Evaluating Microfinance Program Innovation with Randomized Controlled Trials: Examples from Business Training and Group versus Individual Liability[*]

For the workshop titled, "Using Research Findings to Improve Design of Products and Services" at the 2006 Microcredit Summit

Dean Karlan, Yale University
Tomoko Harigaya, Innovations for Poverty Action
and
Sara Nadel, Innovations for Poverty Action

July 2006

Abstract

This paper presents an application of the randomized controlled trial methodology to evaluating modifications to the design of microcredit programs. As microfinance becomes an even more popular tool for fighting poverty, institutions innovate in their products and programs at a rapid pace. Policymakers and practitioners should know the relative impact of different designs, both to the client (in terms of welfare) and to the institution (in terms of financial sustainability). We discuss the current approach to evaluating product or program changes, and the reasons why more rigorous evaluations are necessary. We then discuss why randomized controlled trials can prove vital to microfinance institutions in identifying effective program designs in different environments. In this paper, we focus on the choice of lending methodologies – credit with education versus credit only, and group versus individual liability -- to illustrate the benefits of randomized controlled trials as a business tool for measuring impact and learning how to improve sustainability and growth. This methodology can be employed for a plethora of program design issues, such as timing of payments, loan size, interest rates, term, and other services such as insurance and savings.

1. Introduction

In the last decade, microfinance institutions have experienced a boom in innovations of lending products, partly fueled by donors who see microfinance as the next promise to alleviate poverty. Examples of these new products are the combination of credit with health or life insurance, business and health education, savings products, and the adoption of (or conversion to) individual loan liability. The add-in features generally aim at reducing the vulnerability of clients while contributing to asset creation, hence improving their repayment rate and the sustainability of the service. The product innovations typically result from organizations striving to extend outreach, increase impact, and promote sustainability. As in other industries, microfinance institutions (MFIs) typically decide whether to adopt new strategies based on other MFIs' success with the innovations. Many new micro-lending products and approaches continue to be developed. However, MFIs must generally rely on qualitative and descriptive case studies and anecdotal evidence on the effectiveness of these innovations to decide whether to implement the new strategies. The usual case study approach does not provide tangible evidence that can enable other organizations to know what changes can be expected if they were to adopt similar product changes.

In this paper, we discuss how randomized controlled trials can help test the effectiveness of new lending products. Similar to biomedical studies that test new pharmaceuticals, randomized controlled trials in the field of microfinance isolate the effect of a chosen innovation by assigning a random selection of individuals or villages to the innovation (the treatment group), and another equivalent selection of individuals or villages to maintain the status quo (the control group) and comparing results between the groups. Given the growing innovation of lending product designs among microfinance institutions, it is critical to establish a systematic and reliable evaluation method which measures the impact of specific characteristics of a lending product. Throughout the paper we present as an example an ongoing randomized control evaluation of group-versus individual-liability loans in the Philippines and of a business-development training that was offered in conjunction with a credit program. Many of the issues discussed in this example, however, apply to evaluations of a wide variety of microlending product innovations. We discuss a few further examples at the end of the paper.

Many microfinance institutions test new product designs by allowing a few volunteer clients to use a new lending product, or by offering to a small group of particularly chosen clients (often, their best) a new product. Alternatively, a microfinance institution can implement a change throughout one branch (but for all clients in that branch). We argue that such approaches are risky for lenders, and inferences about the benefits of changes evaluated in such a manner can be misleading. As explained in Section 2, one cannot conclude from such non-experimental approaches that the innovation or change *causes* an improvement for the institution (or the client). Establishing this causal link should be important not only for the microfinance institution implementing the change, but also for policymakers and other MFIs which want to know whether they should implement similar changes. This is a situation in which randomized controlled trials are

a win-win proposition: less risky (and hence less costly in the long run) from a business and operations perspective, and optimal from a public goods perspective, in that through research the lessons learned can be disseminated to other MFIs.

The primary operational differences between experimental and typical non-experimental evaluations are two-fold: First, experimental evaluations include random assignment (rather than self-selected or MFI-selected) of individual clients (or groups of clients) to different program designs or products. This eliminates the chance that results will be confounded by factors not causally related to the intervention (such as one's entrepreneurial spirit which led them to join the program and improve their business, irrespective of the credit). Second, experimental evaluations are *prospective (i.e., both participants and the control group are randomly assigned at the outset of the study)*, whereas typical (but not all) non-experimental evaluations are *retrospective (i.e., a comparison group of non-participants who are selected to be similar to participants is chosen after the treatment)*. The prospective nature of randomized evaluations makes planning *before the innovation is launched* the most important stage of the evaluation. This paper hopes to shed insights into the motivations for and possibilities of using experimental evaluations to assess different microfinance product designs. The paper is written primarily for microfinance practitioners.

The rest of the paper is organized as follows. In Section 2, we discuss the problems with non-experimental approaches usually used in microfinance to evaluate program innovation and the reasons why the random control trial methodology is preferable. In Section 3, we introduce experimental pilot approach to product innovation and the steps to design a randomized controlled trial. In Section 4, we discuss some key issues needed to be considered when designing an experiment. In Section 5, we present an example of a randomized controlled trial on credit with education versus credit only in Peru.. Section 6 presents an example of testing group versus individual liability in the Philippines. Section 7 provides further examples of different microfinance programs in which randomized controlled trials could be employed to evaluate the program impact. Finally, Section 8 concludes.

2. Why do we need a control group and how can we get it

In evaluating a lending product innovation, we typically discuss an existing loan program in which some change is being made. Therefore an existing client base already exists. Program innovation evaluations seek to compare the actual outcomes of an innovation to a program with the outcomes that would have resulted in the absence of the innovation. Because a potential client can be either borrowing in the program with the new lending feature or not, and cannot do both, the outcomes in the absence of the new lending feature are unobservable for those who receive the new product. Any evaluation then amounts to establishing the counterfactual outcome: *what program client's outcomes would have occurred had the new lending feature not been introduced?*

a. Why do we need a control group?

In the field of microfinance, practitioners frequently evaluate new lending products by using non-experimental designs. Most often, they let a few volunteer clients use the new lending product under study, or offer it to a small group of selected clients (usually their best). Alternatively, they introduce a product change throughout an entire branch of a lending institution with all clients in that branch using it. The evaluator then attributes the observed change in the clients' outcome indicator to the product change introduced, without explicitly constructing what would have happened had the change not taken place (the "counterfactual" outcome).

This type of evaluation contains a strategic error. The problem is that in addition to the introduced product change there may be other factors that also contribute to the changes in clients' outcomes. These other factors may come from the environment in which the clients live or may be peculiar to the clients. For example, suppose we are interested in the change in the clients' income. The observed increase in the clients' income may be due to several factors: (i) the product change introduced; (ii) general economic improvement in the region; (iii) new income-generating opportunities (e.g., a new factory in the region); (iv) the ability of the borrower to use the loan effectively; and so forth.

Consider a farming community that enjoyed unusually favorable weather conditions at the onset of the introduction of a new product. It is observed that clients' income rose during the study time. Given only this observation, an evaluator cannot be sure if the rise in income was completely attributable to the new product, or is mostly due to better harvest that results from good weather. The estimation error (also called bias) in this case is the growth in income due to better harvest brought about by favorable weather.

Accurate evaluations must control for these external intervening factors. If we could observe the same client at the same point in time both borrowing and not borrowing the new loan product, this would effectively account for any other observed and unobserved intervening factors. Since this is impossible, to isolate the effect of the product change from effects caused by other intervening factors, a control group is necessary. We need a comparator group of clients not availing the new lending product but have similar characteristics as those borrowing. Simply observing the change in clients' outcomes without a control group makes it impossible to assess the true, isolated impact of the product feature being evaluated.

Even evaluating the success of a product change based on the experience of one entire branch to which the innovation was introduced is erroneous. In such evaluations, the evaluator assumes that results of clients in the selected branch would have been similar to results of clients in other branches that did not receive the innovation, had they themselves also not received the product change. This approach is flawed because each branch is unique in its characteristics, with different geography, economic conditions, and human resources of the branch staff. Just as before, when the improvement in clients' incomes might have been caused by favorable weather, two branches with different characteristics can have quite divergent experiences. For example, if the branch with the

product innovation being evaluated happened to have clients with more entrepreneurship, comparing it with other branches could cause the MFI to falsely attribute the difference to the effect of the innovation.

An MFI with a sufficient number of branches could in fact compare *several* branches which receive the innovation to *several* branches which do not. This would require the branches to be randomly assigned to treatment and control groups, as described below, and a sufficient number of branches to allow for an adequate sample size.

b. Why the control group should be randomly chosen

The objective of product change evaluation is to establish a credible control group of individuals who are identical in every way to individuals in the treatment group, except that they are not accessing the new product.

Establishing such a credible control group faces some difficulties in practice. The problem is that in reality borrowers and non-borrowers usually are different. Microfinance programs usually target certain groups of clients, such as women in poor neighborhoods. This endogenous program placement effectively makes borrowers and non-borrowers different in some set of characteristics (e.g., on average borrowers have a lower income than non-borrowers). When participation is voluntary, the fact that clients select themselves into the program indicates differences (observable or unobservable) between borrowers and non-borrowers. For instance, borrowers in microcredit programs designed to promote household businesses may be intrinsically more entrepreneurial than non-borrowers. Or in a program of credit with education designed to promote children's education, borrowers may choose to borrow because they value their children's education more than non-borrowers do.

Because institutions *choose* where to lend and not to lend deliberately (known as "endogenous program placement), and because participants *choose* to borrow or not borrow (known as "endogenous program participation), those who are not borrowing are often not a good comparison group for those borrowing. The reason why an institution enters one area, allows some individuals in, or why some individuals choose to borrow is critical to understand, and yet also difficult if not impossible to fully understand. Merely observing that *some* characteristics are similar is not sufficient. Often the *unobserved* characteristics are believed to be the most important (e.g., entrepreneurial spirit, or expected economic growth in an area). When these selection issues are not dealt with properly, the observed difference in outcomes can be attributed to both the program's impact and the pre-existing differences between the two groups. The comparison between the two groups will yield the accurate program impact only if the two groups have no pre-existing differences other than access to the product change being evaluated.

The key feature in experimental methods is random assignment. Random assignment removes any systematic correlation between treatment status and both observed and unobserved characteristics of clients. Clients (or groups of clients) are randomly assigned to a treatment group (who will borrow the new lending product under study) and

a control group (who will not borrow). By construction, the randomization procedure ensures that the two groups are identical at the outset. Individuals in these groups live through the same external events throughout the same period of time, and thus encounter the same external intervening factors. The only thing different between the two groups is that those in the treatment borrow the new product and those in the control do not. Therefore, any difference in the outcomes between the two groups at the end of the study must be attributable to the product change. Random assignment assures the direction of causality: the product innovation or change *causes* an improvement for the client (or the institution).
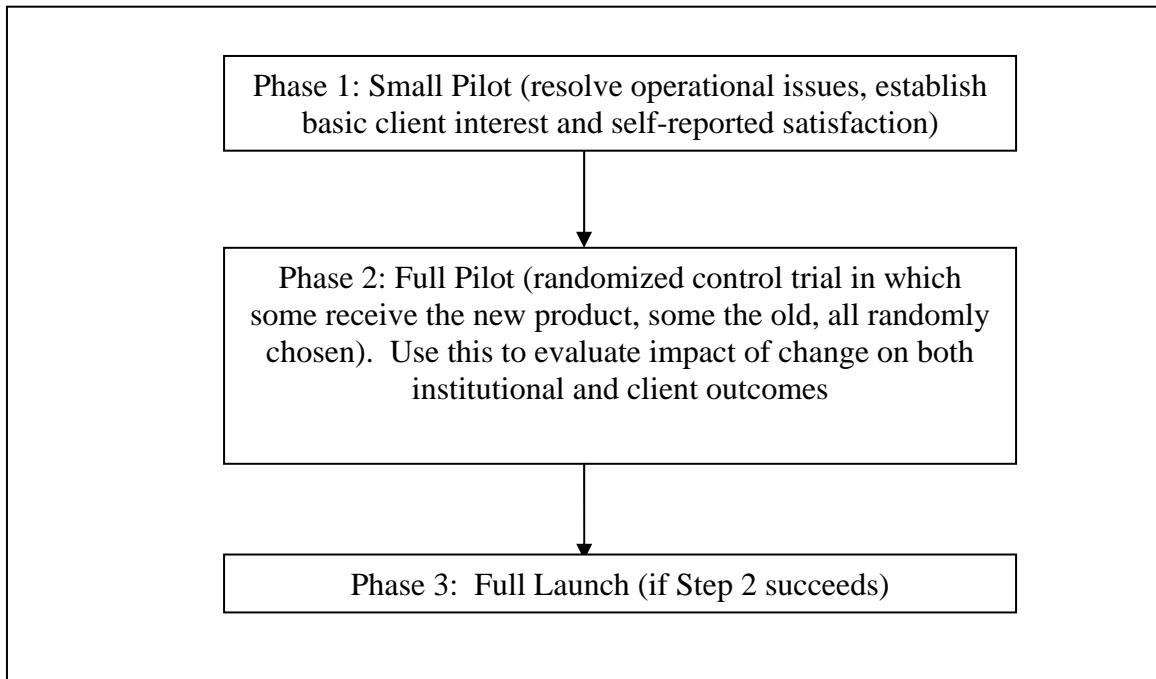
3. Experimental pilot approach to product innovation.

In a randomized controlled trial, one program design is compared to another by randomly assigning clients (or potential clients) to either the treatment or the control group. If the program design is an "add-on" or conversion, the design is simple: The microfinance institution randomly chooses existing clients to be offered the new product. Then, one compares the outcomes of interest for those who are converted to those who remained with the original program. A similar approach is also possible with new clients, although it is slightly more difficult. In this section, we will discuss the logistics of how to change an existing product, where clients already use some service in the program.

The flowchart below presents the basic phases. Often, microfinance institutions innovate by doing a small pilot and the full launch (Phases 1 and 3), but not a full pilot (Phase 2). Hence, this paper focuses heavily on why this second step is important and outlines its basic steps.

a. Identify the problem, potential solution, and conduct a small pilot

Product innovation typically aims at solving a problem of the existing product or improving the impact and feasibility of the product. The first step is to identify the problem of the current product and potential solutions through a qualitative process. This should include examination of historical data, focus groups and brainstorming sessions with clients and staff, and ideally discussions with other microfinance institutions that have had similar problems. Once a potential solution is identified, an operating plan and small pilot should be planned.

```
┌─────────────────────────────────────────────────────────────┐
│   ┌─────────────────────────────────────────────────────┐   │
│   │  Phase 1: Small Pilot (resolve operational issues,  │   │
│   │  establish basic client interest and self-reported  │   │
│   │  satisfaction)                                      │   │
│   └─────────────────────────────────────────────────────┘   │
│                          │                                  │
│                          ▼                                  │
│   ┌─────────────────────────────────────────────────────┐   │
│   │  Phase 2: Full Pilot (randomized control trial in   │   │
│   │  which some receive the new product, some the old,  │   │
│   │  all randomly chosen).  Use this to evaluate impact │   │
│   │  of change on both institutional and client         │   │
│   │  outcomes                                           │   │
│   └─────────────────────────────────────────────────────┘   │
│                          │                                  │
│                          ▼                                  │
│   ┌─────────────────────────────────────────────────────┐   │
│   │   Phase 3:  Full Launch (if Step 2 succeeds)        │   │
│   └─────────────────────────────────────────────────────┘   │
│                                                             │
└─────────────────────────────────────────────────────────────┘
```

An operating plan should include specifics on all necessary operations components to introduce the proposed change. This includes, for instance, development of training materials, process for training staff, changes to the internal accounting software, compensation systems, and marketing materials.

In order to resolve operational issues and depending on the complexity of the proposed change, a small pilot implementation should be done next. This can be done on a small scale, and is merely to test the operational success of the program design change. This initial pre-pilot does *not* answer the question of impact to the institution or the client. It instead intends to resolve operational issues so that the full pilot can reflect accurately the impact from a full launch.[†]

 After the proposed solution has been identified and a small pilot is conducted, the "testing" is not over. It is important to know the impact of the product innovation on both the institution (repayment rates, client retention rates, operating costs, etc.) and the client (welfare, consumption, income, social capital, etc.). To measure such outcomes properly, one can not merely follow the participants and report their changes. The flaws of this were discussed in the previous section. One needs a control group.

        b. Identify treatment assignments

[†] This paper does not elaborate on this step any further, as much has been written on it already by organizations such as Micro-Save Africa. In this paper we put forth a process that begins where such organizations stop.

Often a proposed solution has a main change, but many minor issues that need to be decided. For instance, when testing Credit with Education in the FINCA program in Peru (Karlan and Valdivia 2006), we had to select which type of education modules to offer, and when testing group versus individual liability (Giné and Karlan 2006), we needed to determine how to set the savings policies. A careful experimental design can include tests of such sub-questions. Specific examples will be provided below when we discuss testing group versus individual liability. These questions often arise naturally through the brainstorming questions. Any contentious decision is perfect for such analysis, since if the decision is contentious then the answer is not obvious!

c. Sample frame and sample size

The sample frame is the pool of clients (or potential clients) who are included in the impact study. One will assign clients (or potential clients) randomly to "treatment" or "control" groups (that is, clients will be divided randomly into at least two groups. Members of one group will get the innovation and members of the other will not). Two types of sample frames should be considered: existing clients and new clients. When the innovation is a change to an existing product, an initial test can consist of existing clients. Defining a sample frame of potential clients can be more difficult. The following section shows how this is being done with the group versus individual liability evaluation in the Philippines.

Determining necessary sample size is also key to a successful evaluation. To calculate the necessary sample size, one needs to consider (a) what a "successful" outcome looks like (e.g., if repayment rates are 90%, would increasing them to 94% be considered satisfactory enough to then warrant a full conversion to a new product?), (b) what the current level is for the outcome measure, and (c) if the outcome measure is not a binary variable (e.g., being in default), then one needs to know the typical variation (i.e., the standard deviation) of the outcome of interest.[‡]

4. Issues to be considered when designing an experiment

a. Spillovers

The validity of experimental designs rests on integrity of the data from treatment and control groups. The results are improved when treatment and control groups remain intact throughout the study.[§] However, in microfinance programs, this cannot always be guaranteed and spillovers may arise. With proper care and information about non-

---

[‡] We recommend the free software *Optimal Design* for helping to determine sample sizes. It can be downloaded from http://sitemaker.umich.edu/group-based/optimal_design_software.

[§] In some experimental designs, treatment is not mandatory for the treatment group, and/or control group are permitted to get treated. These are called *encouragement* designs, in which everyone receives the treatment but only the treatment group is given an encouragement to participate in the treatment (but is not required to participate). The control groups are not given the encouragement (but are allowed to participate in the program if they choose so.) These work as long as the encouragement leads to a higher enough take-up rate in the treatment group than the control group. See Ashraf, Karlan and Yin (2006c; 2006b) for an example of such a design.

compliance, this can be dealt with in the analysis (although if the non-compliance is severe, this could irreparably damage the study).

There are two types of spillovers to discuss. One merely affects the experimental design, and involves what to do when someone from the treatment (or control) group learns about the existence of the other group and asks why they are not receiving what the other person is receiving. We will call this the "experimental spillover." Experimental spillovers are often more of a concern in theory than in practice. However, this does not mean they should be ignored. They need to be minimized, and also should be carefully recorded because they may affect the results of the evaluation.

For instance, in the group versus individual liability experiment with the Green Bank of Caraga in the Philippines (discussed in more detail in the next section), we identified "sibling" barangays (neighborhoods or villages) as those which border each other and for which there is much social interaction. We treated them as one barangay for the sake of the randomization, thus ensuring that no "sibling" barangays were split whereby the barangay received different program designs. In the credit with education experiment with FINCA Peru, both treatment and control groups actually met in the same location. Although there was much concern about whether jealousy would cause problems, proper training of the staff to explain what was happening was effective. The few individuals that asked were told that FINCA Peru was rolling out the education slowly in order to learn whether it is effective or not.

Still, all experiments must be prepared for the groups to learn about each other. Staff must be trained in how to deal with these questions. We have found that the truth works best when clients ask "why did I receive X when my cousin, who is also a client, is receiving Y?" The truth is that the MFI is considering making a change and is testing it out carefully on a subset of clients. Clients had an equal and fair chance at being selected for the change, it was not done preferentially. If it works well, then it will be expanded fully. Ideally, the MFI can record information about all such inquiries, because learning about such interest (or disinterest) can help when evaluating the outcome and deciding whether to proceed with a full launch of the change.

The other type of spillover has to do with the indirect effects brought about by the program—not on clients but on others, including clients' families, neighbors, or community members. We will call this second type "impact spillovers." Impact spillovers can be both good and bad. A "good" spillover refers to the effect on other people of providing one person with a particular service or product. By only treating one person, often times you treat many more. De-worming interventions are a perfect example of this. In a study in Kenya, researchers found that de-worming school-aged children did *not* pass a cost-benefit analysis relative to other interventions when you only consider the *direct* effect. However, when you take into account the *indirect* effects as well (the "spillovers" in this situation take place because worms are passed from one child to another through the dirt in communal play areas), the intervention does indeed pass a cost-benefit analysis (Miguel and Kremer 2004). In microcredit, several examples exist for spillovers, both positive and negative. For credit with education programs,

clients may share what they learn with others in their community or family. For credit itself, the increased business of one client may create employment in the community. For group lending, it may help build social capital among the group members, which may influence others to form similar bonds (due to observing the success of the group members). A bad spillover may come from competitive pressures: if the MFI funds an individual to start a particular type of business, this may adversely affect other similar businesses in the community (although it might increase aggregate welfare for the community by lowering prices or improving product quality for the consumers in the community).

b. Ethical considerations

Stakeholders sometimes have ethical arguments about randomization, as some perceive them as arbitrarily and "unethically" depriving the controls from positive benefits. While ethical concerns are important to consider in any intervention, the presumption that randomized controlled trials introduce additional ethical considerations rests on two assumptions that typically are flawed.

First, this concern is based on the assumption that the program change is unequivocally good. If there is no doubt that the change should occur, that it not only will improve the situation *for everyone* but that it will do so more than any other change, then indeed testing the change would be a waste of resources. Such situations are rare, however. More often than not policy changes are debated and although strong hypotheses may exist, there is not adequate evidence to know unequivocally that the change will yield positive results for everyone. The MFI initiating a product change must decide the amount of resources it is willing to invest in testing the change based on how much uncertainty there is regarding the consequences of the change. If there is doubt about the efficacy of the change, then the experimental test may indeed be the most reasonable choice, so that the organization learns whether to implement the project further.

Second, this ethical criticism assumes unlimited resources for the change to reach everyone in the program. In many cases, this is not true for either budgetary or logistical reasons. For example, if the intervention is credit with education, the training of staff to provide the education modules is both costly and time consuming. Large organizations cannot do this all at once, but rather usually stagger the training of their employees on how to teach the material to the clients. In this way, a randomized rollout of the product can be offered to just as many clients as the organization has the capacity to reach, with or without the experiment.

c. Cost of randomized experiments

Experimental methodologies are often perceived as more costly than non-experimental methodologies. Relative to no evaluation at all, certainly an experimental evaluation costs more in the short run. Yet an experimental evaluation may be *less* costly in the long run if the results from the evaluation help to guide the long-term decisions and planning for the institution. For example, an MFI is considering whether or not to increase the

interest rates of its loans. An increase in interest rates may meet the MFI's financial targets in the short run. However, raising interest rates too high will drive away customers and reduce the loan repayment rates. This in the long run may erode the surpluses generated by the program's clients. It is therefore critical for the MFI to understand the net effect of an interest change to set the most desirable rate. Thus, spending some money now to have a credible assessment on the client's response to the proposed interest increase is by far less costly than saving the money and making a wrong decision.[**]

Relative to non-experimental evaluations, experimental evaluations are often less costly in the short run, and certainly less costly in the long run, when the benefits of more accurate results are factored in. When household surveys are used, they encompass the largest component of the budget, typically. However, the cost of collecting data for a non-experimental evaluation is often more expensive than for an experimental evaluation because non-experimental evaluations usually require larger datasets to process complicated econometric models. The analysis for an experimental evaluation, if designed correctly, is quite simple: one can obtain the answer simply by comparing mean outcomes between treatment and control groups.

The bottom line is that the cost of random experiments must be judged within particular contexts. The literature on microfinance provides no specific information on the overall or unit costs of evaluations but a vast array from a few thousand to multiple million dollars, depending on the questions studied and the number of MFIs involved (Hulme 1997).[††] A simple experiment, such as an evaluation of a program innovation, may indeed not require surveys, but rather just the MFI administrative data (e.g., repayment rates), to measure the efficacy of the change. In this case, the data can be retrieved at no cost from the MFI's accounting software or management of information system. The cost of the experiment is merely the management time required to design the experiment and train and motivate the staff in why the program innovation is being tested in this manner, as well as to analyze the data. If the organization is undergoing change to its products or processes, then enacting this change with an experiment rather than ad-hoc may not even add any further costs.

5. Credit with Business Education: Peru

In this project, we designed a randomized control experiment to evaluate the impact of adding a business training component to a traditional community bank microlending project. Many microfinance organizations have begun to include trainings in conjunction with their credit program. As this practice grows, and is adopted in various forms—the topic, training method, and whether the training is required or optional vary widely across organizations—it is useful and important to evaluate these programs to determine their

---

[**] More generally, findings from good random experiments can help avoid costly mistakes. For example, Duflo and Hanna (2005) find that in an education program in India adding a second teacher to the classroom makes no improvement in students' test scores and this helps redirect the funds to other more effective initiatives. Such decisions can have vast financial implications for programs at a national level.

[††] One example of the cost of a comprehensive non-experimental evaluation, reported by Montgomery et al. (1996), is the 1994 impact evaluation of BRAC's credit program which cost US$250,000.

effectiveness and the value the training adds to the microcredit product which is usually the central mission of the delivering organization.

The study described below examines a particular business training that was conducted in conjunction with a regular community bank microfinance product. The randomized control method applied here provided important information about the success of the training in meeting a variety of goals, and offers useful information about future implementations of similar products.

     a. Motivation of study

Business training is one of the many add-on training programs that is gaining popularity globally. Microfinance organizations like Pro Mujer have made trainings a central part of their credit model, and NGOs such as Freedom From Hunger specialize in working with microfinance organizations to adopt training sessions in conjunction with their credit meetings. Previous studies have suggested that members of community banks which include training about important health practices are more likely to know which nutritious foods to feed their children, the importance of breast feeding, and how to treat a child with diarrhea (MkNelly, Watetip, Lassen and Dunford 1996). However, these studies were conducted post-facto and thus selection biases may have persisted in the way the comparison group was formulated (i.e., the comparison group did not actively decide to join a credit with education program, whereas the treatment groups did).

Indeed, one can identify both positive and negative potential effects on a microfinance organization that also offers or requires training, particularly business training. Without conducting a rigorous evaluation of a credit-with-training product, it is impossible, in practice, to determine whether the positive effects outweigh the negative effects.

The potential advantages to pairing business training with credit programs include the following:

- As clients improve their business practices, the chance of default decreases
- As clients' businesses grow, they borrow larger loans
- Many clients are small-business owners with little education. Teaching basic business practices could have huge effects on their business for relatively low cost
- Client retention could increase as a result of a greater feeling of indebtedness to the lending institution for providing valuable training
- Potential clients may choose to enter the loan program of the financial institution offering training in part because they want to receive that training

However, business training also has potential disadvantages. Such disadvantages include:
- The length of communal bank meetings are often already a source of discontent among clients, who may be displeased to be forced to sit through an additional training, and eventually leave the program
- Clients may blame the lending institution for downturns in their business, and therefore become more likely to default in such a situation

- Talented credit officers may be less talented as teachers of business practices, resulting in poor quality of training
- As opposed to the teaching of good health practices, business acumen could be an inherent skill which is impossible to teach

The benefit of including business trainings in credit is debatable and many persuasive arguments can be made both in favor of and against it. It is therefore extremely valuable for an MFI to test the impact of such a practice before fully incorporating business training into its credit model. In a worst-case scenario, adding business training to a credit program could lead to increased default and dropout rates without impacting the business practices of loan clients. On the other hand, the potential benefits of including business training in a credit program—increased retention, increased repayment rates, and larger loans for clients who succeed in increasing the size of their business—are so great that many organizations may benefit from such a product.

As with the transition from a group-liability model to an individual-liability model, credit-with-education provides an ideal opportunity for a randomized control evaluation. There are convincing arguments both in favor and against the practice, so the value of implementation is not obvious. Furthermore, since such programs are growing in popularity, it is important that the microfinance community have a thorough understanding of both the positive and negative impact that such a product could have on credit programs.

b. Objective of the study and hypotheses

This study was implemented with FINCA Peru, a long-standing microfinance organization that operates in Lima and Ayacucho, a provincial capital in the Andean Region. We worked with them as they implemented a business-training program in both regions. In particular, we designed a study that considered the following potential impacts:
- Loan repayment
  - If businesses generate increased revenues, repayment may improve
  - The additional training may engender feelings of goodwill and reciprocity, causing loan clients to be more likely to back loans, even during difficult periods
  - If a clients' business falters, particularly because of a risky business decision, she may be blame the business training, and hence the credit organization and default on loan
- Loan sizes and savings volumes
  - If clients manage cash flows better, they may need less credit, and loan sizes could decrease

Community Bank Lending Programs in FINCA Peru

Foundation for International Community Assistance (FINCA) is a small, non-profit (but financially sustainable) microfinance institution (MFI) that has been operating in Peru since 1993. In addition to operations in Ayacucho and Lima, FINCA has operations in Huancavelica, another Andean province. FINCA's mission is to improve the socio-economic situation of the poor and to empower women through the promotion of the village-banking methodology.

Typically, FINCA provides loans in cycles of 4 months to groups of 30 women to help them expand their individual small businesses. Every loan is guaranteed by the entire community bank, meaning that all loan clients are responsible for paying off a defaulted loan. In addition to providing credit, FINCA teaches its clients to save by requiring regular savings deposits that correspond to the size of their loans and by encouraging additional voluntary savings for which they receive market interest rates.

In Ayacucho, community banks meet weekly, and in Lima, community banks meet bi-weekly. Each weekly (bi-weekly) payment includes interest, $1/16^{th}$ ($1/8^{th}$) of the original loan principal, and a mandatory savings deposit of $1/80^{th}$ ($1/40^{th}$) of the original loan principal. Clients are also encouraged to add to their voluntary savings. All savings, mandatory and voluntary, function as collateral for loans. FINCA further empowers clients by giving them the opportunity to run their banks through rotating participation on the village-bank board.

FINCA members are relatively young and have little formal education, particularly those in Ayacucho. FINCA clients each hold, on average, US$ 233 in savings whereas the average loan is US$ 203, with a recovery rate of 99 percent. FINCA charges sufficient interest to be self-sustainable. Its sustainability indicator (Total income / Total Expenses) was 99 percent in 1998; 105.5 percent in 1999; and 132.2 percent in August, 2000.

- On the other hand, if clients increase their businesses activities, they may need more credit, and loan sizes will increase
- Household decision-making
  - Improved businesses could empower female microentrepreneurs with respect to their husbands/partners in business and family decisions, leading to greater control over their finances
- Child labor
  - Changes to the household enterprise could cause an increase in family requirements that children contribute their labor to the family business
  - Conversely, businesses could generate more income and result in a shift in family priorities to the education of children
- Client retention
  - If the client's business grows enough, her credit requirements may exceed the capacity of the microfinance organization, and she may leave to seek a loan with a formal lending institution
  - The training could be perceived as an additional benefit of membership in the community bank, and client retention could improve
  c. Experimental design

The implementation of this experiment followed the following design:

Design of training materials

The training materials were designed by Atinchek, a firm specializing in business training materials for microentrepreneurs, and by Freedom From Hunger, a US-based NGO which works with MFIs globally to develop training in microfinance products, capacity building, and education in health and business development.

The training materials differed between Lima and Ayacucho, due to the distinct characteristics of the clientele in each district. In Lima, the training materials were in Spanish, and clients were expected to do homework between meetings. In Ayacucho, the training materials were occasionally in Quechua, and due to the higher illiteracy rate among those clients, the training materials relied heavily on visual aids and clients did not have homework.

Pilot implementation and Credit Officer Training

Before the full experiment was implemented, the training materials were piloted in 2-4 community banks in Lima and Ayacucho to determine the feasibility of the program and evaluate the training materials. This was a useful step not only for FINCA to finalize the program itself but also for the study because we could conduct focus group sessions with one or more of the banks used in the pilot to gather qualitative information about the program.

During this phase, all FINCA credit officers were taught how to conduct the trainings. These sessions took place over four months before the launch of the program.

Randomization

All FINCA banks that did not take part in the pilot trainings were randomly assigned, by computer, to one of the following groups:

- **Mandatory treatment** (Lima: 49 banks; Ayacucho: 55 banks): All clients were required to arrive early or stay after the meeting to receive training. Clients were fined for missing training, and continued absence could lead to expulsion from the bank.
- **Optional treatment** (Ayacucho: 34 banks): In this group, which included banks only in the Ayacucho location, attendance at training was voluntary; clients could leave after the regular meeting before the training began.
- **No training/Control** (Lima: 50 banks; Ayacucho: 51 banks): The control group clients did not receive training, but continued to have their regular credit meetings as they had in the past.

The treatment group in which the credit officer conducted the training at the end of the meeting and clients were not required to stay was included only in Ayacucho. This occurred because FINCA did not think that optional trainings were feasible in Lima for operational reasons.

We conducted the randomization at the community bank level. The randomization was designed so that the selection of banks included in the control was similar to the selection of banks included in each of the treatment groups. In addition to conducting the randomization so that there were no discernible differences between the banks of either of the treatment groups or the control group, each credit officer had a proportionate number of banks randomized into each group. This prevented the possibility that the quality of the credit officer herself could skew the outcomes of one particular group of banks.

Baseline Survey

Before the implementation of the project, we conducted baseline surveys in all banks, both treatment and control. The survey collected information about business processes, knowledge, and savings practices of the client, her business activities, and information about her household, in particular, decision-making processes with her spouse.

Every treatment group was surveyed before beginning training, which coincided with the beginning of a new loan cycle. We were careful to conduct surveys in control groups over the same period of time as we conducted surveys in treatment groups, although we were not bound by the same restrictions of completing the surveys before the beginning of the following loan cycle.

Program Implementation: Training conducted in credit meetings

During the phase of program implementation, the trainings were conducted until all sessions had been delivered, or until the end of two years. Some banks delayed in completing all twenty-two sessions within the two-year time frame. However, these banks were considered treatment-group banks during analysis. Had we only included those banks that completed the training in our analysis, we would have failed to consider the impact of the trainings on all banks, but rather would have only learned the impact of the training on banks that were able to complete the training in the allotted time frame.

During the study, there were some problems with experimental spillover: since all credit officers had learned to administer the training, one challenge we faced was to prevent them from training certain individuals or entire banks in the control group who either asked for the training, or for whom they believed that the trainings could be particularly beneficial. This problem, however, was resolved by explaining the purpose of the study and the study design to credit officers. Once the officers understood that the control group was only temporary and was necessary to asses the success of the training, and that once the study was complete, all clients and banks would be eligible to receive training, they accepted their directive to only train treatment group banks.

Follow-up Survey

At the end of the two-year implementation period, follow-up surveys were administered. Follow-up surveys, in addition to collecting information similar to that of the baseline survey, also asked specific questions about the allocation of profits and record-keeping of business financial information. In Lima, the follow-up survey included questions about child labor practices within the family.

These surveys were administered to everyone who was interviewed in the baseline survey whom we were able to locate who were willing to conduct the interview, *regardless of their current member status in FINCA* (76% of the clients surveyed in the baseline survey). We did not interview women who had become members of FINCA after the baseline survey, and therefore had not been interviewed at the start of the project.

d. Measuring impact

We measure the impact of the program by comparing data from the baseline survey, the follow-up survey, and FINCA financial-transaction data, which includes information about loan payments, interest, mandatory and voluntary savings and some socio-economic characteristics.

This data was analyzed in two ways. In the first type of analysis, in indicators of interest, a comparison was made between the control and treatment groups. Secondly, in indicators of interest, a comparison was made between the *change* exhibited by treatment groups during the length of the study, and the *change* exhibited by control groups during the same time period.

e. Results

Our analysis was divided into the following four outcome variables:

- *Institutional outcomes*: Analysis of FINCA financial-transaction data revealed that repayment among treatment groups is 3% higher than among control groups, and that clients in treatment groups were 4-5% less likely to drop out than clients in control groups. However, treatment group clients were more likely to name the length of the meeting as a reason for dropout. This supports the hypothesis that the clients find the training useful and valuable, and either they are more able to pay back their loans or prioritize repayment more than non-treatment clients. There was no difference in loan size or in cumulative savings between treatment and control groups.

  Interestingly, these results are the strongest for clients with larger businesses and clients who had expressed the least interest in the program during the baseline survey. This suggests that a market approach, in which the additional service is offered optionally, on a pay-basis, may not be the most successful approach since the clients who reacted most favorably to the training did not want it at first.

- *Business skills and practices:* In the follow-up surveys, clients were asked questions about their business practices. The questions were designed to specifically test information which had been covered in the business training sessions. Results indicate that treatment groups in general demonstrated greater business knowledge and better business practices, but only in limited areas. Clients in treatment banks were more likely to reinvest profits in their business, maintain sales records for their business, maintain withdrawal records for their business, and be able to point to changes or innovations that they have made in their business over the past year.

- *Business Results*: In addition to business practices, surveys collected information from clients about their actual businesses. Again, results indicate that the business training did impact the success of the clients' business. For example, treatment groups had sales in the month prior to the survey of 16% higher than control groups and reported sales during their worst month that were 28% higher than the sales of control group during their worst month. Despite an increase in overall sales, there was no change in the profit margin on the most common product sold.

- *Household outcomes*: Household information collected from the surveys, focusing on decision-making processes and child labor practices within the household suggested that there was no impact on decision-making processes with respect to use of loan, nor was there change in whether clients separated money from their husband or partner or methods of tracking family finances. However, this could be because the FINCA model already emphasizes female empowerment in household decision-making, so the business training could add little value in that area. With respect to child labor, there was no overall effect when male and

female children are looked at together, but there is a positive treatment effect for number of house that female children dedicate on average to school and schoolwork.

In summary, the evaluation of FINCA's business-training with credit program found that many of the anticipated benefits of the project did, in fact, occur. Clients in the control group demonstrated greater business savvy, and the institution benefited from an increased repayment rate and greater retention. The concern over increasing the length of the meeting by adding a training session also proved to be accurate in that a greater percentage of treatment group dropouts complained about meeting length. However, since retention was higher in treatment groups, the increased length of the meeting was, for the most part, outweighed by the perceived benefit of the training.

Replications of the study

As more microfinance organizations include trainings in their lending model, the types and methods of implementing the trainings become more and more diverse. For this reason, this study should be replicated before making broader general statements about the impact of credit-with-education programs. Factors that are unique to FINCA that may have influenced the outcome of the education product include:

- *Location of meeting:* In Ayacucho, FINCA holds meetings at the FINCA center near the downtown, although many of the clients live outside the town. Since clients already have to travel long distances to come to the meetings, it could be a bigger burden for clients to spend an additional 30-60 minutes there during the training than it is for members of community banks that meet in the neighborhoods where the clients live.
- *Credit officers as trainers:* FINCA trained their credit officers to conduct trainings, but some other organizations hire professional educators to facilitate trainings. Since credit officers may not be natural or interested teachers, it is possible that expecting them to double as educators could results in less interesting lessons and thus less interested clients.
- *Type of clients*: As in many, but not all, microfinance organizations, FINCA targets female entrepreneurs as clients. Since the clients are already entrepreneurs, they many react more favorably and absorb more easily the content of the business training than clients with other types of income sources.
- *Context:* Both Lima and Ayacucho are saturated with credit options. Particularly in Ayacucho, where FINCA's program is very well-known and highly regarded and the addition of a training program could either cause FINCA's program to stand out as unique among the options, or as burdensome because of the additional time commitment. In either case, clients are aware that they can receive credit from other sources if the FINCA model no longer suits them. This could result in different reactions from clients than one would see in an area with fewer loan options.

6. Evaluating Group- versus Individual-Liability Loans: the Philippines case

A randomized control experiment was designed in the Philippines to evaluate the impact of group- versus individual-liability lending programs. While group-lending programs are still prominent in microfinance practice, a small but increasing number of microfinance institutions are expanding rapidly using individual lending. As these institutions explore the benefits of individual-liability loans for the poor, there is an opportunity to apply randomized controlled trials to evaluate rigorously the impact of the innovation compared to the group-liability program.

Indeed, given the popularity and apparent success of the two methodologies, as well as the lack of rigorous evaluations of both of them, it is difficult to know the real advantages and disadvantages of each—and therefore to formulate policies on this matter. An example like the one proposed can fill this void and provide useful guidance to the microfinance industry at large.

a. Motivation of the study

Unlike individual liability, under which each borrower is only responsible for her own loan, joint liability requires members of a defined group to help repay the debt of other members when they cannot repay. Unless the group as a whole repays every member's amount due, no member will be granted another loan. The Grameen Bank in Bangladesh developed a lending methodology based on joint liability that is now employed by many NGOs and microfinance institutions around the world. The success and popularity of this approach can be linked to its numerous perceived advantages. (Some of the advantages, while associated with group liability, are not inherent to group lending alone, as will be shown below.) Such oft-cited advantages include:

- Clients face both peer and legal pressures to repay their loans.
- Clients have incentives to screen other clients so that only trustworthy individuals are allowed into the program.
- Low transaction costs as clients meet and pay at the same time and location.
- Cheaper training costs as clients all gather periodically.
- Clients have incentives to market the program to their peers, thereby helping to bring in more clients.
- Group process may help build social and business relationships.

As is the case with most methodologies, joint liability is not without potential disadvantages. These include:

- Clients' dislike of the tension caused by the peer pressure could lead to lower client satisfaction and hence higher dropout.
- Older clients tend to borrow significantly more than newer clients, and this heterogeneity often causes tension within the group, because new clients do not want to be responsible for others' much larger loans.

- Group lending could be more costly for clients since they are often required to repay the loans of their peers.
- Clients dislike the longer meetings typically required for group lending.
- Default rates could be higher than if there were no group liability because bad borrowers can bring down good borrowers (i.e., once your peer has gone into default, you have less incentive to pay back the loan yourself).
- Default rates could be higher than if there were no group liability because clients can "free ride" off of good clients. In other words, a client does not repay the loan because the client knows that another client will pay it for them, and the bank will not care because they still will get their money back.
- Villagers with fewer social connections might be hesitant (or even unwelcome) to join a borrower group.

Given the existence of these potential negative aspects and the fact that the last three advantages listed can be obtained without resorting to group liability,[‡‡] there is a strong case to be made for an MFI to experiment with offering individual loans to their clients. This concern over the excessive tension generated among members by imposing group liability is precisely the main motivation for the shift from group to individual-liability loans. Practitioners worry that the conflict among members could not only lead to high dropout rates and affect the sustainability of the program, but also potentially harm social capital so valuable to the poor who lack economic security.

Two features of this innovation make it a perfect case for a randomized control evaluation. First, there are conflicting arguments for and against individual liability loans, and the net impact of such programs compared to group-liability lending programs is not clear. Besides the obvious benefit of removing group liability for the clients (reducing pressure and tension among members), the individual liability loans may also benefit the lending institution by increasing the client retention rate (because clients prefer individual liability) and thereby the MFI's portfolio. However, the lender will lose a crucial enforcement mechanism when group liability is removed. It would negatively affect the repayment rate if none of the group members is willing to make a voluntary contribution to cover the repayment of defaulted members. Using a randomized controlled trial, the relative merits of group- versus individual-liability loans for both clients and institutions can be evaluated.

Secondly, in recent years individual-liability loans in the microfinance community have gained popularity around the world. Although replication of the study is necessary to generalize the results of this particular evaluation, it will help identify the effective environment and design of the program, benefiting not only the lending institution and its clients, but also the entire microfinance community. As such, it can play an important role in both policymaking and product design.

[‡‡] For instance, under the methodology employed by the MFI ASA, clients still meet together but are individually liable for their loans.

b. Objective of the study and hypotheses

We collaborate with Green Bank, a commercial bank based in Mindanao, as it expands its microfinance operation in Leyte and Samar islands in the Philippines, to conduct a pilot-testing experiment to evaluate individual-liability loans. In this experiment, we seek to evaluate the following impacts:
1. Relative impact of group versus individual liability on clients and their communities
2. Relative cost and benefit of group- versus individual-liability loans for Green Bank
3. Impact of credit on individuals and their communities

Specifically, we pose the following questions:
1. How does group relative to individual liability affect institutional outcomes such as repayment, client retention, loan size, and operating (labor) costs?
2. Does group liability motivate peers to monitor and/or enforce repayment of loans?
3. Does group liability motivate peers to select less risky clients for a bank?
4. How does selection on other dimensions (e.g., poverty, social connectedness) differ under group versus individual liability? Are those less-connected (hence perhaps less likely to have good informal social safety nets) less likely to participate in group lending than individual lending programs?
5. What is the impact on the household, enterprise and community from a microfinance institution offering credit in their community? How does this impact differ for group versus individual liability loans?
6. What are the impacts, positive and negative, on social networks from group versus individual liability loans?

c. Experimental design

The experimental design employs one strategy for "existing areas" and one for "new areas." The "existing areas" strategy involves converting existing centers to individual-liability loans. The results of this initial study can be found in Gine and Karlan (2006). The advantage of this approach is that one can attribute the differences between group and individual liability to differences in the loan liability, and not to differences in the individual *characteristics* of the clients per se. This is true because all existing clients joined the program under a joint liability scheme. Thus there is no selection bias as would be inherent in comparing the outcomes of clients who have chosen group liability to the outcomes of clients who have chosen individual liability. The disadvantage is that there may be differences between clients who have enrolled in a group-liability program and the borrowers that would enroll in an individual-liability program. Therefore while the results from the "existing areas" strategy will be accurate for those who are willing to sign up for group liability, we cannot say from this strategy alone how the product will work among clients who know from the outset they are joining an individual-liability program.

It is then important to understand these potential differences among borrowers, especially when generalizing the results of the cost and benefits of joint liability. For this reason, the study includes a "new areas" strategy by working with Green Bank as it expands to new areas in the eastern coast of Leyte (Tacloban) and the neighboring islands of Cebu and Samar.

This expansion also provides a unique opportunity to test the impact of the credit itself. A randomized program placement strategy is employed to assign barangays to either individual or group liability, and also to a control group. This allows us to test the impact on household, enterprise and community outcomes from receiving either group or individual liability loans.

Group vs. Individual Loan Programs in GREEN BANK

Green Bank is a for-profit commercial bank which was established in 1975 and currently operates in northern Mindanao and the Visayas. Its microfinance department started in the late 1990s, and their group-liability lending program BULAK in 2000.

BULAK follows a modified version of the Grameen approach. In BULAK there are four different units, which from smallest to largest are: individuals, groups, centers and branches. Up to five low-income women come together to form a group. The group is formed by them and not by the bank. Then, three to six groups come together to form a center. The center is where all of the groups jointly hold their weekly meetings and collect payments. Typically a barangay (sub-municipality) will have one center. In total, Green Bank has over 12,000 clients.

All loans given under the BULAK program are to be used for expanding the client's microenterprise. The initial loan is between 1,000 - 5,000 pesos (roughly $18 - $90) and increases by 5,000 pesos after every cycle, such that the maximum loan size in the 5th cycle is 25,000 pesos. However, the loan size is a function of the repayment of their last loan, attendance at meetings, business growth, and contribution to their personal savings. Loans are charged an interest rate of 2% per month over the original balance of the loan. The client has between 8-25 weeks to repay the loan, but payments must be made on a weekly basis.

As part of the BULAK program, clients are also required to make mandatory savings deposits at each meeting. Each member has 100 pesos ($1.80) deducted from every loan release. In addition, 10% of their weekly due amount (principal plus interest) is deposited in their individual savings account. Member savings may be used to repay debts and may also be used as collateral, although in this last case there are no fixed rules. Finally, 10 pesos ($.36) per meeting are required for the group and center savings. These center savings cover mostly the construction of the center meeting place, and are only used as a last resort to repay member loans.

The individual-liability program (BULAK II) being pilot tested in the experiment has all the features of BULAK program, including weekly repayment meetings and consolidation of repayment by center groups, except the following two features: first, no client is liable for her group members' loans, and second, there are no longer mandatory center and group savings. All center activities will now be paid individually on a per-activity basis.

Since this change to individual liability is significant, careful testing is required before the full launch of the experiment can implemented. For this reason, a small pilot test was conducted in Leyte, which will also serve as the location of the full study. Green Bank has 186 lending centers in Leyte, with an average membership of 25 individuals (or 5 groups) per center. For the pilot phase, one center from each credit officer's portfolio was randomly chosen, 11 centers in all, to convert to the new individual-liability methodology. This random selection of centers is critical. If, for instance, one were to pick only the best centers, then one would not know whether the results were generalizable to the inferior centers. One might falsely conclude that individual liability is better, when it in fact is perhaps only good for the best groups. This pilot phase began in August 2004 and proceeded with little to no difficulties.

Existing Areas Full Launch ((Group versus Individual Liability Experiment #1)

In early November 2004, 24 more centers were randomly converted. The full pilot phase as of May 2005 included 93 converted centers and 93 original (group liability) centers.

The findings are striking: there is no change in repayment among the Green Bank centers shifted to individual liability, relative to its centers which remained in group liability. We also find that the individual liability centers attract more new clients, and lose fewer clients to dropout. We also find no change in the time required of the credit officer, so there is no indication that this shift will reduce the capacity of an individual credit officer to manage centers. The full results can be found in Gine and Karlan (2006).

New Area Plan (Group versus Individual Liability Experiment #2)

Evaluating the relative impact of group- versus individual-liability loans poses a challenge in conventional non-experimental evaluation method because the two programs attract different types of clients—unobservable heterogeneity between the two groups of clients may confound the results. In a randomized controlled trial, random selection of the sample allows you to compare between the two groups. The procedure to start operations in new areas is novel and another contribution of the study. It consists of two parts, the identification of eligible barangays and of potential clients through a marketing meeting.

- **Identification of the Barangays:** The first step is to gather basic information about the barangays from the municipality office. This information is mainly used to exclude barangays with low population density as it is deemed too costly to start operations in these areas. The credit officer visits the selected barangays and conducts a survey to verify the following criteria: (i) the number of microentreprises, (ii) the residents' main sources of income, (iii) the barangays' accessibility and security, and (iv) the perceived demand by the residents for microcredit services. The survey is administered to the secretary of the barangay,

typically the person with the most information about the administrative aspects of the barangay.

- **Census of Microentrepreneurs:** The purpose of the census is to construct the sample framework to assess which businesses are interested in credit and could eventually be clients of Green Bank. The census records basic information regarding the size of their business and their credit history. While it is being conducted, they are told about the marketing meeting.

The sample villages identified are randomly assigned to the following four groups.

1. *BULAK*: Green Bank will offer group-liability loan program.
2. *BULAK to BULAK II*: Green Bank will offer group-liability loans and remove group liability after the first loan cycle.
3. *BULAK II*: Green Bank will offer individual-liability loan program.
4. *NO CREDIT*: Green Bank does not offer their services (control group).

It is important to note that our sample in Groups 1, 2, and 3 is NOT the actual borrowers, but rather the "potential clients." This is because if we were to compare those who choose to participate in the program in the areas in which the program is offered to those in the control group, our estimate of impact will suffer from self-selection bias. We would capture, in addition to the true effect of the program, the extra motivation of the clients who decide to enroll. However, instead of watering down our estimate of average impact (calculating the average outcomes among those who do participate, as well as all those who do not) we can improve our estimate—and keep it unbiased—by employing a technique called propensity score matching (PSM) and weighting the impact estimate by the likelihood that each individual becomes a client. The key in this sample formation is to identify those who "would" receive a loan from Green Bank if Green Bank were to operate in the village. PSM uses the baseline characteristics of the potential clients to statistically identify those most likely to participate in the program. We measure the impact on each client by comparing their outcomes to the outcomes of those in the control group with a similar propensity to participate.

Because the sample selection in the four groups is consistent, sample bias in sub-sets from these groups is consistent, and we can compare the impact between any of the four groups. This experimental design provides a unique opportunity to measure the clean impact of credit by comparing Groups 1, 2, and 3 with Group 4.

d. Measuring impact

We measure the impact by comparing different outcome measures between the treatment groups and control groups. The impact of the program can be measured at three different levels: individual client, community, and institutional. By looking at the impact not only at the client and institutional levels but also at the community level, we can evaluate the broader implication of the program and how it could affect the local economic status. In

order to make necessary comparisons, the data will be collected in three different methods:

- *Baseline Survey*: the information on sample villages and clients is collected *before* the experiment takes place. This information is used in validating the randomization as well as in analyzing the *post*-experiment impact. In the Green Bank study, we collected information on loan history, business status, household well-being (economic and psychological), social networks, and risk preferences of the sample individuals. By definition, randomization will create comparable treatment and control groups; however, it is always a good idea to validate the random assignment by checking some key variables from the baseline survey before the launch of the experiment (comparing the means of the variables for treatment and control groups and ensuring they do not differ significantly).
- *Follow-up Survey on clients*: The survey conducted after the study period will be used to evaluate the program impact. The information collected will include clients' performance in the Green Bank program and clients' business performance as well as their household welfare.
- *Activity-based Cost Exercise*: This exercise records all activities of development (loan) officers. By comparing the total time spent on BULAK II versus BULAK centers, we will be able to calculate the cost for the institution of the individual-liability program relative to the group-liability program.

Replication of the study

Given the decision by several MFIs to employ individual-liability loans, it is not only in Green Bank's interest, but also in the interest of the microfinance community as a whole to learn the impact of group- versus individual-liability programs. However, we cannot draw a general conclusion from the result of this specific program evaluation in the Philippines. Only after replications of the evaluation, with different MFIs in different places and with different clients, can we make more general statements about the impact of group- versus individual-liability loans.

Many factors may make the results of the evaluation unique to Green Bank and its context. The following are some of such factors:

- *Initial Social Network*: The importance of social networks among program members depends on many exogenous factors: culture, the size of the village, and its economic activities. The more economically vulnerable clients are, the more they rely on their social networks for support. If this is the case, removing group liability among uncollateralized clients may result in better repayment performance among lower-income groups than among those with more stable income flows.
- *Type of Clients*: For example, Green Bank targets small female entrepreneurs in rural areas. There is a large volume of literature that concludes that female borrowers repay better than male borrowers. The impact of group- versus individual-liability loans could well be different between the gender groups.

- *Type of institution*: Green Bank is a commercial bank; thus the financial sustainability of its microfinance programs is a critical part of its operational goal. The implications of cost-benefit analysis would be different for Green Bank than for subsidized institutions.
- *Context*: In most areas where Green Bank operates, it competes for clients with other lenders. For the most part, these tend to offer group-lending loans, so the impact of introducing joint liability will be affected by the presence of other lenders and their specific products.

7. Pilot Experimental Approach for Other Lending Product Innovations:

In the previous section we used an example from a randomized controlled trial designed to evaluate the impact of group- versus individual-liability programs with Green Bank in the Philippines. This experiment pilot approach is applicable to many other innovations whose net impact on clients and benefit for the institution is not known. Below are some examples of such cases.

a. Mandatory/Voluntary savings rules for lending programs

Savings schemes in lending programs aim to reduce clients' vulnerability to unexpected negative economic shocks, as well as to improve clients' financial management skills by encouraging them to make small regular savings. However, if clients lack the discipline to save, they might view mandatory savings merely as an additional burden, reducing the number of borrowers.

b. Savings products with commitment features

Due to self-control or household (e.g., spousal) control issues, some people prefer to have commitment savings products in which deposits are withheld from their access until a specific savings goal is reached. Such products take on many forms, but little empirical evidence of their effectiveness currently exists (Ashraf, Gons, Karlan and Yin 2003; Ashraf, Karlan and Yin 2006c; a; b).

c. Frequency of payments

Frequency of payment varies from program to program. MFIs generally demand relatively frequent repayment schedules (often weekly) while clients often prefer less-frequent payment. Particularly for those who have inconsistent income flows, a frequent repayment schedule could increase the default rate.

d. Health/life/disability insurance

Insurance offered with credit aims at reducing vulnerability of clients. Clients as well as microfinance institutions may benefit from the insurance services as they are insured for certain types of economic shocks. However, insurance services may cause *adverse*

selection by attracting riskier clients to the program, which could lead to higher default rates. Or insurance could cause *advantageous selection* by attracting risk-averse clients, which could lead to lower default rates.

e. Local public goods (community "empowerment" training)

The mission of some microfinance institutions is not merely increasing credit access for the poor, but also to empower the economically/socially marginalized sector of population. Empowerment training may increase impact on clients by improving women's mobility and ability to make economic decisions; or it could increase client exit if the clients do not have an interest in the training.

f. Human resource policies (e.g., credit officer incentives)

Providing incentives for credit officers could improve repayment rates if they use enforcement power appropriately. However, the incentive schemes could cause conflicts between the officers and clients because the officers now have a personal stake in better repayment rates. Such friction between the credit officers and clients may affect the retention rate.

g. Interest rate policies

Little is known empirically about the elasticity of demand with respect to interest rates (the extent to which clients are willing to accept higher interest rates, and the extent to which demand for loans increases at lower interest rates). Furthermore, much economic theory has been written about how higher interest rates might drive down repayment rates through information asymmetries such as adverse selection and moral hazard. Some authors try to examine these issues using survey data, see for example Dehejia et al. (2005) and Gross and Souleles (2002). Experimental studies can be done to study the relationship between interest rates, demand for credit and repayment rates. See Karlan and Zinman (2006a; 2006c) for an example of such a study.

h. Credit scoring and credit evaluation decision-making

Many financial institutions are employing credit scoring to facilitate the credit granting decision. Questions remain, particularly for typical microfinance clients, how to balance objective and subjective inputs into the credit approval process, and specifically how to balance human versus computer decision-making. Furthermore, depending on how the credit scoring is implemented, one can use the process to create randomly treatment and control groups to assess the impact of receiving credit on the welfare of the client (e.g., see Karlan and Zinman 2006b).

7. Conclusion

In this paper we have examined the flaws in methods commonly used to assess the impact of microfinance programs and showed that modifications to the design of

microfinance programs may be best evaluated through randomized controlled trials. Randomized evaluations can be performed ethically and cost-effectively, and the accuracy of their results makes them valuable both to the institution implementing the evaluation and to the microfinance community at large. Through the example of individual-liability loans in the Philippines we showed the steps involved in performing an experimental evaluation. Many questions remain, however, and until an evaluation has been replicated in a variety of settings, it remains unknown whether a particular innovation is likely to work for other programs. This is the nature of *all* evaluative work, regardless of the methodology employed. To stimulate the experimental evaluations of more program innovations we have provided a list of several modifications which could be tested using similar methodology.

Bibliography

Ashraf, N., N. Gons, D. Karlan and W. Yin (2003). "A Review of Commitment Savings Products in Developing Countries." Asian Development Bank Economics and Research Department Working Paper Series 45.

Ashraf, N., D. Karlan and W. Yin (2006a). "Deposit Collectors." Advances in Economic Analysis & Policy 6(2): Article 5.

Ashraf, N., D. Karlan and W. Yin (2006b). "Household Decision Making and Savings Impacts: Further Evidence from a Commitment Savings Product in the Philippines." working paper.

Ashraf, N., D. Karlan and W. Yin (2006c). "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." Quarterly Journal of Economics 121(2): 673-697.

Dehejia, R., H. Montgomery and J. Morduch (2005). "Do Interest Rates Matter? Credit Demand in the Dhaka Slums." working paper.

Duflo, E. and R. Hanna (2005). "Monitoring Works: Getting Teachers to Come to School." M.I.T. Poverty Action Lab working paper.

Giné, X. and D. Karlan (2006). "Group versus Individual Liability: Evidence from a Field Experiment in the Philippines." Yale University Economic Growth Center working paper 940.

Gross, D. B. and N. S. Souleles (2002). "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data." The Quarterly Journal of Economics 117(1): 149-85.

Hulme, D. (1997). "Impact Assessment Methodologies for Microfinance: A Review." University of Manchester Institute for Development Policy and Management working paper.

Karlan, D. and M. Valdivia (2006). "Teaching Entrepreneurship: Impact of Business Training on Microfinance Institutions and Clients." Yale University Economic Growth Center working paper.

Karlan, D. and J. Zinman (2006a). "Credit Elasticities in Less Developed Economies: Implications for Microfinance." working paper.

Karlan, D. and J. Zinman (2006b). "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." working paper.

Karlan, D. and J. Zinman (2006c). "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment." working paper.

Miguel, E. and M. Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." Econometrica 72(1): 159-217.

MkNelly, B., C. Watetip, C. A. Lassen and C. Dunford (1996). "Preliminary Evidence that Integrated Financial and Educational Services can be Effective against Hunger and Malnutrition." Freedom from Hunger Research Paper Series 2.

Montgomery, R., D. Bhattacharya and D. Hulme (1996). Credit for the poor in Bangladesh. Finance Against Poverty. Hulme D and M. P. London and New York, Routledge. 2: 94-176.