Impact Assessment Methodologies for Microfinance: Theory, Experience and Better Practice

DAVID HULME *
University of Manchester, Manchester, UK

Summary. — Microfinance programs and institutions are increasingly important in development strategies but knowledge about their impacts is partial and contested. This paper reviews the methodological options for the impact assessment (IA) of microfinance. Following a discussion of the varying objectives of IA it examines the choice of conceptual frameworks and presents three paradigms of impact assessment: the scientific method, the humanities tradition and participatory learning and action (PLA). Key issues and lessons in the practice of microfinance IAs are then explored and it is argued that the central issue in IA design is how to combine different methodological approaches so that a "fit" is achieved between IA objectives, program context and the constraints of IA costs, human resources and timing. The conclusion argues for a greater focus on internal impact monitoring by microfinance institutions. © 2000 Elsevier Science Ltd. All rights reserved.

Key words — methods, microfinance, credit, impact assessment, monitoring and evaluation, poverty reduction

1. INTRODUCTION

In recent years impact assessment has become an increasingly important aspect of development activity as agencies, and particularly aid donors, have sought to ensure that funds are well spent. As microfinance programs and institutions have become an important component of strategies to reduce poverty or promote micro and small enterprise development then the spotlight has begun to focus on them. But knowledge about the achievements of such initiatives remains only partial and is contested. At one end of the spectrum are studies arguing that microfinance has very beneficial economic and social impacts (Holcombe, 1995; Hossain, 1988; Khandker, 1998; Otero & Rhyne, 1994; Remenyi, 1991; Schuler, Hashemi & Riley, 1997). At the other are writers who caution against such optimism and point to the negative impacts that microfinance can have (Adams & von Pischke, 1992; Buckley, 1997; Montgomery, 1996; Rogaly, 1996; Wood & Sharrif, 1997). In the "middle" is work that identifies beneficial impacts but argues that microfinance does not assist the poorest, as is so often claimed (Hulme & Mosley, 1996; Mosely & Hulme, 1998).

Given this state of affairs the assessment of microfinance programs remains an important

field for researchers, policy-makers and development practitioners. [1] This paper reviews the methodological options for assessing the impacts of such programs drawing on writings on microfinance and the broader literature on evaluation and impact assessment. Subsequently it explores ways in which impact assessment practice might be improved. It views impact assessment (IA) as being "...as much an art as a science..." (a phrase lifted from Little, 1997, p. 2). Enhancing the contribution that impact assessment can make to developmental goals requires both better

science and better art. The scientific improvements relate to improving standards of measurement, sampling and analytical technique. Econometricians and statisticians are particularly concerned with this field. Improving the ''art'' of impact assessment has at least three strands. One concerns making more systematic and informed judgements about the overall design of IAs in relation to their costs, specific objectives and contexts. The second is about what mixes of impact assessment methods are most appropriate for any given study. The third relates to increasing our understanding of the ways in which the results of IA studies influence policymakers and microfinance institution (MFI) managers.

2. IMPACT ASSESSMENT: OBJECTIVES

Impact assessment studies have become increasingly popular with donor agencies and, in consequence, have become an increasingly significant activity for recipient agencies. In part this reflects a cosmetic change, with the term IA simply being substituted for evaluation. But it has also been associated with a greater focus on the outcomes of interventions, rather than inputs and outputs. While the goals of IA studies commonly incorporate both ''proving'' impacts and ''improving'' interventions, IAs are more likely to prioritize the proving goal than did the evaluations of the 1980s. A set of factors are associated with the extreme ''pole'' positions of this continuum and these underpin many of the issues that must be resolved (and personal and institutional tensions that arise) when impact assessments are being initiated (Figure 1).

Behind the shift from ''evaluation'' to ''IA'' are a number of factors. These are not explored in any detail in this paper but they form an essential element for the understanding of IA and its potential contributions. Explicitly, IAs are promoted by both the sponsors and implementers of programs so that they can learn what is being achieved and improve the effectiveness and efficiency of their activities. Implicitly, IAs are a method by which sponsors seek to get more information about program effectiveness than is available from the routine accountability systems of implementing organizations. IAs are also of significance to aid agencies in terms of meeting the ever increasing accountability demands of their governments (in this era of ''results'' and ''value for money'') and for contesting the rhetoric of the anti-aid lobby. While recipient agencies benefit from this, they are one stage removed, and many are likely to see donor-initiated IA as an activity that has limited practical relevance for program activities. To quote the director of a large Asian microfinance institution that has received substantial amounts of aid financed IA consultancy and internal IA-capacity building ''...impact assessment studies keep donors happy... we don't use them very much.''

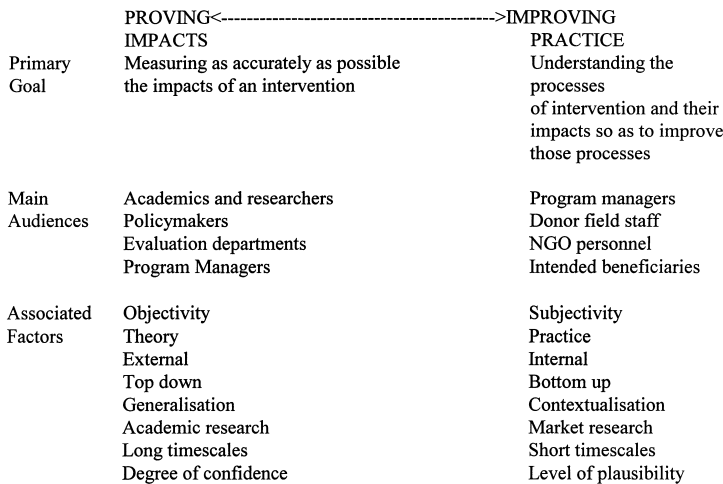|	PROVING<-->IMPROVING	
	IMPACTS	PRACTICE
Primary Goal	Measuring as accurately as possible the impacts of an intervention	Understanding the processes of intervention and their impacts so as to improve those processes
Main Audiences	Academics and researchers Policymakers Evaluation departments Program Managers	Program managers Donor field staff NGO personnel Intended beneficiaries
Associated Factors	Objectivity Theory External Top down Generalisation Academic research Long timescales Degree of confidence	Subjectivity Practice Internal Bottom up Contextualisation Market research Short timescales Level of plausibility

Figure 1. *The goals of impact assessment.*

A final issue to raise in this section is whether the expectations of OECD-based agencies about the feasibility of the accurate measurement of impacts in the difficult contexts of developing countries (limited numbers of professional researchers, few written records, illiteracy, communication problems etc.) are higher than in their own countries. My professional experience of EU-financed "small enterprise development" projects in Manchester has revealed a startling lack of concern with impacts: this is in marked contrast to my consultancy work in Bangladesh where donors criticize nongovernment organizations (NGOs) for failing to make impact assessment a priority! If recipients perceive that the IA standards expected of "them" are higher than donors expect of themselves then IA will be seen as an external imposition rather than a shared opportunity.

3. ASSESSING IMPACT: THE CHOICE OF CONCEPTUAL FRAMEWORKS

All impact assessment exercises have a conceptual framework at their heart. In well-planned and well-resourced IAs with long "lead-in" times such frameworks are usually explicitly identified (e.g., Khandker, 1998; Sebstad, Neill, Barnes & Chen, 1995; Schuler & Hashemi, 1994). By contrast, in many smaller scale exercises the framework is implicit and may be seen as "common sense." There are three main elements to a conceptual framework:

—a model of the impact chain that the study is to examine,
—the specification of the unit(s), or levels, at which impacts are assessed, and
—the specification of the types of impact that are to be assessed.

(a) Models of impact chains

Behind all microfinance programs, and indeed virtually all aid financed initiatives, [2] is the assumption that intervention will change human behaviors and practices in ways that lead to the achievement (or raise the probability of achievement) of desired outcomes. IAs assess the difference in the values of key variables between the outcomes on "agents" (individuals, enterprises, households, populations, policymakers, etc.) which have experienced an intervention against the values of those variables that would have occurred had there been no intervention. The fact that no agent can both experience an intervention and at the same time not experience an intervention generates many methodological problems. All changes are influenced by mediating processes (specific characteristics of the agent and of the economic, physical, social and political environment) that influence both behavioral changes and the outcomes in ways that are difficult to predict (Sebstad et al., 1995).

The impact chain is very simply depicted in Figure 2. A more detailed conceptualization would present a complex set of links as each "effect" becomes a "cause" in its own right generating further effects. For example, in a
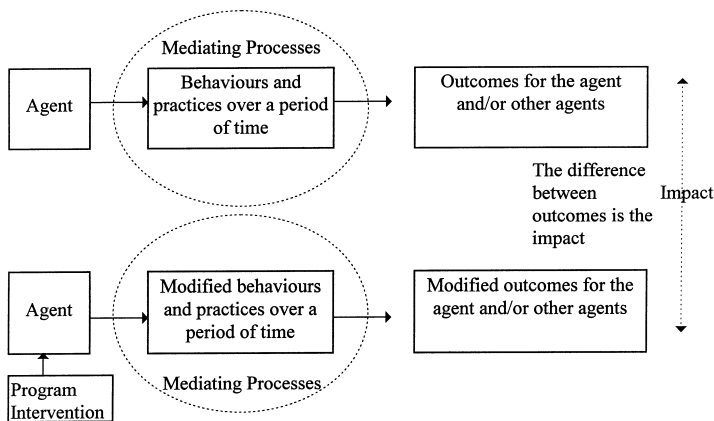

Figure 2. *The conventional model of the impact chain.*

conventional microfinance project a package of technical assistance and capital changes the behavior (and products) of a microfinance institution (MFI). The MFI subsequently provides different services to a client, most commonly in the form of a loan. These services lead to the client modifying her/his microenterprise activities which in turn leads to increased or decreased microenterprise income. The change in microenterprise income causes changes in household income which in turn leads to greater or lesser household economic security. The modified level of household economic security leads to changes in the morbidity and mortality of household members, in educational and skill levels and in future economic and social opportunities. Ultimately, perhaps, these changes lead to modifications in social and political relations and structures. The complexity of such chains provides the assessor with a range of choices about which link (or links) to focus on. For microfinance, it is useful to distinguish between two main schools of thought with regard to which link(s) in the chain to focus on. For convenience, these are termed the ''intended beneficiary'' [3] school and the ''intermediary'' school.

The intended beneficiary school, building on the ideas of conventional evaluation, seeks to get as far down the impact chain as is feasible (in terms of budgets and techniques) and to assess the impact on intended beneficiaries (individuals or households). The intermediary school focuses purely on the beginning of the chain and in particular on changes in the MFI and its operations. Its roots are closely associated with the Ohio State University School's analyses of rural finance. Generally, two key variables are focused on: institutional outreach and institutional sustainability [4] (Yaron, Benjamin & Piprek, 1997). If both outreach and sustainability have been enhanced then the intervention is judged to have a beneficial impact as it has widened the financial market in a sustainable fashion. This is based on the assumption that such institutional impacts extend the choices of people looking for credit and savings services and that this extension of choice ultimately leads to improved microenterprise performance and household economic security. While this assumption can be supported by theoretical frameworks (if a set of further assumptions are made about perfect competition and other factors) it is an assumption which has proved invalid in a number of experiences. [5] In addition, it will not reveal borrower ''crossfinanc-

ing'' of loans (Wiig, 1997) which may threaten the long-term viability of an MFI.

While the choice between these two schools can ultimately be seen as an ideological choice (does one prioritize contributions to improved welfare or to more efficient markets?) it is possible to recognize different strengths and weaknesses. The intended beneficiary school makes fewer assumptions about the impact chain and is better able to distinguish ''who'' benefits and ''how.'' It is, however, demanding in both methodological and cost terms. The intermediary school usefully incorporates notions of sustainability and provides an IA methodological framework that can be operated largely with pre-existing data. It is, though, very weak on ''who'' benefits and ''how'' (as illustrated by assessments of the USAID-financed APPLE program). [6] Possible ways of strengthening the intermediary school approach have been suggested by Feinstein (1997) through the analysis of borrower transaction costs. He proposes the collection of longitudinal data on borrowers transaction costs (p. 5) to assess whether an MFI has benefited borrowers, i.e., has reduced their total costs for accessing finance. This offers a potential ''bridge'' between the two main ''schools,'' if data on ''who'' borrowers are also collected.

(b) *Units of assessment*

Following on from the design of a model of the impact path comes the choice of the unit(s) of assessment (or levels of assessment). Common units of assessment are the household, the enterprise or the institutional environment within which agents operate. Occasionally studies have attempted to assess impact at an individual level (e.g., Goetz & Sen Gupta, 1996; Peace & Hulme, 1994), but this is relatively rare and has to take a qualitative focus. More recently some studies have attempted to assess impacts at a number of levels, such as Hulme and Mosley (1996) who looked at microenterprise, household, community and institutional levels and USAID's Assessing the Impact of Microenterprise (AIMS) Project. Through a household economic portfolio model (HEPM) the latter seeks to assess impacts at household, enterprise, individual and community levels and thus produce a fuller picture of overall impacts (Chen & Dunn, 1996).

The relative advantages and disadvantages of different units of assessment are summarized in

Table 1. *Units of assessment and their advantages and disadvantages*

Unit	Advantages	Disadvantages
Individual	—Easily defined and identified	—Most interventions have impacts beyond the individual —Difficulties of disaggregating group impacts and impacts on "relations"
Enterprise	—Availability of analytical tools (profitability, return on investment etc)	—Definition and identification is difficult in microenterprises —Much microfinance is used for other enterprises and/or consumption —Links between enterprise performance and livelihoods need careful validation
Household	—Relatively easily defined and identified —Permits an appreciation of livelihood impacts —Permits an appreciation of interlinkages of different enterprises and consumption	—Sometimes exact membership difficult to gauge —The assumption that what is good for a household in aggregate is good for all of its members individually is often invalid
Community	—Permits major externalities of interventions to be captured	—Quantitative data is difficult to gather —Definition of its boundary is arbitrary
Institutional impacts	—Availability of data —Availability of analytical tools (profitability, SDIs, transaction costs)	—How valid are inferences about the outcomes produced by institutional activity?
Household economic portfolio (i.e. household, enterprise, individual and community)	—Comprehensive coverage of impacts —Appreciation of linkages between different units	—Complexity —High costs —Demands sophisticated analytical skills —Time consuming

Table 1. As can be seen, a focus purely on the "individual" or the "enterprise" has such drawbacks that they could be viewed as discredited. The household economic portfolio model has much to recommend it—especially if institutional impacts are incorporated in the community level analysis. It does have the profound disadvantage, though, of making assessment demanding in terms of costs, skilled personnel and time. If used with limited resources it risks sacrificing depth for breadth of coverage of possible impacts.

(c) *Types of impact*

An almost infinite array of variables can be identified to assess impacts on different units. To be of use these must be able to be defined with precision and must be measurable.

Conventionally, economic indicators have dominated microfinance IAs with assessors particularly keen to measure changes in income despite the enormous problems this presents. Other popular variables have been levels and patterns of expenditure, consumption and assets. A strong case can be made that assets are a particularly useful indicator of impact because their level does not fluctuate as greatly as other economic indicators and is not simply based on an annual estimate (Barnes, 1996, p. v).

The social indicators that became popular in the early 1980s (e.g., educational status, access to health services, nutritional levels, anthropometric measures and contraceptive use) have recently been extended into the socio-political arena in an attempt to assess whether microfinance can promote empowerment (Mayoux, 1997; Goetz & Sen Gupta, 1996; Schuler &

Hashemi, 1994; Hashemi, Schuler & Riley, 1996). This has led to the measurement of individual control over resources, involvement in household and community decision-making, levels of participation in community activities and social networks and electoral participation. The bulk of this work has focused on changes in gender relations, but there are sometimes partially-formulated assessments of class relations within it (Fuglesang & Chandler, 1993). These extensions to the types of impact assessed permit IAs to be more sophisticated and to shed light on developmental impacts at a time when the goals of development have also been extended. They do add, however, to the complexity of IA work and require the skills of assessors who are experienced at making judgements on social relations.

Sebstad *et al.* (1995) usefully distinguish between "domains of change" (e.g., household income) and the specific "markers of change" (e.g., amount of income, number of income sources and seasonality of income) within each domain. While not comprehensive, the detailed sets of domains and markers, produced in their paper provide an excellent checklist for impact assessors to consider at the IA design stage. [7] Often the exact markers used will be shaped by the methodology that is selected. This can cause problems for multi-method IAs which may not be able to apply a single definition for a marker for each of the methods used. In addition, impact assessors should always seek to keep the number of variables they measure to a manageable number and not be tempted to go for a comprehensive approach that will impact adversely on data quality and study relevance.

4. THE THREE PARADIGMS OF IMPACT ASSESSMENT: PROBLEMS OF ATTRIBUTION AND FUNGIBILITY

The major methodological problems that confront the IA of microfinance relate to attribution and fungibility. At the heart of impact assessment is the attribution of specific effects (i.e., impacts) to specific causes (i.e., interventions). From the vast literature on microfinance IA it is possible to draw out three very different paradigms by which authors seek to demonstrate attribution. The first is the conventional scientific method with its origins in the natural sciences. The second has its roots in the humanities and focuses on making a

reasoned argument supported by theory and specific pieces of evidence. Although the former has tended to dominate discussions about microfinance impact assessment (see, for example, the studies reviewed by Gaile & Foster, 1996) the latter tradition is being increasingly used by MFIs and researchers (Bouman & Hospes, 1994; Ardener & Burman, 1995; Remenyi, 1991; Rutherford, 1999). The third part of this section explores a recent entrant to the field—participatory learning and action (PLA)—which offers a radical challenge to both conventional IA and to "science" itself. Although these three approaches can be separated for analytical purposes, in recent practice many studies have woven elements of these approaches together (see Section 5 for a discussion).

(a) Scientific method

Scientific method seeks to ensure that effects can be attributed to causes through experimentation. [8] A particular stimulus to a particular object in a rigorously controlled environment is judged to be the cause of the observed effect. The experimental approach is virtually infeasible in the social sciences, because of the nature of the subject matter, and so the approach has been adapted into quasi-experiments (Casely & Lury, 1982). Quasi-experiments seek to compare the outcomes of an intervention with a simulation of what the outcomes would have been, had their been no intervention. One method for this is multiple regression, but this has rarely been used in microfinance IA because of its enormous demands for data on other possible causal factors and its assumptions (Mosley, 1997, pp. 2–3). A second approach is the *control group* method which has been widely used. This requires a before and after comparison of a population that received a specific treatment [9] (i.e., a microfinance program) and an identical population (or as near as possible) that did not receive the treatment. While this idea is elegantly simple a number of "elephant-traps" may befall its user. In particular problems of sample selection bias, misspecification of underlying causal relationships and respondent motivation (see later) must be overcome.

Selection bias may occur because of:
(i) difficulties in finding a location at which the control group's economic, physical and social environment matches that of the treatment group,

(ii) the treatment group systematically possessing an "invisible" attribute which the control group lacks (most commonly identified as entrepreneurial drive and ability),

(iii) receiving any form of intervention may result in a short-term positive response from the treatment group (the Hawthorne effect),

(iv) the control group becoming contaminated by contact with the treatment group (though this could be a long-term program goal!), and

(v) the fungibility of the treatment (e.g., when a loan is transferred from a borrower to someone else or when the loan is not used in the planned way).

Problems (i) and (iv) can be tackled by more careful selection of the control group. This applies particularly to controlling for access to infrastructure (which has a key influence on input and output prices as well as other variables) and ensuring that the control group is located far away from the treatment group. Problems (ii) and (iii) are more intractable, but in many cases they can be tackled by using program-accepted "clients-to-be," who have not yet received microfinance services, as the control group (Hulme & Mosley, 1996, chapter 4). It must be noted, however, that this approach will not be valid when the take up of microfinance services is based on diffusion through a heterogeneous population. [10]

This leaves the problem of *loan fungibility*. This can be seen as an intractable problem as "...no study has successfully controlled for the fungibility of resources between the household and the assisted enterprise" (Gaile & Foster, 1996, p. 24). Using case study materials to crosscheck actual loan use against intended loan use and thus estimating "leakage" is one possible approach to controlling for fungibility (Pulley, 1989; Mosley, 1997). But for all studies except those that focus exclusively on "the enterprise," then a concern about fungibility may be irrelevant. For studies looking at the household, the community or the household economic portfolio (see Section 3b) fungibility is not a problem for the assessor, rather it is a vital strategy for the client. The best investment returns may be on "consumption" (in terms of developing or maintaining human capital through school fees and doctors' bills, or buying food at a time of crisis when the credit terms on "in-kind" borrowing from traders may be exceptionally high). From this perspective the task of the assessor is not to pretend that microenterprises are "firms" whose inputs and outputs can be precisely identified and measured but to recognize that the impacts of microfinance must be assessed at a variety of levels. The assessor attempting to control for fungibility (to prove impact) has failed to recognize that fungibility is a process to be encouraged (to improve impact)!

The misspecification of underlying causal relationships arises most commonly because of the assumption that causality is a one-way process (Figure 2). This may be a reasonable assumption in the physical sciences (though it does not go unchallenged by contemporary philosophers of science). For human activity it is commonly invalid, as causation may also run from impact back to intervention. Mosley (1997, p. 6) illustrates this with the example of a program whose field staff put pressure on a borrower to repay her loan; this may succeed in the short term but may induce the borrower to sell assets (machinery, land, trees) which reduce the probability of repayment in the longer term. Such reverse causation need not necessarily be negative and, from the perspective of more process-oriented analytical frameworks, is essential if programs are to continually learn from their experience and improve (rather than prove) their impact.

Such problems can be overcome by the adoption of models that conceptualize causation as a two-way process by the use of two-stage least squares technique and regression analysis (Mosley, 1997, p. 7). Such an approach is enormously demanding in terms of data requirements, technical expertise and costs. It will only be feasible on very rare occasions (for example, see Khandker, 1998). For most researchers adopting the scientific method, reverse causality is a problem to be coped with rather than overcome. The main means of dealing with it are tracing dropouts from both the treated and control groups; only conducting IAs on relatively mature programs; interim impact monitoring activities to gather qualitative information about the complexity of causality; and retrospective in-depth interviews with clients (Mosley, 1997, p. 6).

(b) *The humanities tradition* [11]

The broad set of approaches that fall under this heading have their roots in the humanities. Originally geography and rural sociology were the "lead" subjects, but over the last 20 years

anthropology has become most important. Its main features are an inductive approach, a focus on key informants, recording by notes or image, and the data analyst is usually directly (and heavily) involved in data collection. [12] This tradition does not try to "prove" impact within statistically definable limits of probability. Rather, it seeks to provide an interpretation of the processes involved in intervention and of the impacts that have a high level of plausibility. It recognizes that there are usually different, and often conflicting, accounts of what has happened and what has been achieved by a program. The validity of specific IAs adopting this approach has to be judged by the reader on the basis of the logical consistency of the arguments and materials presented; the strength and quality of the evidence provided; the degree of triangulation used to crosscheck evidence; the quality of the methodology; and the reputation of the researcher(s). Whether "standards" could be specified for such work—to help its users appreciate how rigorously designed they are—is an important issue that merits attention. [13] Commonly the bulk of data generated by such an approach is "qualitative," although at later stages of analysis such work often quantifies some data. The main types of methods used have been discussed in Section 4 (in particular see Table 2).

Although such work has been common in development studies for decades, it is only during the 1980s that its relevance for IA has been recognized. [14] This recognition has arisen partly because of the potential contribution of qualitative approaches (especially in understanding changes in social relations, the nature of program staff-beneficiary relations and fungibility) and partly because of the widespread recognition that much IA survey work was based on inaccurate information collected by questionnaire from biased samples (Chambers, 1993). Low-budget and low-rigor IAs claiming to adopt the scientific method were at best pseudo-science, but more often simply bad science, despite the sophisticated analytical tools that were applied to poor datasets.

IAs with their roots in the humanities have considerable difficulties with regard to the attribution of cause and effect. Such studies cannot usually demonstrate the causal link as they are not able to generate a "without program" control group (although at times some researchers neglect to mention this to the reader and simply assume causality). Instead, causality is inferred from the information about the causal chain collected from intended beneficiaries and key informants, and by comparisons with data from secondary sources about changes in out-of-program areas. Problems also arise because not infrequently the labels "rapid appraisal," "mini-survey" and "case study" are applied to work which has been done in an *ad hoc* manner and does not achieve a minimum professional standard in terms of informant selection and the rigor of data collection and analysis. Examples of this include: basing data collection only in program areas that are performing well, and surveying best clients; and inferring that the data collected in one area apply to all clients without explaining this assumption.

While such studies cannot provide the degree of confidence in their conclusions that a fully

Table 2. *Common impact assessment methods*

Method	Key features
Sample surveys	Collect quantifiable data through questionnaires. Usually a random sample and a matched control group are used to measure predetermined indicators before and after intervention
Rapid appraisal	A range of tools and techniques developed originally as rapid rural appraisal (RRA). It involves the use of focus groups, semi-structured interview with key informants, case studies, participant observation and secondary sources
Participant observation	Extended residence in a program community by field researchers using qualitative techniques and mini-scale sample surveys
Case studies	Detailed studies of a specific unit (a group, locality, organisation) involving open-ended questioning and the preparation of "histories"
Participatory learning and action	The preparation by the intended beneficiaries of a program of timelines, impact flow charts, village and resource maps, well-being and wealth ranking, seasonal diagrams, problem ranking and institutional assessments through group processes assisted by a facilitator

resourced scientific method approach can yield, my personal judgement is that in many cases their conclusions are more valid than survey based IA work that masquerades as science but has not collected data with scientific rigor. It is nonetheless becoming increasingly common to combine "scientific" and "humanities" approaches so as to check the validity of information and provide added confidence in the findings (e.g., Hashemi *et al.*, 1996; Hulme & Mosley, 1996; Schuler & Hashemi, 1994). In the future, dealing with attribution by multimethod approaches seems the way forward.

(c) *Participatory learning and action (PLA)*

In the last five years participatory approaches to development planning and management have moved from being a fringe activity to center stage. While many donor agencies have simply added a bit of PLA to their existing procedures, it can be argued that this is inappropriate as conceptually participatory approaches challenge the validity and utility of the scientific method as applied to developmental problems (Chambers, 1997). According to this line of argument the scientific method fails as: it ignores the complexity, diversity and contingency of winning a livelihood; it reduces causality to simple unidirectional chains, rather than complex webs; it measures the irrelevant or pretends to measure the immeasurable; and, it empowers professionals, policy-makers and elites, thus reinforcing the *status quo* and directly retarding the achievement of development goals. At heart, PLA theorists do not believe that ultimately there is one objective reality that must be understood. Rather, there are multiple realities and before any analysis or action is taken the individuals concerned must ask themselves, "whose reality counts?" (Chambers, 1997). Their answer is that the perceived reality of the poor must take pride of place as, if development is about "empowering the poor" or "empowering women" (as virtually all development agencies now say), then the first step toward empowerment is ensuring that "the poor" or "women" take the lead in problem identification and analysis and knowledge creation. [15]

For impact assessment the purist PLA line is damning—"...conventional baseline surveys are virtually useless for impact assessments... The question now is how widely local people can be enabled to identify their own indicators, establish their own participatory baselines, monitor change, and evaluate causality..." (Mayoux, 1997, p. 123). By this means two objectives may be achieved: better impact assessments, and intended beneficiaries will be "...empower[ed] through the research process itself" (Mayoux, 1997, p. 2). In practice, the art of participatory impact assessment (PIA) is in its infancy and a pragmatic rather than a purist approach has been common. Agencies such as Proshika in Bangladesh have begun to use PLA methods extensively for their assessment and planning exercises. [16]

The reliability of participatory methods varies enormously, as with "scientific" surveys, depending "...largely on the motivation and skills of facilitators and those investigated and the ways in which informants" perceptions of the consequences of research are addressed" (Mayoux, 1997, pp. 12–13). Nevertheless, it is argued that "...a number of rigorous comparative studies have shown that, when well-conducted, participatory methods can be more reliable than conventional surveys" (Mayoux, 1997 and also Chambers, 1997, pp. 141–146).

To date the literature on PLA and PIA has only partially addressed the issue of attribution. From a scientific perspective PIA has grave problems because of the subjectivity of its conceptualizations of impact; the subjectivity of the data used to assess impact; the variables and measures used vary from case to case and do not permit comparison; its pluralist approach may lead to a number of mutually conflicting accounts being generated about causality; and, the assumption that because lots of people are taking part in an exercise means that all are able to "voice" their concerns (so that opinions are representative) is naive about the nature of local power relations. From the perspective of a "new professional" (Chambers, 1997) then such a set of accounts is unproblematic, as it reflects the complexity and contingency of causality in the real world. In addition, it can be argued that PIA contributes to program goals (perhaps particularly in terms of empowering women (Mayoux, 1997) and the poor) by not facilitating the continued dominance of target groups by powerful outsiders. Why dwell on issues of attribution when efforts to overcome such problems require the adoption of IA methods that actively undermine the attainment of program goals?

Table 3. *Comparative strengths and weaknesses of different methods*[a]

Method criteria	Surveys	Rapid appraisal	Participant observation	Case studies	Participatory learning and action
1. Coverage (scale of applicability)	High	Medium	Low	Low	Medium
2. Representativeness	High	Medium	Low	Low	Medium
3. Ease of data standardisation, aggregation and synthesis (e.g., quantification)	High	Medium	Medium or Low	Low	Medium or low
4. Ability to isolate and measure nonproject causes of change	High	Low	Low	Low	Low
5. Ability to cope with the attribution problem	High	Medium	Medium	Medium	Medium
6. Ability to capture qualitative information	Low	High	High	High	High
7. Ability to capture causal processes	Low	High	High	Medium	High
8. Ability to understand complex processes (e.g., institution building)	Minimal	Medium	High	Medium	Medium
9. Ability to capture diversity of perceptions	Low	High	Very high	Medium	High
10. Ability to elicitviews of women and disadvantaged groups	Low	Medium	High	High (if targeted)	Medium
11. Ability to capture unexpected or negative impacts	Low	High	Very high	High	High
12. Ability to identify and articulate felt needs	Low	High	High	Medium (due to low coverage)	High
13. Degree of participation encouraged by method	Low	High	Medium	Medium	Very high
14. Potential to contribute to stakeholder capacity building	Low	High	Low	Medium to low	Very high
15. Probability of enhancing downwards accountability	Low	High	Medium	Medium	High
16. Human resource requirements	Specialist supervision, large numbers of less qualified field workers	High skilled practitioners, who are able to write-up and analyse results	Medium skilled practitioners, with good supervision, who are prepared to commit for lengthy period	Medium skilled practitioners with good supervision	High skilled ctitioners
17. Cost range	Very high to medium	High to medium	Medium to low	Medium to low	High to Medium
18. Timescale	Very high to medium	Medium to low	High	High to medium	Medium to low

[a] Source: Adapted from Montgomery *et al.* (1996).

5. THE PRACTICE OF MICROFINANCE IMPACT ASSESSMENT

(a) Knowledge creation: the methodological menu [17]

Over the last decade microfinance impact assessment studies have increasingly moved away from single method approaches (e.g., Hossain, 1988; Fuglesang & Chandler, 1986) to multimethod or pluralist approaches (e.g., Hulme & Mosley, 1996; Mustafa et al., 1996). The introduction of participatory approaches to impact assessment has extended the methodological menu for data collection and knowledge creation. While sample surveys remain a common model, rapid appraisal, participant-observation and participatory learning and action are increasingly used (Table 2). Each of these methods has a different pattern of strengths and weaknesses (Table 3) and this has led to a growing consensus among impact assessors that the central methodological question is no longer "what is the optimal method for this study?," but "what mix of methods is most appropriate for this study and how should they be combined?" Depending on the level of resources available and the context impact studies increasingly seek to combine the strengths of different approaches and, in particular, seek to combine the advantages of sample survey and statistical approaches (representativeness, quantification and attribution) with the advantages of humanities or participatory approaches (ability to uncover processes, capture the diversity of perceptions, views of minorities, unexpected impacts etc.). In well-resourced studies with long time scales (e.g., Mustafa et al., 1996) all of these different methods may be utilized in a comprehensive fashion. In cases where a high degree of statistical confidence is required (for example, when it is desired to "prove" impact for policy or major investment purposes) then a large-scale, longitudinal sample survey must be mounted, preferably supported and triangulated by the use of their methods on a limited scale. By contrast, if an IA is required to provide independent corroboration of the impact of a small-scale program and strengthen aspects of its implementation then a mix of rapid appraisal and small scale survey is likely to be appropriate. The Appendix A provides a summary of the conditions under which different methods are and are not appropriate. [18]

(b) Costs and confidence

The design of an IA must be very closely related to the budget available: this may be a platitude but over ambitious designs continue to lead to poor quality studies or delays that make findings irrelevant. Interestingly, in this age of cost consciousness, the literature on microfinance provides no specific information on the overall or unit costs of IA studies of microfinance; "high," "medium" and "low" are about as good as the data get! [19]

From verbal reports it is clear that IAs adopting the scientific method and seeking to "prove" impact cost the earth (probably US$500,000 to US$5 million depending on the number of MFIs studied). At the other extreme high quality, rapid appraisals of the impact of individual schemes by gifted and knowledgeable individuals can produce useful findings on "improvement" for relatively small sums (around US$5,000 to US$10,000). Between these two extremes are a vast array of different options. A reading of the contemporary literature produces the following findings.

(i) Studies intended to produce authoritative evidence of microfinance impact using the scientific method will be rare exceptions in the IA field. Their costs are so great that few agencies can fund them and their time-scales so long that the agencies studied are likely to treat them as "historical" rather than being of operational relevance.

(ii) The idea that "qualitative" and "participatory" assessments methodologies are cheap needs to be challenged (Mayoux, 1997). While such approaches are much cheaper than large-scale surveys, rigorous qualitative IAs will require the use of high calibre staff who are given time to prepare properly. Costs of tens of thousands of dollars, rather than thousands, should be anticipated.

(iii) For studies of moderate budget (i.e. most studies) the best approach to ensuring the validity of findings will be through triangulation and using a mix of survey, qualitative and participatory techniques. The alternative, of trying to achieve a representative sample size on a limited budget, is likely to lead to severe losses in the quality of data and/or the representativeness of the sample.

(iv) Limited investments in project monitoring by program staff make moderate cost impact assessment at high levels of quality

much more feasible as less primary data collection is necessary (see Montgomery, 1996 and later parts of this paper).

(c) *Human resources for impact assessment*

In many, if not most, developing countries recruiting IA personnel who have the skills and qualities to interview, collate, analyze and write up findings is a key problem at both consultant and fieldworker levels. Commonly, different studies find themselves competing for the same small pool of people which, while it may usefully raise payments for scarce skills, puts these individuals under great strain and does not appear to stimulate a "supply-side response". It is beyond the capacity of this paper to explore this issue, but it must be recognized as a key constraint and that efforts to build "impact assessment" capacities professionally and institutionally should be a priority for development agencies if they intend to continue to emphasize the need for IA.

(d) *Respondents: motivation and representation*

A "rational actor" confronted by an impact assessor asking standard IA questions ("what is your income? how do you spend your money on? how do you get on with your husband?") would soon tell the interviewer where to put his/her survey instrument. Fortunately, in the world of practice, more polite responses are the norm but the issue of how to persuade respondents to spare the time for an interview, and provide accurate and honest answers, is an important one that is rarely mentioned in IA methodological statements. Different strategies are needed for different types of respondent— program beneficiary, control group and program drop out. As a rule of thumb many researchers suggest that interviews should be concluded within one hour and that one and a half hours should be seen as the absolute maximum for an interview.

Beneficiaries are the easiest group to approach as generally they accept "answering questions" as one of the unavoidable transaction costs of being in a program or dealing with an MFI. Motivation can be enhanced by having interviewers introduced by program officers: but, this has the danger of linking the assessor with field-level staff and encouraging the recounting of "the right answers". [20] For both data quality and ethical reasons the personal introductions that interviewers make prior to interview need to be carefully worked out so that respondents understand why they are being interviewed and have an opportunity to ask their own questions before the interview begins.

Motivation is a more difficult issue with control groups as, having by definition no connection with a program, they have no incentive to cooperate. In many cases, however, the novelty and amusement value of being interviewed is sufficient encouragement (though *expatriates* should note that when they are working at a field site the willingness of people to be interviewed may be higher than is the norm because of the rarity value of foreigners). The problems of response increase significantly if longitudinal data are collected, as second and third interviews have much less amusement value. In such cases rewarding interviewees should be considered to promote data quality and for ethical reasons (what right have impact assessors to assume that the opportunity costs of an interview, particularly for poor people, are zero?). This can take the form of a social reward, such as bringing soda water and snacks to share with respondents (this works well in East Africa), or "bribery" (Mosley, 1997, p. 8) where the interviewee is paid cash for surrendering her/his time. [21]

Program dropouts represent a particular problem, and a failure to pursue dropouts may have led to some IAs underestimating the negative impacts of microfinance (e.g., Hulme & Mosley, 1996). When the dropout is traceable then significant effort is merited to obtain an interview/re-interview. Where dropouts cannot be traced, or death has occurred, then a replacement respondent sampled at random from the original population, and preferably from the same stratum, should be interviewed (see Mosley, 1997, pp. 7–8).

Participatory and rapid appraisal methods that work with groups generally manage to muster respondents because of the social interaction they create. Care needs to be taken, however, to observe who has turned up and, perhaps more significantly, who has not come to the meeting (Mayoux, 1997). The assumption that participants in a PLA exercise represent "the community" will commonly not be valid (Mosse, 1994). Additional interviews or focus groups may be necessary to collect information from people who do not turn up for communal PLA or RRA sessions.

(e) *The problem of "low impact" impact assessments* [22]

A final problem of IA concerns the impact of IAs on policy and practice. This depends in part on the original objectives of a study. It applies to both "proving" and "improving" IAs. The evaluation literature of the 1980s bemoans the limited influence of evaluation on subsequent decision-making. IA has inherited this problem, as illustrated by the very limited influence of large-scale impact assessment studies (Mustafa *et al.*, 1996) on the microfinance activities of the Bangladesh Rural Advancement Committee (BRAC) and the long time lag between the World Bank's excellent studies of MFIs in Bangladesh in the early 1990s and the dissemination of findings (Khandker, 1998) to policy-makers in Europe and MFI managers in Bangladesh.

A number of ways of ameliorating this problem can be identified.

(i) Impact assessors need to devote more time to the "use" of their studies (and perhaps a little less time to the product itself!). Their focus must go beyond "the report" into a dissemination strategy aimed at decision-makers: bullet point summaries, short user-friendly papers, snappy presentations and strategic cups of coffee are the key to this environment.

(ii) The timing of findings needs to be carefully considered. As a general rule of thumb the longer the length of time between data collection and findings presentation, then the lower the impact for IAs focused on "improving" practice. The common response to initial findings presented more than nine months after completion of fieldwork is "our program has already been redesigned so your findings have little relevance."

(iii) Program managers often regard impact assessors as impractical people who have lots of time on their hands. For high-cost approaches pursuing the scientific method this will be of only limited significance as the people to whom one's results must be credible are in Washington and European capitals. For the vast majority of IA studies, however, the issue of how to develop constructive relationships with program staff requires careful thought and action. Efforts to achieve co-ownership of findings by involving program staff in IA design, showing respect for their ideas and opinions, and discussing interim findings are possible ways of making influence more probable.

6. EFFECTIVE IMPACT ASSESSMENT: ACHIEVING "FIT"

The key task for the IA designer is to select an approach that can meet the objectives of the specific assessment at an acceptable level of rigor, that is compatible with the program's context, that is feasible in terms of costs, timing and human resource availability and that avoids the problems identified in earlier sections. Wherever possible an IA methodology should be piloted before full implementation. The questions that s/he must answer can be summarized as follows.

—What are the objectives of the assessment?
—How is the information to be used and by whom?
—What level of reliability is required?
—How complex is the program, what type of program is it, what is already known about it?
—What resources (money, human and time) are available?

The range of specific responses to these questions is infinite, but for the purposes of this paper they are grouped into four categories. These categories are based on Little (1997) but their characteristics have been substantially modified. These range from impact monitoring and validation, through simple and moderate approaches to more complex approaches. These can be viewed as a hierarchy, but there is a great danger in this as this may seem to infer that complex approaches are best!

(a) *Impact monitoring and validation (or, do not do an impact assessment!)*

Commonly the answer to the above questions should be "don't proceed with an impact assessment," as a program's emphasis on "institution building" will be undermined by IA and/or sufficient resources are not available. [23] Instead, donors could focus on strengthening the internal impact monitoring capacities of the microfinance institution and occasionally checking the quality of this information by using external monitors for validation purposes. The greater the involvement of staff in assessing program achievements then the greater is the likelihood of findings being used (Hyman & Dearden, 1998, p. 275).

Contrary to common practice, and donor preference, building internal impact monitoring capacities does not mean creating a large impact assessment unit within an MFI. Rather, it means helping the MFI develop its MIS and the work of its pre-existing internal monitoring and research units to collect readily available data (outreach, repayments, dropout rates etc.) alongside "simple to gather" types of data on who is using services, what for, why and what they like or dislike about the services. Much of this work can be done by focus groups, short interviews and rapid appraisal. [24] It is more akin to the market research that private business uses than the academic research that dominates aid-financed development.

These systems already operate formally and informally in some of the large Asian MFIs and are the basis upon which their directors take many "improvement" decisions. [25] Strengthening these systems and occasionally verifying them—rather than financing complex impact assessments by visiting consultants—is probably the best way to achieve the "improving" goals of IA. The types of verification process used in Social Audits (New Economics Foundation, 1996; Zadek & Gatward, 1996) provide a model for ensuring that internal impact assessments are valid.

(b) A simple approach

This seeks to provide timely information at relatively low cost about program impacts. These are the most common forms of IAs. Reliability is moderate, at best (and based mainly on triangulation), and the major objective is to test the existing understanding of impacts and contribute to improvements in program operation. The main audiences are program managers and donor "country-based" staff. The central methodological feature of such an approach is the use of a variety of methods. Usually this involves a small-scale client survey, compared with a comparison group that could be rapidly identified (e.g., approved clients who have not yet received services), and crosschecked by rapid or participatory appraisal methods. If a baseline study is not available then a recall methodology would be utilized. The key variables to be studied would depend on program objectives, but for income and assets the focus would be on ordinal and nominal measurements (see Little, 1997, p. 17). For programs prioritizing empowerment goals and local institutional

development, then participatory methods would be highlighted and the survey work might be dropped altogether.

Several participants in a CGAP Impact Assessment Group virtual meeting (Gaile, 1997, p. 5) argued that improving the credibility, utility and cost effectiveness of simple approaches was where the greatest gains in IA could be realized. Characteristics to enhance the effectiveness of simple IAs include: (i) focusing on a small set of key hypotheses; (ii) using variables that have a "track record" from earlier IA studies; (iii) clearly documented use of triangulation; (iv) methods applied consistently over time; (v) use of small sized comparison group; (vi) careful training of IA staff.

(c) A moderate approach

The moderate approach involves substantially more costs than the simple approach, yields higher levels of reliability (statistical inference rather than triangulation) and is likely to take two or three years before it delivers findings. Its focus is on both proving impact and improving programs. Its audiences would include policy-makers (looking for reassurance about their agency's investments) and the senior managers of programs. The methodological "mix" would center on a significant survey that would stratify clients and compare them with a carefully matched control group. The survey would involve at least two visits with a minimum of 12 months between them and recall techniques would not be used. Contextual and crosschecking materials would be produced by rapid appraisal techniques and carefully planned participant observation and case studies might also be commissioned. While the selection of variables would depend on program objectives the income and assets data would be extended (Gaile, 1997, p. 20) and measurement would focus on interval and nominal scales.

(d) A complex approach

The complex approach focuses on ensuring high levels of reliability with regard to the attribution of causality and has an exclusively "proving" orientation. Its main audiences are policy-makers and researchers and it is likely to be four to six years after launch before findings are available. The central method in such an approach is a large-scale sample survey very carefully constructed to represent all key

features of the client population. This is compared against a carefully selected control group, so that the number of households surveyed is likely to be between 750 and 1,500. At least three interviews will be conducted with each household over a period of two to three years. A much wider set of income and asset variables will be measured (Gaile, 1997, p. 23) and the focus will be on high precision through interval measurements. A set of related studies on institutional performance would be conducted, but the heart of the study would be the statistical and econometric analysis of survey findings. The budgets for such approaches are likely to exceed a million dollars as the survey costs are high, data processing and analysis inevitably generate problems and significant amounts of high-powered and high-cost econometrician and/or statistician time is needed.

7. CONCLUSION

In recent years donors have been keen to assess the impact of their programs. The initial emphasis on "scientific" sample surveys and statistical analyses has shifted as multimethod impact assessment studies and most recently participatory approaches have been utilized. Microfinance programs and institutions have experienced these shifts and examples of IAs on this topic provide a resource from which this paper has sought to draw out lessons for future practice. Much further work will be needed as the claims that microfinance is a panacea for poverty-reduction (most publicly through the Micro-Credit Summit and its follow up), and the counterclaims that caution against such enthusiasm (Rogaly, 1996; Wood & Sharrif,

1997) demand rigorous empirical testing to find out what is being achieved and how more might be achieved.

The desire of MFIs, donors and impact assessors themselves to produce results that will verify findings about impact at high levels of statistical confidence has too often driven the design of IA studies. This can compromise quality (with small sample surveys claiming exaggerated levels of representativeness) and impact (with external, data-extraction approaches making MFI staff unwilling to use findings and contradicting the "empowerment" goals of many programs). This paper has argued that IA effectiveness should not be automatically equated with the level of scientific "proof" that a study can claim. While all studies must pursue rigor, and this applies equally to quantitative and qualitative work, the effectiveness of an IA will depend on how well it achieves a fit between its objectives, the financial and human resources it can command and its context. There is no optimal model and different designs—characterized in this paper as "low," "moderate" and "complex" and combining scientific, humanities and participatory approaches—will be appropriate for different studies. All too often, however, donor desires for objective and external IAs (to meet their domestic accountability requirements) lead to the neglect of a key alternative: strengthening the impact monitoring capacity of the MFI itself. While striving for technical best practice should be a key goal for all in this field it would be foolish not to recognize that IA is a "battlefield of knowledge" (Long & Long, 1992) in which different actors seek to influence the knowledge creation process so that it meets their needs.

NOTES

1. For a discussion of impact assessment for other forms of services to microenterprise see Hyman and Dearden (1998).

2. The impact chain model not only underpins *ex post* assessment but is central to most aid financed activity in the form of a variety of "logical frameworks" that donors use to design projects and programs.

3. I use the term intended beneficiary here, rather than client, as most MFIs utilize (or have utilized) aid funds that are intended, at least in part, to benefit poor or vulnerable people and not purely self-selected clients.

This is an important point as, (a) some MFIs present their client populations as intended beneficiaries when many clients are known to be nonpoor, and (b) some agencies (e.g., the Consultative Group to Assist the Poorest) present the client populations of MFIs as "the poorest" when they know full well that only a proportion of clients are poor and that few if any are "the poorest" i.e. social outcasts, destitute, disabled, refugees, widows or elderly.

4. The assessment of institutional sustainability has been greatly advanced by Yaron's subsidy dependence index (SDI) which permits the assessor to move beyond

simple statements of profit and loss qualified by footnotes about changes in levels of subsidy.

5. The grandest invalidation is probably the United Kingdom in the mid to late 1980s when an unprecedented increase in outreach and profitability of institutions—providing financial services for households and small enterprises created a ''bubble'' which ultimately destroyed many enterprises, led to some losing their homes and impoverished hundreds of thousands of households. Also note that Mayoux (1997) has found that programs performing well in terms of outreach and repayment rates can have negative impacts in terms of women's empowerment.

6. I cannot locate my copy of a study on the USAID Anti-Poverty Lending Program! This study ably computed almost every measure of outreach, outreach growth and institutional financial health possible however, it failed to contain any information on ''who'' clients were, but simply assumed they were ''the poor.''

7. Gaile and Foster (1996, Annex 1) provide a vast list of all the variables measured in 11 recent studies of microfinance impacts. The reader will also find this very useful.

8. This section draws heavily on Mosley (1997) excellent paper.

9. This method is used widely in the medical and agricultural sciences.

10. We were not able to use this method for the Thrift and Credit Cooperatives in Sri Lanka, as the ''early'' members of a cooperative are generally drawn from higher income/assets/status groups while ''late'' joiners are from lower income/assets/status groups. In this case the use of a ''clients-to-be'' control group would have led to an exaggerated assessment of the economic impacts of microfinance.

11. Commonly such work is referred to as ''qualitative,'' but the quantitative/qualitative dichotomy is a false dichotomy. Most quantitative studies extract qualitative data from the respondent, have the interviewer immediately codify or quantify it and then use only numerical analysis. Many ''qualitative'' studies transform their data into quantities at later stages of analysis (Moris & Copestake, 1993, p. 4).

12. For a classic example of such work see Rutherford (1999).

13. This issue is raised in Hulme (1997) and was debated at the CGAP Impact Assessment Group virtual meeting.

14. As Moris and Copestake (1993, p. 1) point out ''the much recommended text on data collection by Casely and Lury (1982)... included two cursory paragraphs on ''quick and dirty'' techniques... almost half of the World Bank's publication that superseded it (Casley & Kumar, 1988) is concerned with qualitative methods.'' Studies using this approach include Hulme and Mosley (1996), Rutherford (1993), Fuglesang and Chandler (1986 and 1993). For details of many recent examples see Mayoux (1997).

15. The reader wishing to explore PLA and PRA (participatory rural appraisal) is referred to Chambers (1997) and the references he provides as there is not space to more fully explore these ideas in this paper. See Mayoux (1997) for a discussion of empowerment as a program goal, with particular reference to gender.

16. See Chao-Beroff (1997) for an example of an NGO's use of participatory methods and Martyn-Johns (1996) for a comprehensive review of PIA.

17. This section draws heavily on the work of Montgomery, Davies, Saxena and Ashley (1996).

18. For a full discussion see the excellent study by Montgomery *et al.* (1996).

19. The only report to hand that explicitly identifies absolute costs and relative costs (as a percentage of program budgets) is Montgomery *et al.* (1996) mainly in relation to natural resource and social programs. It reports that the 1994 IA of BRAC's credit program cost US $250,000.

20. At one large Asian MFI (Hulme and Mosley, 1996, Vol. 2), program field-staff visited villages that had been randomly selected for survey and told borrowers to make sure they gave interviewers ''the right answers.'' Fortunately, qualitative research revealed this and other villages were selected for survey!

21. This method is widespread in the United States and United Kingdom when market researchers convene focus groups to test new products. It should be noted that once interviewees in an area are paid for interviews

then the likelihood of noncooperation in the future, unless fees are paid, is greatly increased.

22. An anonymous reviewer pointed out that, like most IA theory and practice, this paper is dominated by supply-side issues and neglects ''use'' i.e., the demand side. This is a valid criticism especially given my earlier work which indicated that some development agencies actively ''do not learn'' (Hulme, 1988) and my continued belief that the World Bank has a learning disability (Hulme, 1992 and 1994).

23. Unfortunately donor agencies generally lack the courage to reach this decision and consultants (*mea culpa*) may have a vested interest in not promoting this option.

24. See Montgomery *et al.* (1996) for an excellent discussion of the role of impact monitoring vis á vis impact assessment.

25. For example, for BRAC in 1994 and 1995 vast amounts of technically valid data collection and analysis were occuring for the Impact Assessment Study (Mustafa *et al.*, 1996) and the World Bank-BIDS Bangladesh MFI study. The main source of information and ideas for BRAC's five-year plan (1995–2000) came from a rapid informal and focus group research exercise carried out by research officers from the Research and Evaluation Department who were ''pulled out'' of the technical studies, along with discussions with BRAC staff at headquarters and in the field (F. Abed, personal communication).

REFERENCES

Adams, D., & von Pischke, J. D. (1992). Microenterprise credit programs: Deja vu. *World Development*, *20*, 1463–1470.

Ardener, S., & Burman, S. (1995). *Money-go-rounds: the importance of rotating savings and credit associations for women*. Oxford: Berg.

Barnes, C. (1996). *Assets and the impact of microenterprise finance programs*. Washington, DC: Management Systems International.

Bouman, F. J. A., & Hospes, O. (1994). *Financial landscapes reconstructed*. Boulder, Co: Westview.

Buckley, G. (1997). Microfinance in Africa: Is it either the problem or the solution? *World Development*, *25* (7), 1081–1094.

Casely, D., & Lury, D. A (1982). *Monitoring and evaluation of agricultural and rural development projects*. Baltimore: Johns Hopkins University Press.

Casley, D., & Kumar, K. (1988). *The collections, analysis and use of monitoring evaluation data*. Baltimore: John Hopkins University Press.

Chambers, R. (1993). *Challenging the professions: frontiers for rural development*. London: IT Publications.

Chambers, R. (1997). *Whose reality counts? putting the first last*. London: IT Publications.

Chao-Beroff, R. (1997). *Impact assessment seen from an NGO practitioner's scope mimeo*. Washington, DC: Consultative Group to Assist the Poorest (CGAP).

Chen, M. A., & Dunn, E. (1996). *Household economic portfolios*. Washington, DC: Management Systems International.

Feinstein, O. (1997). *A transaction costs approach for microfinance project assessment mimeo*. Washington, DC: Consultative Group to Assist the Poorest (CGAP).

Fuglesang, A., & Chandler, D. (1986). *Participation as process: what we can learn from the Grameen bank*. Oslo: NORAD.

Fuglesang, A., & Chandler, D. (1993). *Participation as process – process as growth*. Dhaka: Grameen Trust.

Gaile, G. L. (1997). *Highlights and recommendations from the microenterprise impact virtual meeting*. Washington DC: Management Systems International.

Gaile, L., & Foster, J. (1996). *Review of methodological approaches to the study of the impact of microenterprise credit programs*. Washington, DC: Management Systems International.

Goetz, A. M., & Sen Gupta, R. (1996). Who takes the credit? gender, power and control over loan use in rural credit programs in Bangladesh. *World Development*, *24* (1), 45–64.

Hashemi, S. M., Schuler, S. R., & Riley, A. P. (1996). Rural credit programs and women's empowerment in Bangladesh. *World Development*, *24* (4), 635–654.

Holcombe, S. (1995). *Managing to empower: the Grameen Bank's experience of poverty alleviation*. London: Zed Press.

Hossain, M. (1988). *Credit for alleviation of rural poverty: the Grameen Bank in Bangladesh*. Washington, DC: IFPRI.

Hulme, D. (1992). Enhancing organizational effectiveness in developing countries: the training and visit system revisited. *Public Administration and Development*, *12* (5), 433–445.

Hulme, D. (1994). Does the World Bank have a learning disability? *Public Administration and Development*, *14* (1), 93–97.

Hulme, D. (1988). Learning and not learning from experience in rural project planning. *Public Administration and Development*, *9* (1), 1–17.

Hulme, D. (1997). Impact assessment methodologies for microfinance: A review. A paper prepared for the virtual meeting of CGAP Working Group on Impact Assessment Methodologies, April 17–19.

Hulme, D., & Mosley, P. (1996). *Finance against poverty, Vols. 1 and 2*. London: Routledge.

Hyman, E., & Dearden, K. (1998). Comprehensive impact assessment systems for NGO microenterprise

development program. *World Development*, 26 (2), 261–276.

Khandker, S. R. (1998). *Fighting poverty with micro-credit: experience in Bangladesh*. Washington, DC: Oxford University Press for the World Bank.

Little, P. (1997). *Assessing the impact of microfinance programs on incomes and assets mimeo*. Washington, DC: Consultative Group to Assist the Poorest (CGAP).

Long, N., & Long, A. (1992). *Battlefields of knowledge*. London: Routledge.

Martyn-Johns, E. (1996). Participatory Impact Assessment: *Putting Theory into Practice*. MA dissertation, IDPM, University of Manchester, UK.

Mayoux, L. (1997). *Impact assessment and women's empowerment in microfinance programmes: issues for a participatory action and learning approach mimeo*. Washington, DC: Consultative Group to Assist the Poorest (CGAP).

Montgomery, R. (1996). Disciplining or protecting the poor? avoiding the social costs of peer pressure in microcredit schemes. *Journal of International Development*, 8 (2), 289–305.

Montgomery, R., Davies, R., Saxena, N.C., & Ashley, S. (1996). *Guidance materials for improved project monitoring and impact review systems in India*. Centre for Development Studies, University College, Swansea, UK.

Moris, J., & Copestake, J. (1993). *Qualitative enquiry for rural development: a review*. London: IT Publications.

Mosely, P., & Hulme, D. (1998). Microenterprise finance: is there a conflict between growth and poverty alleviation. *World Development*, 26 (5), 783–790.

Mosley, P. (1997). *The use of control groups in impact assessment for microfinance mimeo*. Washington, DC: Consultative Group to Assist the Poorest (CGAP).

Mosse, D. (1994). Authority gender and knowledge: theoretical reflections on the practice of participatory rural appraisal. *Development and Change*, 25 (3), 497–525.

Mustafa, S. *et al.*(1996). *Beacon of hope: an impact assessment study of BRAC's rural development programme*. Dhaka: BRAC.

New Economics Foundation (1996). *A social audit manual for small organizations*. London: New Economics Foundation.

Otero, M., & Rhyne, E. (1994). *The new world of microenterprise finance.* London: IT Publications.

Peace, G., & Hulme, D. (1994). Microenterprise and children what are the intra household impacts of income-generating programmes. *Small Enterprise Development*, 5 (1), 21–29.

Pulley, R. (1989). Making the poor creditworthy: a case study of the integrated rural development program in India. World Bank Discussion Paper 58. Washington, DC: World Bank.

Remenyi, J. (1991). *Where credit is due*. London: IT Publications.

Rogaly, B. (1996). Micro-finance evangelism, destitute women and the hard selling of a new anti-poverty formula. *Development in Practice*, 6 (2), 100–112.

Rutherford, Stuart (1999). The Poor and Their Money, *Finance and Development Working Paper* No. 3. IDPM: University of Manchester.

Rutherford, Stuart (1993). Learning to lend: informal savings and credit schemes in Bangladesh. SCF Working Paper No. 5, London: Save the Children Fund.

Schuler, S. R., Hashemi, S. M., & Riley, A. P. (1997). The influence of women's changing roles and status in Bangladesh's fertility transition: evidence from a study of credit programs and contraceptive use. *World Development*, 25 (4), 563–576.

Schuler, S. R., & Hashemi, S. M. (1994). Credit programs women's empowerment and contraceptive use in Rural Bangladesh. *Studies in Family Planning*, 25 (2), 65–76.

Sebstad, J., Neill, C., Barnes, C., & Chen, G. (1995). *Assessing the impacts of microenterprise interventions: a framework for analysis*. Washington, DC: USAID.

Wiig, A. (1997). *Credit expansion in microcredit programmes: dilemmas and feasible methods for studying them mimeo*. Washington, DC: Consultative Group to Assist the Poorest (CGAP).

Wood, G., & Sharrif, S., (1997). *Credit where credit is due*. Dhaka and London: UPL and Zed Press.

Yaron, J., Benjamin, M., & Piprek, G. (1997). *Rural finance: issues, design and best practices*. Washington, DC: Agriculture and Natural Resources Department, World Bank.

Zadek, S., & Gatward, M. (1996). Transforming the Transnational NGOs: Social Auditing or Bust? In M. Edwards (Ed.), *Beyond the magic bullet: NGO performance and accountability in the post-war world* (pp. 226–240). West Hartford: Kumarian Press and London: Earthscan.

APPENDIX A. CONDITIONS IN WHICH KEY DATA COLLECTION METHODS ARE, AND ARE NOT, APPROPRIATE

Sample surveys are appropriate when:	Rapid appraisal and/or PLA are appropriate when:	Participant observation and/or case studies are appropriate when:
A project affects large numbers of beneficiaries	The project is adopting or promoting participatory principles in (re-)planning, implementation, monitoring and evaluation	An understanding of the motivations and perceptions of project clientele is a priority
Policy-makers require accurate estimates of project impacts	An understanding of the motivations and perceptions of project clientele is a priority	Other methods (surveys and rapid appraisals) are unlikely to capture the views of minorities or women
Statistical comparisons must be made between groups over time and/or between locations	One of the purposes of the study is to assess whether or not felt needs are being addressed by the project	One of the purposes of the study is to assess whether or not felt needs are being addressed by the project
Project delivery/implementation mechanisms are operating well, thereby justifying investment in the assessment of impacts	The impact of community-based organizations or other institution-building activities are of importance	The impact of community-based organizations or other institution building activities are of importance
The target population is heterogeneous and it is difficult to isolate the influence of factors unrelated to the project (e.g., contextual variables, other programs etc.)	There is a need to understand the quality of other data collected through surveys	There is a need to understand the quality of other data collected through surveys or rapid appraisals (e.g. causal processes)
	There is a need for contextual studies before designing more complex monitoring or impact assessment exercises (e.g., case studies, or surveys)	There is a need for contextual studies before designing more complex monitoring or impact assessment exercises (e.g., before carrying out rapid appraisals, or before designing a survey)

Table—*Continued*

Sample Surveys are usually not appropriate when:	Rapid Appraisal and/or PLA are not appropriate when:	Participant Observation and/or Case Studies are usually not appropriate when:
A project affects small numbers of beneficiaries	Projects are relatively uncomplicated in which bounded locations are not units of analysis (e.g., Health centers serving a wide catchment area, or large schools serving a variety of communities)	The project is small and "uncomplicated," providing a specific service or limited intervention which is unlikely to affect community dynamics beyond a few specific effects (e.g. disease-specific health facilities or campaigns)
Policymakers are mainly concerned with project outcomes. (Was infrastructure completed on time and within budget? How many people us the health clinics?)	Indicators of project impact are uncontroversial, and negative impacts are unlikely	Bounded locations are not units of analysis (e.g., health centers serving a wide catchment area, or large schools serving a variety of communities)
Project implementation is recent and untested, and it is likely that the way in which the project is implemented will have little impact at the present time	Standardized and statistically representative generalizations for a large and diverse project population are regarded as the sole priority	Indicators of project impact are clear and easily measurable or assessable (by survey or rapid appraisals)
The purpose of the assessment is to study and evaluate complex activities or processes (e.g., the development & operation of community-based organizations; qualitative use of skills as a result of training programs)	Participation of clientele is not a priority (e.g., in public administration or power sector reform, or an organizational change program—in these types of projects more limited focus group discussions with staff may be more appropriate)	Indicators of project impact are uncontroversial, and negative impacts are unlikely
The purpose of the assessment is to document easily observable changes in the physical environment or other tangibles (which can be assessed through simpler, structured visits)		Information is needed quickly, and standardized, representative (statistical) generalizations are regarded as the sole priority
The purpose of the assessment is to understand whether or not the project is meeting the felt needs of the project clientele		

Source: Adapted from Montgomery *et al.* (1996).