draft.

De Haan, A. (1999) '*Social exclusion: Towards an holistic understanding of deprivation'*. Department for International Development: London.

Department for International Development (DFID). (2005) *Poverty and social impact analysis: Principles for good practice*. DFID, London.

Department for International Development (DFID). (2007) *Results action plan*. http://www.dfid.gov.uk/Pubs/files/results-action-plan08.pdf [13 October 2008].

Department for International Development (DFID). (2008) *'(Draft) Policy statements on impact evaluation*—September version.' Internal Draft.

Department for International Development (DFID). (2008b) *'(Draft) Gender and Social Policy Exclusion Analysis*'. How to note. Internal Draft, October 2008.

Department of International Development (DFID). Project memorandum: DFID support to the international initiative on impact evaluation (3ie).

Earl, S., Carden, F. and T. Smutylo. (2001) O*utcome mapping building learning and reflection into* development *programs.* IRDC http://www.idrc.ca/en/ev-28377-201-1-DO_TOPIC.html [4 December 2008].

Gacitua-Mario, E. and Q. Woden. (2001) '*Measurement and meaning: Combining quantitative and qualitative methods for the analysis of poverty and social exclusion in Latin America'*. The World Bank: Washington D.C.

Grootaert, C. and van Bastelaer, T. (2001) *Understanding and measuring social capital: A synthesis of findings and recommendations from the social capital initiative*. Social Capital Initiative. Working Paper No. 24. The World Bank, Washington D.C.

Gunning, J. W. (2006) 'Aid evaluation: Pursuing development as if evidence matters' in *Swedish Economic Policy Review* 13, pp: 145-163.

Heller, P. (1996) Social capital as a product of class mobilization and state intervention: Industrial workers in Kerala, India, in *World Development* 24 (6), pp: 1055-1071.

Hentschel, J. (1999) 'Contextuality and data collection methods: A Framework and application to health service utilisation' in *The Journal of Development Studies* 35, pp: 64-94.

Holland, J. (2007) *Tools for Institutional, Political and Social Analysis of Policy Reform (TIPA): A Sourcebook for Development Practitioners*, The World Bank: Washington D.C.

Holland, J. with J. Campbell (eds). (2005) '*Methods in development research: Combining qualitative and quantitative approaches.* ITDG Publications: London.

Holland, J., Brook, S., Dudwick, N., Bertelsen, M. and G. Yaron. (2007) *Monitoring empowerment in policy and programme interventions:* Combining Qualitative and Quantitative Approaches', Q-Squared Working Paper No. 45, University of Toronto.

Independent Evaluation Group. (nd) *Impact evaluation—the experience of the Independent Evaluation Group of the World Bank*. World Bank, Washington D.C. http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/35BC420995BF58F8852571E00068C6BD/$file/impact_evaluation.pdf [20 October 2008].

Jones, V. N. and M. Woolcock. (2007) '*Using mixed methods to assess social capital in low income countries: A practical guide*'. BWPI Working Paper 12. Brooks World Poverty Institute. Manchester.

Kabeer, N. (2000) '*Social exclusion, poverty and discrimination: Towards an analytical framework'*, IDS Bulletin, 31(4), Institute of Development Studies, Brighton.

Kanbur, R. (ed) (2003) *Q-Squared: Qualitative and quantitative methods of poverty appraisal*, Permanent Black: Delhi.

Knack, S. and P. Keefer. (1997) 'Does social capital have an economic payoff? A cross-country investigation'. *The Quarterly Journal of Economics*, MIT Press, vol. 112(4), pp. 1251-1288.

Kuenzi, T. (2008) '*Social capital and political trust in West Africa'*. Afrobarometer Working Paper No. 96.

McGee, R. (2003) 'Qualitative and quantitative poverty appraisal workshop: Some reflections and responses', in Kanbur R (ed) *Q-Squared: Qualitative and Quantitative Methods of Poverty Appraisal*. Permanent Black: Delhi.

Moser, C. and C. McIlwaine. (2004) '*Encounters with Violence in Latin America: Urban Poor Perceptions from Colombia and Guatemala*.' Routledge: New York and London.

Narayan, D. and L. Pritchett. (1997) *Cents and sociability: Household income and social capital in rural Tanzania.* Policy Research Working Paper 1796. The World Bank: Washington D.C.

Network of Networks for Impact Evaluation (NONIE): http://www.worldbank.org/ieg/nonie/index.html

No author (2007) *Impact evaluation of new aid instruments and country programs*. Discussion paper (second draft, December 12, 2007).

OECD (nd). Evaluation and aid effectiveness: Glossary of key terms in evaluation and results-based management, OECD DAC.

Prowse, M. (2007) *Aid effectiveness: the role of qualitative research in impact evaluation*. Background note. Overseas Development Institute. London.

Putnam, R. D. (1993) Making democracy work: Civic traditions in modern Italy. Princeton University Press: Princeton, New Jersey.

Q2: http://www.q-squared.ca/papers.html

Ravallion, M. (2008) *Evaluation in the practice of development*. Policy Research Working Paper 4547, The World Bank: Washington D.C.

Riemenschneider, N. and J. Holland (forthcoming). '*Combining qualitative and quantitative results: do they tell the same story? Experience from using Community Scorecards and Citizen Report Cards in the Maldives'*, *OPM Briefing Note*, Oxford Policy Management: Oxford.

Schnell, S., Poulsen, P., Condy, A., Tertsunen, M. and J. Holland. (2005) *Principles for PSIA process in policy cycles and stakeholder participation:* A document produced jointly by GTZ and DFID for sharing with the PSIA network. Department for International Development: London.

Sen, A. (2000) '*Social exclusion: Concept, application, and scrutiny*'. Asian Development Bank, Manila.

Spencer, L., Ritchie, J., Lewis, J. and L. Dillon (2004) '*Quality in qualitative evaluation: A framework for assessing research evidence'*, Cabinet Office, Government Chief Social Researcher's Office, 2nd Revised Edition.

UNDP (1996) *UNDP's 1996 Report on Human Development in Bangladesh*, Volume 3 Poor People's Perspectives, UNDP: Dhaka.

UNDP (2000) *UNDP strategy note on governance for human development.* UNDP: New York.

UNDP (nd) Governance indicators: A Users' Guide, Oslo: UNDP.

Venables, T. and S. Holloway (2008) *Using number to improve decision taking*. PS/PUSS Malik. DFID internal note.

White, H. (2002) 'Combining quantitative and qualitative approaches in poverty analysis', in *World Development* 30(3)

White, H. (2005) '*Challenges in evaluating development effectiveness'*. Working Paper 242, Institute of Development Studies: Brighton.

White, H. (2006). *DFID Economists' Manual: Impact Evaluation*, Independent Evaluation Group, World Bank, Washington D.C.

White, H. (2008) *Of probits and participation: The use of mixed methods in quantitative impact evaluation*. Working Paper No. 6, Network of Networks on Impact Evaluation (NONIE). http://www.worldbank.org/ieg/nonie/docs/WP7_White.pdf [14 October 2008].

White, H. (nd). Making impact evaluation matter. International Initiative for Impact Evaluation. http://www.ideas-int.org/Documents/Howard_White.doc [accessed: 2 December 2008].

White, H., Sinha, S. and A. Flanagan. (nd) A review of the state of impact evaluation. Independent Evaluation Group, World Bank.

World Bank (2006a) Conducting quality impact evaluations under budget, time and data constraints, Independent Evaluation Group, The World Bank: Washington D.C.

World Bank (2006b*) A decade of measuring the quality of governance. Governance matters 2006.* Worldwide governance indicators. The World Bank: Washington D.C.

World Bank (2008) *Good practice note: Using Poverty and Social Impact Analysis to support development policy operations*. The World Bank, Washington D.C.

Yates, J., Cooper, R. and J. Holland (2006) 'Social Protection and Health: Experiences in Uganda', in *Development Policy Review* 24 (3) 339-356.

Annex A Measuring qualitative impacts

Given the potential for qualitative approaches to add to the measurement and analysis in impact evaluation, in this section we review existing non-monetary or non-material dimensions of poverty that have been subject to quantification. These are qualitative, often relational, states that have been measured using perception scoring and/or by capturing observable changes in behaviour. In some cases this measurement has been integrated with contextual methods that provide in-depth understanding of the poverty dynamics underpinning the numbers.

In this section we categorise and briefly review a non-exhaustive range of non-monetary or non-material poverty dimensions that have been measured using quantified indicators or indexes in some cases integrated with qualitative diagnostic research.

A.1 Governance

Governance is an important qualitative area of measurement for donors seeking to improve – and measure their impact on -- the institutions that govern the design and delivery of public policy. Governance 'is the system of values, policies and institutions by which a society manages its economic, political and social affairs through interactions within and among the state, civil society and private sector' (UNDP, nd).

Governance indicators have been developed by many agencies and have covered different dimensions of governance, including corruption, openness, rule of law, accountability, electoral freedom and executive performance. The UNDP Governance Centre (UNDP, nd) has produced a User's Guide to Governance Indicators. Notable amongst these are:

- The World Governance Indicators compiled in the World Bank's Annual Governance Matters report (e.g. Kaufmann, Kraay and Mastruzzi, 2005) are probably the most widely used and comprehensive international comparative measures of governance. They are based on a composite of 31 sources from 25 separate organisations, which are grouped into six categories: (i) Voice and accountability; (ii) political stability; (iii) government effectiveness; (iv) regulatory quality; (v) Rule of law; and (vi) Control of corruption.
- Afrobarometer, which uses satisfaction surveys to produce comparative measures of national public attitudes on democracy and economic conditions.
- Freedom House index which uses a survey of 'expert' ratings to measure progress in developing political freedoms, comprising scores for political rights and civil liberties.
- Transparency International Bribe Payers Index which use perception scoring to rank leading exporting countries in terms of the degree to which international companies with their HQ in those countries are likely to pay bribes to senior public officials
- Transparency International Corruption Perceptions Index, which is a composite index drawing on 12 polls and surveys from 9 independent institutions, which gather the opinions of business people and country analysts
- The World Bank Country Policy and Institutional assessment which uses Bank staff to assess country performance against four clusters (economic management, structural policies, policies for social inclusion and equity and public sector management and institutions) in order to measure how conducive that framework is to fostering poverty reduction, sustainable growth and the effective use of development assistance.
- The UNDP Gender Empowerment Measure (GEM) which uses countable proxy indicators to capture gender inequality in three key areas: The extent of women's political

participation and decision making, economic participation and decision making power, and the power exerted by women over economic resources.

- The Global Integrity Index (produced by Global Integrity) which measures the opposite of corruption by combining qualitative journalistic reporting with an in-depth quantitative scorecard approach to assess the institutions and practices that citizens can use to hold their governments accountable. Using 300 disaggregated indicators, the Global Integrity Index measures the following dimensions of governance: civil society, public information and media; elections; government accountability; administration and civil service; oversight and regulatory mechanisms; and anti-corruption mechanisms and rule of law.

A.2 Empowerment

Recent attempts to measure empowerment have drawn on the work of Sen and have drawn on social theory to look at the relationship between individual or group capabilities and institutional contexts (see for example Alsop et al, 2006).

Poverty reduction focuses on stimulating growth (with the assumption that growth will increase employment) and providing resources and services to address needs and enhance material well-being. A focus on empowerment brings an additional emphasis on people's choices and opportunities. An empowerment focus, articulated using a simple analytical framework, also provides an entry point into identifying tools and indicators for measuring changes in empowerment. Although empowerment is now seen as a legitimate developmental goal in its own right, there is a growing body of anecdotal and case study evidence to suggest that empowerment also brings improved poverty reduction outcomes.

A recent World Bank project managed by PRMPR has developed and tested empowerment measures in Jamaica, Bangladesh (see Box A.1) and Ghana. The project defines empowerment as 'the process of increasing the assets and capabilities of individuals or groups to make purposive choices and to transform those choices into desired actions and outcomes' (World Bank, 2007). The framework rests on the concept of 'asset-based agency,' (which relates to stocks of assets) and 'institution-based opportunity structure' (relating to how the 'rules of the game' operate).

Measures of empowerment are important both as indicators of country performance and of individual or group opportunity. At the macro level, empowerment indicators contribute to the process of strengthening governance and reducing corruption, both important donor priorities. Measuring empowerment can be considered a complementary approach to top-down reforms because it addresses the capacity of citizens to reinforce change by demanding better governance and hold officials to account. 'Empowerment' as a concept moves away from treating people primarily as 'beneficiaries,' by treating them as agents capable of effecting institutional change. Empowerment as a means (if not an end in itself) is thus implicit in successful accountability initiatives.

The project developed mixed method diagnostic tools that used scorecards sequenced with ethnographic research to measure changes in the capacity of citizens effectively to demand service improvements, and in some cases the capacity of service providers to actually provide these services. Identifying these changes required measurement and analysis of empowerment outcomes (changes in power relations attributable to policy or programme intervention) and empowerment processes (changes in perceptions and behaviours indicative of empowerment and attributable to policy or programme intervention).

Box A.1 Measuring and analysing empowerment impacts: Social Safety Net Programs in Bangladesh

A World Bank study measured the empowerment impacts of seven social safety net programs (SSNP) in Bangladesh using a rich model of empowerment and an innovative combination of qualitative and quantitative methods. Using an empowerment analytical framework that conceptualised empowerment as the interaction of asset-based agency and institution-based opportunity structure, the team developed questions and research instruments. The research involved: adding an empowerment module to an existing national survey of 2,040 beneficiary and 701 non-beneficiary households to evaluate the most important SSNP; adding a short women's questionnaire administered by female enumerators to the wife or female household head; using community questionnaires, which asked questions about local facilities as well as social norms and practices; and undertaking 72 focus group discussions (FGD) to investigate SSNP impacts in more depth.

Two econometric methods were used to analyze the data. Ordered logit analysis was used to analyze findings from the household survey's empowerment module, while propensity score matching (PSM) was used to analyze a data set drawn from those parts of the household survey asked to female beneficiaries and non-beneficiaries, a community survey and the women's questionnaire. Results were triangulated with FGD findings to understand the empowerment impacts of the SSNP in the economic, social and civic domains of empowerment.

Combining data from all four instruments (the household, women's and community questionnaires and the focus group discussions) revealed that the SSNP help to address extreme poverty, but did not help those women who receive program benefits in terms of social or civic empowerment. Women gained assets—and self confidence—but found that conservative norms either prevented them from changing their behaviour or from successfully transforming their choices into desired outcomes?

Source: Gil Yaron, pers comm[15]

The working hypothesis underpinning the design of mixed-method diagnostic tools was that dysfunctional relations between service providers and users or between government officials and citizens may be symptomatic of deeper, embedded institutional norms that are characterised by inequalities in power. These social structures perpetuate and are in turn reinforced by everyday interactions and negotiations around service delivery and policy implementation. The result is that improvements in policy or programme outcomes will be hard to achieve and sustain without interventions that attempt to tackle institutional norms.[16] Empowerment and social change require a level of transformation in critical consciousness that challenges habitual[17] or everyday interaction and decision making.

The policy implication of an impact evaluation focus on measuring empowerment was therefore that policies and programmes must take into account and respond effectively to this challenge by building individual and group capabilities and by finding ways to influence and change the formal and informal institutional structures that govern people's behaviour and influence the success or failure of the choices they make.

[15] Gil Yaron (gil_yaron@gya.co.uk) is Director of GY Associates.

[16] Alsop et al, op cit. See also World Bank, 2005. *World Development Report 2006: Equity and Development*, New York: Oxford University Press

[17] See Bourdieu, P. 1977. *Outline of a Theory of Practice.* Cambridge: Cambridge University Press.

A.3 Social capital

Theoretical and conceptual debates on the definition and measuring of social capital in developing and developed countries are ongoing and this short section does not intend to give an exhaustive overview over the debate but attempts to give some insights on the value of combining quantitative and qualitative to measure social capital. This section will take the following World Bank definition as starting point:

> 'Social capital refers to the institutions, relationships, and norms that shape the quality and quantity of a society's social interactions. Increasing evidence shows that social cohesion is critical for societies to prosper economically and for development to be sustainable. Social capital is not just the sum of the institutions which underpin a society—it is the glue that holds them together.'[18]

In order to apply the concept of social capital at a practical and operational level, the World Bank further suggests five key dimensions of social capital: (1) groups and networks—collections of individuals that promote and protect personal relationships which improve welfare; (2) trust and solidarity—elements of interpersonal behaviour which fosters greater cohesion and more robust collective action; (3) collective action and cooperation—ability of people to work together toward resolving communal issues; (4) social cohesion and inclusion—mitigates the risk of conflict and promotes equitable access to benefits of development by enhancing participation of the marginalized; and (5) information and communication—breaks down negative social capital and also enables positive social capital by improving access to information.

While there have been innovative attempts to quantity social capital direct measurement is difficult or impossible. The measurement challenge is to identify a contextually relevant indicator of social capital and to establish an empirical correlation with relevant benefit indicators.

There are a number of studies that used quantitative methods to measure social capital. Knack and Keefer (1997) use indicators of trust and civic norms from the World Values Survey for a sample of 29 market economies. They use these measures as proxies for the strength of civic associations in order to test two different propositions on the effects of social capital on economic growth, the 'Olson effects' (associations stifle growth through rent-seeking) and 'Putnam effects' (associations facilitate growth by increasing trust). Narayan and Pritchett (1997) construct a measure of social capital in rural Tanzania, using data from the Tanzania Social Capital and Poverty Survey (SCPS). This large-scale survey asked individuals about the extent and characteristics of their associational activity, and their trust in various institutions and individuals and linked social capital to household income. Recently, Kuenzi (2008) has used survey data from the Afrobarometer for Ghana and Nigeria to explore the variables that influence social capital and political trust.

Putnam (1993), in his research comparing north and south Italy, examines social capital in terms of the degree of civic involvement, as measured by voter turnout, newspaper readership, membership in choral societies and football clubs, and confidence in public institutions.

[18] Definition used by the World Bank. See http://go.worldbank.org/K4LUMW43B0

Qualitative research on social capital ranges from historical records, key informant interviews, focus groups discussion, ethnographic investigations and we only want to highlight a few examples of the wide range of available literature on industrialised and developing countries. Heller (1996) examines the case of the south Indian state of Kerala, where literacy rates, longevity, and infant mortality rates have long been among the most favourable. Tracing the history of state-society relations in Kerala, Heller shows how the state has played a crucial role in bringing about these results, by creating the conditions that enabled subordinate social groups to organize in their collective interest.

Given that it is most frequently defined in terms of groups, networks, norms of reciprocity, cooperation, and trust, research on social capital must be able to capture this multi-dimensionality. In order to make use of social capital findings to improve development processes and outcomes, it is also necessary to understand the dynamic nature of interpersonal and group relations in the context in which it is being studied. As such, social capital readily lends itself to a mixed-methods research approach (Jones and Woolcock 2007).

Table A.1 gives examples of open and closed questions—lending themselves to qualitative and quantitative research respectively—to assess different dimensions of social capital using the five dimensions suggested by the World Bank to operationally apply the concept.

Table A.1 Examples of open and closed questions to research social capital along five dimensions

Dimension of social capital	Closed questions	Open questions
1. Groups and networks	If you suddenly needed a small amount of money, how many people beyond your immediate household could you turn to? About how many close friends do you have these days? These are people you feel at ease with, can talk to about private matters, or call on for help.	Are informal groups based on occasions (e.g., weddings, births, or deaths)? What other triggers bring members of a group together? What is exchanged (e.g., goods, services, favours, information, goods, moral support, etc.) in community groups or networks? Who plays a leadership or mobilising role in the groups or networks? How are they selected?
2. Trust and solidarity	If you were caring for a child and needed to go out for a while, would you ask a neighbour for help? Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?	How long have people in a given neighbourhood or community lived together? How well do they know one another? On whom do people rely for different kinds of assistance (e.g., goods, labour, cash, finding employment, entering university, etc.)? Do patterns of mistrust and suspicion exist between households or among groups?
3. Collective action and cooperation	Do you help out a local group as a volunteer? Have you attended a local community event in the past 6 months (e.g. church fete, school concert, craft exhibition)?	Describe recent examples of collective action that have taken place in the community (or a segment of the community). What was the course and outcome of these activities? Who initiated the activities? How were people mobilised? Do social, cultural, or legal constraints limit the participation of specific groups (e.g., women, young people, poor people, minorities, etc.)?
4. Social cohesion and inclusion	Are there any services where you or members of your household are occasionally denied service or have only limited opportunity to use? There are often differences in characteristics between people living in the same village/neighbourhood. For example, differences in wealth, income, social status, ethnic or linguistic background/ race/ caste/ tribe. There can also be differences in religious or political beliefs, or there can be differences due to age or sex. To what extent do any such differences characterise your village/neighbourhood? Use a five point scale where 1 means to a very great extent and 5 means to a very small extent.	Are there recurring disagreements in networks and groups, or even demonstrated conflict? How often do people from different social groups intermarry? What prevents public services and expenditures from reaching the poorest and most vulnerable groups? Are the reasons related to ethnicity, gender, a political agenda, or geographic isolation?

5. Information and communication	In the past month, how many times have you made or received a phone call?	What information is available through different networks? To different households and/or groups?
	What are your three main sources of information about what the government is doing (such as agricultural extension, workfare, family planning, etc.). Options can include relatives, neighbours, friends, NGOs, work etc.	What are the preferred local sources and channels of information?

Source: Authors (based on Jones and Woolcock 2007)

A.4 Social exclusion

The concept of social exclusion originated in France and spread quickly to other parts of Europe during the 1980s in a climate of political social and economic crisis. Only more recently the term has been applied to developing countries and much of the initial literature discussed if and how you can apply the concept to the South where poverty is often a mass phenomenon.[19] People may be born into an excluded group (in the case of ethnicity, caste, or being born with a disability), or may become excluded either due to changes in circumstances (such as migration, acute illness, or disaster) or to chronic processes (such as long-term unemployment). People may also be excluded because of where they live (for example, people who live in remote areas or slums). But social exclusion is a concept that does go beyond the definition of excluded groups. De Haan (1999) discusses social exclusion as framework for understanding deprivation with its value lying in focusing on two central elements of deprivation: its multidimensionality and the processes and social relations that underlie it. Sen (2000) has placed social exclusion in the general perspective of capability failure with an emphasis on the relational features in the deprivation of capability. He further notes that being excluded can be in itself a deprivation (e.g. not being able to take part in the life of the community is a loss on its own) or have instrumental importance (e.g. landlessness can lead to economic and social deprivations).

DFID is currently drafting a guidance note on how to undertake a Gender and Social Exclusion Analysis (see Box A.2). In preparing the GSEA, the country offices are expected to draw on a range of data sources—qualitative and quantitative; international and local; expert and citizen-led; aggregate and specific. The guidance for the GSEA highlights that both, qualitative and quantitative data sources will be needed to understand the causes and effects of gender and social exclusion.

While some of the earlier research on social exclusion in developing countries has focused on the concept and some qualitative assessments, more recent research has attempted to measure social exclusion, often through a combination of qualitative and quantitative research methods, some of which we would like to highlight in the following.

Gacitua-Mario and Woden (2001) summarise the experience of combining qualitative and quantitative methods from three case studies in Latin America. The social exclusion study in urban Uruguay complements quantitative analysis with interviews and focus groups to obtain a better understanding of key vulnerable groups exposed to processes of exclusion. Quantitative methods provide the basis for the selection of the areas where the qualitative study is done and for the identification of specific groups of interest. The qualitative analysis

[19] For example see the discussion in Kabeer 2000.

explores the dimensions of exclusion more in-depth and collects information on existing perceptions of exclusion in various settings, including labour markets.

Carraro et al (2005) combines quantitative and qualitative methods to measure social exclusion in Pakistan. Based on a review of secondary data, including the Pakistan Participatory Poverty Assessment, key variables relevant to social exclusion in Pakistan were defined: religion, language, gender, disability, access to land/occupation and zaat/kinship/caste-like structures. Subsequently, quantitative survey data was used to statistically describe and econometrically model the level of deprivation of the groups indentified through the literature review. The analysis included MDG indicators for socially excluded groups, the impact of multiple exclusions (e.g. gender and language or gender and access to land) and the evolution of inequality over time. [20]

Box A.2 DFID's Gender and Social Exclusion Analysis (GSEA)

Following from DFID's Social Exclusion Policy (2005) the Gender and Social Exclusion workstream, which has been merged from separate departments on gender and social exclusion, is currently drafting a practice note for its country offices how to undertake a GSEA as part of the country planning process and to inform programme design.

The GSEA covers (1) who is excluded; (2) the processes that cause and perpetuate gender inequality and social exclusion, and; (3) the impacts of social exclusion on poverty reduction and the MDGs, economic growth and conflict.

In line with the five key questions in the Gender Equality Action Plan, the GSEA asks:

- Have we counted all excluded groups, women and men?
- Have excluded groups, women and men been consulted?
- Have we invested equally in excluded groups?
- Do excluded groups, women and men have a fair share?
- Do staff have the skills, knowledge and commitment to make a lasting change?

Source: Draft DFID (2008b)

Smith et al (2005) assess the potential for developing an 'index of multiple deprivation' in Bangladesh. The Index of Multiple Deprivation (IMD) is used across the United Kingdom and was applied to South Africa and refers to a standard set of indicators measuring deprivation at the local level. In the Bangladesh scoping study multiple deprivation refers to the combination of economic poverty, poor health, poor education, inadequate housing/shelter, high risk environment and high levels of crime/social disorder in particular areas. The study concludes that developing a set of measures or indices of multiple deprivation related to the MDGs would be highly desirable and feasible in Bangladesh.

[20] A similar analysis was undertaken by Bennett (2005) in form of the Nepal Gender and Social Exclusion Evaluation (GSEA), a collaborative policy research initiative of the World Bank and DFID.