MINISTRY OF FOREIGN AFFAIRS OF DENMARK
Danida

SYNTHESIS OF IMPACT EVALUATIONS OF MICROCREDIT

EVALUATION STUDY 2009/4

Synthesis of impact evaluations of microcredit

August 2009

Jens Kovsted, PhD
Centre for Economic and Business Research
Porcelænshaven 16A, DK-2000 Frederiksberg

Associate Professor Thomas Barnebeck Andersen, PhD
Institute of Economics, University of Copenhagen
Studiestræde 6, DK-1455 Copenhagen K

Andreas Kuchler, PhD student
Department of Business and Economics, University of Southern Denmark
Campusvej 55, DK-5230 Odense M

Table of contents

List of abbreviations

ADB	Asian Development Bank
AIMS	Assessing the Impact of Microentreprise Services (USAID)
BRDB	Bangladesh Rural Development Board
CGAP	Consultative Group to Assist the Poor
CGD	Centre for Global Development
DFID	Department for International Development
GTZ	Deutsche Gesellschaft für Technishe Zusammenarbeit
IADB	Inter American Development Bank
IFPRI	International Food Policy Research Institute
MDG	Millennium Development Goals
NGO	Non Governmental Organisation
NWEP	Northwest Microfinance Expansion Project (BRAC programme)
RDP	Rural Development Programme (BRAC programme)
SIDA	Swedish International Development Cooperation Agency
UNDP	United Nations Development Programme
USAID	United States Agency for International Development
WB	World Bank

1. Introduction

There is widespread belief that microcredit can be effective in reducing global poverty. At the same time, the fact that microcredit relies on private enterprise, as well as the gender bias, appeals strongly to donors and NGOs alike. In conjunction, these factors have resulted in a significant growth in microcredit funding and outreach. According to the Microcredit Summit Campaign, there were more than 150 million microcredit clients in the world (more than two thirds were women) as of December 2007 (Banerjee et al., 2009).

Notwithstanding these developments, thoroughly tested knowledge about the impact and effect of microcredit is partial and contested. The knowledge base mostly consists of anecdotal evidence and case studies, and research has generally failed to account for the potentially severe endogeneity problems (see below) and/or has not measured associated costs and benefits properly. Yet, this has not held the World Bank's Consultative Group to Assist the Poor (CGAP) back from announcing that there is mounting evidence that microcredit can help achieve the Millennium Development Goals (MDGs) (Banerjee et al., 2009).

The objective of this synthesis study is to answer the following key question: What can be said about the causal impact of a microcredit programme on the basis of existing evaluations? Second, and related to the above-mentioned question, the study will also evaluate the methodology used in current microcredit evaluations undertaken by donors.

It is customary to distinguish between microfinance and microcredit. Microfinance covers a range of financial services (credit, savings, insurance) whereas microcredit is the provision of small loans.[1] Here focus is on the microcredit tool. Consequently, the evaluations and studies included in this review are of programmes offering microcredit, although a number of the evaluated programmes also offer savings and other services.

---

[1] In addition to programmes focusing on the provision of credit, microcredit is sometimes offered in combination with savings, education etc. and is also sometimes offered as a part of development programmes with a wider scope than merely providing credit.

The emergence of microcredit[2] is often associated with loans to individuals in solidarity groups with joint liability, where the group members essentially substitute for collateral.[3] Yet group lending is one among several microcredit lending technologies.[4] Individual loan contracts with dynamic incentives and public repayment sessions are also widespread. The evaluations considered in this review represent microcredit programmes utilizing both group- and individual lending technologies.

So far, empirical studies of microcredit programmes have focused on estimating microcredit impact on consumption and income and on financial indicators such as loan recovery rates. A recent trend among microcredit providers has been to include social impacts (e.g. gender empowerment effects) in their internal impact evaluations. To assess the strength of microcredit as a development tool surely all dimensions are important, but if focus is on poverty alleviation then the impact on consumption and income are prime indicators. Consequently, the studies considered in this synthesis study focus on the impact on consumption and income from participation in a microcredit programme.

The evaluations reviewed in this synthesis study are published by both bi- and multilateral donors, by international NGOs and in peer-reviewed international journals. Most of the evaluations have been published during the last decade. The fact that microcredit has been developed and operated on the largest scale in Asia and Latin America is reflected by the large number of evaluations in this synthesis study focusing on microcredit in these areas.

An evaluation of the impact of a microcredit programme essentially needs to ask the question: What would have happened in the absence of the programme? This question is by nature impossible to answer as one cannot simultaneously observe the same individual both exposed and not exposed to the programme. A simple and frequently used solution is to observe an individual before and after entering the programme and observe if any changes arise. However, changes can rarely be ascribed solely to the programme as other factors may change over the same period. Much effort by evaluators has therefore been put into the construction of relevant control groups. That is: groups of individual who resemble those exposed to a microcredit programme without being exposed to it. Much, however, depends upon

---

[2] Chapter Three in Armendariz and Morduch (2005) provide a historical overview of microcredit developments and trends.

[3] Group lending mitigates the problems of *adverse selection* (i.e. that the individuals with a risk-taking behaviour are more likely to demand a loan in the absence of liability) and *moral hazard* (i.e. that incentives to make the project yield high returns may be less in the absence of liability).

[4] Only 9% of the 890 microbanks that reported to the "MicroBanking Bulletin" (autumn 2008) exclusively rely on group lending. A total of 49% use a combination of group and individual lending.

the similarity between the individuals (in the treatment group) who have received microcredit and the individuals (in the control group) who were not exposed to microcredit. This is a challenge for every (microcredit) evaluation, and one independent contribution of this report is to critically assess applied methodologies and their impact on the reported results.

Even though microcredit advocates claim that microcredit is solely about improving access to capital, microcredit programmes can affect existing credit markets, changing, for example, the level of competition in the local credit market. Evaluations are seldom concerned with these interaction effects among the different suppliers of credit.

From a practitioner's point of view, assessing the potential for microcredit as a poverty reduction tool is the key criteria. In addition, practitioners worry whether microcredit is sustainable (i.e. profitable for the lender) in the longer run, and whether supporting microcredit is cost-effective relative to other interventions. Little is known on this subject and it is consequently outside the scope of this study to address it. A related question also outside the scope of the present review is whether funding and implementing agencies should primarily focus on developing the microcredit infrastructure or provide direct operational funding. This question is obviously related to the sustainability issue, because as long as microcredit to the poor is not financially sustainable, outside funding is needed.

Following the Introduction, Chapter 2 presents methodological considerations used to identify key evaluation questions needed for scoring the different studies according to how well they identify the causal impact of a given microcredit programme. Chapter 3 presents and discusses the available research on this issue, whilst the evaluations and studies conducted by donors are presented in Chapter 4. Donor evaluations are ranked according to how well they can be used as basis for a decision to scale up the programme in question. Scoring methodology and rationale as well as the results of the scoring exercise is presented and discussed in Chapter 5, before Chapter 6 concludes, summarizing the main lessons of the study.

2. Impact evaluation methodology[5]

In order to investigate whether receiving a microcredit loan raises borrower income, we face a rather difficult counterfactual problem. That is, we must ask: How would the individuals, who received a microcredit loan, have fared in the absence of the loan? The difficulty is immediate: At a given point in time, an individual is either exposed to the microcredit loan or not exposed. Comparing the same individual over time will rarely provide a reliable estimate of programme impact, the reason being that other things are rarely equal. For this reason we cannot obtain an estimate of the impact of the loan on each and every individual. What we can do, however, is to estimate an average impact of microcredit loans on a group of individuals by comparing them with a similar group that did not receive loans.

Consequently, a key step in impact evaluation is the creation of a credible control group. This is a group of individuals who, in the absence of the loan, would have had outcomes broadly similar to those who received the loan (the treatment group). The hope is therefore that the control group will provide a good gauge of what would have happened to the treatment group if it had not received the loan. Consequently, comparing the two groups by looking at the difference between the treatment group and the control group provides the estimated average impact on the group in question.

However, if the people/households that receive microcredit (the treated[6]) and those who do not (the untreated) are different, just looking at the difference between the two groups will lump together pre-existing differences (selection bias) and the impact of the programme.[7] Because we have no way of estimating the size of the selection bias with certainty, we cannot decompose the observed difference into an average treatment effect and a bias term. We therefore have to ask: Under what conditions will the difference between the treated and the untreated consistently estimate the average effect of the programme? A sufficient condition for consis-

---

[5] This section builds mostly on Duflo (2004) and Winship and Morgan (1999).

[6] The origin of a considerable part of evaluation terminology is medical/clinical experiments – a context where the meaning of terms like "treatment" and "control" may be more intuitive.

[7] In microcredit programmes, which often rely on solidarity groups, micro-entrepreneurs have a clear incentive to form groups with other good micro-entrepreneurs so as to avoid cross-subsidizing bad micro-entrepreneurs (Ghatak, 1999). Technically, there is assortative matching in group formation. This means that those who receive a loan are more "able" than those who do not.

tency of the estimate is that the assignment of individuals into treatment (i.e. receiving a microcredit loan) is random. If this is the case, it can be shown that the difference between the treated and the untreated is a consistent estimator[8] of the effect of the programme (the treatment). For this reason, randomization is often seen as "the gold standard" research design.

Consider now an evaluation of the effect of a microcredit programme on the income levels of poor female micro-entrepreneurs. The programme participants receive a small microcredit loan, whereas the control group does not receive a loan. Now, if the people who choose (or are chosen) to participate in the programme systematically differ from those who do not participate, the assignment into treatment and control groups will be non-random. This will, for example, be the case if participants are more entrepreneurial and/or wealthier than non-participants. Assume now that after the loan is repaid (one year later, say) we observe that the female micro-entrepreneurs in the treatment group have experienced an increase in levels of income, whereas those in the control group have not. Does this mean that the given microcredit programmes have raised income levels of the female micro entrepreneurs? Without further analysis, we don't know!

From the above discussion, it is possible to identify three possible explanations for the raised income levels of the female micro-entrepreneurs:

i) Microcredit increases income levels by relieving credit constraints among the poor female micro entrepreneurs. This is the true average treatment effect of the programme, which we wish to estimate.

ii) Female micro-entrepreneurs, who received microcredit in the example provided, are more motivated and more entrepreneurial than those who did not receive microcredit. This can be thought of as a baseline difference between programme participants and non-participants, which originates from the non-random selection of programme participants. The notion that the programme appeals to/attracts a certain type of people/household is also known as assortative matching; a name reflecting that participation and the match between programme and participants is not random.

iii) Finally, the entrepreneurial ability of those who received microcredit may increase more than it would have done for the control group even if they had received microcredit. This will occur if the effect of participation differs across the treated and untreated, for example if participants (due to their inherent characteristics/ability) gain more from participation.

---

[8] Technically, consistency means that the estimator gives us the "true" average treatment effect as the number of program participants grows without bound.

Points ii) and iii) are the biases that any impact study must aim to neutralise in order to estimate the effect of the programme. Failure to do so convincingly weakens any conclusions drawn. In what follows, we will simply refer to points ii) and iii) as endogeneity[9] bias.

In conclusion, this brief introduction to the evaluation problem has provided a basic insight into the complexity of both the problem and the potential solutions when seeking to answer the seemingly simple question of whether a micro loan has had an effect on household income or not.

---

[9] In this context endogeneity means that unobserved characteristics and abilities of the participants affect participation and the outcomes of participation.

3. Peer-reviewed impact evaluation studies

In theory, randomized assignment into treatment eliminates any bias. Unfortunately, there is a dearth of impact evaluation relying on this methodology.[10] Impact studies mostly use observational data and try to eliminate bias by recourse to more or less sophisticated statistical methods and/or quasi-experimental[11] designs. In what follows we discuss the few peer-reviewed impact evaluation studies.

3.1 Microcredit in Bangladesh

The leading study on microcredit in Bangladesh is Pitt and Khandker (1998 published in the *Journal of Political Economy*, a top five economics journal. Pitt and Khandker evaluate three group-based Bangladeshi credit programs: the Grameen Bank, BRAC[12], and the Bangladesh Rural Development Board (BRDB). Their point of departure is that previous studies are unsound as they only compare outcomes between participating and non-participating households. That is, previous studies simply employ the standard estimator on observational data. This means that they are tainted by endogeneity bias, and therefore uninformative about true average treatment effects.

Pitt and Khandker identify three main sources of potential endogeneity bias, which they aim to tackle. These are non-random programme placement, unmeasured village-level attributes that affect both the demand for programme credit and economic outcomes, and unmeasured household-level attributes that affect both the demand for programme credit and economic outcomes. In order to eliminate the resulting endogeneity bias, Pitt and Khandker provide identification of the treatment effect through a quasi-experimental design.

More specifically, they include in the sample households who live in treatment villages but who are excluded from participating in the programme based on an ex-

---

[10] We know of only two studies using randomization, Karlan and Zinman (2008) and Banerjee et al. (2009).

[11] As implied by its name quasi-experiments are almost true experiments – often missing the randomized allocation into a treatment and a control group to account for both selection and placebo effects. While there are many varieties of quasi-experimental evaluations designs, they all seek to overcome this basic lack of randomly allocated treatment and control groups.

[12] BRAC was founded in 1972 as the Bangladesh Rural Advancement Committee. It is today mostly known by its acronym BRAC.

ogenous[13] rule imposed by the NGOs offering the microcredit. The rule in question is enforced to ensure the poverty orientation of the microcredit programmes imposing that only households owning less than half an acre of land can participate in the three microcredit programs. Pitt and Khandker compares economic outcomes between households who can participate and households who cannot. The comparison is conditional on village-level fixed effects (to tackle unobserved village-level attributes) and observed household and individual attributes. Moreover, since Pitt and Khandker have included households who live in treatment villages but who are excluded from participating in the programme, they can alleviate concerns about non-random programme placement. That is, by comparing the difference between treated and untreated in programme villages with the difference between eligible-untreated and ineligible-untreated in control villages, concerns about non-random programme placement are alleviated. The key assumption when it comes to identification of the effects of the microcredit is therefore that landownership is exogenous to the allocation of credit. Pitt and Khandkers results depend crucially on the assumption that landownership is uncorrelated with (unobserved) household characteristics that affect either the likelihood of being enrolled in the microcredit programmes and/or the effect of getting access to microcredit. This assumption is, for example, violated if the household who owns land are more entrepreneurial and therefore more likely to access credit.

Data are from a multipurpose household survey conducted in 87 villages of 29 sub-districts in rural Bangladesh during 1991/92. The 29 sub-districts were chosen with equal probability from among 391 sub-districts. In the sample, 24 of the sub-districts had a microcredit programme running, whereas 5 did not. The total number of households is 1,798, and among these 905 participated in a microcredit program.

The principal finding of the evaluation is that an additional taka of credit provided to women adds 0.18 taka to total annual household expenditure, as compared to 0.11 for men. These results are statistically significant.[14]

The findings by Pitt and Khandker, however, remain open to debate. In an unpublished but widely cited paper,[15] Morduch (1998) has argued that the exogenous eligibility rule based on landholdings is frequently violated. It is, for example, noted that 30% of Grameen borrowers own more land than the half-acre cut off, with landholdings as large as 14 acres. Morduch moreover showed that an alternative estima-

---

[13] Exogenous refers to the rule being set in order to target the programme to the poor and not to affect performance and related effects of participation.

[14] Please note, that the effect measured here is the marginal impact of programme – i.e. the effect of providing an additional taka of credit. This should not be confused with the total effect of the programme.

[15] In a recent book published by MIT Press, a leading academic publisher, Armendáriz and Morduch (2005) raise the Morduch (1998) criticisms anew.

tor failed to detect a positive impact of microcredit. In a rejoinder to Morduch, Pitt (1999) argues that Morduch has misunderstood and mischaracterized the methods of Pitt and Khandker and has applied incorrect methods to obtain his new evidence. Moreover, Khandker (2005) extends the Pitt and Khandker (1998) study by using, in addition to the 1991/92 data, follow-up data collected in 1998/99. This allows Khandker to employ panel data methods, which provides a range of econometric benefits. Khandker's (2005) findings confirm the original Pitt and Khandker (1998) results. Moreover, using the Pitt-Khandker data, Mckernan (2002) finds that micro-credit increased the profit generated by household enterprise activities.

3.2   Microcredit in Thailand

The leading study on microcredit in Thailand is Coleman (1999). The study is published in the *Journal of Development Economics*, the top field journal in development economics. Coleman relies on a survey of 445 households in 14 villages Northeast Thailand in 1995/96. Eight of the villages are supported by the Rural Friends Association (RFA), and six are supported by the Foundation for Integrated Agricultural Management (FIAM).[16] The unique feature of the Coleman study (and crucial for his identification strategy) is that that the RFA and FIAM in January 1995 pre-identified four villages and two villages, respectively, where they would begin supporting with village bank loans in 1996. In February and March 1995, field staff organized the villagers into new village banks, allowing them to self-select according to the standard procedures normally used to organize new village banks, the only difference being that the villagers were told that loans would begin approximately one year later. By implication, a perfect control group of would-be village bank members was identified. Indeed, this unique setup allows Coleman (via a refined difference-in-difference approach) to control for both non-random member selection and non-random programme placement.

Coleman finds two interesting results. First, when one does not address the endogeneity issue, i.e. when one relies on the standard estimator, programme impacts are significantly overestimated. Second, when one does take endogeneity into account, the effect of microcredit on women's income is not significantly different from zero. Because of the unique design, there are no obvious threats to internal validity of results. However, Coleman himself argues that there are threats to external validity. Thailand is a relatively wealthy country with annual GDP growth at close to 10% for the two decades leading up to the survey, and many villagers already had access to low-interest credit from financial institutions such as the Bank for Agricul-

---

[16] RFA and FIAM are Thai NGOs who have promoted village banks since 1988. Both receive financial and technical assistance from the American NGO Catholic Relief Services (CRS). The operation of both NGOs' village banks is virtually identical as both follow the FINCA village banking methodology.

ture and Agricultural Cooperatives. In such an environment, it should not be surprising that small loans would have a negligible impact. Indeed, a common complaint of women surveyed was that the size of village bank loans was far too small for them to be productive. Consequently, as stressed by Coleman, caution is in order before extrapolation of these results.

Another study of microcredit in (rural and semi-urban) Thailand is Kaboski and Townsend (2005). This study is published in the *Journal of the European Economic Association*, a top 20 economics journal. In contrast to previous work, this study examines a large set of heterogeneous village-level microfinance institutions.[17] It links impacts on households to variation in the characteristics and policies of these institutions, and evaluates whether the observed impacts of these types of intermediation are consistent with what theories predict. The theories in question are structural general equilibrium models of growth with either endogenous or exogenous financial intermediation assuming that households have limited access to credit and/or savings services.

More specifically, the paper relies on household and institution level data from a survey conducted in May 1997 (before the Asian financial crisis) in four Thai provinces (changwats). The survey design was based partly on results of prior field research. The survey has three subcomponents: the institutional module, the household module,[18] and the key informant module. The institutional survey was given to all known microfinance institutions that were encountered in the villages at the time of the household survey. In total, records for 161 institutions were obtained across 108 of the villages. Survey questions focused on both the individual policies and the experiences of the institutions, including their founding, membership, and saving and lending services.[19] The most common type of institution encountered is production credit groups (CDD), whose members are mostly the better-off women. CDD mostly lend cash. Another common institution is rice banks, which usually make small-term, emergency consumption loans intended for consumption smoothing purposes. The form or type of the institution is not the only dimension of variation among the entities in the institutional survey. The survey also contains data on services, policies and characteristics of the institutions, which offers much variation.

---

[17] Kaboski and Townsend argue that the fact that their institutions are operated at the village level is a virtue. These village-level entities are promoted by a variety of different agencies and ministries, and Kaboski and Townsend rely on this variation as an instrument to identify impacts.

[18] The household survey was administered to 2,880 Thai households: 15 households in each of 192 survey villages. Households provided an extensive array of demographic and socioeconomic information.

[19] These institutions are all quasi-formal: that is, they keep records and often have bank accounts, but do not in general have their own office. Most also have some relationship to the Thai government, despite being administered at the local level. The government agencies also offer various types of assistance such as accounting assistance.

The Kaboski-Townsend study asks whether microfinance produces the impacts of financial intermediation predicted by theory. Their empirical strategy is to estimate impact using variables associated with financial intermediation, whose variation is either exogenous, or endogenous in ways that can be controlled. The variables they use are the presence of (or membership in) the institution, the different types of institutions, and the different policies. They examine two sets of policies: one set of policies associated with successful financial intermediation in the data; and a second set involving policies such as group liability, dynamic incentives, or better monitoring technologies, policies predicted to be important by theory. For policies, Kaboski and Townsend lack independent membership data, and so can only look at the effect of institutions with these policies on outcomes of the average villager, not on individual members. For institutions overall, and each of the different types of institutions, they have membership data, and so can address individual villager impact.

The focus on membership introduces the issue of household-level selection bias. Households that are members of village institutions (in villages with institutions) may differ systematically from non-member households in the same villages. If these differences are the result of biased selection into the institutions (whether on the part of the household demand or the institution supply) they should not be attributed to the impact of the institution. To address this, Kaboski and Townsend use the presence of the institution in the village (a village-level variable) as instrument for membership. Finally, they address endogenous programme placement problems using a wide range of village-level controls.

Two very interesting results emerge from the analysis: First, households in villages that hold specialized institutions offering emergency loans were 10-29 percentage points less likely to reduce consumption in years with negative income shocks. This result does not, however, hold for the "average" institution, where in fact there is no risk alleviating effect; if anything there is a troubling risk-increasing effect. Second, microfinance institutions reduce reliance on moneylenders, and this result is very robust. The effect on the average villager is to reduce the probability of becoming a moneylender costumer by 8 percentage points. Kaboski and Townsend interpret this as microfinance institutions loosening households' constraints on formal credit, at least to credit that could be acquired alternatively from moneylenders. This effect is particularly manifest for women's groups.

3.3 Microcredit in South Africa

While the above studies have focused on credit to micro-entrepreneurs, Karlan and Zinman (2008) experimentally explore the impact of consumer credit. The study is forthcoming in the *Review of Financial Studies*, which is a leading field journal in finance. The experiment was implemented by a consumer lender in a high-interest

rate, high-risk South African loan market where credit constraints appeared to be binding. Karlan and Zinman address the identification issue by making the lender introduce exogenous variation in the loan approval process.[20] Their "treatment" randomly encouraged loan officers to approve some marginal applications. Specifically, the lender added three additional steps to its normal process for new loan applicants: (i) Loan officers were told to mark rejected applications as either extremely or marginally uncreditworthy (a total of 786 applicants were deemed marginally uncreditworthy). (ii) A computer then instructed the officer to re-evaluate the marginal applications in real-time by randomly producing a message to "approve" or "still reject." Neither the treatment (computer said "approve") nor the control (computer said "reject") groups were informed by the lender that a component of the loan decision was randomized. Officers were told by management to follow the computer's instructions in all cases. (iii) Loan officers had financial incentives to be conservative; they approved the loan only in roughly half of the cases where the computer instructed them to approve. Consequently, Karlan and Zinman "only" study the impacts on marginal applicants – that is the effect on the participants who where almost eligible for participation in the programme. This should not be confused with measuring the effect on those who where found eligible to participate in the first round. This is important to note because the treatment effect may differ between the two categories.

The outcome data used were from the lender's records on repayment and profitability, from credit-bureau reports over two years after the start of the experiment, and from household surveys conducted by an independent firm at the home or workplace of the marginal applicants six to twelve months after the start of the experiment. The survey measured borrowing activity, loan uses, and a range of proxies for household well-being.

Results indicate that expanded access to credit considerably enhanced outcomes in a variety of dimensions. In particular, Karlan and Zinman found that over a 6 to 12 month horizon after the experiment ended (i.e., the initial loan repayments were due), applicants in the treatment group experienced higher consumption. Households randomly assigned a loan were an estimated 5.8 percentage points less likely to report hunger (p-value of 0.03), and 3.7 percentage points more likely to report a food quality improvement, but this effect was statistically insignificant (p-value of 0.32).

---

[20] The particular Lender was merged with a larger bank in 2005 and thus no longer exists as a separate entity.

3.4 Microcredit in India

In a new paper from the MIT Poverty Action Lab, Banerjee et al. (2009) report results from the first randomized evaluation of the impact of the canonical group-lending microcredit model. This study is at present *unpublished*. However, given the professional standing of the authors (two of them, Abhijit Banerjee and Esther Duflo, are leading development economists); we have chosen to include it.

In 2005, 52 of 104 neighbourhoods[21] in Hyderabad (the fifth largest city in India, and the capital of Andhra Pradesh, the Indian State where microcredit has expanded the fastest) were randomly selected for opening of a microfinance branch by one of the fastest growing microfinance institutions in the area, Spandana, while the remainder were not.[22] Fifteen to 18 months after the introduction of microcredit in each area, a comprehensive household survey was conducted in an average of 65 households in each slum, a total of 6,850 households.

Turning to the results, Banerjee et al. report that there is no significant difference in total household expenditure per adult equivalent between treatment and comparison areas.[23] The average household in a comparison area has expenditure of Rs (Indian rupees) 1,420 per adult equivalent per month; in treatment areas the number is 1,453, not statistically different. They also report that women in treatment areas were no more likely to be make decisions about household spending, investment, savings, or education. There is no effect on health or education outcomes, either. Households in treatment areas spend no more on medical and sanitation than do comparison households, and among households with children, households in treatment areas were no less likely to report that a child had a major illness in the past year. Among households with school aged children, households in treatment areas are not more likely to have children in school. Looking just at girls' school enrolment gives the same conclusion. Treatment households also do not spend more on

---

[21] These areas were selected based on having no pre-existing microcredit presence, and having residents who were desirable potential borrowers: poor, but not "the poorest of the poor". While those areas are commonly referred to as "slums", these are permanent settlements, with concrete houses, and some public amenities (electricity, water, etc.). The population in the neighborhoods selected for the study ranges from 46 to 555.

[22] The basic Spandana product is the group loan product. A group is comprised of six to 10 women, and 25-45 groups form a "centre". Women are jointly responsible for the loan of their group, and of the centre. The first loan is Rs. 10,000 (about $200 at market exchange rates, or $1,000 at PPP-adjusted exchange rates). It takes 50 weeks to reimburse principal and interest rate; the interest rate is 12% (non-declining balance; equivalent to a 20% APR). If they all reimburse they are eligible for second loans of 3 Rs. 10,000-12,000; loans amounts increase up to Rs. 20,000. Unlike other microcredit organizations, Spandana does not require its clients to borrow to start a business: the organization recognizes that money is fungible, and clients are left entirely free to chose the best use of the money, as long as they repay their loan. Eligibility is determined using the following criteria: (a) female, (b) aged 18 to 59, (c) residing in the same area for at least one year, (d) has valid identification and residential proof (ration card, voter card, or electricity bill), (e) at least 80% of women in a group must own their home. Groups are formed by women themselves, not by Spandana.

[23] Due to the presence of spill over effects, they rely on what they denote the intent to treat estimator (ITT): ITT compares averages over treatment and control *areas*.

schooling: spending on tuition, school fees and uniforms is the same in treatment and comparison areas.

3.5 Summing up

The different studies covered above do not provide an unambiguous depiction of the impact of microcredit on key variables of interest such as income and consumption. Morduch (1998), Coleman (1999) and Banerjee et al. (2009) found no effect, while the other studies found mildly positive effects.

4. Donor impact evaluations of microcredit

Over the past decade impact evaluations have, as pointed out by Khalily (2004), come to dominate both the research and evaluation agendas within microcredit. According to Hulme (2000) this was caused by the advent of a "value for money", results-oriented political agenda in especially donor countries, leading to an increased focus on proving and documenting the outcome(s) and effectiveness of microfinance programmes.

The assessments of the implications and benefits associated with this emphasis on impact evaluations, however, vary considerably. Karlan and Goldberg (2007) find impact evaluations to be "akin to good market and client research" that may provide valuable information and lessons to both practitioners and policymakers. Hulme, at the other end, express fear that the managers of the evaluated programmes will perceive impact evaluations to be a donor-driven exercise – "an external imposition rather than a shared opportunity." This is partly because Hulme claims that impact evaluations tend to focus on proving the impact of programmes rather than improving the programmes. Although the latter objective is not incompatible with conducting an impact evaluation, it tends to be crowded out by attempts to establish the impact. Finally, Hulme points to the considerable lags between the initiation of an impact assessment and the dissemination of the results, potentially causing practitioners to perceive impact evaluations to be historical rather than relevant studies.

In this context, the evaluations presented here represent an opportunity to address the following three key questions:

   i) The review of peer-reviewed studies presented in Chapter 3 indicates that microcredit has a weak positive effect. Can this finding be corroborated by the evaluations conducted by donors and international organisations?

   ii) The peer-reviewed studies go to great lengths to minimize the potential biases resulting from non-random programme placement, non-random participation and non-random participant performance. Has the above-mentioned overlap in agendas and objectives between researchers and (donor) policymakers resulted in the adaptation of similar methods and considerations in the donor initiated evaluations?

iii) What about the practical relevance of impact assessments. Can the present review of evaluations assist in assessing the opposing views represented by Karlan & Goldberg and Hulme?

The basis for answering these questions is a review of 23 impact evaluations and 5 surveys conducted by bilateral donors, multilateral donors and international NGO's over the past decade. The list of relevant studies was identified in January 2009 through a criteria-based search of the websites of potentially relevant agencies and organisations.[24] As even a cursory reading of other existing surveys of microfinance studies (e.g. Goldberg (2005) and Hermes and Lensink (2007)) will indicate, the list of included evaluations is not complete. We are, however, confident that the screening process as well as consultations of previous surveys has resulted in a representative list of relevant studies.

The three key questions listed above will be addressed as follows:

i) Summary findings of the 28 evaluations and surveys are presented in an abbreviated form in a separate report[25], which will form the basis for the assessment of the results.

ii) An index score of how well evaluation methodologies and approach is suited to maximise the internal and external validity and relevance of an evaluation has been developed. A total of 21[26] of the 23 evaluations will, as a consequence, receive a score between 1 (being the highest) and 5, enabling a systematic survey of potential differences and trends in methodological "qualities".[27]

iii) The above-mentioned survey of the evaluations will also provide the basis for the ensuing discussion of the operational and practical relevance of impact evaluations – a discussion that was already initiated (albeit from a methodological perspective) in Chapter 2.

---

[24] The search criteria as well as the agency web-sites consulted is described in a separate note that is available upon request.

[25] The report is almost 50 pages long and is, to conserve the brevity of this report, not included as an appendix. It is, however, also available upon request.

[26] Two evaluations do not receive a score: Europeaid (2000) and Hussain et al. (1988). The latter is excluded because the terms of reference only cursorily include impact on household income and the former because it was published a full decade before the period considered in this survey.

[27] The details of the scoring methodology are presented in Appendix 2. In this context it suffices to note that the inclusion and weighting of both internal and external validity is an attempt to adopt a policymaker rather than an academic perspective in the scoring.

Following this introduction, the evaluations conducted by bilateral donors, multilateral donors and NGOs will be presented in separate sections. The issue of whether impact evaluations provide relevant input to a discussion of whether to scale up a given programme is postponed to the subsequent Chapter 5.

4.1 Evaluations conducted by bilateral donors

Bilateral donors have undertaken a number of microcredit impact evaluations, both with the purpose of assessing supported programmes and in order to gain knowledge on whether microcredit can be efficient as a development tool.

USAID was an early and substantial contributor in the field of microfinance evaluations and the accumulation and dissemination of microfinance experiences and practices. This resulted in a number of USAID-sponsored programmes aimed at providing practitioners and donors with a better understanding of how microfinance might reach and assist the poor. This includes Microenterprise Best Practices, MicroServe, Assessing the Impact of Microenterprise Services (AIMS), and Growth and Equity through Microenterprise Investments and Institutions.

Especially the AIMS programme generated several interesting and relevant studies, including the baseline and impact evaluations of microfinance programmes in Peru (Dunn & Arbuckle, 2001); Uganda (Barnes et.al, 2001a), Zimbabwe (Barnes et.al, 2001b) and India (Chen and Snodgrass, 2001) as well as several position papers about measuring and interpretations of the impact of microfinance.

The above-mentioned evaluations from Peru, Zimbabwe and India sought to determine the nature, extent and distribution of impacts from participation in microfinance programmes in three different parts of the world. A related objective was to obtain a better understanding of the role of credit within the household and the circumstances under which microfinance programmes can reach the poor. All three evaluations use the same approach, namely a comprehensive quasi-experimental[28] survey with clients and control groups of non-clients, supplemented with case studies of a smaller number of households to validate and (possibly) explain the mechanisms underlying the evaluation findings. Though a substantial effort has been put into identifying control groups, bias from self-selection and programme placement cannot be ruled out.

---

[28] The quasi-experiment undertaken is essentially the identification of a suitable control group <u>after</u> the programme has been implemented. Hence, the objective is to identify a group of people/households who did not participate in the programme, but who, at the time of programme enrolment, resembled programme participants. Subsequently, the current situation of the two groups – the participants and the constructed control group of non-participants are compared – this is also known as "matching".

The three evaluations analyse impacts on several dimensions of enterprise, household and individual welfare, and significant impacts from participation is found on many dimensions, e.g. income, assets, schooling (especially boys' schooling) and diversification of the sources of income (reduced vulnerability). The detailed level of the analyses makes it difficult to extract general lessons from the evaluations. But an important conclusion is that microcredit tends to benefit relative wealthy households the most[29]. Still, the three programmes evaluated have a substantial outreach. The Indian SEWA Bank, for example, reports that only 13% of its borrowers earned an income of more than 2 USD per day. The SEWA Bank evaluation furthermore finds that in some cases, clients who only have a savings account (i.e. clients who do not borrow) experienced larger impacts than clients who borrowed – suggesting that microcredit may not be the only financial service needed by poor clients.

The fourth USAID sponsored study considered here (Barnes et.al, 2001b) was an impact evaluation of three microfinance programmes in Uganda. The programmes – in addition to the provision of credit – have compulsory savings, weekly group meetings, and offer insurance. A random sample of clients in rural, suburban and urban areas is drawn, along with a control group consisting of recently active micro-entrepreneurs with the same gender distribution as the treatment group. The individuals are surveyed in 1997 and again in 1999, where 28% of the initially interviewed individuals could not be retraced. Comparisons across groups and time suggest that clients benefit from participation through an increase in income, expenditure and sales volume. Furthermore, the results suggest that loans can be used for other purposes than the planned within the household. The study does not, however, address possible additional differences between the control and treatment groups, and the high level of missing follow-up should also be a cause for concern when interpreting the results as self-selection and attrition may bias the results.

In 2004 the German development aid agency GTZ published an evaluation (Jansen, 2004) of their support to the microfinance Civil Society Organization (Calpiá) in El Salvador. Here, impact was defined as impact on rural clients' income situation and corresponding poverty measures from 1997 to 1999. To complement the monetary-based poverty assessment, a basic needs index was also estimated. The evaluation team conducted two surveys on the same households in 1997 and 1999 to form a panel of borrowers as well as a control group. The evaluation team was not able to observe whether control group members obtained loans from other sources, resulting in a possible bias of the results as the implicit assumption that the programme

---

[29] A similar regressive effect is found in other studies, e.g. Coleman (2002) who finds that relatively wealthy microfinance participants tend to gain the most from participation.

under evaluation addresses a binding credit constraint may not be true. Levels and differences between the two survey rounds are compared between the treatment and control group.

Results of the analysis suggest that Calpiá participation reduces income for all but the wealthiest households – a development, which is attributed to the accompanying shift in economic activities towards micro-entrepreneurial activities. This shift was, however, also found to be associated with a reduction in income variability (vulnerability). However, the study does not attempt to assign outcomes to the programme, so self-selection and differences between treatment and control group may also have exerted a significant impact upon observed outcomes.

In the following year GTZ published another study of interest (Hannover, 2005), which reviews the impact of microfinance linkage banking[30] in India – summarizing the results from three already existing studies.[31] Focus is on the use of microfinance to achieve the Millennium Development Goals, which is interpreted as identifying positive effects on income, assets, poverty levels, and a number of socioeconomic variables. Overall, significant positive impacts on income, savings and poverty levels are found for both small-scale farmers, rural micro entrepreneurs and landless labourers. The conclusions are, however, based on simple comparisons of before and after values of the relevant variables, i.e. no control group or other methodological issues are addressed in the study, invoking fears that self-selection and programme placement effects may drive the results.

Finally, GTZ published a collection of microfinance client stories about the social and economic impact of microfinance in the rural areas in Northern Namibia (Polzin, 2006). All entrepreneurs are members of a local microfinance associations network providing basic financial services – no indications are provided as to whether the clients are representative or how they were selected. For each of the clients a detailed story of their background, economic history, strategies and plans for the future is provided. All client stories state very positive impacts (in terms of e.g. increased income) from participation in the microfinance programme. The outcomes are detailed for each client, but the absence of information about sampling strategy

---

[30] Linkage banking is a partnership between a regulated financial institution (e.g. a bank or credit institution) and one or many independent, non-regulated institution(s), such as Savings and Credit Cooperatives (SACCOs) or credit-only microfinance institutions. Unlike a merger, linked institutions retain their independence.

[31] Two evaluations were carried out by the Department of Economic Analysis and Research of the Indian National Bank for Agricultural and Rural Development, whilst the third evaluation was undertaken by a local NGO called MYRADA. The emphasis of the latter was on the empowerment effects of programme participation.

and alternative outcomes makes it difficult to validate and (possibly) transfer experiences to other people/contexts.

Also in 2006, Oxford Policy Management published an evaluation of seven microfinance programmes in Pakistan for the Department for International Development (DFID). In addition to addressing a number of issues at the institutional level (e.g. financial sustainability and self perceptions) the evaluation also addressed programme impacts at the individual client level. Two earlier studies dating from before policy reforms eased the provision of credit in Pakistan are reanalysed, and this is argued to act as a baseline for the evaluation. Clients and non-clients are interviewed, and the clients reported substantial benefits from participation, which is also seen in relatively high profits on the investments generated from the loans. It should, however, also be noted that the surveyed programmes all appear to hold a non-market based (charitable) vision of microcredit operations, as the institutions are reported to be averse to charging sustainable (higher) interest rates for fear of being perceived as usurious and in conflict with their poverty alleviation mission. Furthermore, the results (not surprisingly) indicate considerable interchangeability of funds within households, rendering the interpretation of client based outcomes uncertain. Finally, the evaluation fails to address participant self-selection and possible programme placement effects, which may bias the results substantially.

To investigate whether support to microfinance is a good poverty alleviation strategy, the Swedish International Development Cooperation Agency (SIDA) have published two reviews of the impact of microfinance; one based on already published studies and research, primarily from Asia (Bali Swain, 2004), and another based on SIDA's experience with microfinance projects (Birgegård, 2004). Bali Swain reports that the evidence from existing impact studies is mixed. Limited evidence is found in favour of microfinance having an impact on consumption or income, but the positive impacts are mainly found for female borrowers. In addition, the review finds that many studies report a reduction in vulnerability, but that households near the poverty line are likely to benefit the most from microfinance, whereas benefits are almost non-existing for the extremely poor. Bali Swain also finds that in absence of economic growth, microfinance may result in redistribution of income rather than an increase of total income. And finally, some evidence of microcredit being a relatively cost-effective vehicle for poverty reduction is presented.

The review of Sida's experience in microfinance (Birgegård, 2004) briefly touches upon the limited number of impact evaluations made of the Swedish microfinance programmes. These evaluations are mainly case-studies, and in general reach the conclusion that participation in a microfinance programme reduces vulnerability and

that credit is most likely to benefit households that are relatively wealthy (own some assets) prior to entrance in the microfinance programme. The purpose of the study is broader than the present, focusing on policy options and their likely consequences.

4.2 Evaluations conducted by multilateral donors

Compared to many of the bilateral donors international organisations and multilateral donors have greater research and evaluation departments capable of lifting complex tasks as microcredit impact evaluation. It should therefore come as no surprise that the multilateral donors have produced more evaluations of this type, but it should be noted that the majority of these originate from a few, highly specialized and highly capable organisations like the Asian Development Bank and the World Bank.

Whether it is due to the regional dispersion of microfinance programmes or the presence of a highly skilled evaluation department, the Asian Development Bank (ADB) has so far undertaken far more microcredit impact evaluations than the other regional development banks.[32]

The Asian Development Bank undertook its first evaluation of a microcredit programme in 2001. The programme under evaluation was a training and credit project in Bangladesh, designed to alleviate rural poverty. The objectives of the programme were to provide livelihood skills training and microcredit to the landless rural poor and to improve the socioeconomic status of the programme beneficiaries. Finally, an additional objective of the programme was to strengthen the training and supervision capacity of the Executing Agency—the Ministry of Youth and Sports.

The basis for the evaluation of the effects upon participant's socioeconomic status was a retrospective household survey of 606 randomly selected households. The households were chosen from 18 centres to yield an appropriate geographical spread. The surveyed households were visited in 2001 and the study did not identify a control group or tracked performance over time. Instead, programme beneficiaries were asked to report the impact of microcredit upon enterprise turnover and household income. The self-reported effects were sought corroborated by the recording of household assets and housing status. Overall, the evaluation finds improvements in household income, housing conditions and household enterprise turnover, but acknowledges that the conducted before-and-after recall survey can-

---

[32] The African Development Bank has, for example, only conducted a single study, which furthermore was not found to be relevant for this survey.

not be used to identify programme specific-effects as it fails to counter potential self-selection and programme placement biases.

Another ADB evaluation of interest was undertaken by Gine et al. (2006) who cite two impact studies of microfinance institutions in the Philippines. Both impact studies seek to counter participant self-selection – one by using an instrument-variable approach and the other by using prospective clients as a control group. Despite being unable to meet the stated objective of reaching the poorest of the poor several positive impacts of the programmes are identified as participants experience an increase in household expenditures, incomes and assets. In addition, the case-studies corroborate the positive impacts as participants report greater flexibility and higher income. However, despite extensive and systematic triangulation of results the present evaluation set-up cannot completely rule out a potential programme placement bias.

The ADB Operations Evaluation Department, however, revisited the above-mentioned programmes in the Philippines in 2007 as part of a special regional survey.[33] The objective of the survey was to assess the extent to which selected ADB microfinance programmes have reduced the poverty of rural poor households and improved the socioeconomic status of women. The Philippines, Bangladesh and Uzbekistan were selected for the study, which conducted a nationwide household survey to measure the impact in the Philippines, whilst smaller sample surveys aimed at estimating socioeconomic profile of target groups were undertaken in Bangladesh and Uzbekistan.

In the Philippines the evaluation design was quasi-experimental as programme villages were matched to control villages scheduled for programme inclusion. In total 2200 households (10 participants and 10 non-participants[34] from each village) were included as the basis for a diff-in-diff evaluation aimed at eliminating unobservable fixed effects at the household and village levels. Moreover, significant efforts were exerted to trace and include programme dropouts in order to minimise potential attrition bias. In Bangladesh and Uzbekistan sample surveys were less ambitious and were undertaken to generate a socioeconomic profile of the target groups of the

---

[33] Kondo (2007) and Kondo et al. (2008) also present the evaluation undertaken in the Philippines, adding only details and few supplementary analyses compared to the original ADB presentation.

[34] It should be noted that eligible non-participants were identified by village leaders or microfinance institution personnel, constituting a potential source of bias (of unknown direction and magnitude).

respective projects.[35] Finally, approximately 200 women participated in 27 focus group interviews in the three countries.

In the Philippines the positive programme impact on per capita income was found to be significant (at a 10% level), whilst observed differences in total expenditure and food expenditure were also (mildly) significant. Differences in other impact measures were not statistically significant, including the impact on household asset holdings and human capital investments. The detailed and extensive data set collected enabled an analysis of impact across different levels of initial household wealth. This analysis indicated that impact was regressive – being positive only for the wealthier households within the sample, while the impact on the poorest households were found to be negative.[36]

In Bangladesh and Uzbekistan participants' self-assessment indicated that household income and food consumption rose following the receipt of a micro loan. The focus group interviews moreover indicated that microcredit had a positive effect on the status of women through a greater role in the generation of cash, greater involvement in decision making and acquisition of skills and assets.

Turning next to the Inter-American Development Bank (IADB) that played a pivotal part in establishing microfinance programmes and organizations in Latin America an evaluation of two microcredit programmes in Chile and Brazil was conducted by Aroca in (2002). In Chile, bank and NGO clients were compared to non clients in two regions; whilst the case study from Brazil was based on clients (two different banks and three different NGOs) who were compared to non-clients in five different regions. Very little information is provided on the different contexts and client groups whilst propensity score matching[37] is used to identify controls and correct for observed (and unobserved) differences. Results are equally perplexing as observed differences in income are found to be statistically insignificant in Chile, whilst they are found to be highly significant and very large (more than 100%) in Brazil. Unfortunately, the very limited information about potential differences in context, micro credit conditions and client characteristics as well as the potential programme

---

[35] In Bangladesh, a total of 200 respondents were identified from the Rural Livelihood Project and the Participatory Livelihood Project, whilst 84 respondents from various types of financial institutions were included in Uzbekistan.

[36] This corroborates the findings of Coleman (1999) and (2002) and Hulme and Mosley (1996).

[37] Matching is typically based on one or several characteristics that distinguish treatment and control groups. Propensity score matching uses several characteristics to estimate a predicted probability of, for example, obtaining a loan. This probability is usually obtained from logistic regression and is calculated for both treatments and controls that are subsequently matched based on their predicted probability of getting a loan.

placement and selection biases render interpretations and assessment of relevance very difficult.

The World Bank is given its scope, influence and resources almost always a significant contributor to the knowledge within a field, and microcredit impact evaluations are no exception. One of the first comprehensive (World Bank) studies on the impact of microcredit was Khandker (1998)[38] – a study that Khalily (2004) refers to as "the mother of all surveys in microfinance." It included almost 1.800 randomly selected households from 86 villages in Bangladesh. Households were clients in the Grameen Bank, BRAC or RD-12 and were selected on the basis of pre-survey censuses and the criteria that household landholdings could not exceed 0.5 acres.

As mentioned in Chapter 3.1 the (Pitt and) Khandker study was among the first evaluations to convincingly address potential selection and programme placement biases to establish a significant positive effect from microcredit programme participation and impact on household income and welfare. Khandker (1998) moreover studied village level programme effects as well as analysed the financial sustainability of the three organisations under review. Overall, the key conclusion of this expanded version of the WB-sponsored evaluation is that microcredit programmes can be an effective policy instrument for reducing poverty among the sub-group of poor people who have the skills to become self-employed. In addition, microcredit programmes are found to be cost-effective and of particular importance to women.

World Bank analysts returned to Bangladesh a few years later to re-analyse the poverty impact of BRAC programmes (Zaman, 2001). In total, 500 BRAC clients and 500 controls were given a microcredit questionnaire as part of the 1995 national household survey to test the dual hypothesis that microcredit mitigates vulnerability for all whereas impact on income is dependent on loan size (larger is better) and initial household characteristics (wealthier households are better suited to take advantage). A Heckman two-step estimation procedure is employed to account for problems related to self-selection, resulting in an analysis that to some extent corroborates the initial hypothesis. More specifically, the analyses indicate the existence of a threshold cumulative loan size over which microcredit can result in a significant poverty reduction. In addition, the estimations suggest that microcredit reduce vulnerability through increased options for diversification, consumption smoothing and/or empowerment.

---

[38] This is an expanded "book" version of Pitt and Khandker' s (1999) seminal research article presented in Chapter 3.1.

Overall, the decade from the middle of the 1990's to 2005 was characterized by high (and increasing) World Bank confidence in benefits and use of microcredit as an effective tool in the fight against poverty. The operational confidence in microfinance as a development tool was, however, shaken by the CGAP evaluation of the World Bank portfolio of microcredit programmes (Thomas, 2006). Commissioned in 2005 the evaluation was conducted by a panel of three microfinance experts, who independently graded 66 United Nations Development Programme (UNDP) microcredit programmes and 69 World Bank programmes based on field visits and their loan repayment rates and on overall cost recovery. The results were, to say the least, disappointing, as less than a quarter of the programmes which funded microcredit were judged to be successful. The rest failed, or appeared unlikely, to produce long-lasting, sustainable results, implying that future infusions of outside funding would be needed. Overall, the report described performance as "unacceptably low" classifying both agencies' programmes as being at "at the lower end of weak."

This no doubt provided the World Bank with further incentives to become better at assessing what programmes actually work and not least why. This is to a high degree reflected in the more recent World Bank studies. Most notably, Karlan and Goldberg (2007) provides a detailed overview of the methodological issues and challenges associated with doing impact evaluations in microfinance. The survey ends by concluding that randomized controlled trials are "*the most promising means to allow microfinance institutions to assess reliably the effectiveness of their operations on poverty alleviation and for investors and donors to learn which types of programmes produce the strongest welfare improvements.*"

Finally[39], Diagne and Zeller (2001) of the International Food Policy Research Institute (IFPRI) is another multilateral donor/organisation that provides a novel perspective to microcredit impact evaluation. Evaluating four microcredit programmes operating in the five districts of Malawi the basic assumption underlying this evaluation is that the impact of the credit programmes includes the effect on potential client's perceived ability to lend rather than just the actual loans provided. If households think that they in times of crisis will receive a loan they will, according to Diagne and Zeller, be willing to change risk taking behaviours. As a consequence, the analysis departs from standard practice and makes the distinction between access to credit (formal or informal) and participation (in formal credit programmes or in the informal credit market). A household has access to a particular source of

---

[39] EuropeAid (2000) is another multilateral donor evaluation of microcredit. It does, however, focus on microfinance institution institutional performance rather than on individual impact, implying that it was not considered relevant for this survey.

credit if it is able to borrow from that source, although for some reasons it may choose not to.

The data to test this hypothesis come from a three round study of 404 households in 45 villages in five districts of Malawi. The study took place in 1995 and involved asking respondents quite abstract (and for some hypothetical) questions about their perceived ability to obtain a loan from one (or more) of the four microfinance institutions under evaluation. A two-stage (maximum likelihood based) estimation method, similar to Heckman's two-step procedure for Tobit models, is used to estimate the effects of access to or improvements in access to credit, but the study finds no significant effect of access to credit on net crop income, per capita incomes, food security and nutritional status of credit programme clients. This leads to the conclusion that the "*contribution of rural microfinance institutions to the income of smallholders can be limited or outright negative if the design of the institutions and their services does not take into account the constraints on and demands of their clients.*"

4.3   *Evaluations conducted by international NGOs*

International NGO's (many of them Bangladeshi[40]) have played pivotal roles in the development and dissemination of microcredit innovations and lessons, making it natural to consider the impact evaluations conducted by them. As a consequence, a number of evaluations conducted by some of the most well known and resourceful international NGOs are described here.

BRAC is one of the largest microfinance institutions in Bangladesh officially targeting households with less than 0.5 acres of land and whose main occupation is manual labour. The former criterion is, however, in practice more adhered to than the latter. Still, BRAC has been found to recruit a significant proportion (50-60%) of its clients among the ultra-poor. Earlier studies have moreover found that 80% of BRAC credit is invested in income generating activities by the borrowing household.

BRAC conducted two large impact evaluations of its Rural Development Programme (RDP) in the 1990's (Mustafa et al. (1996) and Husain et al. (1998)). The first evaluation by Mustafa et al. undertook a large collection of data that included: (i) a household survey of 2125 BRAC and non-BRAC members over two rounds to capture seasonal variability, (ii) village profile surveys of 225 villages of which two-thirds had a BRAC programme, and (iii) 15 Village Organisation Case studies undertaken to contribute to the understanding and analysis of the socio-economic context

---

[40] The Bangladesh microfinance industry is described and analysed in detail by several studies including: Khandker (1998), Matin (2003), and Khalily (2004).

of BRAC members. The objective of the evaluation was to assess the impact of the RDP programme at household and individual lender level.

Unfortunately, the evaluation did not fully utilize the wealth and details of the data collected. The analysis was, as pointed out by Khalily (2004), by and large descriptive in nature, resorting to OLS[41] regression analysis and comparison of average outcomes across members and non-members, not testing the findings of the descriptive analysis. In particular, considerable attrition coupled with using length of relationship with BRAC-RDP as a key impact outcome could result in biased results.

Subsequently, Husain et al. (1998) conducted a follow-up study, which in addition to the parameters considered in the 1996 study also looked at poverty reduction and empowerment (in its various forms). In total, 1700 households (of which 1250 were BRAC beneficiaries) were included in the follow-up study. Unfortunately, the possibility of using the resulting panel data to assess the effect(s) of RDP credit on households and individuals was not utilized in the study. Overall, the findings of the first study was confirmed as RDP credit appeared to have a significant and positive impact on non-land assets, net worth, savings and children's schooling. The follow-up study, however, also raised a question about the sustainability of the impact given that household assets were found to decline for members associated longer than four years. Again, however, reservations about the robustness and validity of the results must be voiced given the descriptive and non-systematic nature of the analysis.

In another recent impact evaluation of another BRAC programme Barua and Suleiman (2007) evaluate the Northwest[42] Microfinance Expansion Project (NWEP) providing financial and non-financial services to poor women. Survey data were collected in 2003 and 2007 from eight BRAC branch offices in six districts. In total 1100 households were surveyed at the baseline – all of them beneficiaries of the BRAC project. The 2007 follow-up was able to find 92% of the original cohort. In addition to using the panel structure of the data collected IV[43] estimation was used to correct for biases resulting from participant variation in initial stock of assets and/or capabilities.

---

[41] Ordinary Least Squares (OLS) is a frequently employed method to estimate an underlying model thought to be responsible for a set of observations. The basic principle is to estimate the model that minimizes the aggregated distance between the observed value and the value given by the model.

[42] The North-western provinces of Bangladesh are poorer and relatively underserved with microcredit institutions compared to the rest of the country. For these reasons BRAC initiated the Northwest Expansion Programme.

[43] Variables related to the programme officers of the respective branch offices were used as instruments.

Overall, Barua and Suleiman find that the NWEP membership resulted in a significant improvement in household stock of assets (measured as quality of housing and ownership of durable goods). In addition, household poverty levels (measured by a poverty scorecard) were significantly reduced over the period.

Another well known and influential Bangladeshi microfinance institution is, of course, the Grameen Bank, which (in addition to being evaluated by, for example, Khandker, 1998) was subject to an early evaluation of its effects upon employment (Hossain, 1988). For the evaluation a survey was conducted of 280 households in five Grameen Bank villages and two control villages. After regression-based comparisons of the average treatment household and the average control household indicated reasonable levels of similarity between treatments and controls, impact was assessed as average differences in outcome between treatments and controls. The fact that self-selection was not taken into account in analysis of the data collected can perhaps explain the very large effects found. More specifically, Hossain finds that obtaining a Grameen Bank loan adds an extra 12 days/month to household employment. The extent to which this represents a reallocation from previously unrecorded activities is, however, also unclear.

The Grameen Bank also sponsored the survey of microfinance impact evaluations conducted by Goldberg (2005). The overall conclusion from this survey is one of optimism. Goldberg concludes that despite heterogeneous programmes and programme contexts microcredit appears to be able to lift participants' income and/or improve their livelihood. In addition, Goldberg calls for more (and better) research and that an "incontrovertible study" (which is synonymous with a randomized control trial) showing a positive impact will be of "enormous benefit" to the microfinance industry.

4.4 Summing up

With the exception of one study (Diagne and Zeller, 2001) the impact evaluations conducted by bilateral and multilateral donors as well as by a number of high profile international NGOs find (often significant) positive effects of micro-credit on household income or expenditures. This finding is irrespective of evaluation methodology and of the scoring presented in the subsequent Chapter 5. Hence, in contrast to the peer-reviewed evaluations the donor sponsored impact evaluations find almost unanimous support for a positive effect of microcredit.

Many evaluations contain lengthy considerations as to whether the microcredit programme under evaluation reached a stated target audience of the poorest of the poor – they never did. This leads some to conclude that the effect of microcredit is regressive, benefiting the wealthier clients relatively more. Although some evalua-

tions point in this direction no systematic evidence was found to support this – only a general agreement about the difficulties associated with reaching and assisting the poorest of the poor. Hence, despite the general agreement that "microfinance works" the aforementioned ambiguity should give rise to a disclaimer noting "… but not necessarily for all."

5. Scaling-up as a basis for scoring the evaluations?

The scoring methodology has been developed to summarize and synthesize the overall assessments of the surveyed microcredit evaluations. It is, as any composite index, a simplified representation dependent on both implied definitions and the ultimately *ad hoc* weighting of underlying factors.[44] Value added will consequently stem from the identification of trends.

The perspective adopted in the construction of the methodology is that of a policy maker facing a scaling-up decision. The question of scaling-up has preoccupied governments and donors since the first success stories of microcredit came out of Bangladesh in the early 1980s. Addressing to what extent a program can be scaled-up is, however, surprisingly difficult. It will among other things involve balancing the internal and external validity of the evaluation. Internal validity of an evaluation is concerned with estimating the true average treatment effect of the group affected by the programme, whilst the external validity deals with whether the causal effect can be generalized to other contexts and groups (see e.g. Meyer, 1995). To be sure, internal validity—the focus of academic studies—is not a "sufficient statistic" for policymakers interested in scaling up; they need to worry about external validity as well (see e.g. Rodrik, 2008).

As an example, consider a setting with two types of microcredit borrowers, safe and risky. Assume that the local credit market is being served by a profit-maximizing informal lender. If we introduce a subsidized microcredit lender relying on the group-lending technology into this environment, the new lender will attract the safe borrowers. This increases the average riskiness of the remaining borrowers (a sorting effect), causing the informal lender's average credit risk to rise (Bose, 1998) and scale advantage to fall (Hoff and Stiglitz, 1998).[45] As a result, the introduction of a subsidized microcredit lender worsens credit conditions for risky borrowers, i.e. those without access to microcredit. This may leave the economy as a whole worse off.

---

[44] The scoring methodology is described in detail in Appendix 2. The individual factors underlying the composite score is made available in Appendix 3, enabling still critical readers to form alternative composite scores.

[45] Recall from the Kaboski and Townsend (2005) study that microcredit institutions reduce borrower reliance on informal moneylenders.

Consequently, an impact evaluation focusing solely on the group served by a subsidized microcredit lender may (by the above logic) find unambiguous positive effects. However, the evaluation misses general equilibrium effects, which (by the same logic) can be negative. Put differently, an evaluation in this case would have no external validity. Even randomized evaluations are problematic unless accompanied by complementary studies.

Philosophers of science (e.g., Cartwright, 2007) thus argue that randomization will not even necessarily ensure internal validity. The problem is that evaluators are unlikely to rely exclusively on pre-defined rules and instructions when implementing an intervention as complex as a randomized experiment. This introduces a risk of subjective judgements into the experiment, rendering randomization no different from other econometric techniques. That is, if the premises on which the method rests are fulfilled, valid causal inference is ensured; if not, the conclusions derived from a randomized experiment may be as wrong as the results of any method of analysis.

In this light, case studies and even anecdotal evidence should not be discarded up front due to a "lack of rigor". Such "soft" evidence may, if collected according to transparent and representative criteria, be both internally and externally valid. The validity and precision of the conclusion is not assured by choice of a specific methodology. In this context, it is interesting to note that the Chinese, who pioneered experiment-based policymaking, never relied on randomized experiments. China successfully scaled up local policy innovations that appeared to work in specific contexts (Heilmann, 2008), relying mostly on soft evidence (Ravallion, 2009). As noted by Ravallion, it appears that reasonable lessons could be distilled from such "experiments".

The above discussion suggests that a basic premise for the evaluation scoring methodology presented here is that no method a priori is perceived to be superior, at least when scaling up is the aim.

The following three dimensions have been identified as relevant when designing the scoring methodology to accommodate the question of scaling-up:

i) Evaluation prerequisites: This dimension includes information about the programme subject to evaluation as well as the context under which it operates. In addition, the question of data and information availability and quality is also addressed here. Are conclusions based on sufficient data, enabling proper identification of programme effects? This dimension is thus included to survey how well the evaluation provides information about the content and context of the programme subject to evaluation.

ii) Addressing non-randomness: This dimension is included to assess the extent to which the evaluation succeeds in negating the effects due to non-random programme placement, non-random programme participation, and non-random programme participant performance. In addition, the fungibility of funds must be addressed – at minimum with the inclusion of a household level of measurement.

iii) Robustness and relevance: This dimension is based on an assessment of the extent to which evaluation conclusions are supported by the application of different methodologies and/or sources. Furthermore, the degree to which the external validity of evaluation results is addressed by evaluators is also considered in this dimension.[46]

The three dimensions are by choice non-substitutable and assigned equal weight. By implication, an evaluation that successfully addresses potential biases due to non-random placement, participation and performance will only score high if details of the programme content and context is provided alongside considerations of the generalizability and robustness of the findings.

A total of 21 evaluations[47] are scored on a scale from 1 to 5, where "1" is the highest score, reflecting an evaluation that comprehensively address all three dimensions, whilst a score of "5" represents an evaluation that fails to address one of the three dimensions and only addresses the other two in a rudimentary and incomplete fashion. As mentioned previously, all 21 evaluations have been published over the past decade by bilateral and multilateral donors (8 and 9 evaluations, respectively) and large NGOs. Table 1 below provides a break-down of the distribution of scores according to type of institution:

Score	5	4	3	2	1	Average score
Multilateral donors	0	2	4	2	1	2.8
Bilateral donors	1	1	3	3	0	3.0
International NGOs	0	1	1	2	0	2.8

Table 1: Scoring of micro credit evaluations

---

[46] It could be considered to include programme cost and cost-effectiveness under this dimension. This would, however, be a symbolic and non-distinguishing gesture as only the evaluations that already receive the highest score of 1 include any information and analyses related to programme costs.

[47] A list of the 21 evaluations and their scores is provided in Appendix 3

As is apparent from both the distribution of scores and average scores there is virtually no difference across multilateral and bilateral agencies. Only Khandker (1998) – at the time denoted the "mother of all surveys" (Hulme, 2000) – stand out with a collection of data that spanned years, surveying a large number of households in multiple locations significantly raising the costs and technical skills imposed upon the implementing agency.

According to Effron (2007) multilateral development banks and bilateral donors spend approximately the same share (1-2%) of their administrative budgets on monitoring and evaluation, but as budgets are larger in multilateral development banks this would enable them to undertake such larger evaluations as well as to build the necessary in-house capacity and skills to undertake these more demanding evaluations. Whether this is the explanation is, of course, open to speculation as the choice of evaluation methodology may also depend on the demand for knowledge, the presence of dedicated and skilful individuals and the level of institutional preparedness (Levine 2006).
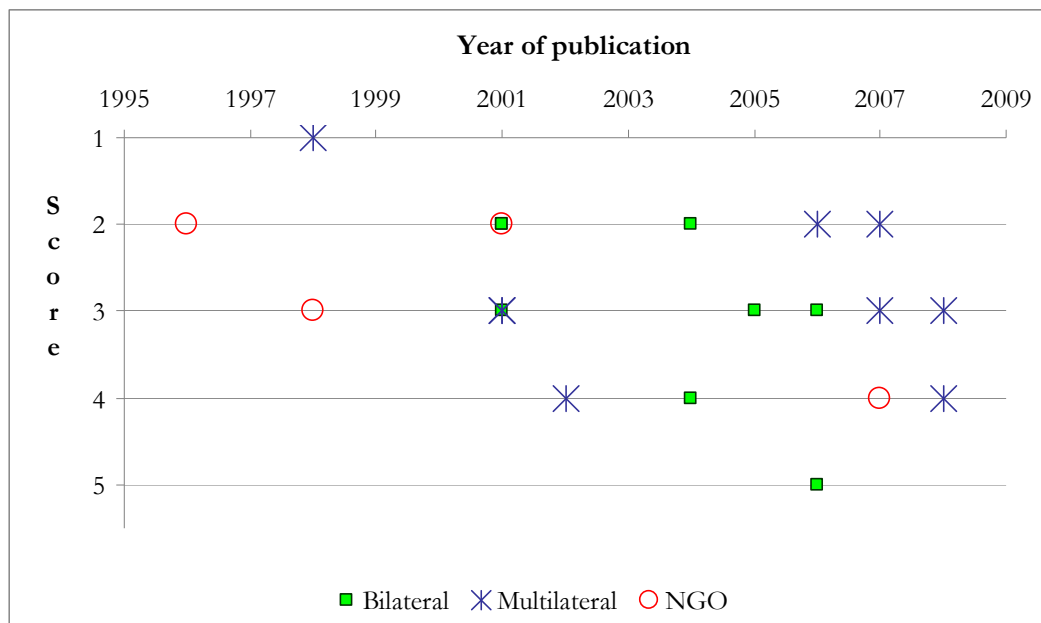

Figure 1: Scoring of micro credit evaluations by year

Looking at whether a trend can be observed Figure 1 depicts how scores evolve over time. Of the 21 evaluations undertaken over the past decade 11 have been published in last five years. Average score obtained over the last five years was 3.2 which is worse than the 2.8 average (remember: 1 is the best score) obtained in the first part of the decade. The absence of a clear trend over time also becomes appar-

ent looking at Figure 1, depicting scores for the three types of evaluating agencies according to publication year.

Attempting to deconstruct the scores, there appears to be a tendency towards the older evaluations being better at addressing relevance and robustness of the results, whilst the more recent evaluations perform better in terms of addressing the biases that may arise from non-randomness. This is, however, a tentative conclusion based on few observations.

Overall, the application of the scoring methodology indicates that:

1. Multilateral and bilateral donor evaluations (with the exception of large "flagship" evaluations) are approximately equally well suited as part of the basis on which to make a decision of whether to scale-up microcredit programmes or not.

2. No clear trend emerges over time. If anything, the potential usefulness of the evaluations in terms of deciding whether or not to scale-up has declined. This conclusion, however, rests on very few evaluations.

In terms of the future, programme statements and survey papers from the respective agencies indicate an inclination towards becoming better at addressing the internal validity of evaluations. Whether or not this will compromise the external validity of the same evaluations remains to be seen.

6. Conclusion

The key objective of this study has been to address the following question: On the basis of existing evaluations, what can be said about the causal impact of microcredit programmes on participant household consumption and/or income?

According to impact evaluations conducted by International Development Cooperation Agencies, the causal impact of microcredit on consumption and/or income is almost always significantly positive. This holds across different types of microcredit programmes and contexts. Unfortunately, these evaluations do not meet appropriate methodological standards in terms of addressing non-random selection issues. Peer-reviewed evaluations do not suffer from these methodological weaknesses but generally fail to find similar positive results. In fact, focusing only on internal validity, the most convincing studies find no impact of microcredit programmes on the level of income. Therefore, as also noted by Banerjee et al. (2009), microcredit does not appear to be the miracle it is often claimed to be.

The implications for donor support to microcredit programmes are not straightforward. Although peer-reviewed studies tend to find no effect, they have limited external validity. An educated decision about the appropriate scale of programmes must balance external and internal validity. Contexts and programmes differ, and it would be reckless to extrapolate from only a handful of studies, rigorous or otherwise. The conclusions emerging from the peer-reviewed studies, however, should serve to make us more sceptical about what can realistically be achieved through microcredit. Moreover, microcredit programmes continue to rely on donor subsidies for their continued existence. Hence, even an unequivocal positive impact would not be sufficient for continued funding. Sufficiency would require both a positive impact and that microcredit is the most cost effective policy among all policies with this positive impact. In sum, there are no magic bullets!

7. References

ADB (2007) *Effect of microfinance operations on poor households and the status women*, Asian Development Bank, Operations Evaluation Department Special Evaluation Study, Manila, The Philippines.

ADB (2006) *Performance evaluation report: The Rural Microenterprise Finance Project in the Philippines*, Asian Development Bank, Operations Evaluation Department, Manila, The Philippines.

ADB (2001) *Project performance audit report on the rural training project (loan 1066-ban[SF]) in the Peoples Republic of Bangladesh,* Asian Development Bank, Manila, The Philippines.

Armendariz B and Morduch J (2005) *The Economics of Microfinance*, MIT Press, Cambridge MA, USA.

Aroca P (2002) *Microcredit impact assessment: The Brazilian and Chilean cases*, The Inter-American Development Bank, Washington DC, USA.

Bali Swain R (2004) *Is microfinance a good poverty alleviation strategy? Evidence from impact assessments*, Swedish International Development cooperation Agency, Stockholm, Sweden.

Banerjee A et al (2009) *The miracle of microfinance? Evidence from a randomized evaluation*, MIT Department of Economics Working Paper, Massachusetts Institute of Technology, USA.

Barnes C et al. (2001a) *The impact of three microfinance programs in Uganda*, USAID AIMS core evaluation, United States Agency for International Development, Washington DC, USA.

Barnes C et al (2001b) *Microfinance program clients and impact: an assessment of the Zambuko Trust, Zimbabwe*, USAID AIMS core evaluation, United States Agency for International Development, Washington DC, USA.

Barua P and Sulaiman M (2007) *Impact evaluation and client satisfaction of Northwest Microfinance Expansion Project*, BRAC Research and Evaluation Division, Dhaka, Bangladesh.

Birgegård L-E (2004) *A review of SIDA's experience in microfinance*, Swedish International Development cooperation Agency, Stockholm, Sweden.

Bose P (1998) Formal - informal sector interaction in rural credit markets, *Journal of Development Economics* Vol. 56, 265–280.

Cartwright N (2007) Are RCT's the Gold Standard?, *Biosocieties*, 2, 11-20.

Chen M and Snodgrass D (2001) *Managing resources, activities and risk in urban India: The impact of SEWA bank*, USAID AIMS core evaluation, United States Agency for International Development, Washington DC, USA.

Coleman B E (2002) *Microfinance in North East Thailand: Who benefits and how much?* ERD Working Paper No. 9, Asian Development Bank, Manila, The Philippines.

Coleman B E (1999) The impact of lending in Northeast Thailand, *Journal of Development Economics*, 60, 105-141.

Commission of European Communities (2000) *Cross-evaluation of EC-financed projects involving micro-credit in Bangladesh*, The EU Commission, Brussels, Belgium.

Diagne A and Zeller M (2001) *Access to credit and its impact on welfare in Malawi*, The International Food Policy Research Institute, Washington DC, USA.

Duflo E (2004) *Scaling Up and Evaluation*, Annual Conference on Development Economics 2004, The World Bank.

Dunn E and Arbuckle JG (2001) *The impacts of microcredit: A case study from Peru*, USAID AIMS core evaluation, United States Agency for International Development, Washington DC, USA.

Effron L (2007) *Monitoring and evaluation in the World Bank and other donors*, Presentation given in Prague, February 2007.

Europeaid (2000) *Cross-evaluation of EC-financed projects involving micro-credit in Bangladesh*, The Commission of European Communities, Brussels, Belgium.

Ghatak M (1999) Group lending, local information and peer selection, *Journal of Development Economics* Vol. 60, 27–50.

Gine X, Jakiela P, Karlan D and Morduch J (2006) *Microfinance games*, Yale University Economic Growth Centre Discussion paper No. 936, Yale University, New Haven, USA.

Goldberg N (2005) *Measuring the impact of microfinance: Taking stock of what we know*, The Grameen Bank Foundation, Washington DC, USA.

Hannover W (2005) *Impact of microfinance linkage banking in India on the Millennium Development Goals (MDG): Summary of major results from existing studies*, Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), Berlin, Germany.

Heilmann S (2008) Policy Experimentation in China's Economic Rise, *Studies in Comparative International Development*, 43, 1-26.

Hermes N and Lensink R (2007) The Empirics of Microfinance: What Do We Know?, *The Economic Journal*, 117, 1–10.

Hoff K and Stiglitz J (1998) Moneylenders and Bankers: Price-Increasing Subsidies in a Monopolistically Competitive Market, *Journal of Development Economics*, 55 (2), 485-518.

Hossain M (1988) *Credit for alleviation of rural poverty: The Grameen Bank in Bangladesh*, IFPRI Research Report No. 65, Washington DC, USA.

Hulme D (2000) Impact assessment methodologies for microfinance: Theory, experience, and better practice, *World Development*, 28(1), 79-98.

Hulme D and Mosley P (1996) *Finance against poverty*, Routledge, London, England.

Hussain A M et. al. (1998) *Poverty alleviation and empowerment: The second impact assessment study of BRAC's Rural Development Programme*, BRAC Research and Evaluation Division, Dhaka, Bangladesh.

Imbens G and Wooldridge J (2009) Recent developments in the econometrics of program evaluation, *Journal of Economic Literature*, 47(1), 5-86.

Jansen S (2004) *Measuring the impact of microfinance: The Case of Financiera Calpiá, El Salvador*, Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), Berlin, Germany.

Kaboski J & Townsend R (2005), Policies and Impact: An Analysis of Village-Level Microfinance Institutions, *Journal of the European Economic Association* 3 (1), 1-50.

Karlan D and Goldberg N (2007) *Impact Evaluation for Microfinance*, Doing Impact Evaluation No. 7, World Bank, Washington, D.C.

Karlan D and Zinman J (2009) Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts, *Review of Economic Studies (forthcoming)*.

Karlan D and Zinman J (2008) *Expanding credit access: Using randomized supply decisions to estimate the impacts*, CGD Working Paper No. 108, Center for Global Development, Washington DC, USA.

Khalily MAB (2004) Quantitative approach to impact analysis of microfinance programmes in Bangladesh – what have we learned? *Journal of International Development*, 16, 331-353.

Khandker S (2005) Microfinance and Poverty: Evidence using Panel Data from Bangladesh, *World Bank Economic Review*, 19(2), 263-286.

Khandker S R (1998) *Fighting poverty with microcredit: Experience in Bangladesh*, The World Bank, Washington DC, USA.

Kondo T et al (2008) *Impact of microfinance on rural household in The Philippines*, The Philippine Institute for Development Studies, Manila, The Philippines.

Kondo T (2007) *Impact of microfinance on rural households in the Philippines: A case study from the Special Evaluation Study on the Effects of Microfinance Operations on Poor Rural Households and the Status of Women*, Asian Development Bank, Operations Evaluation Department, Manila, The Philippines.

Levine R (2006) *Multilateral development banks: Promoting effectiveness and fighting corruption*, Testimony to the Senate Committee on Foreign Relations on March 28 2006, Center for Global Development, Washington DC, USA.

Matin I (2003) *Poverty impact and outreach of microfinance in Bangladesh: Current knowledge, practices and moving ahead, issues and options*, Background Paper prepared for the review of the microfinance sector in Bangladesh, BRAC Research and Evaluation Division, Dhaka, Bangladesh.

McKernan S M (2002) The impact of microcredit programs on self-employment profits: Do non-credit program aspects matter? *Review of Economics and Statistics*, 84(1), 93-115.

Meyer B (1995) Natural and Quasi-Experiments in Economics, *Journal of Business and Economic Statistics*, 13, 151-161.

Morduch J (1998) *Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh*, Mimeo (New York University).

Mustafa S et al. (1996) *Beacon of hope: an impact assessment study of BRAC's Rural Development Programme*, BRAC Research and Evaluation Division, Dhaka, Bangladesh.

Oxford Policy Management (2006) *Poverty and social impact assessment: Pakistan microfinance policy*, Department for International Development assistance, London, England.

Pitt M and Khandker S (1998) The impact of group-based credit programs on poor households in Bangladesh: Does the gender of participants matter? *Journal of Political Economy*, 106(5), 958-996.

Pitt M (1999) *Reply to Jonathan Morduch's "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh"*, Mimeo (Brown University).

Polzin C (2006) *Koshi Yomuti – Banking under the tree: Client stories about the social and economic impact of microfinance in the rural areas of Northern Namibia*, Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), Berlin, Germany.

Ravallion M (2009) Are There Lessons for Africa from China's Success against Poverty? *World development*, 37, 303-313.

Rodrik D (2008) *The New Development Economics: We Shall Experiment But How Shall We Learn?* Mimeo (Harvard University).

Thomas J (2006) *Bold moves by World Bank and UNDP to improve poor performance in microfinance credit lines*, CGAP Press Release, Washington DC, June 9, 2006.

Winship C and Morgan S L (1999) The Estimation of Causal Effects from Observational Data, *Annual Review of Sociology*, 25, 659-707.

Zaman H (2001) *Assessing the poverty and vulnerability impact of microcredit in Bangladesh: A case study of BRAC*, The World Bank, Washington DC, USA.

Appendix 1.   Terms of reference

December 2008

Danida reference: 104.A.1.e.100

Introduction and background

Microcredit is regarded as a key development tool when it comes to reaching the poor and vulnerable. The (still growing) appeal of microcredit rests on its reliance on private enterprise as well as the (perhaps perceived) notion that channelling money to the poor through microfinance can be effective, sustainable and profitable.

As a consequence of the growing recognition and use of microcredit, the UN designated 2005 to be the International Microcredit Year. In addition Bangladeshi economist and microcredit pioneer, Mohammad Yunus and the high-profiled Grameen Bank (founded by Yunus) shared the 2006 Nobel Peace Prize for their efforts to create economic and social development from below.

Microcredit has in effect become part of the mainstream development toolkit, which has resulted in a parallel and very significant growth in funding and outreach. The increase in outreach is, for example, reflected in the 2008 annual report of the Microcredit Summit Campaign. The report summarizes data collected from more than 3,300 microcredit institutions reporting to reach 133 million clients by the end of 2006. This is up from approximately 600 institutions reporting having reached 13 million clients just nine years earlier.

Despite the broad recognition (and increasing commercialization) of microcredit, thoroughly tested knowledge about its impact and effect arguably remains partial and contested. The knowledge base has mostly consisted of anecdotal evidence, case studies and research that did not account for the potentially severe problems related

to endogeneity (see below) and/or failed to measure associated costs and benefits properly.

This is, however, beginning to change as recent and emerging research address key questions of whether and why microcredit works. The recent contributions can be categorized into:

- Controlled and quasi experiments: Gine et al (2006), Karlan and Zinman (2007), and Cason et al (2008) all attempt to untangle causal relationships within microcredit by conducting controlled experiments (microcredit games), in which a few parameters are randomly varied and the effects measured. Coleman (1999) is an example of a quasi-experimental approach.

- New estimation techniques: The research conducted by Kaboski and Townsend (2005) represent a novel estimation techniques using variations in policies and institutional characteristics to evaluate the impacts of village-level microcredit institutions in rural Thailand. Recent donor-based evaluations also occasionally rely on propensity score matching to create synthetic control groups.

- Comprehensive surveys: Hermes and Lensink (2007) present research filling an entire issue of the Economic Journal (volume 117, issue 517) devoted to presenting evidence on the impact of microcredit. Goldberg (2005) summarizes and synthesizes the evidence emerging from case studies, whilst Armendariz and Morduch (2005) is a thorough and comprehensive analysis of the available economic research.

Overall, the research based evidence base have thus been both growing and developing in new directions over the past five years. Whether this is also the case when it comes to the donor-based evaluations is not entirely clear at this stage. Some development cooperation agencies appear to have conducted comprehensive evaluations and analyses of microfinance programmes, including: The International Food Policy Research Institute (IFPRI), and the Asian Development Bank (ADB). It is, however, not clear whether this is a trend.

The context of a growing evidence base, substantive developments in evaluation methods and the growing mainstream adoption (and financing) of microcredit has prompted the Evaluation Department of Danida to commission a synthesis of the recent research and available finalized evaluations of support to microcredit.

Following this introduction the next section will outline the proposed approach, whilst the subsequent section describes the setup and practical details concerning the synthesis.

Approach

The key objective of the synthesis is to address the following key question: What works in microcredit? What can be said about the causal impact of a microcredit programme once an experimental/quasi-experimental evaluation design is applied?

Answering this question will be the key focus of the synthesis report and will take up the majority of the consultants' time and the final report. Secondary, and related to the above-mentioned key question, the consultants will also look into the methodology used in current microcredit evaluations undertaken by International Development Cooperation Agencies. This includes assessing the strengths and weaknesses associated with current evaluation methodologies?

The work preparing the synthesis will fall in three phases: literature survey and identification of relevant studies, drafting the report, presenting the report and getting feedback from external reference group before finalising the report.

Phase 1

A literature survey will be undertaken to identify a gross list of studies and evaluations that should be included. The identification of studies on the gross list will be based on the following criteria:

o   Definition: As mentioned previously microfinance in principle covers all types of financial services (credit, savings and insurance), whilst microcredit is the provision of small loans. The latter will be the focus of the studies and evaluations considered for this synthesis evaluation.

o   Period: Pitt and Khandker (1998) was an influential study because it was among the first attempts to use statistical methods to generate a truly accurate assessment of the impact of microcredit. Hence, most relevant studies will by implication be from the last decade, rendering attempts to define any other period than this unnecessary.

o   Geographical coverage: Other synthesis evaluations published by the Danida Evaluation Department focus on sub-Saharan Africa, reflecting the overall distribution of development challenges and aid across the major regions. The

concentration and influence of microcredit institutions and programmes in South East Asia and Latin America will, however, make it difficult to maintain (and justify) a focus on SSA when it comes to microcredit. Hence, although no specific geographical orientation will be imposed, the basis for the synthesis is likely to be dominated by studies of South East Asian or Latin American origin.

o Type of studies/evaluations included: The focus on methodological rigour and empirical testing is expected to result in (but will not be restricted to) a focus on empirical studies presented in peer-reviewed journals.

In addition, answering the question of what methodology is used in current microcredit evaluations undertaken by International Development Cooperation Agencies will, of course, also necessitate a survey of evaluations conducted by bi- and multilateral organisations.

o Type of impact considered: Here focus will be on studies that attempt to measure the causal impact of microcredit on economic outcomes such as consumption and income. Although of obvious and central importance this is, as pointed out by Armendariz and Morduch (2005), not sufficient basis to act.

Funding and implementing agencies should also consider and assess the cost-effectiveness of the programmes under evaluation. This will, however, be outside the scope of this synthesis evaluation. In addition, evaluations and studies that consider support given to develop microcredit infrastructure and/or tools (like, for example, customer or institutional rating schemes) will not be considered either.

Having identified studies that are relevant and meet the above mentioned criteria, the consultants will present a list of evaluations and studies, which will also be annexed to the final report. Based on the list the consultants will propose a sample of evaluation and studies to be analysed in more detail in the report. The sample, which will be approved by the Evaluation Department of Danida, should be selected based on the above mentioned criteria.

Phase 2

Based on the list of selected studies, the consultants will complete a draft report that seeks to answer the two key questions: (i) what can be said about the causal impact of a microcredit programme once a proper evaluation methodology is applied? And (ii) what type of evaluation methodology is applied by development cooperation agencies when it comes to assessing impact?

The studies and evaluations contributing to answering the question of "what works?" must to some extent have addressed the potential problems such as endogenous programme participation and placement, which may contaminate the evaluation design and result in biased causal inference.

<u>Phase 3</u>

The consultants will present a draft report subject to discussion and feed-back from both Danida staff and a reference group of external experts established for this study.

Suggested members of the external reference group include:

- o Nikolaj Malchow-Møller, Department of Business and Economics, University of Southern Denmark

- o Yet to be decided

The input collected from Danida and the reference group will be used to finalise the report. The final report will be a synthesis report of not more than 30 pages plus appendices.

The final synthesis report will be published and made available on the Internet by the Evaluation Department of Danida.

Study setup

The team of consultants will consist of:

- o Team leader Jens Kovsted, Ph.D., Centre for Economic and Business Research, CBS,

- o Thomas Barnebeck, Associate Professor, Ph.D., Department of Economics, University of Copenhagen,

The Synthesis Evaluation will be managed by the Danida Evaluation Department, but officials from other departments in Danida can be consulted during the elaboration of the Synthesis Evaluation.

The work of the consultants will commence at 1. December, 2008. Total working time allocated to this project is 12 weeks of which 6 are reserved for Thomas Barnebeck.

The proposed set of evaluations and studies to be included in the sample will be presented to Danida not later than 5. January 2009.

A draft report will be submitted to the Evaluation Department of Danida not later than 3. March 2009, and a final report not later than two weeks after comments to the draft report have been received from the consultants undertaking the study.

Appendix 2.　　　　Scoring methodology

The scoring of evaluations has been based on the degree of compliance with the following requirements:

Prerequisites

1.1.　A detailed and thorough description of the programme under study, enabling the reader to assess the context and potential consequences for the external validity

1.2.　Surveys include a sufficient number of observations and are carefully designed to enable identification of treatment effects

Non-random effects

2.1.　Non-random programme participation (i.e. self-selection of individuals into the programme) has been addressed

2.2.　Non-random programme placement has been taken into account

2.3.　Non-random programme performance (differences in treatment effects and attrition between groups) have been taken into account – unless the interest is explicitly stated to be in the "treatment effect on the treated"

2.4.　Loan fungibility between the assisted household member/enterprise and the entire household is addressed

Robustness and relevance

3.1.　The extent to which programme insights and conclusion is of relevance to policy and programme design in other contexts (the question of external validity) is discussed

3.2.　A number of techniques (e.g. surveys, case studies etc.) have been used to complement the analysis and test for robustness (triangulation)

Each of the relevant evaluations has been assigned a score on a scale ranging from 1 to 5, where 1 is given to a study which complies with all or almost all of the above

requirements and 5 is given to a study which complies with none or very few of the above requirements as indicated below:

Score	Prerequisites	Non-random effects	Robustness and relevance
1	All issues are addressed	3 of 4 issues are addressed	All issues are addressed
2	All issues are addressed	2 of 4 issues are addressed	1 issue is addressed
3	1 of 2 issues are addressed	2 of 4 issues are addressed	1 issues is addressed
4	1 of 2 issues are addressed	1 of 4 issues is addressed	1 issue is addressed
5	1 dimension is not addressed, and the others are addressed at a rudimentary level		

Please note that:

- The listed scoring criteria are non-substitutable across the three dimensions. Hence, an evaluation that only address one of the four issues raised under the heading "non-random effects" will receive a score of 3, regardless of how the same evaluation score under the other two dimensions listed. As consequence, the table is applicable to combinations of scores that are not listed in the table above.

- How detailed and comprehensive an evaluation will have to be in order to meet one of the nine criteria listed is based on our (subjective) assessment. It will, however, not suffice to merely acknowledge the existence of a potential problem of non-random programme participation – the evaluation will have to actively address the problem.

- Some of the criteria may be considered interdependent. The application of appropriate statistical methods will, for example, depend on the existence of a sufficient number of observations. Complementarities and overlap of this nature is, however, impossible to avoid, implying that the reviewer will have to take it into account in a systematic and consistent manner.

Appendix 3.   Evaluation scores

The 21 evaluations that were scored according to the just presented scoring methodology were:

Name of study	Score
Polzin (2006) *Koshi Yomuti – Banking under the tree*	5
Hannover (2005) *Impact of microfinance linkage banking in India*	3
Jansen (2004) *Measuring impact of microfinance: The case of Financiera Calpiá, El Salvador*	4
Oxford Policy Management (2006) *Poverty and social impact assessment: Pakistan microfinance policy*	3
Barnes et al. (2001) *The impact of three microfinance programs in Uganda*	3
Dunn and Arbuckle (2004) *The impacts of microcredit: A case study from Peru*	2
Barnes et al. (2001) *Microfinance program clients and impact: Zambuko Trust, Zimbabwe*	2
Chen and Snodgrass (2001) *The Impact of SEWA Bank, India*	2
ADB (2001) *Project performance audit report on the Rural Training Project, Bangladesh*	3
ADB (2006) *The Rural Microenterprise Finance Project in the Philippines*	2
ADB (2007) *Effect of microfinance on poor rural households and the status of women*	2
Kondo (2007) *Impact of microfinance on rural households in the Philippines: Empowerment*	3
Kondo et al. (2007) *Impact of microfinance on rural households in the Philippines*	4
Aroca (2002) *Microcredit impact assessment: The Brazilian and Chilean cases*	4
Khandker (1998) *Fighting poverty with microcredit*	1
Zaman (2001) *Assessing the poverty and vulnerability impact of microcredit in Bangladesh*	3
Karlan & Zinman (2007) *Expanding credit access using randomized supply decisions to estimate the impact*	3
Diagne and Zeller (2001) *Access to credit and its impact on welfare in Malawi*	3
Mustafa et al (1996) *Beacon of hope: an impact assessment of BRAC's rural development programme*	2
Barua and Sulaiman (2007) *Impact evaluation and client satisfaction: NME Project, Bangladesh*	4
Husain et al. (1998) *The second impact assessment study of BRAC's Rural Development Programme*	3

Other Evaluation Studies:
2007/1: Assistance to Development Research
2007/2: Synthesis of Evaluations on Technical Assistance
2007/3: Evaluating the Impact of Rural Roads in Nicaragua
2008/1: Synthesis of Evaluations of HIV/AIDS Assistance
2008/2: Synergies between Bilateral and Multilateral Activities
2008/3: Assessing Multilateral Organisation Effectiveness
2008/4: Economic Empowerment of Women
2008/5: Public Private Partnership Programme
2009/1: Evaluating Aid Effectiveness in the Aggregate: Methodological Issues
2009/2: Addressing the Underlying and Basic Causes of Child Undernutrition in Developing Countries: What Works and Why?
2009/3: Experiences with Conducting Evaluations Jointly with Partner Countries