

Chapter 17

A Review of Impact Assessment Methodologies for Microenterprise Development Programmes

by

Gary Woller,

*Associate Professor Romney Institute of Public Management,
Brigham Young University Utah, USA*

Introduction

Over the last several years, OECD governments have invested millions of dollars in microenterprise development programmes in OECD and lesser developed countries (LDCs). Microenterprise development is based on a couple of underlying premises: 1) self-employment is a key component in creating economic opportunities for low-income persons with otherwise limited employment or earning options, and 2) the primary constraints to productive self-employment among low-income persons are access to capital (loans) and training. For OECD governments, self-employment expands the range of policy options to combat poverty in its many manifestations. For the poor, self-employment expands the range of livelihood and coping options. Self-employment policies appear to offer particular benefits in economies characterised by chronic under or unemployment or by high levels of informal economy participation by the poor.¹

In theory, programme participants leverage loans and training to start and expand micro and small enterprises thereby generating higher levels of enterprise returns; higher enterprise returns translate into higher household income; and higher household income is in turn invested in improved household socioeconomic well-being. In practice, the specific impacts of microenterprise development programmes are hard to pin down and harder still to measure. Impact assessments require adoption of research methodologies capable of isolating specific impacts out of a complicated web of causal and mediating factors and high decibels of random environmental “noise”, as well as attaching specific units of measurement to tangible and intangible impacts that may or may not lend themselves to precise definition or measurement. It is not an easy task.

Nonetheless, microenterprise development competes with other development-employment policies for scarce public funds, and it is reasonable that policymakers should want to know whether microenterprise development is a good investment relative to other policy options. Fortunately, methodologies to assess programme impact do exist. Drawing on principles and experience in the natural and social sciences, impact assessment (IA) methodologies are well-developed and are well-known to researchers. The same methodologies, however, are much less well-known to policymakers, with important implications for policy analysis.

The validity of the findings of any impact assessment is in direct proportion to the validity of the IA methodology used. There exists substantial variation in the methodologies used by IA researchers, but not all IA methodologies are equally valid. As a result, the quality of IA studies runs the gamut. Methodological variation reflects a number of factors, such as researcher skill and inclination, the purpose of the assessment, and resource or environmental constraints. The truth is that most IA researchers work under constraints that require them to make trade-offs between methodological precision and methodological feasibility. The validity of a particular IA study often turns on the validity of the trade-offs made.

Given the methodological issues that inevitably arise during any impact assessment, the ability of policymakers to reach informed decisions regarding the impact of self-employment policies (or any public policy, for that matter) arguably depends to a large degree on their ability to make informed judgments about the validity of the assessment methodologies used and the justifications for methodological tradeoffs made. To further this end, this study examines IA methodologies used in 67 IA studies of 90 microfinance programmes in 31 LDCs and 20 IA studies of 20 microenterprise programmes in 2 OECD countries (19 of them in the United States, see Tables 17.1-17.3).²

The rest of the study proceeds as follows. The following section describes the conceptual foundations to IA, followed in the next section by a discussion of the two dominant methodological paradigms and the five methodological approaches that fall within them. The fourth and fifth sections discuss, respectively, the major methodological pitfalls bedevilling IA and other miscellaneous methodological shortcomings common to IA studies. The sixth section offers additional thoughts about judging the methodological rigor of IA studies. The final section offers some policy recommendations.

Before going farther, one clarification is in order. From here on, the term *microfinance* is used to connote microenterprise development in lesser developed countries (LDCs), and the term *microenterprise* to connote the same in OECD countries. The different terminology reflects important distinctions between the two. In LDCs, the microfinance industry is evolving into a full-fledged financial service industry for the poor. Increasingly, microfinance institutions (MFIs) are offering a range of financial services – such as savings, consumption loans, or insurance – in addition to enterprise loans. Historically, the poor in LDCs have not had access to formal financial services of any kind. Notwithstanding, MFIs are discovering a large latent demand for a diversified set of formal financial services among the poor, and they are evolving to meet that demand. Primary among the goals of many microfinance advocates is “financial deepening”, or the creation of a system of sustainable financial intermediation for the poor.

Table 17.1. Summary of reviewed IA assessments

Reviewed impact assessment ^a	Year	Country	Methodology ^b	Control group ^c	Timeframe ^d	Comment ^e
LDCs						
Ashe and Parrot	2001	Nepal	S, Q	N	CS	
Barnes	2001	Zimbabwe	S, Q	Y	L	
Barnes, Morris, and Gaile	1999	Uganda	S, Q	Y	L	
Bolnick and Nelson	1990	Indonesia	S	Y	CS	
Buckley	1996a	Kenya	S	Y	CS	
Buckley	1996b	Malawi	S, Q	Y	CS	
Buvinic, Berger, and Jarmillo	1989	Ecuador	S	Y	L	
Chen and Snodgrass	2001	India	S, Q	Y	L	
Churchill	1995	So. Africa	S, Q	Y	CS	
Coleman	1999	Thailand	S, Q	Y	L	SB, DS, CL
Coleman	2001b	Thailand	S, Q	Y	L	SB, DS, CL
Copestake, Bhalotra, and Johnson	2001	Zambia	S, Q	Y	CS	
Creevey, Ndour, and Thiam	1995	Guinea	S, Q	Y	CS	
Deardon and Khan	1994	Bangladesh	S	Y	L	
Diagne	1998	Malawi	S	Y	L	F
Dunn and Arbuckle	2001	Peru	S, Q	Y	L	
Goetz and Gupta	1996	Bangladesh	Q	N	CS	
Gupta and Davalos	1993	Jamaica	S	N	CS	
Hashemi, Schuler, and Riley	1996	Bangladesh	S, Q	Y	L	SB, DC, CL
Hulme, Montgomery, and Bhattacharya	1996	Sri Lanka	S	Y	CS	
Karlan and Alexander	2002	Peru	S	Y	CS	SB, DO
Kevane and Wydick	2001	Guatemala	S	Y	CS	
Khandker	1996	Bangladesh	S	Y	CS	SB, DS, CL
Khandker	2001	Bangladesh	S	Y	L	SB, DS, CL
Khandker, Samad, and Khan	1998	Bangladesh	S	Y	CS	SB, DS, CL
Kilby and D'Zmura	1985	Brazil	S	N	CS	BC, CL
		Upper Volta	S	N	CS	BC, CL
		Honduras	S	N	CS	BC, CL
		Dom. Rep.	S	Y	CS	BC, CL
		Peru	S	N	CS	BC, CL
		Peru	S	N	CS	BC, CL
Lapar, Graham, Meyer, and Kraybill ^f	1995	Philippines	S	Y	CS	SB
Lapar, Graham, and Meyer ^f	1995	Philippines	S	Y	CS	SB
McKernan	1996	Bangladesh	S	Y	CS	SB, DS
MkNelly and Dunford	1999a	Bolivia	S, Q	Y	L	SB

Table 17.1. **Summary of reviewed IA assessments (cont.)**

Reviewed impact assessment ^a	Year	Country	Methodology ^b	Control group ^c	Timeframe ^d	Comment ^e
MkNelly and Dunford	1999a	Bolivia	S, Q	Y	L	SB
MkNelly and Dunford	1999b	Ghana	S, Q	Y	L	SB
MkNelly, Watetip, and Lassen	1996	Thailand	S, Q	Y	CS	
Montgomery, Bhattacharya, and Hulme	1996	Bangladesh	S, Q	Y	CS	
Morduch	1998	Bangladesh	S	Y	CS	SB, DS, CL
Mosely	1996a	India	S	Y	CS	
Mosely	1996b	Indonesia	S	Y	CS	
Mosely	1996c	Bolivia	S	Y	CS	
Mosely	2001	Bolivia	S, Q	Y	CS	
Mosely and Hulme	1998	Bangladesh	S	Y	CS	
		Bolivia	S	Y	CS	
		India	S	Y	CS	
		Indonesia	S	Y	CS	
		Kenya	S	Y	CS	
		Malawi	S	Y	CS	
		Sri Lanka	S	Y	CS	
Mustafa, Ara, Banu, Hossain, Kubir, Moshin, and Yusuf	1995	Bangladesh	S, Q	Y	CS	DS
Neill, Davalos, Kiiru, and Sebstad	1994	Kenya	S	N	CS	
Nelson	1984	Indonesia	S	Y	CS	
Nelson and Bolnick	1986	Indonesia	S	Y	CS	
Oldham, Hadidid, Hussein, Aziz, and Sakr	1994	Egypt	S	N	CS	
Park and Ren	2001	China	S	Y	CS	
Pitt and Khandker	1996	Bangladesh	S	Y	CS	SB, DS
Pitt and Khandker	1998	Bangladesh	S	Y	CS	SB, DS
Pitt, Khandker, Chowdhury, and Millimet	1998	Bangladesh	S	Y	CS	SB, DS
Pitt, Khandker, McKernan, and Latif	1999	Bangladesh	S	Y	CS	SB, DS, CL
Pulley	1989	India	S	Y	L	
Schuler and Hashemi	1994	Bangladesh	S, Q	Y	L	SB, DS, CL
Schuler, Hashemi, and Badal	1998	Bangladesh	Q	Y	L	
Schuler, Hashemi, and Riley	1997	Bangladesh	S, Q	Y	L	SB, DS, CL
Sebstad	1992	So. Africa	Q	N	CS	
Sebstad and Cohen	2002	Bangladesh	S, Q	Y	CS	
		Bolivia	S, Q	Y	CS	
		Philippines	Q	N	CS	
		Uganda	Q	N	CS	

Table 17.1. **Summary of reviewed IA assessments (cont.)**

Reviewed impact assessment ^a	Year	Country	Methodology ^b	Control group ^c	Timeframe ^d	Comment ^e
Sebstad and Loza	1993	Egypt	S, Q	N	CS	
Sebstad and Walsh	1991	Kenya	S, Q	N	CS	
Smith	2002	Ecuador	S	Y	L	
		Honduras	S	Y	L	
Steele, Amin, and Naved	2001	Bangladesh	S	Y	L	
Sutoro	1990	Indonesia	S, Q	N	CS	
Todd	1996	Bangladesh	Q	Y	CS	
Vengroff and Creevey	1994	Senegal	S, Q	Y	CS	
Woller and Parsons	2002	Ecuador	S	N	CS	CL
Wydick	1999a	Guatemala	S	Y	CS	
Wydick	1999b	Guatemala	S	Y	CS	
Zaman	2001	Bangladesh	S	Y	CS	SB, F
Zeller, Ahmed, Babu, Broca, Diagne, and Sharma	1996	Cameroon	S	Y	L	CL
		Mali	S	Y	L	CL
		Ghana	S	Y	CS	CL
		Nepal	S	Y	L	
		Pakistan	S	Y	L	CL
		China	S	Y	CS	CL
		Bangladesh	S	Y	L	CL
		Madagascar	S	Y	L	CL
		Malawi	S	Y	L	
OECD Countries						
Ashe and MacIntyre	2002	USA	S, Q	N	CS	
Benus, Wood, and Grover	1994	USA	S	Y	L	SB
Blair and Klein	2001	USA	S, Q	N	L	
Clark and Huston	1993	USA	S, Q	N	L	
Clark, Kays, Zandniapour, Soto, Doyle	1999	USA	S, Q	N	L	
Drury, Walsh, and Strong	1994	USA	S	N	L	
Dumas	2001	USA	Q	N	CS	
Else and Clay-Thompson	1998	USA	Q	N	CS	BC
Himes and Servon	1998	USA	S, Q	N	CS	DO
Kosanovich and Fleck	2002	USA	S	Y	L	BC
Institute for Social and Economic Development	1994	USA	S	Y	L	
Mt. Auburn Associates	1998	USA	S	N	CS	DO
Raheim	1996	USA	S	N	CS	
Raheim and Friedman	1999	USA	S, Q	N	L	
Sekkesaeter	2002	Norway	S, Q	N	CS	
The Roberts Foundation	1995	USA	S	N	CS	

Table 17.1. **Summary of reviewed IA assessments (cont.)**

Reviewed impact assessment ^a	Year	Country	Methodology ^b	Control group ^c	Timeframe ^d	Comment ^e
US Department of Health and Human Services	1994a	USA	S	Y	L	
US Department of Health and Human Services	1994b	USA	S	Y	L	BC
US Department of Health and Human Services	1994c	USA	S	Y	L	SB
US Department of Health and Human Services	1994d	USA	S, Q	Y	CS	

a) Several reviewed impact assessments were based in full or in part on the same programme assessment or assessment data. These include Coleman (1999, 2001b); Nelson (1984), Nelson and Bolnick (1986), Bolnick and Nelson (1990); Khandker (1994), Pitt and Khandker (1994, 1998), McKernan (1996), Pitt, Khandker, Chowdhury, Millimet (1998), Khandker, Samad, and Khan (1998), Pitt, Khandker, McKernan, and Latif (1999), and Khandker (2001); Schuler and Hashemi (1994), Hashemi and Riley (1996), Schuler, Hashemi, and Riley (1997), and Schuler, Hashemi, and Badal (1998); Wydick (1999a, 1999b) and Kevane and Wydick (2001).

b) S = Impact survey; Q = Qualitative assessment.

c) Y = Yes; N = No.

d) L = Longitudinal; CS = Cross sectional.

e) SB = Controls for selection bias for unobservable characteristics; DS = Administers survey at different seasons/times during year; CL = Assesses community-level benefits; F = Controls for loan fungibility; BC = Performs benefit-cost analysis.

f) Not a programme assessment, but a study of rural non-farm enterprises.

The same is less true in OECD countries. Relative to LDCs, OECD countries have highly developed financial services markets to which the poor enjoy access (if at times limited). There is, moreover, relatively little discussion among microenterprise advocates of financial deepening; instead, the industry's objectives tend to be defined more narrowly within the context of self-employment as an alternative to formal sector employment. Another important difference between microfinance and microenterprise is that the latter emphasises business training to a much greater degree. Integration of loans with business training is common within the microenterprise industry, but is rarer and is the subject of much dispute within the microfinance industry.

Given the differences that exist between the two, therefore, care is taken here to differentiate between them where relevant. That said, the issues discussed in this study apply more or less equally to impact assessments of both types of programmes. Moreover, the need for sound impact assessments applies equally to both as do the methodological principles for conducting them.

Table 17.2. Summary of IA studies

	LDC	OECD	Total
Published impact studies	67	20	87 ^a
Impact assessments performed	90	20	110 ^b
Methodology			
Survey only	58	10	68
Survey + Qualitative	26	8	34
Qualitative only	6	2	8
Control group			
Yes	73	7	81
No	17 ^c	13	28
Timeframe			
Longitudinal	27	11	38
Cross-sectional	63	9	72
Control for selection bias	19	2	21
Surveys at different seasons	14	0	14
Assess community-level impacts	24	0	24
Controls for loan fungibility	2	0	2
Performs benefit-cost analysis	6	2	8
Includes programme dropouts	1	2	3
Number of countries	31	2	33

- a) Refers to IA studies published, regardless of the number of separate impact assessments covered in each study.
- b) Refers to the number of separate impact assessments found in each published IA study. Impact assessments of two or more programmes in the same country as part of an integrated impact assessment were counted as a single assessment. Impact assessments of programmes in different countries, but published in the same IA study, were counted as separate assessments.
- c) Control groups were not relevant to the community economic impact methodology used by Woller and Parsons (2002).

Conceptual foundations to impact assessment

Impact theoretically occurs at four levels: the individual, the enterprise, the household, and the community.³ In theory, the impact causal chain works something like this: 1) loans and training lead to increased enterprise formation and expansion and to increased investment in working capital and productive assets; 2) increased enterprise formation, expansion, and investment lead to increased enterprise returns; 3) increased enterprise returns lead to increased job creation and increased household income; 4) increased household income leads to higher levels of household consumption, asset accumulation, human resource investment, and physical asset investment. Increased household income and asset accumulation, together with increased access to financial services, in turn expand poor households' *ex ante* and *ex post* coping and livelihood strategies, thereby making them less vulnerable to risk.

Table 17.3. **Countries in which impact assessments performed**

	#
Bangladesh	22
Bolivia	5
Brazil	1
Cameroon	1
China	2
Dominican Republic	1
Ecuador	3
Egypt	2
Ghana	2
Guatemala	3
Guinea	1
Honduras	2
India	4
Indonesia	6
Jamaica	1
Kenya	4
Madagascar	1
Malawi	4
Mali	1
Nepal	2
Norway	1
Pakistan	1
Peru	3
Philippines	3
Senegal	1
South Africa	2
Sri Lanka	2
Thailand	3
Uganda	2
United States	20
Upper Volta	1
Zambia	1
Zimbabwe	1

Moving down to the individual level, access to financial services, control over financial resources, enterprise ownership and operation, increased household income contribution, and group networking and mutual support lead to higher levels of personal and social empowerment, especially among female programme participants. At the community level, benefits created at the other three levels create positive externalities that diffuse through local and surrounding communities. (Table 17.4 lists common indicators used to measure impacts at each of the four levels of impact.)

Table 17.4. **Impact Indicators at the individual, enterprise, household and community levels**

Level of Impact	Indicators
Individual Level	<ul style="list-style-type: none"> • Intra-household decision making (participation in household decision making on issues such as finances, schooling, healthcare, family planning, etc.) • Control over financial and other resources • Contribution to household income • Contraceptive usage • Self-esteem • Attitudes about self, life and the future • Political and social awareness • Participation in social and political spheres • Spousal abuse
Enterprise Level	<ul style="list-style-type: none"> • Sales • Profits • Net worth • Asset ownership and acquisition • Jobs created • Product diversification • Business diversification • Business practices adopted
Household Level	<ul style="list-style-type: none"> • Income <ul style="list-style-type: none"> – Expenditures – Expenditures on food – Expenditures on specific types of foods (<i>e.g.</i> fruits, vegetables, meats, dairy) – Expenditures on medicine and health care – Expenditures on children's schooling • Asset ownership and acquisition • Savings • Investment in housing and home improvements • Investment in land • Access to and use of medicines and healthcare • Knowledge and use of simple hygiene practices • Knowledge and use of simple medical interventions/health practices (<i>e.g.</i> oral rehydration therapy, breast feeding) • Children's school attendance • Types and frequencies of foods consumed • Incidence and duration of "hungry seasons" • Anthropomorphic measures of children • Response to and impact of economic and other shocks
Community Level	<ul style="list-style-type: none"> • Children's school attendance • Contraceptive usage • Jobs created/Employment • Income • Expenditures • Net worth • Production • Wages • Prices • Participation in social and political spheres • Contribution to families' support • Poverty

In practice, the impact causal chain is more complex than depicted above. For one thing, a host of mediating factors influence impact. Examples include programme attributes, client characteristics, geography, social structure and power relationships, the physical and economic infrastructure, and the macro economy. Another thing is the reciprocal relationship between cause and effect in which impacts become causes, causes become impacts, impacts become causes again, and so on, such that it becomes increasingly difficult to distinguish between the one and the other. Finally, loan fungibility makes tracing through the exact sequence of cause and effect virtually impossible (see discussion below), thus creating something more akin to a causal web than a causal chain. The end result is that participants experience impacts differently, and no two programmes create the same impacts in the same way. Thus it is perhaps not surprising that the only consistency in IA findings is their inconsistency – a typical impact assessment yields mixed findings, and findings vary considerably from study to study. This inconsistency highlights the inherent dangers in assuming a causal link between programme participation and any specific policy outcome.

Methodological approaches to impact assessment

Conceptual frameworks in hand, IA researchers may choose from among several methodological approaches for conducting impact assessments. Depending on the purpose, approaches fall within one or both of two methodological paradigms: the “proving” paradigm and the “improving” paradigm. Within the proving paradigm, the purpose of IA is to attribute causality of observed outcomes to programme participation. The proving paradigm adopts the language, methodology, and worldview of the physical and social sciences. Its audience is primarily external – donors, policymakers, and academics – for whom methodological rigor and scientific validity are prime virtues.

Within the improving paradigm, the objective of IA is to improve the impact of financial services on programme participants through improving products and policies. The improving paradigm adopts the language, methodologies, and worldview of management. Its audience is primarily internal – board, management, staff, and clients – for whom usefulness, timeliness, and cost are prime considerations.

Within the context of the proving and improving paradigms, five methodological approaches can be identified:

1. The scientific method, which is based in the natural sciences.
2. The humanities tradition, which uses theory and corroboration of evidence to make reasoned judgments.

3. Midrange assessments, which explicitly take into account constraints imposed by “field realities.”
4. Participatory Learning and Action (PLA), which facilitates subjective articulation of participants’ “reality” to arrive at informed conclusions.⁴
5. Market research, which emphasises the collection and use of market intelligence to inform management decision making.

If we assume a continuum, with the proving paradigm at one extreme and the improving paradigm at the other, the scientific method and market research would lie at or near the respective extremes. The remaining three methods would tend to cluster near the middle of the continuum. The scientific method relies primarily on quantitative evidence and the humanities tradition and PLA primarily on qualitative. Midrange assessments and market research tend to rely more on varying combinations of quantitative and qualitative evidence. In practice, all five approaches may be used for either proving or improving. Moreover, it is not uncommon for researchers to use a combination of approaches, along with combinations of quantitative and qualitative methods, so as to crosscheck data and add greater depth, confidence, or relevance to the findings.

Scientific method

In the classic scientific experiment, study subjects are drawn from the same population, share common characteristics, and are randomly selected into either the treatment group (those receiving the intervention) or the control group (those not receiving the intervention). The intent of the classic experiment is to control for all mediating factors so as to be able to attribute any observed differences in outcomes between treatment and control groups to the intervention. Unfortunately, in the social sciences the conditions for a classic experiment rarely exist, and the ability of the researcher to control for all mediating factors is severely limited, if not impossible.

A step down from the classic experiment is the “quasi-experiment”, which attempts to replicate the conditions of the classic experiment to the extent possible within existing constraints and typically using survey-based research instruments. Another way to think of the quasi-experiment is as a “with-without” test, in which the researcher attempts to establish the counterfactual of what would have happened had the treatment group not received the intervention.

Multiple regression is a form of quasi-experiment. It tests for the impact of certain explanatory variables on observed outcomes, while holding other mediating factors constant. Multiple regression is not widely used in impact assessments, however, owing to the large data demands necessary to account for all relevant mediating factors, the technical expertise required to perform

and interpret it, and the tenuousness of the econometric assumptions necessary to validate it.

The most common form of quasi-experiment is the control group study. As the name suggests, control group studies compare outcomes in a treatment group of programme participants to a control group of non-participants. If all goes well, the control group matches the treatment group on key characteristics, and the researcher can reasonably attribute observed differences in outcomes to programme participation. All does not always go well, however. Control group studies are fraught with several potential methodological pitfalls. Stumbling into any of the pitfalls can seriously compromise the study's validity.

Next in order of rigor are impact assessments that aspire to the scientific method, but are more properly characterised as pseudo-scientific, because they violate certain key scientific principles. A common example is the “before-after” assessment, which compares outcomes among programme participants at one point (*e.g.*, after joining the programme) to that of an earlier point (*e.g.*, before joining the programme), but which does not use control groups. The absence of control groups in the before-after assessment makes any attribution of impact to programme participation statistically invalid.

Humanities tradition

The humanities tradition is an inductive approach to learning that encompasses a broad set of tools developed and refined by social scientists in such fields as sociology and anthropology. Borrowing liberally from ethnography, the humanities approach involves the study of a small group of subjects in their own environment. Rather than looking at a small set of variables and a large number of subjects, it attempts to get a detailed understanding of the circumstances of the few subjects being studied. It is descriptive and interpretive – descriptive because detail is so crucial and interpretive because the researcher must determine the significance of what he or she observes without gathering broad, statistical information. Methods include key informant and in-depth participant interviews, case studies, and participant observation (*e.g.*, extended residence in programme communities by field researchers).

The humanities tradition does not attempt to prove impact in any statistical sense, but to offer an interpretation of the relationship between programme participation and outcomes that achieves high levels of plausibility, which in turn permits inference of causality between observed outcomes and programme participation. Plausibility itself depends on factors such as evidence of sound methodology, logical consistency, quality of evidence and reasoning, extent of confirmatory evidence via triangulation and secondary information sources, and the reputation of the researcher.

Relative to the scientific method, the humanities tradition lacks a set of well-defined methodological standards. Whether such standards could be specified to guide researchers *ex ante* and to help policymakers gauge the quality of work *ex post* is an as-of-yet unanswered question. Nonetheless, experience has shown that this approach can yield results of reasonably high reliability and provide a range and depth of insights not always obtainable through scientific methods. It may even at times produce conclusions of greater validity than scientific methods, particularly in the case of “survey based IA work that masquerades as science, but has not collected data with scientific rigor” (Hulme, 2000, p. 87).

Participatory Learning and Action

Participatory Learning and Action (PLA) encompasses a wide variety of methodologies, including Participatory Rural Appraisal (PRA), Rapid Rural Appraisal (RRA), Participatory Learning Methods (PALM), Participatory Action Research (PAR), or Farming Systems Research (FSR). PLA assumes the existence of multiple subjective realities. It requires first that researchers answer the question “whose reality counts”? (Chambers 1997). The job of the researcher is next to elucidate the shared reality of the target group through the process of “knowledge creation”, defined as the full participation of the target group in problem identification, analysis, and action planning. PLA methods include activities such as visualisation, seasonal calendars, historical timelines, Venn diagrams, transect walks, focus group discussions, relative preference ranking, and semi-structured interviews.

PLA offers the most radical and serious challenge to the scientific method as applied to international development. PLA advocates pull no punches in listing what they see as the flaws of the scientific method (Hulme, 2000):

- It ignores the complexity, diversity, and contingency of poor households’ livelihood and coping strategies.
- It conceives causality as a simple, unidirectional chain and not as the complex web that it is.
- It measures the trivial or pretends to measure what cannot be measured.
- It is extractive and exploitative.
- It reinforces the status quo through empowerment of technocrats, experts, professionals, policy-makers and elites.
- It does not lead to purposive action by or on behalf of poor groups.

If true, these are damning criticisms. Whether these allegations are true is perhaps a matter of one’s own subjective reality, although they probably do contain a good deal of truth. At the very least, they call for a certain level of

humility and critical self-reflection among practitioners of the scientific method, two traits that are not always abundantly evident.

PLA has its own set of weaknesses. Among them are its inherent subjectivity, its lack of standardisation (which makes comparisons difficult), its pluralism (which produces conflicting perspectives about impact), its naïve assumption that participation equals representation given local power relations, its assertion that participation is tantamount to empowerment, and its lack of transparency (which makes *ex-post* evaluations of methodological rigor difficult).

A final issue is attribution. Much like the humanities tradition, PLA neither can nor does claim causality on purely scientific grounds. It relies on triangulation of evidence, depth of knowledge, quality of methods and information, and skill of the researcher to establish a plausible case for causality. Such limitations, however, do not deter its advocates from arguing that well-conducted participatory studies can produce more reliable results than conventional surveys (see, for example, Chambers, 1997, pp. 141-146.)

As in the humanities tradition, the quality and reliability of PLA studies vary widely, depending on factors such as the skill of facilitators, the motivation of the target groups, the applicability of tools to situations, or the degree of participation.⁵ In practice, PLA advocates have been remarkably pragmatic. They are hesitant to prescribe specific best practices, preferring to rely on practitioners' best judgments to adapt the methodology and tools to the circumstances.

Midrange assessments

Midrange impact assessments stem from microfinance practitioners' widespread perception that the scientific method is disconnected from both the realities of the field and the needs of management. Practitioners have long complained that survey-based impact assessments are costly, lengthy, burdensome, and not timely; they require technical expertise and resources beyond institutional capabilities; and they are targeted to external audiences, with little attention to managerial usefulness. Nor do funding agencies or external evaluators invest time or money building institutional IA capacity. Throw into the mix the pressure from donors to achieve institutional self-sufficiency, and IA comes to be seen primarily as yet another line item on the expense report. Lacking the resources, technical skills, and material incentive to conduct IA, most programmes do not do it, the result being that most have little to no idea what their programme impact is.

Midrange assessments are the product of a practitioner-led effort, coordinated by the AIMS project (see footnote 2), to correct the deficiencies of the scientific method, bridge the proving and improving paradigms, and build

institutional IA capacity. The end result of this effort was the creation of a set of “practitioner-friendly” IA tools designed to account for field realities, produce managerially useful results, and be implemented by programme staff. The SEEP/AIMS tools, as they have come to be called,⁶ incorporate methodologies from each of the other four IA methodological approaches. They consist of two quantitative and three qualitative tools: the impact survey, the client exit survey, client satisfaction focus groups, in-depth empowerment interviews, and savings and loan use over time interviews.⁷ Midrange assessments do not aspire to proof of impact but instead aim to establish “plausible association” between observed outcomes and programme participation.

The SEEP/AIMS tools have enjoyed respectable legitimacy among practitioner organisations since their introduction. To date, dozens of MFIs have received formal, intensive training in use of the tools, and several have in turn successfully implemented them in the field. Midrange assessments, however, are not limited to the SEEP/AIMS tool, nor are all practitioners satisfied with them. Work continues to refine the SEEP/AIMS tools, adapt them to local contexts, or develop yet even more practitioner-friendly tools (e.g., “AIMS-Lite”).

For the most part, midrange assessments adhere to the standards of methodological approaches from which they borrow. The major concessions they make are to recommend the use of programme staff to conduct research and, related to the impact survey, to conduct cross-sectional (as opposed to longitudinal) assessments using so-called “pipeline” clients – clients recruited through normal programme operations but who have not yet received loans – as the control group. (See the Appendix for a list of recommendations for conducting midrange assessments.)

Market research

Integral to the improving paradigm is that impact assessment and market research are inextricably intertwined: timely knowledge about impact tells programme management how effective its products and policies are; market knowledge in turn allows programme management to design products and policies that improve impact. Market research (defined as the collection, analysis, and use of market intelligence) therefore plays a central role within the improving paradigm. The emergence of market research on the microfinance agenda is due to several market trends: 1) competition in the industry is increasing and is expected to increase yet more; 2) microfinance consumers are becoming more knowledgeable, discerning, and assertive; 3) clients are deserting microfinance programmes at often alarming rates, and, consequently, 4) MFIs are adopting more commercial strategies and practices.

Market research uses a variety of quantitative and qualitative tools, including surveys, focus group discussions, in-depth interviews, and participatory assessments.⁸ Market research includes the occasional assessment activity and proceeds on up to the integration of client and market information into programmes' operational and management information systems. Market research makes little pretence to scientific validity. Lack of attribution is a particular weakness, particularly given the logistical difficulties of integrating a non-client control group into routine data-gathering systems. From management's perspective, however, the loss in statistical certainty is more than made up for by gains in speed, cost, and usefulness. Lagging behind tool development in market research is the development and establishment of a set of methodological standards and guidelines to implementation of market research tools in the field (for more on the relationship between IA and market research, see Cohen 1999 and Copestake 2000.)

Among the five methodological approaches reviewed above, the scientific method and midrange assessments have the most clearly articulated standards to guide methodology *ex ante* and to evaluate the quality of methodology *ex post* (given midrange assessments' aspiration to scientific plausibility, they are subject to many of the same methodological standards as the scientific method). Moreover, survey-based scientific methods dominate IA studies. In most IA studies, qualitative methods (whether grounded in the humanities tradition or PLA) are used as a supplement to survey-based scientific research. Of the reviewed IA assessments, 68 used surveys as the sole research instrument and 34 used surveys as the primary research instrument, complemented by qualitative methods. Only 8 reviewed assessments relied solely on qualitative methods, including Goetz and Gupta (1996), Todd (1996), Schuler *et al.* (1997), Schuler *et al.* (1998), Dumas (2001), and the programme assessments in the Philippines and Uganda summarised in Cohen and Sebstad (2000). In light of the dominance of scientific IA and its relatively clear methodological standards, the rest of this study focuses on methodological issues relevant to scientific IA.

Methodological pitfalls of scientific IA

As mentioned above, ideal conditions to conduct scientific IA rarely exist. IA researchers must therefore often settle for second best, or worse. In the face of ever-present field constraints, IA researchers have to ask themselves to what degree are they willing to compromise accepted methodological principles to accommodate these constraints. In other words, "What cost in scientific precision are IA researchers willing to accept in exchange for a corresponding gain in implementation feasibility?"

In answering the question about acceptable tradeoffs, it is useful to know what the major pitfalls are to conducting scientific IA. This section reviews these pitfalls. In order of presentation, they are 1) construction of valid control groups, 2) selection bias stemming from observable individual characteristics, unobservable individual characteristics, failure to account for programme dropouts, and non-random programme placement; 3) control group contamination; 4) recall bias; 5) loan fungibility; and 6) IA timeframe (longitudinal vs. cross-sectional studies).

Valid control groups

Any scientific impact assessment claiming to infer causality or plausible association between outcomes and programme participation requires comparison to a valid control group of non-clients. Constructing a valid control group, however, can be difficult. There are many reasons why construction of a valid control group may not be feasible, primary among them being binding resource or environmental constraints (common issues for practitioner-led impact assessments) and, as explained below, challenges for control group construction and tracking posed by longitudinal (time-series) impact assessments.

Whether because of binding constraints or other reasons, several impact assessments reviewed for this study did not use control groups. The absence of control groups was particularly conspicuous among microenterprise programme assessments; only 7 of 20 microenterprise programme assessments used a control group compared to 73 of 90 of microfinance programme assessments.

The inability to infer causality does not mean that impact assessments lacking valid control groups have no value, but that their value lies elsewhere. They can, for example, be valuable tools for monitoring client progress, assessing relative outcomes among different market segments, measuring outcomes against programme objectives, or calculating per unit costs for specific programme outcomes or outputs, all of which are useful information for programme management, donors, and policymakers. They are not useful, however, for determining whether programme participation is causally linked to desired policy outcomes, such as poverty reduction or job creation.

Basically, control group selection requires identification of a population of persons not participating in a credit or training programme and sharing similar characteristics as the treatment group and then randomly selecting from among them. Not everyone who belongs to the population of non-clients, however, is a legitimate control group candidate. Construction of a valid control group requires that the control group match the treatment group on key observable and unobservable characteristics. Failure to do so creates so-called

“selection bias”, which is the major methodological pitfall bedevilling control group studies. Depending on its seriousness, selection bias renders attribution of observed impacts anywhere from problematic to wholly invalid. Twenty of 90 LDC programme assessments attempted to control for selection bias, compared to 2 of 20 OECD programme assessments.

Selection bias

Selection bias stems from four principal sources: 1) failure to match treatment and control groups on observable individual characteristics, 2) failure to match treatment and control groups on unobservable individual characteristics, 3) failure to account for programme dropouts and 4) non-random programme placement. These are considered in turn below.

Failure to match on observable individual characteristics

Outcome differences between treatment and control groups may be as much, if not more, a function of differences in observable characteristics, such as gender, age, education, self-employment status, enterprise type, or geographic location, than programme participation. Construction of a valid control group thus requires that control group members share similar observable characteristics as treatment group members. For example, if the profile of the treatment group is 85 per cent female, 25-50 in age, 0-5 years of formal education, self-employed, rural, and more or less evenly distributed among manufacturing, retail, and services, a valid control group will match these observed characteristics as closely as possible.

Matching treatment and control groups on observable characteristics can be challenging, but it is by no means insurmountable. A good researcher should be able to avoid this pitfall with comparative ease. To ensure that the control group closely matches the treatment group, sample stratification – random sampling within specifically selected groups among the target population – may be necessary. Whether the treatment and control groups reasonably match on observable characteristics can easily be determined by comparing group means on key observable characteristics and testing whether the differences in means are statistically significant.

Failure to match on unobservable individual characteristics

A yet more bedevilling source of selection bias is failure to match treatment and control groups on unobservable individual characteristics that might also have an impact on outcomes. One might for example ask, “Why does one person join a microfinance programme and another not?” Or “Why are some people early joiners and other people late joiners?” The answers to these questions are probably multifaceted, but a reasonable hypothesis is that,

on balance, joiners and early joiners possess some unseen characteristics that non-joiners or late joiners do not, whether those be entrepreneurial drive, willingness to assume risk, a supportive home environment, or simple determination to improve one's life. The answer may also reflect expected net benefits of programme participation. Joiners and early joiners arguably have higher *ex ante* expected net benefits of participation than non-joiners or late joiners. They will, as a result, also tend to enjoy higher *ex post* net benefits. The point is that unobservable individual characteristics are quite possibly key in determining the impacts of programme participation.

The best method to control for selection bias stemming from unobservable characteristics is through random assignment of study participants into treatment and control groups. Random assignment among IA studies, however, is rare, suggesting that IA studies routinely overstate programme impact. The few IA studies using random assignment methods include Benus *et al.* (1994), US Department of Health and Human Services (1994c), MkNelly and Dunford (1999a, 1999b), and Coleman (1999b, 2001).

A good example of random assignment is Benus *et al.*'s (1994) assessment of self-employment demonstrations in Washington state and Massachusetts. Researchers invited unemployment claimants interested in self-employment to attend an information session that explained basic information about the risks and rewards of self-employment and the key features of the demonstration. At the conclusion of the session, claimants still interested were given an application for the programme. Those who completed the application on time and met eligibility requirements were then randomly assigned either to a treatment group eligible to receive business development services and financial assistance or to a control group that was not.

A simpler method to control for selection bias is to use pipeline clients as the control group, as recommended by the SEEP/AIMS tools. Because pipeline clients have self-selected themselves into the programme, they are presumed to share the same unobserved characteristics as existing clients. Pipeline clients have also been used as a control group in several published academic IA studies (Buckley 1996a, 1996b; Montgomery *et al.* 1996; Mosely 1996a, 1996b, 1996c, 2001; Mosely and Hulme 1998; and Copestake *et al.* 2001).

The use of pipeline clients, however, suffers from some important methodological weaknesses. First, it is most appropriate for cross-sectional studies. A longitudinal study requires that the control group, or at least part of it, not receive the treatment during the entire period of the study. It is probably not operationally feasible in most cases to withhold loans for the duration of a longitudinal study from new clients recruited through day-to-day programme operations. The exception is pipeline clients recruited specifically as part of a controlled, longitudinal IA study.

Second, pipeline clients do not account for why some people join early and some join late. Arguably, someone who joins the programme at or soon after its inception is different from someone who joins two years later. This is less of a problem for those pipeline clients who did not have the option to join earlier, who did not join earlier for reasons unrelated to unobservable determinants of success, or who are drawn from communities where the programme is not currently operating.⁹ The point is that in most cases, we do not know why late joiners have waited to join, and almost certainly the explanation frequently involves one or more unobservable traits. In summary, pipeline clients are a far from ideal control group; most will concede that a control group of non-clients is strongly preferable. Advocates of this approach, however, argue that it is a practical solution to situations in which significant or binding constraints exist.

A minimal approach to account for selection bias requires that control group members be drawn from the population of microentrepreneurs who are candidates to join the programme. Better yet is that control group members are drawn from the population of microentrepreneurs who satisfy specific programme eligibility requirements. The weakness of these approaches of course is that they assume incorrectly that all microentrepreneurs or those microentrepreneurs eligible to join the programme would self-select into the programme. Still, the probability that microentrepreneurs eligible to join the programme would self-select into the programme is greater than non-eligible microentrepreneurs. Thus while this approach will not eliminate selection bias, it will at least tend to reduce its prevalence.¹⁰

Given the large number of self-employed poor toiling in the informal sector in LDCs, there exists a large pool of microentrepreneurs who are legitimate candidates or who satisfy eligibility requirements to join microfinance programmes. Thus there is little justifiable reason for IA researchers in LDCs not to select the control group from this pool. In contrast, OECD countries have much lower incidence of self-employment or informal sector activity among the poor, which makes control group selection from among the pool of self-employed poor that much more difficult. In fact, none of the microenterprise assessments reviewed here drew their control group from this pool. Instead control group members were drawn from Temporary Aid for Needy Families (TANF) recipients (Raheim and Salome 1999), welfare recipients (Raheim and Friedman 1999), unemployment benefit recipients (Kosanovich and Fleck 2002), food stamp recipients (US Department of Health and Human Services 1994b), and Aid for Dependent Children (ADFC) recipients (Institute for Social and Economic Development 1994, US Department of Health and Human Services 1994c). Comparison of microenterprise programme participants to a random sample of government aid recipients almost certainly produces selection bias and overstatement of programme impact.

Failure to account for programme dropouts

Of the reviewed impact assessments, only Karlan and Alexander (2002), Himes and Servon (1998), and Mt. Auburn Associates (1998) included programme dropouts among survey respondents (neither Himes and Servon nor Mt. Auburn Associates, however, used a control group of non-clients.) Omitting programme dropouts from the treatment group introduces two potentially serious sources of selection bias, what Karlan (2001) refers to as *incomplete sample bias* and *attrition bias*. Incomplete sample bias arises because dropouts presumably have fared differently, and quite possibly worse, than those who remain. In contrast, the control group (whether non-clients or pipeline clients) includes some who will succeed and some who will fail. Thus IA studies that omit dropouts from the treatment group compare the programme's successes to a control group of both successes and failures. Consequently, incomplete sample bias produces systematic overstatement of programme impacts.

Attrition bias arises if dropouts are systematically different from those who remain, regardless of impact. If, for example, richer members tended to drop out more than poorer members, then the treatment group would include a higher percentage of poorer members than the control group, and *vice versa* if poorer members tended to drop out more than richer members. The result is systematic understatement of programme impact in the first case and systematic overstatement of programme impact in the second case.

To test the effect of incomplete sample and attrition bias on impact assessment findings, Karlan and Alexander (2002) compared a treatment group from a Peruvian MFI minus dropouts to a control group of the MFI's pipeline clients (as per SEEP/AIMS recommendations) and found statistically significant evidence of positive impacts. After adding dropouts back into the treatment group and comparing the two groups again, they found that the positive impacts disappeared.

Non-random programme placement

Programme placement is not random. Programme managers presumably base programme placement on a variety of strategic criteria, for example, consistency with institutional mission (*e.g.*, high density of very poor), logistical feasibility (*e.g.*, within reasonable distance of programme headquarters), financial attractiveness (*e.g.*, high density of self-employed), or likelihood of successful implementation (*e.g.*, relatively well-developed infrastructure or better access to markets). It is reasonable, moreover, to assume that MFIs will begin in and expand first to those locations that best satisfy the strategic criteria. Such locations arguably share a set of observable and unobservable characteristics that would not be present to the same

degree in a random sample of other locations. These characteristics in turn are potentially significant in explaining relative outcomes from programme participation. Thus, controlling for selection bias also requires selection of a control group from communities that share similar observable and unobservable characteristics as the treatment group community.

The best way to control for non-random placement bias is to randomise programme placement. This approach may appear administratively impractical, but it need not be, particularly if designed to exploit natural limits to programme expansion. Consider, for example, an MFI that plans to expand to x number of locations over the next two years. All x locations satisfy the programme's placement criteria. Due to resource constraints, however, the programme cannot expand to all x locations at once, so its plans call for it to expand to y locations this year and $x-y$ locations next year. Since the MFI is largely indifferent to the order of expansion, randomizing the selection process both satisfies the programme's expansion criteria and controls for non-random placement bias.

The only examples of randomised programme placement among the IA studies reviewed here are MkNelly and Dunford (1999a, 1999b) in assessments of microfinance programmes in Bolivia and Ghana. Both assessments followed the same methodology. Programme management selected communities to which it would expand over the next two years. Programme staff next visited each community to recruit participants into the programme. In each instance, the programme staff made clear that the community might be assigned to the control group that would not receive the programme for two years. Programme staff then collected baseline data from all eligible persons who elected to join the programme. Following baseline data collection, researchers stratified the study communities according to key community characteristics such as size, access to the main road, distance from a market, and access to water. Finally, researchers randomly assigned communities to control and treatment communities in a way that minimised the differences between key community characteristics.

Coleman (1999, 2001b) used a somewhat different approach in his assessment of two microfinance programmes in Thailand. The treatment community included eight villages with access to the programmes for two to four years. The six control communities were pre-selected to receive the programme one year after they were identified. Villagers in the control communities self-selected whether to participate in the programme. To account for the possibility that the order in which the fourteen villages received the programme was not random, Coleman collected a third sample of non-participants in each of the villages. Coleman found that naïve estimates that did not account for self-selection or non-random programme placement significantly overestimated programme impact on several outcome variables,

including overall wealth, land holdings, non-land farm assets, savings, and household income. He concluded that unobservable characteristics, not participation in microfinance programmes, were the most significant determinants of small business income.

Econometric techniques may also be used to control for selection bias. An example is the approach used originally by Pitt and Khandker (1994) to assess the impact of three Bangladesh programmes (and subsequently used by McKernan 1996; Pitt *et al.* 1998; Pitt and Khandker 1998; Khandker *et al.* 1998; Pitt *et al.* 1999; Morduch 1998; and Khandker 2001). Their approach exploited programme rules that excluded households with more than a fixed amount of assets from programme participation. In effect, they compared outcomes between eligible and ineligible households in programme villages and outcomes between eligible and ineligible households in non-programme villages and then compared the two differences to each other. They attributed any difference between these two differences to programme participation. Like Coleman, their study showed that naïve estimates that fail to account for selection bias significantly overestimate impact (other IA studies using econometric techniques to control for selection bias include Lapar, Graham, and Meyer 1995; Lapar *et al.* 1995; Zeller *et al.* 1996; and Zaman 2001.) The downside of econometric approaches like Pitt and Khandker's is the sophistication of their econometric methods, which are accessible only to a relatively small group of equally sophisticated methodologists, but not accessible to most policymakers and most certainly not to the average practitioner.¹¹

Contamination bias

Contamination bias occurs when the control group becomes contaminated by contact with the treatment group. Contamination can occur in several ways; for example, if control group members are acquainted with treatment group members, members of the two groups share acquaintances in common, the programme initiates contact with control group members or *vice versa*, or knowledge of the programme spreads through formal or informal social networks. Contamination also occurs when programme participation creates positive or negative externalities that influence the behaviour or outcomes of non-participants.

Once control group members become contaminated, the researcher can never be certain whether and how the control group's behaviour and other outcomes have been influenced as a result. To the extent contamination produces better or worse outcomes among treatment group members, it will create systematic understatement or overstatement of actual programme impacts. Contamination bias may be dealt with easily enough by locating the control group away from the treatment group. The farther away the two, the

less the chance of contamination there is (although this has to be weighed against the increase in cost and logistical hassle).

Recall bias

Survey research requires that respondents recall information about their socioeconomic conditions, behaviours, attitudes, and social relationships. Recall responses may be biased, for several reasons. Survey respondents display natural and reasonable tendencies to 1) want to please the interviewer, cast themselves in a good light, or avoid revealing embarrassing information, 2) suspect researchers' motives, 3) game the process, 4) make wild guesses, or 5) give any answer to avoid prolonging unwanted intrusions on their time. Generally, the more personal or intrusive the questions (*e.g.*, sexual practices) or the more difficult to estimate (*e.g.*, household income), the more researchers can expect inaccuracies and biases to creep into the responses.

Using programme staff to interview clients, as recommended by the SEEP/AIMS tools, substantially raises the risk of response bias, and using staff to interview their own clients practically guarantees it. This is not to suggest that programme staff should not be used. In fact, using programme staff to conduct research is often a practical and necessary concession to programme constraints. Moreover, integrating IA or market research into programme systems or weaning one's self from external evaluators may very well require use of field staff to collect impact data. Using field staff to interview their own clients, however, should be done only as a last resort. Programmes using field staff to conduct impact research need to be fully apprised of the risks it poses and be prepared to do what it can to mitigate those risks. The best approach to mitigate these risks is to mix in heavy doses of training, monitoring, and cross-checking of data.

Even where respondents are inclined to give good faith responses, they may or may not be able to recall information with reasonable accuracy. The longer the time period elapsed, the more difficult to estimate accurately. Seasonality can also play havoc in that responses differ depending on the time of year. Strategies to mitigate this source of bias are to use shorter time periods and to survey at different times of the year to reflect key seasons or crop cycles (see, for example, Coleman 1999, 2001b; Pitt *et al.*, 1998; and Mustafa *et al.* 1995).

Loan fungibility

Loans received by programme participants are typically intermingled with other sources of household income, to be spent according to the household's livelihood needs and spending priorities. In other words, loans are not necessarily earmarked for investment in participants' enterprises. In

similar manner, funds for loan repayment do not necessarily come from enterprise cash flows but often come from other sources of household income. Loan fungibility greatly complicates the ability of IA to make a direct link between receipt of a loan and changes in household income, consumption, asset accumulation, individual empowerment, etc. The current state of impact assessment methodology includes no established procedure to account for loan fungibility.

Loan fungibility becomes less a problem, however, if the focus of IA is the household economic portfolio (Chen and Dunn 1996). The concept of household economic portfolio explicitly recognises the fungibility of money as a vital component of poor households' livelihood and coping strategies. It is less interested in the route of the causal chain of impact per se than in the how the impacts ultimately manifest themselves at the different levels of analysis.¹²

Another problem caused by loan fungibility is to make even honest responses to survey questions misleading. To illustrate this point, Coleman (2001a) gives the example of a programme participant who uses her programme loans to pay her children's school fees in place of selling assets to pay the fees, as is her normal practice. In this case, the true incremental benefit of programme participation to the woman is the preservation of assets, which is not measured, and not payment of school fees, which is measured.

Of the studies reviewed here, only Diagne (1998) and Zaman (2001) explicitly deal with loan fungibility. Diagne's approach was to circumvent the problem by making the relevant treatment access to credit rather than receipt of credit based on the reasoning that changes in outcomes because of access to credit were easier to isolate and identify than changes in outcomes because of receipt of credit. Zaman in contrast used an econometric approach based on a household economic portfolio model. His model assumes that money borrowed is spent as needed by utility maximizing households and that by controlling for other factors through application of econometric procedures it is possible to attribute specific outcomes to receipt of the loan. Another possible approach to account for loan fungibility is to collect information on use of loan funds and sources of repayment through additional survey questions or qualitative research.

Longitudinal vs. cross-sectional assessment

Presumably, impacts occur over time, and longitudinal assessments that track outcomes over different points in time shed more light on this process and how it unfolds than do cross-sectional assessments performed at a single point in time. If a longitudinal study is not possible, researchers can proxy a time-series with a cross-sectional assessment by purposefully selecting participants with specific years of experience in the programme and

comparing them either to a control group of non-clients or to pipeline clients. The SEEP/AIMS tools, for example, recommend that researchers select participants with one and two years of programme experience as treatment group members under the assumption that these represent useful points in time where significant impacts might be observed.¹³

The virtues of cross-sectional impact assessments are relatively low cost and low data collection demands. In longitudinal assessments, panel attrition (treatment and control group members dropping out of the study) can also be a problem.¹⁴ Typical ways to deal with panel attrition are tracking down panel dropouts and/or sampling a larger number than otherwise needed, both of which add yet more costs relative to cross-sectional studies.

Despite preference for longitudinal assessments, several cross-sectional impact studies have been published by well-respected researchers or in well-respected academic journals, and cross-sectional assessments make up the bulk of microfinance programme assessments; 63 microfinance assessments were cross-sectional compared to only 27 longitudinal assessments. In contrast, 11 of the reviewed microenterprise assessments were longitudinal and 9 were cross-sectional. Where longitudinal assessments are not feasible, case cross-sectional assessments are an acceptable second best alternative. Best practice holds, however, that where longitudinal assessments are feasible, they should be done.

Other methodological shortcomings

The previous section reviewed the methodological pitfalls common to survey-based impact assessments. Understanding these pitfalls, their causes, and their cures is essential for policymakers to evaluate the validity of IA studies so as to determine their usefulness in informing public policy. In addition to these pitfalls, impact assessments suffer from a variety of other methodological shortcomings that also affect their usefulness for public-policy decision making. These methodological shortcomings have less to do with issues of scientific precision than with providing policymakers with a thorough assessment of the benefits and costs of programme participation.

Ideally, scarce public funds are allocated to those social programmes that yield the highest net social welfare, where net social welfare is defined as the present value of programme benefits minus the present value of programme costs. Estimating net social welfare in turn implies the following: 1) identification of all relevant programme benefits and costs, 2) estimation of relevant programme benefits and costs, 3) conversion of relevant benefits and costs into standardised units so as to permit comparisons within and across programmes, and 4) weighting of relevant programme benefits and costs to reflect social values and priorities. In contrast, most IA studies give only a very

limited picture of programme benefits and virtually no information on programme costs; they do not attempt to convert benefits into standardised units, and they treat all benefits and all costs as equal. Each of these shortcomings, and its implications, is discussed briefly below.

Omission of programme benefits

All impact assessments make choices about which benefits to examine. Most examine some combination of enterprise and household level benefits, many also examine individual level benefits, but nearly all omit community-level benefits, even though evidence suggests that they can be significant in the aggregate.¹⁵ The choice of benefits examined may be based on a number of criteria, for example, donor or programme priorities, industry convention, or personal interest. But whatever the criteria, the end result is that important benefits are inevitably omitted from programme assessments

Omission of important programme benefits gives policymakers an incomplete picture of programme impact. Absent this, the only way for policymakers to form a complete picture of programme impact is to cobble together findings from assorted programme assessments that examine different types of benefits at different levels of analysis. While this approach can be helpful, it can at best give only a very broad picture of impact, and its usefulness in assessing the impact of a particular programme is limited, given the significant contextual disparity in which programmes operate.

Omission of programme costs

Most impact assessments do not mention, let alone estimate, programme costs. Relevant programme costs include the present value of administrative costs and the monetary and opportunity cost of donated/invested funds, soft liabilities, grants-in-kind, price and non-price transaction costs borne by programme participants, and displacement costs (benefit incurred by programme participants at the expense of non-participants). Calculating direct programme costs and price costs to participants is relatively straightforward, but estimating grants-in-kind, participant non-price transaction costs, and displacement costs will tax even the most conscientious researcher, which perhaps explains why so few have done it.

Take non-price transaction costs as an example. Group lending programmes typically require participants to form groups, participate in weekly or biweekly group meetings, and monitor and enforce group loan performance. Such non-price transaction costs are hard to measure, even though they impose significant burdens on programme participants. Harder to measure yet are displacement costs, which can occur, for example, when programmes draw large numbers of the self-employed into sectors attractive

to low-skilled microentrepreneurs and characterised by low barriers to entry, high competition, and low profits. Displacement costs can be significant, reaching as high in some cases as one-half the net benefits accruing to programme participants (Bendick and Egan, 1987). Of the reviewed impact assessments, only Kosanovitch and Fleck (2002) consider displacement costs.

To the extent programme costs are available or can reasonably be estimated, they should arguably be included in impact assessments, if only to provide some baseline for comparison, such as cost per outcome. If costs cannot reasonably be estimated, researchers might at a minimum be expected to identify potential costs, perhaps give some estimate of their order of magnitude, and explain how they might affect the analysis were they to be included.

Kilby and D'Zmura (1985) is the sole microfinance programme assessment reviewed to conduct a benefit-cost analysis. Kilby and D'Zmura assess direct and indirect benefits primarily as measured by value added to assisted enterprises and value added to enterprises outside the project. The former consist of wages, rent, interest, and profit adjusted for the opportunity cost of labor,¹⁶ and the latter of purchases of factors of production made by assisted enterprises and purchases of consumer goods made with direct factor income earned by assisted enterprises. Other benefits considered include training, price reduction, diversion benefits (the benefits derived from diverting a microenterprise loan to another purpose), and weighted wages for the very poor. Costs assessed include all administrative expenditures, bad debt, and capital erosion (the effective interest rate below the rate of inflation).

Several reviewed microenterprise programme assessments conducted at least some form of benefit-cost analysis. Else and Clay-Thompson (1998) calculated the cost per unit of output for clients served, jobs and businesses created, businesses assisted, and loans made. Similarly, the US Department of Health and Human Services (1994) compared the increase in food stamp earnings of demonstration participants to programme administrative costs and the increases in public assistance payments to demonstration participants.

The most complete benefit-cost analysis performed among reviewed microenterprise assessments is Kosanovich and Fleck's (2002) assessment of Self-Employment Assistance Programmes in Maine, New Jersey, and New York. The authors evaluated benefits and costs to programme participants, state governments, and non-participants. Participant benefits assessed included the income gain from self-employment or wage/salary employment, professional development, work satisfaction, and community economic development. Government benefits assessed included increased tax revenue and the reduction in welfare transfers. Participant costs assessed were financial costs of programme administration, training, counselling and the opportunity costs borne by participants who forgo work search and possible

reemployment opportunities while pursuing self employment. Government costs assessed were unemployment insurance payments, programme administration, training, and counselling. Finally, displacement impacts assessed included changes in unemployment insurance benefits and tax payments benefiting participants but paid for by non-participants and the reduction in long-term welfare or unemployment insurance payments benefiting non-participants but at the cost of participants.

Lack of common standards for comparison

Comparing benefits and costs within and across programmes (whether to another microfinance or microenterprise programme or to some other type of development or employment programme) requires that benefits and costs be converted into standardised units. Typically, this entails conversion into monetary units. Without some common standard, policymakers have no way other than their own subjective guesstimates to compare, say, the value of a job created to an increase in a participant's self-esteem in one programme or the value of jobs created by the same programme to the value of increased participant access to health care in another programme. Standardisation of benefits and costs has the added advantage of forcing researchers to be explicit about their assumptions and judgments, thereby both improving analysis and facilitating *ex post* assessments of methodological rigor.

Failure to weigh benefits and costs

Impact assessments do not distinguish between the relative worth of outcomes, instead treating all programme benefits and all programme costs as equal. Benefits as disparate as increased household consumption and increased participant self-esteem, for example, are treated as equal, as are similar, but clearly distinguishable, benefits, such as full-time and part-time jobs created. In fact, neither all benefits nor all costs are equal. Society attaches greater value to some programme benefits and greater cost to some programme costs than to others. Arguably, therefore, impact assessments should reflect social values via some kind of weighting scheme. Short of an explicit weighting scheme, but still helpful, would be some discussion of relative weights and how they might affect the analysis. Nonetheless, neither weighting schemes nor any discussion of the relative weight of programme outcomes can be found in reviewed impact assessment.

Miscellaneous other methodological shortcomings

Impact assessments are subject to a number of other miscellaneous methodological shortcomings. These methodological shortcomings pertain for the most part to choices made by IA researchers about which information

to report and how to report it. Choices made about the presentation of information can materially influence assessment findings and how findings are interpreted.

Choices made about presentation of information reflect a number of factors, including the skill of the researcher, the objectives of the assessment, subjective decisions about the relevance of information, or the biases of the researcher or the funder. While we cannot assume that choices about presentation of information are necessarily influenced by the biases of the researcher or funder, neither can we assume that biases never figure in the choices. This issue takes on greater relevance when we realise that in many cases impact assessments are funded or conducted by people or organisations that openly advocate microenterprise development. This combined with all the other factors that influence choices about presentation of information suggest that it is useful for policymakers to be familiarised with the ways in which such choices can influence assessment results. Schreiner (2002) catalogues several of the more common examples found in assessments of US microenterprise programmes. The same practices, however, can be found to greater or lesser degrees in other microfinance and microenterprise programme assessments, such that it is useful to summarise Schreiner's arguments below.

Selective Presentation of Programme Outcomes. Impact assessments at times selectively present information in ways that bias the analysis and conclusions. One example is the failure to distinguish between stocks (*e.g.*, loans outstanding or current trainees) and flows (*e.g.*, loans disbursed or people trained). Flows aggregate past performance with current performance, thereby producing a distorted picture of current performance. Flows also exceed stocks at any point in time, such that reporting flows rather than stocks gives a more favourable view of programme performance. Whether it is appropriate to report the one or the other depends on the question asked. Regardless, researchers should report the unit being used and why.

A similar practice is to report aggregate figures as opposed to aggregate figures adjusted for the number of programme participants. A finding, for example, that sales of programme participants totalled \$3.5 million sounds more impressive than a finding that average sales per programme participant totalled \$12 000.

Another case of selective presentation is the use of half-statistics. Interpretation of findings can be influenced by how they are "spun". For example, reporting that "over one-half" of programme participants increased enterprise returns is a positive spin to the more negative finding that "nearly one-half" of programme participants did not increase enterprise returns. Reporting broad summary statistics when disaggregated statistics are

available is yet one more practice found in impact assessments. For example, the loan repayment rate is a commonly reported measure of portfolio quality; however, this is a broad summary statistic that hides crucial information discernible in far better measures, such as aged portfolio at risk.¹⁷

A final example of selective presentation is the failure to report programme dropouts. Dropouts are perhaps the simplest, most effective way to gauge whether the programme creates value for participants. It is reasonable to conclude that programmes with high rates of client turnover are less effective at creating value and have less impact than programmes with relatively low rates of client turnover. Moreover, exclusion of dropouts from the impact analysis almost certainly produces systematic overstatement of programme benefits.

Misestimation of Programme Benefits. It is not unusual for impact assessments to misestimate programme benefits. Schreiner points out, for example, that some microenterprise assessments report new businesses starts as if all are attributable directly to programme participation, overlooking that some participants enter the programme with pre-existing enterprises, others start businesses after dropping out, and the majority of new enterprises fail within a few years.

Another example is the misreporting of enterprise income, such as the practice of 1) reporting income levels instead of changes in income, 2) reporting enterprise income instead of enterprise returns, and 3) failing to define income clearly. In the first case, reporting income levels overstates programme benefits unless income was zero prior to joining the programme. In the second case, enterprise income gives a distorted picture of impact because it does not account for returns to time worked and capital invested. Absolute income levels that appear high may not appear so high once adjusted for time worked and capital invested. Finally, lack of clarity about the definition of income allows researchers to report a variety of outcomes as enterprise income, some more favourable than others. A prime example is the practice of reporting unadjusted enterprise revenues as income or as a proxy for income, which inevitably skews findings upwards.

Judging IA methodology

Knowledge of the methodological pitfalls and shortcomings of impact assessments is useful for judging the quality of programme assessments. Before casting judgment, however, it bears repeating that while methodological purity is a worthy ideal, it is rarely achieved in practice. Virtually all impact assessments suffer from one methodological shortcoming or another. Some are blatant, and some require a more careful reading to catch. Some are the result of environmental or resource constraints, some are

the result of subjective choices made by researchers, and some are the result of both.

In practice, most researchers are forced to make concessions of one kind or another to resource or environmental constraints, and they must make subjective choices in how to deal with them. More broadly, all researchers must make a myriad of subjective choices about how to design and implement assessment studies, analyze assessment data, and report assessment findings. While there is little disagreement among researchers on what major methodological issues and pitfalls are or what constitutes ideal methodology, there is substantial disagreement on what constitutes acceptable methodology, what acceptable tradeoffs between rigor and cost are, what methodological concessions one might legitimately make to resource or environmental constraints, or which subjective choices are valid and which not.

Publication in peer-reviewed academic journals by no means implies methodological seal of approval. Editorial boards of academic journals have different methodological standards, as do the reviewers they use. Few published peer-reviewed academic impact assessments satisfy everyone's standards of methodological rigor.¹⁸ But if the methodological gulf between academic researchers is wide, the gulf between academics and practitioners is a positive chasm. Not that practitioners do not care about scientific validity, they do to a degree. It is just a luxury that most of them feel they cannot afford. They are content to let academics hash out methodological niceties, while they concentrate on running programmes. Besides, what manager in any organisation has ever made key programme decisions with a ± 0.05 degree of certainty?

Given the disagreement about appropriate methodological rigor, policymakers should not be expected to sort out what professional researchers and practitioners cannot. What policymakers can do, however, is to insist on full disclosure of methodological approaches and shortcomings, constraints faced, and tradeoffs and subjective choices made. Full disclosure promotes transparency, and transparency promotes informed judgments about methodological appropriateness and informed interpretation of assessment findings. Full disclosure is an easy, yet critical, objective to which all impact assessments should aspire. As Schreiner (2002) has noted, "The heart of the social-scientific method is not experiments but explicitness." (p. 69)

Recommendations

Taking all of the above into account, the following recommendations are proposed to help policymakers judge the quality/rigor of impact assessments and interpret their findings and to guide their efforts at improving IA practice and usefulness.

Encourage the use of mixed method approaches. The heavy reliance on impact surveys probably reflects as much or more the training, inclinations, and biases of IA researchers than any inherent methodological superiority. Qualitative methods (whether based in the humanities tradition or PLA) have a great deal to offer, particularly in terms producing a deeper and more nuanced understanding of impact. Qualitative methods also can do much to overcome the problems of loan fungibility, and they have been used too infrequently for this purpose. One of the most insightful impact assessments reviewed here was Todd's (1996) ethnographic study of Grameen Bank members; it provided a richness of understanding missing from most survey-based studies. Where time or resources do not permit ethnographic studies of this sort, PLA or other rapid assessment methods can similarly yield insightful and useful information about the process through which impact occurs and how it is manifested at different levels of analysis. The impact survey should remain a primary tool, but it need not be as dominant as it has been to date.

Use valid control groups. If the purpose of a survey-based assessment is to attribute impact to programme participation, a valid control group must be used. It makes little sense to invest scarce public funds in an impact assessment that cannot hope to answer the questions asked of it. At a minimum, a valid control group should consist of microentrepreneurs eligible to join the programme. Pipeline clients satisfy this minimum requirement where binding constraints exist, but efforts should be made to select pipeline clients who have not had the opportunity to join the programme earlier, such as pipeline clients from communities into which the programme has only recently expanded.

Control for selection bias. All reasonable effort should be made to control for all forms of selection bias. Random assignment and random programme placement should be used where possible. Random assignment need not be costly or overly intrusive. Coleman (2001a), for example, estimates that replication of his methodology could easily be implemented at a cost of only \$25 000-\$50 000, which is close to the minimum range that a high-end impact assessment would cost anyway. The study by Benus et al. (1994) demonstrates that random assignment is possible in an OECD context as well. Even were controlling for selection bias to cost more, it is worth an extra increment of spending. Again, it makes little sense to invest money in a programme assessment that can be predicted ahead of time to yield questionable or invalid findings, particularly when valid findings could be produced for only a slightly larger investment. Researchers should also show evidence of controlling for selection bias and be candid in discussing its implications.

Include programme dropouts in the treatment group. The nearly universal exclusion of programme dropouts from treatment groups is a serious methodological shortcoming that has almost assuredly produced systematic overstatement of impact. Including programme dropouts requires additional

investment of money and effort, but it is certainly doable in most cases, and the return in terms of statistical precision is likely to be more than worth the additional investment.

Locate the control group a sufficient distance from the treatment group. To avoid contamination bias, the control group should be located far enough away, within reason, from the control group to minimise the probability of contamination bias.

Allocate sufficient money to conduct methodologically sound impact assessments. If policymakers or funding agencies require programmes to conduct impact assessments, they should allocate sufficient funding to implement them and to implement them in a methodologically sound way. Too often funding agencies expect convincing evidence of impact but do not allocate enough money (or at times any money) to implement a valid impact assessment. All this tends to produce is half-hearted effort and cynicism about impact assessment.

Perform longitudinal assessments where feasible. Cross-sectional assessments are acceptable, but they should be the clear second choice and reserved for those situations in which longitudinal assessments are not feasible.

Agree ahead of time on a clear and complete scope of work with programme management and IA researchers. Methodological problems can arise because of differing expectations or misunderstandings among funding agencies, programme management, and researchers. Thus many methodological problems can probably be avoided if funding agencies work more closely with programme management and researchers to define a scope of work that covers, at a minimum, a) the objectives of the assessment, b) methodological options consistent with the assessment objectives, c) levels of impact to be assessed and corresponding indicators to be used, d) field constraints (objectives and methodology may need to be negotiated in light of field constraints), and e) reporting requirements (what is to be reported and how).

Perform more rigorous benefit-cost analyses. Whether microenterprise development warrants large expenditures of public money has yet to be determined owing to the dearth of good benefit-cost analyses. While it is helpful to know whether and to what extent microfinance and microenterprise programmes benefit participants and non-participants, this information does not necessarily allow informed choices about whether and to what degree to fund the programmes relative to other policy options. This requires rigorous programme assessments that assess a variety of benefits and costs. It also requires some method to standardise and weight findings.

Schreiner (2002) points out that emphasis on rigorous benefit-cost analysis creates a potential prisoner's dilemma for microfinance and microenterprise programmes in that greater rigor (and disclosure) places it at a disadvantage among policymakers relative to competing policies that do

not adhere to correspondingly high levels of assessment rigor. Thus a subsidiary recommendation is that policy makers hold all programme assessments to high standards of methodological rigor. Regardless, poor assessment methodology in competing social programmes does not relieve microfinance or microenterprise programmes of the ethical responsibility to make reasonable and valid efforts to justify the scarce public funds they receive.

Insist on full and candid disclosure. IA researchers should be held accountable to disclose completely and candidly all information relevant for non-specialists to understand and interpret assessment findings. Things to be disclosed include, at a minimum, a) methodologies chosen, why they were chosen, and their implications, b) weaknesses of methodologies chosen and their implications, c) constraints faced and tradeoffs made and their implications, d) subjective decisions made and their implications, e) methodological shortcomings and their implications, and f) biases or conflicts of interest (real or potential) that may or may not influence methodology, analysis, or interpretation and presentation of findings

Develop IA capacities of practitioner organisations. Any set of recommendations also needs to take into explicit account the drawbacks of the proving approach to impact assessment, principal among them its cost, length, difficulty, and technical requirements, all of which have limited its usefulness to programme management. Given the resource and technical constraints of practitioner organisations, an approach to IA that places scientific validity as the primary criterion is unlikely to be adopted by practitioner organisations to any significant extent. On the other hand, an approach to IA that focuses on the usefulness of IA to programme management, and which can offer reasonable guarantees to management that its benefits in terms of improved programme effectiveness outweigh its costs, is more likely to be adopted by practitioner organisations on a wider scale. This fact needs to be acknowledged at the outset.

Thus if the primary objective of IA is to prove impact, and policymakers are content with assessing a relatively small number of programmes on a sporadic basis, then a reliance on scientific IA is probably appropriate. But if an objective of IA is to improve impact, and policymakers want more or less regular information on programme performance across a wider range of programmes, then a reliance on scientific IA is probably not appropriate. In the latter case, public investment in IA should concentrate increasingly on developing the IA and market research capacity of practitioner organisations, and assisting in the development and implementation of methodologically sound, practical, low-cost, practitioner-friendly, and useful methodologies. This requires in turn that policymakers work with practitioner organisations to understand what their needs and constraints are and work jointly with them to develop a funding and technical assistance policy that passes on critical knowledge and skills so that

programmes can conduct their own assessments in a way that is useful to them and to clients. Moreover, for policymakers wishing to promote programme sustainability, improving impact through IA and market research offers one of the most potentially effective policy approaches.

Notes

1. A high rate of informal sector activity by the poor is a particular characteristic of LDCs.
2. For other reviews of IA methodology, see Gaile and Foster (1996), Khandker (1998), Hulme (2000), Coleman (2001a), and Schreiner (2002).
3. Perhaps the best source for understanding the conceptual foundations of impact assessment and its many complex relationships is the series of conceptual papers commissioned and published by the Assessing the Impact of Microenterprise Services Project (AIMS), funded by the Office for Microenterprise Development at USAID. AIMS publications cover topics such as assessing impact at the enterprise and household levels (Inserra 1996) and at the individual level (Chen 1997), income and assets as impact indicators (Barnes 1996, Little 1997), measuring profits and net worth of microenterprises (Daniels 1999), assessing impacts within a “household economic portfolio” (Chen and Dunn 1996), and microfinance and risk management (Cohen and Sebstad 1999). All AIMS publications can be downloaded at www.usaidmicro.org.
4. The terms “scientific method”, “humanities tradition”, and “PLA” used here to describe IA methodological approaches, as well as the ensuing summaries of each approach, are based on Hulme’s (2000) excellent discussion of IA methodologies.
5. Many “participatory” studies are participatory in name only. The term *participation* is fast reaching cliché status – oft used and oft devoid of substantive meaning.
6. SEEP refers to the Small Enterprise Education and Promotion Network. SEEP is a professional network of North American Private Voluntary Organisations engaged in the promotion of microfinance. SEEP worked in conjunction with the AIMS research team to develop the SEEP/AIMS tools.
7. See Nelson et al. (2001) for an in-depth description of and implementation instructions for each of the SEEP/AIMS tools.
8. MicroSave Africa, for example, has developed a set of qualitative market research tools using largely PLA methodologies. To date, MicroSave has conducted several training workshops around the globe involving dozens of MFIs. Information on the MicroSave market research tools can be viewed at www.microsave-africa.com.
9. For example, a late joiner may only recently have moved into the community, or she may have only recently started a business, or she may not have heard about the programme until recently.
10. An example of how not to select the control group is Kosanovich and Fleck (2002) who selected control group members from among individuals offered enrolment in the microenterprise programme but who declined. This approach virtually guarantees significant selection bias.
11. Pitt and Khandker’s approach was criticised by Morduch (1999), who questioned whether the three programmes assessed actually enforced the asset-based eligibility rules, and who found other problems with Pitt and Khandker’s

econometric methodology. Pitt's (1999) response to Morduch defended the original approach and purported to demonstrate how it was superior to the alternative methodology proposed by Morduch.

12. Hulme (2000) points out that diversion of loans for consumption may in fact produce higher returns than investment in enterprise assets, for example, if spending on consumption takes the form of investment in human capital (*e.g.*, school fees, health care), replaces borrowing from other sources at a higher cost, or is used to acquire basic needs (*e.g.*, food, medicine) necessary to sustain adequate levels of labour productivity.
13. It would also be useful to include three-year clients and on up in the treatment group; however, in many MFIs, three-and-four year clients on up are hard to come by, owing to high client dropout rates.
14. This author, for example, participated in a longitudinal impact assessment of a microfinance programme in Tanzania that suffered from over 80 per cent panel attrition after just over one year into a two-year study.
15. Woller and Parsons (2002), for example, find that microfinance programmes can contribute from hundreds of thousands to millions of dollars to local economies via direct expenditures and income multipliers. Other impact assessments examining community-level impacts include Coleman (1999, 2001b), Khandker (1996, 2001), Khandker *et al.* (1998), Kilby and D'Zmura (1985), Morduch, (1998), Pitt *et al.* (1999), Schuler and Hashemi (1994), Schuler, Hashemi, and Riley (1997), and Zeller *et al.* (1996).
16. The opportunity cost of labour occurs when new employees leave a previous job and are not replaced, or are replaced by less productive workers.
17. For a good critique that demonstrates how loan repayment rates can be misleading, see Rosenberg (1999).

Appendix

Recommendations for mid-range impact assessments

In April 1997 and April 1998 the Consultative Group to Assist the Poorest (CGAP) conducted two virtual meetings of microfinance experts to develop methodological guidelines for midrange assessments (Cohen and Gail, 1998). They reached consensus on the following guidelines:

1. Use some form of time perspective. Allow enough time for impacts to occur (both in terms of client participation and programme maturity).
2. Use some form of comparison group. Non-clients are preferable to pipeline clients where possible. The higher cost and other limitations of this approach, however, are well-recognised.
3. Tailor assessments to the specific context being studied.
4. Begin with a small set of indicators that have demonstrated validity in previous IA studies and that are relatively easy to collect. Incorporate new indicators as appropriate.
5. Collect baseline indicators, if possible when clients enter the programme. If not possible, use retrospective information.
6. Use interval-level data where possible.
7. Make greater use of IA as a management tool for generating information that is useful for programme improvement; for example, incorporate client satisfaction into IA studies.
8. Employ methods to establish plausible association between programme participation and observed outcomes.
9. Incorporate client satisfaction as part of IA.
10. Use a carefully designed IA that ensures transparency and external review/oversight.

11. Incorporate plans for IA into programme design and implementation as early as possible.
12. Measure direction of change where exact change cannot be estimated.
13. Build local capacity to conduct IA, both internal and external to the organisation.

References

- ASHE, Jeffrey and Lisa PARROTT (2001), *PACT's Women's Empowerment Program in Nepal. A Savings and Literacy Led Alternative to Financial Institution Building*. Washington, DC: PACT.
- ASHE, Jeffrey and Michael MACINTYRE (2000), *Working Capital. Building the Grass Roots Economy of Low Income Communities. Part 1: Membership Characteristics and Program Impact*. Cambridge, MA: Working Capital.
- BARNES, Carolyn (1996), "Assets and the Impact of Microenterprise Finance Programs". AIMS Paper. Washington, DC: Management Systems International.
- BARNES, Carolyn (2001), "Microfinance Program Clients and Impact: An Assessment of Zambuko Trust, Zimbabwe". AIMS Paper. Washington, DC: Management Systems International.
- BARNES, Carolyn, Gaile MORRIS and Gary GALE (1999), "An Assessment of Clients of Microfinance Programs in Uganda". *International Journal of Economic Development*, Vol. 1, No. 1, 80-121.
- BENDICK, M. and M.L. EGAN (1987), "Transfer Payment Diversion for Small Business Development: British and French Experience". *Industrial and Labour Relations Review*, Vol. 40, No. 4, 528-542.
- BENUS, Jacob M., Michelle WOOD and Neelima GROVER (1994), *A Comparative Analysis of the Washington and Massachusetts UI Self-Employment Demonstrations*. Bethesda, MD: Abt Associates.
- BLAIR, Amy Kay and Joyce KLEIN (2001), *Microenterprise as a Welfare to Work Strategy: Client Characteristics*. Washington, DC: The Aspen Institute.
- BOLNICK, Bruce R. and Eric R. NELSON (1990), "Evaluating the Economic Impact of a Special Credit Programme: KIK/KMKP in Indonesia". *The Journal of Development Studies*, Vol. 26, No. 2, 299-312.
- BUCKLEY, Graeme (1996a), "Financing the Jua Kali Sector in Kenya. The K-Rep Juhudi Scheme and Kenya Industrial Estates Informal Sector Program". In *Finance Against Policy, Volume II: Country Case Studies*, edited by David Hulme and Paul Mosley, 271-332, London: Routledge.
- BUCKLEY, Graeme (1996b), "Rural and Agricultural Credit in Malawi. A Study of the Malawi Mudzi Fund and the Smallholder Agricultural Administration". In *Finance Against Policy, Volume II: Country Case Studies*, edited by David Hulme and Paul Mosley, 333-408, London: Routledge.
- BUVINIC, Mayra, Marguerite BERGER and Cecelia JARAMILLO (1989), "Impact of a Credit Project for Women and Men Microentrepreneurs in Quito, Ecuador". In *Women's Ventures*, edited by M. Berger and M. Buvinic, 222-246, West Hartford, CT: Kumarian Press.

- CHAMBERS, Robert (1997), *Whose Reality Counts? Putting the First Last*. London: IT Publications.
- CHEN, Martha Alter (1997), "A Guide for Assessing the Impact of Microenterprise Services at the Individual Level". AIMS Paper. Washington, DC: Management Systems International.
- CHEN, Martha Alter and Elizabeth DUNN (1996), "Household Economic Portfolios". AIMS Paper. Washington, DC: Management Systems International.
- CHEN, Martha Alter and Donald SNODGRASS (2001), "Managing Resources, Activities, and Risk in Urban India: The Impact of SEWA Bank". AIMS Paper. Washington, DC: Management Systems International.
- CHURCHILL, Craig F. (1995), "The Impact of Credit on Informal Sector Enterprises in South Africa: A Study of Get Ahead Foundation's Stokvel Lending Programme". MA Thesis, Worcester, MA: Clark University.
- CLARK, Peggy and Tracy HUSTON (1993), *Assisting the Smallest Businesses: Assessing Microenterprise Development as a Strategy for Boosting Poor Communities*. Washington, DC: The Aspen Institute.
- CLARK, Peggy, Amy KAYS, Lily ZANDNIAPOUR, Enrique SOTO and Karen DOYLE (1999), Washington DC: The Aspen Institute.
- COHEN, Monique (1999), "Opening Up the Impact Assessment Agenda". AIMS Paper. Washington, DC: Management Systems International.
- COHEN, Monique and Gary GAILE (1998), "Highlights and Recommendations of the Second Virtual Meeting of the CGAP Working Group on Impact Assessment Methodologies: April 14-28, 1998. Developing Lower Cost Microenterprise Impact Assessment Methodologies for Microenterprise Programs". AIMS Paper. Washington, DC: Management Systems International.
- COHEN, Monique and Jennefer SEBSTAD (1999), "Microfinance and Risk Management: A Client Perspective". AIMS Paper. Washington, DC: Management Systems International.
- COHEN, Monique and Jennefer SEBSTAD (2000), "Microfinance, Risk Management, and Poverty". AIMS Paper. Washington, DC: Management Systems International.
- COLEMAN, Brett E. (1999), "The Impact of Group Lending in Northeast Thailand". *Journal of Development Economics*, Vol. 60, No. 2, 105-142.
- COLEMAN, Brett E. (2001a), "Measuring Impact of Microfinance Programs". *ADB Finance for the Poor*, Vol. 2, No. 4, 5-8
- COLEMAN, Brett E. (2001b), "Microfinance in Northeast Thailand: Who Benefits and How Much?" Mimeo. Asian Development Bank.
- COPESTAKE, James (2000), "Integrating Impact Monitoring and Assessment of Microfinance". *Journal of Development in Practice*, Vol. 10, No. 5, 705-711.
- COPESTAKE, James, Sonia BHALOTRA and Susan JOHNSON (2001), "Assessing the Impact of Microcredit: A Zambian Case Study". *The Journal of Development Studies*, Vol. 37, No. 4, 81-100.
- CREEVEY, LucY E. (1994), "Summary Report: A UNIFEM Comparative Study Assessing Impacts of Projects to Support Microenterprises".

- CREEVEY, Lucy E., Koumakh NDOUR and Abdourahmane THIAM (1995), "Evaluation of the Impacts of PRIDE/VITA (The Guinea Rural Enterprise Development Project)". GEMINI Technical Report, No. 94. Bethesda, MD: Development Alternatives, Inc.
- DANIELS, Lisa (1999), "Alternatives for Measuring Profits and New Worth of Microenterprises". 1999. AIMS Paper. Washington, DC: Management Systems International.
- DEARDEN, Kirk and Nazmul KHAN (1994), "Assessing the Impact of Women's Savings and Credit Programs on Fertility: A Case from Bangladesh". Save the Children Women/Child Impact Program Monograph No. 2, Westport: CT.
- DIAGNE, Aliou (1998), "Impact of Access to Credit on Income and Food Security in Malawi". FCND Discussion Paper No. 46. Washington, DC: International Food Policy Research Institute.
- DUNN, Elizabeth and Arbuckle J. GORDON Jr. (2001), "The Impact of Microcredit: A Case Study from Peru". AIMS Paper. Washington, DC: Management Systems International.
- DRURY, David, Stephen WALSH and Marlene STRONG (1994), *Evaluation of the EDWAA Job Creation Demonstration*. Washington, DC: US Department of Labour, Employment and Training Administration.
- DUMAS, Colette (2001), "Evaluating the Outcomes of Microenterprise Training for Low Income Women: A Case Study". *Journal of Developmental Entrepreneurship*, Vol. 6, No. 2, 97-128.
- ELSE, John F. and Carmel CLAY-THOMPSON, *Refugee Microenterprise Development: Achievements and Lessons Learned*. Iowa City: Institute for Social and Economic Development.
- GAILE, Gary L. and Jennifer FOSTER (1996), "Review of Methodological Approaches to the Study of the Impact of Microenterprise Credit Programs". AIMS Paper. Washington, DC: Management Systems International.
- GOETZ, Anne Marie and Rina SEN GUPTA (1996), "Who Takes the Credit? Gender, Power, and Control Over Loan Use in Rural Credit Programs in Bangladesh". *World Development*, Vol. 24, No. 1, 45-63.
- GUPTA, Surendra and Mario D. DAVALOS (1993), "Midterm Evaluation of the Microenterprise Development Project of Jamaica". GEMINI Technical Report No. 59. Bethesda, MD: Development Alternatives, Inc.
- HASHEMI, Syed M., Sidney Ruth SCHULER and Ann P. RILEY (1996), "Rural Credit Programs and Women's Empowerment in Bangladesh". *World Development*, Vol. 24, No. 4, 635-53.
- HIMES, Cristina and Lisa J. SERVON (1998), *Measuring Client Success: An Evaluation of ACCION's Impact on Microenterprises in the United States*. Cambridge, MA: ACCION International.
- HULME, David (2000), *Impact Assessment Methodologies for Microfinance: Theory, Experience, and Better Practice*. *World Development*, Vol. 28, No. 1, 79-98.
- HULME, David, Richard MONTGOMERY, with Debapriya BHATTACHARYA (1996), "Mutual Finance and the Poor. A Study of the Federation of Thrift and Credit Cooperatives in Sri Lanka (SANASA)". In *Finance Against Policy, Volume II: Country Case Studies*, edited by David Hulme and Paul Mosley, 177-245, London: Routledge.

- INSERRA, Anne (1996), "A Review of Approaches for Measurement of Microenterprise and Household Income. AIMS Paper". Washington, DC: Management Systems International.
- INSTITUTE FOR SOCIAL AND ECONOMIC DEVELOPMENT (1994), "Rivercities of Iowa/Illinois Self-Employment (RISE) Final Report". Iowa City: Institute for Social and Economic Development.
- KARLAN, Dean and Gwendolyn ALEXANDER (2002), "Credit to the Poor: Examining the Attrition Bias of Cross-Sectional Impact Assessments". Mimeo. Princeton University.
- KEVANE, Michael and Bruce WYDICK (2001), "Microenterprise Lending to Female Entrepreneurs: Sacrificing Economic Growth for Poverty Alleviation". *World Development*, Vol. 29, No. 7, 1225-1236.
- KHANDKER, Shahidur R. (1996), "Grameen Bank: Impact, Costs, and Program Sustainability". *Asian Development Review*, Vol. 14, No. 1, 97-130.
- KHANDKER, Shahidur R. (1998), "Micro-Credit Programme Evaluation: A Critical Review". *IDS Bulletin*, Vol. 29, No. 4, 11-20.
- KHANDKER, Shahid (2001), "Does Micro-Finance Really Benefit the Poor: Evidence from Bangladesh". Paper presented at the Asia and Pacific Forum on Poverty: Reforming Policies and Institutions for Poverty Reduction, Manila, February 5-9.
- KHANDKER, Shahidur R., Hussain A. SAMAD and Zahed H. KHAN (1998), "Income and Employment Effects of Micro-Credit Programmes: Village-Level Evidence from Bangladesh". *Journal of Development Studies*, Vol. 35, No. 2, 96-124.
- KILBY, Peter and David D'ZMURA (1985), "Searching for Benefits". USAID Special Study No. 28. Washington, DC: United States Agency for International Development.
- KOSANOVICH, William T. and Heather FLECK (2002), *Comprehensive Assessment of Self-Employment Assistance Programs*. Washington, DC: US Department of Labour.
- LAPAR, MA. Lucila A., Douglas H. GRAHAM and Richard L. MEYER (1995), "The Effect of Credit on Output: Are the Sectoral Differences?" Rural Finance Program. Department of Agricultural Economics. Columbus, OH: The Ohio State University.
- LAPAR, MA. Lucila A., Douglas H. GRAHAM, Richard L. MEYER and David S. KRAYBILL (1995), "Selectivity Bias in Estimating the Effect of Credit on Output: The Case of Rural Non-Farm Enterprises in the Philippines". Rural Finance Program. Department of Agricultural Economics. Columbus, OH: The Ohio State University.
- LITTLE, Peter D. (1997), "Income and Assets as Impact Indicators". AIMS Paper. Washington, DC: Management Systems International.
- MCKERNAN, Signe-Mary (1996), "The Impact of Microcredit Programmes on Self-Employment Profits: Do Non-credit Programme Aspects Matter?" *The Review of Economics and Statistics*, Vol. 84, No. 1, 93-115.
- MKNELLY, Barbara, Chatree WATETIP, Cheryl A. LASSEN and Christopher DUNFORD (1996), "Preliminary Evidence that Integrated and Financial and Educational Services Can Be Effective Against Hunger and Malnutrition". Research paper No. 2. Davis, CA: Freedom from Hunger.
- MKNELLY, Barbara and Christopher DUNFORD (1999a), "Impact of Credit with Education on Mothers and their Young Children's Nutrition: GRECER Credit with Education Program in Bolivia". Research Paper No. 5. Davis, CA: Freedom from Hunger.

- MKNELLY, Barbara and Christopher DUNFORD (1999b), "Impact of Credit with Education on Mothers and their Young Children's Nutrition: Lower Pra Rural Bank Credit with Education Program in Ghana". Research Paper No. 4. Davis, CA: Freedom from Hunger.
- MONTGOMERY, Richard, Debapriya BHATTACHARYA and David HULME (1996), "Credit for the Poor in Bangladesh. The BRAC Rural Development Programme and the Government Thana Resource Development and Employment Programme". In *Finance Against Poverty, Volume II: Country Case Studies*, edited by David Hulme and Paul Mosley, 94-176, London: Routledge.
- MORDUCH, Jonathan (1998), "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh". Cambridge, MA: Harvard University.
- MOSLEY, Paul (1996a), "India. The Regional Rural Banks". In *Finance Against Policy, Volume II: Country Case Studies*, edited by David Hulme and Paul Mosley, 246-270, London: Routledge.
- MOSLEY, Paul (1996b), "Indonesia: BKK, KUCK, and the BRI Unit Desa Institutions". In *Finance Against Poverty, Volume II: Country Case Studies*, edited by David Hulme and Paul Mosley, 32-93, London: Routledge.
- MOSLEY, Paul (1996c), "Metamorphosis from NGO to Commercial Bank: The Case of Bancosol in Bolivia". In *Finance Against Policy, Volume II: Country Case Studies*, edited by David Hulme and Paul Mosley, 1-31, London: Routledge.
- MOSLEY, Paul (2001), "Microfinance and Poverty in Bolivia". *The Journal of Development Studies*, Vol. 37, No. 4, 101-132.
- MOSLEY, Paul and David HULME (1998), "Microenterprise Finance: Is There a Conflict Between Growth and Poverty Alleviation?" *World Development*, Vol. 26, No. 5, 783-790.
- MT. AUBURN ASSOCIATES (1998), *Evaluation of Working Capital Delaware*. Somerville, MA: Mt. Auburn Associates.
- MUSTAFA, SHAMS, ISHRAT ARA, DILRUBA BANU, ALTAF HOSSAIN, AJMAL KUBIR, MOHAMMAD MOHSIN and ABU YUSUF (1995) "Impact Assessment Study of BRAC's Rural Development Programme". Final Report.
- NELL, Catherine, Mario DAVALOS, Washington KIIRU, M. MANUNDU and Jennefer SEBSTAD (1994), "The Kenya Rural Enterprise Programme under Cooperative Agreement No. AID-615-0282-A-00-7026-00: A Final Evaluation". GEMINI Technical Report No. 77. Bethesda, MD: Development Alternatives, Inc.
- NELSON, Eric (1984), "The Economic Impact of KIK/KMKP Credit on Indonesian Small Enterprises". Draft Memo to Kapala Urusan Kredit Koperasi dan Kredit Kecil. Mimeo.
- NELSON, Eric and Bruce R. BOLNICK (1986), "Survey Methods for Assessing Small Credit Programs Evaluating the Economic Impact of KIK/KMPK Credits". Jakarta.
- NELSON, Candace, Barbara MKNELLY, Elaine EDGCOMB, Gary GAILE, Carter GARBER, Nancy HORN, Karren LIPPOLD and Brian BEARD (2001), "Learning from Clients: Assessment Tool for Microfinance Practitioners". AIMS Paper. Washington, DC: Management Systems International.
- OLDHAM, Linda, Hager EL HADIDID, Hussein TAMA, Michele SULIMAN NASHED, SHERIF EL DEWANY, Mahmoud HUSSEIN, Mohammed ABDEL AZIZ and Mohamed SAKR (1994), "Measuring Socioeconomic Impact of Credit on SMI: Assessment of the Monitoring System used by the Alexandria Businessmen's Association", Egypt. GEMINI Technical Report No. 76. Bethesda, MD: Development Alternatives, Inc.

- PARK, Albert and Changqing REN (2001), "Microfinance with Chinese Characteristics". *World Development*, Vol. 29, No. 1, 39-62.
- PITT, Mark M. (1999), "Reply to Jonathan Morduch's 'Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh'", Providence, RI: Brown University.
- PITT, Mark M. and Shahidur R. KHANDKER (1996), "Household and Intrahousehold Impacts of the Grameen Bank and Similar Targeted Credit Program in Bangladesh". World Bank Discussion Paper No. 320. Washington, DC: World Bank.
- PITT, Mark M. and Shahidur R. KHANDKER (1998), "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?" *Journal of Political Economy*, Vol. 106, No. 5, 958-996.
- PITT, Mark M., Shahidur KHANDKER, Osman H. CHOWDHURY and Daniel L. MILLIMET (1999), "Credit Programs for the Poor and Health Status of Children in Rural Bangladesh". PSTC Working Paper #98-01. Providence, RI: Brown University.
- PITT, Mark M., Shahidur KHANDKER, Signe-Mary MCKERNAN and M. ABDUL LATIF (1999), "Credit Programs for the Poor and Reproductive Behaviour in Low-Income Countries: Are Reported Causal Relationships the Result of Heterogeneity Bias?" *Demography*, Vol. 36, No. 1, 1-21.
- PULLEY, Robert V. (1989), *Making the Poor Creditworthy. A Case Study of the Integrated Rural Development Program in India*. Washington, DC: The World Bank.
- RAHEIM, Salome (1996), "Microenterprise as an Approach for Promoting Economic Development in Social Work: Lessons from Self-Employment Investment Demonstration". *International Journal of Social Work*, Vol. 39, 69-82.
- RAHEIM, Salome and Jason FRIEDMAN (1999), "Microenterprise in the Heartland: Self-Employment as a Self-Sufficiency Strategy for TANF Recipients in Iowa 1993-1998". *Journal of Microfinance*, Vol. 1, No. 1, 66-90.
- ROSENBERG, Richard (1999), "Measuring Microcredit Delinquency: Ratios Can Be Harmful to Your Health". Occasional Paper No. 3. Washington, DC: Consultative Group to Assist the Poorest.
- SCHREINER, Mark (2002), "Evaluation and Microenterprise Programs in the United States". *Journal of Microfinance*, Vol. 4, No. 2, 67-91.
- SCHULER, Sidney Ruth and Syed M. HASHEMI (1994), "Credit Programs, Women's Empowerment, and Contraceptive Use in Rural Bangladesh". *Studies in Family Planning*, Vol. 25, No. 2, 65-76.
- SCHULER, Sidney Ruth, Syed M. HASHEMI and Ann P. RILEY (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*, Vol. 25, No. 4, 563-575.
- SCHULER, Sidney Ruth, Syed M. HASHEMI and Shamsul HUDA BADAL (1998), "Men's Violence Against Women in Rural Bangladesh: Undermined or Exacerbated by Microcredit Programmes?" *Development in Practice*, Vol. 8, No. 2, 148-157.
- SEBSTAD, Jennefer (1992), "Get Ahead Foundation Credit Programs in South Africa: The Effects of Loans on Client Enterprises". GEMINI Technical Report, No. 44. Bethesda, MD: Development Alternatives, Inc.

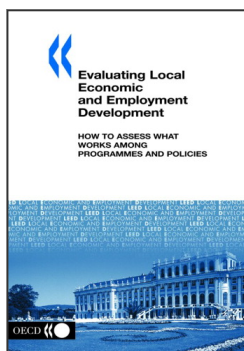
- SEBSTAD, Jennefer and Sara LOZA (1993), "Lending and Learning: Formal Banks and Microenterprise in Egypt". Report to the Ford Foundation, Cairo. Washington, DC: Community Economics Corporation.
- SEBSTAD, Jennefer and Martin WALSH (1991), "Microenterprise Credit and its Effects in Kenya: An Exploratory Study". Report prepared for USAID AFR/MDI and SandT/WID. Washington, DC: Coopers and Lybrand.
- SEKKESAETER, Unni Beate (2002), *Evaluation of Hordaland Network Credit. A Microfinance Programme in Hordaland County, Norway – Part of the "Equal Credit" Project under the EU Receipte II Programme*. Bradford, West Yorkshire: Bradford School of Management.
- SMITH, Stephen C. (2002), "Village Banking and Maternal and Child Health: Theory and Evidence from Ecuador and Honduras". *World Development*, Vol. 30, No. 4, 707-723.
- STEELE, Fiona, SajedA AMIN and Ruchira T. NAVED (2001), "Savings/Credit Group Formation and Change in Contraception". *Demography*, Vol. 38, No. 2, 267-282.
- SUTORO, Ann Dunham (1990), "KUPEDES Development Impact Survey: Briefing Booklet". BRI, Planning, Research and Development Department.
- THE ROBERTS FOUNDATION (1995), *Self Employment and Very Low-Income Women. A Final Report on the Experience of the San Francisco Homeless Women's Economic Development Project*. San Francisco: The Roberts Foundation.
- TODD, Helen (1996), *Women at the Center. Grameen Bank Borrowers after One Decade*. Boulder: Westview Press.
- US DEPARTMENT OF HEALTH AND HUMAN SERVICES (1994a), "Capital Opportunities Expansion Program. Human Resource Development Council of District IX, Inc.: Bozeman, Montana" In *Micro Business and Self-Employment: Summary of Final Evaluation Findings from 1990*, 54-83, Washington, DC: US Department of Health and Human Services.
- US DEPARTMENT OF HEALTH AND HUMAN SERVICES (1994b), "Micro-Business Development Program. Central Vermont Community Action Council, Inc.: Barre, Vermont". In *Micro Business and Self-Employment: Summary of Final Evaluation Findings from 1990*, 1-19, Washington, DC: US Department of Health and Human Services.
- US DEPARTMENT OF HEALTH AND HUMAN SERVICES (1994c), "Microenterprise Development Program (MEDP). Mayor's Office of Community Services: Philadelphia, Pennsylvania". In *Micro Business and Self-Employment: Summary of Final Evaluation Findings from 1990*, 38-53, Washington, DC: US Department of Health and Human Services.
- US DEPARTMENT OF HEALTH AND HUMAN SERVICES (1994d), "Small Enterprise and Family Development. Southeast Iowa Community Action: Burlington, Iowa". In *Micro Business and Self-Employment: Summary of Final Evaluation Findings from 1990*, 20-37, Washington, DC: US Department of Health and Human Services.
- VENGROFF, Richard and Lucy CREEVEY (1994), *Evaluation of Project Impact: ACEP Component of the Community Enterprise Development Project*. Storrs: The University of Connecticut.
- WOLLER, Gary and Robert PARSONS (2002), "Assessing the Community Economic Impact of Microfinance Institutions", *Journal of Developmental Entrepreneurship*, Vol. 7, No. 2, 133-150.
- WYDICK, W. Bruce (1999a), "Credit Access, Human Capital, and Class Structure Mobility". *The Journal of Development Studies*, Vol. 55, No. 6, 131-152.

- WYDICK, W. Bruce (1999b), "The Effect of Microenterprise Lending on Child Schooling in Guatemala". *Economic Development and Cultural Change*, Vol. 47, No. 4, 853-869.
- ZAMAN, Hassan (2001), "Assessing the Poverty and Vulnerability Impact of Micro-Credit in Bangladesh: A Case Study of BRAC". Washington, DC: Consultative Group to Assist the Poorest.
- ZELLER, Manfred, Akhter AHMED, Suresh BABU, Sumiter BROCA, Aliou DIAGNE and Manohar SHARMA (1996), "Rural Financial Policies for Food Security of the Poor. Methodologies for a Multicountry Research Project". Washington, DC: International Food Policy Research Institute.

Table of Contents

Chapter 1.	Introduction and Summary Evaluating Programmes for Local Economic and Employment Development: an Overview with Policy Recommendations by Alistair Nolan and Ging Wong.....	7
Chapter 2.	Policy Learning through Evaluation: Challenges and Opportunities by Ging Wong	49
Chapter 3.	Evaluation: Evidence for Public Policy by Robert Walker	63
Chapter 4.	Evaluating the Impacts of Local Economic Development Policies on Local Economic Outcomes: What has been done and what is doable? by Timothy J. Bartik	113
Chapter 5.	Four Directions to Improve Evaluation Practices in the European Union: A Commentary on Timothy Bartik’s Paper by Daniele Bondonio	143
Chapter 6.	The Evaluation of Programs aimed at Local and Regional Development: Methodology and Twenty Years of Experience using REMI Policy Insight by Frederick Treyz and George I. Treyz	151
Chapter 7.	A Commentary on the Frederik and George Treyz’s Paper and the Workshop “Analysis Policies for Local Development Using Forecasting Models” by Robert Wilson.....	191
Chapter 8.	Area-based Policy Evaluation by Brian Robson	199
Chapter 9.	A Commentary on Brian Rubson’s Paper and the Workshop “Area-based Policy Evaluation” by Jonathan Potter.....	221

Chapter 10.	Evaluating Business Assistance Programs by Eric Oldsman and Kris Hallerg.....	229
Chapter 11.	Evaluating Training Programs: Impacts at the Local Level by Randall W. Eberts and Christopher J. O’Leary	251
Chapter 12.	Evaluating Local Economic Development Policies: Theory and Practice by Jeffrey Smith	287
Chapter 13.	Evaluation and Third Sector Programmes by Andrea Westall	333
Chapter 14.	Methodological and Practical Issues for the Evaluation of Territorial Pacts. The Experience of Italy by Paola Casavola	355
Chapter 15.	Evaluating Territorial Employment Pacts – Methodological and Practical Issues. The experience of Austria by Peter Huber	369
Chapter 16.	A Commentary on the Workshop “Evaluating Territorial Employment Pacts” by Hugh Mosley	381
Chapter 17.	A Review of Impact Assessment Methodologies for Microenterprise Development Programmes by Gary Woller	389
Chapter 18.	An Overview of the Panel Discussion: Evaluating Local Economic and Employment Development by Alice Nakamura	437
	About the Authors and Contributors	443



From:
**Evaluating Local Economic and Employment
Development**
How to Assess What Works among Programmes and Policies

Access the complete publication at:
<https://doi.org/10.1787/9789264017092-en>

Please cite this chapter as:

Woller, Gary (2005), "A Review of Impact Assessment Methodologies for Microenterprise Development Programmes", in OECD, *Evaluating Local Economic and Employment Development: How to Assess What Works among Programmes and Policies*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264017092-18-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.