Testing Microfinance Program Innovation with Randomized Control Trials: An Example from Group versus Individual Lending

Xavier Giné, World Bank
Tomoko Harigaya, Innovations for Poverty Action
Dean Karlan, Yale University
Binh T. Nguyen, Asian Development Bank

November 2005

Fourth Draft

Abstract

This paper presents an experimental methodology for evaluating modifications to the design of microcredit programs. As microfinance becomes an even more popular tool for fighting poverty, institutions innovate 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 discuss the reasons why more rigorous evaluations are necessary. We then discuss why randomized control trials can prove useful to microfinance institutions around the world in identifying effective program designs in different environments. In this paper, we focus on the choice of lending methodologies -- group versus individual liability -- to illustrate the benefits of randomized control trials as a business tool for not only measuring impact, but learning how to improve sustainability and growth.
1. Introduction

In the last decade, microfinance institutions have experienced a boom of innovation in lending products, partly fueled by donors who see microfinance as the next promise to alleviate poverty. Examples of such 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 reach sustainability. As in other industries, microfinance institutions (MFIs) typically decide whether to adopt new strategies based on other MFIs’ success with the innovations. Yet while many new micro-lending products and approaches continue to be developed, MFIs must generally rely on descriptive case studies of the effectiveness of these innovations to decide whether to implement the new strategies. The typical 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 control trials can help test the effectiveness of new lending products. 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. 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 letting a few volunteer clients use a new lending product, or by offering to a small group of particularly chosen clients (i.e., 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 microfinance institutions which want to know whether they should enact similar changes. This is a situation in which randomized control trials are a win-win proposition: safe and 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 microfinance institutions.
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. 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 conducted in the microfinance industry 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 \textit{before the innovation is launched} the most important stage of the evaluation. This paper hopes to shed insight into the motivations for and possibilities of using experimental evaluations to ask important operational questions for microfinance practitioners.

The rest of the paper is organized as follows. In Section 2, we discuss the problems with non-experimental approaches to program innovation. In Section 3 we introduce experimental pilot approach to product innovation and the steps to design a randomized control trial, followed by a discussion of the critiques and myths about the experimental designs in Section 4. In Section 5, we present an example of a randomized control trial on group versus individual liability in the Philippines. Finally, Section 6 provides further examples of different programs in which randomized control trials could be employed to evaluate the impact of the program, and concludes.

2. \textbf{Why Randomize?}

In the context of evaluating a lending product innovation, we are typically discussing 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 observed 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: \textit{what program client’s outcomes would have occurred had the new lending feature not been introduced?}

a. \textbf{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 simply attributes the observed change in the clients’ outcome indicator to the product change introduced, without explicitly constructing 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, observable or unobservable, that also contribute to the changes in clients’ outcomes. 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); and so forth. For instance, 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 are those that control for external intervening factors. If we could observe the same client at the same point in time both accessing and not accessing the new loan product, this would effectively account for any other observed and unobserved intervening factors. Since this is impossible, to disentangle 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.

Let us consider a particular example. In an experiment with FINCA in Peru, we measured the impact of business education by comparing the group that received credit and business education (the treatment group) with the group that received only credit (the control group). Although the program evaluations observed in the field often do not have this ‘un-treated’ group to compare the treatment group with, having a control group is critical in conducting accurate program assessment. Figure 1 illustrates the difference in the outcome between the evaluation methods with and without a control group. Suppose that credit access increases clients’ business income from $I$ to $I'$ during the study period $t$ (along the “Credit Only” line), and credit access offered with business education increases the income from $I$ to $I''$ (along the “Credit with Business Education” line). The program evaluation without a control group will attribute the increase of $I'' - I$ to the provision of business education because they do not have the baseline measure of how much business income would increase without business education. However, the true impact of providing business education with credit should be measured by $I'' - I'$. In this case, the absence of a control group overestimates the impact by failing to isolate the impact of business education from the impact of giving credit. By comparing treatment with the control group, which has all the same characteristics as the treatment group except the specific feature that is being tested (business education in this example), the treatment-control experimental method can measure the independent impact of this feature.
b. Why a single branch cannot simply be compared to another?

In the example above, where the MFI decided to introduce a product change throughout an entire branch and evaluate the success of the innovation based on the experience of this one branch alone, we noted there would be no counterfactual example from which we could determine what clients would have experienced in the absence of the innovation. The MFI might assume that they could create this counterfactual by comparing the outcomes of clients in the branch which receives the innovation to the outcomes of clients in a different branch which is not offered the innovation. This approach would be flawed. 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 can have quite divergent experiences with various sources of macroeconomic shocks. If one branch were to experience a drought or a flood (or more likely, simply more or less rainfall), or the building of a new factory, etc., comparing the two 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.

c. 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 credible control group faces some difficulties in practice. The problem is that in reality borrowers and non-borrowers usually are different. Microfinance institutions usually target certain groups of clients, such as women in poor neighborhoods. In case of a product change, they usually test on a small set of clients (e.g., one branch) selected on the basis of, for example, their previous repayment record. This endogenous program placement effectively makes borrowers and non-borrowers different in some set of characteristics (e.g., on average borrowers of new product have a better repayment record). 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 wherein a new product has been introduced to promote household businesses may be intrinsically more entrepreneurial than non-borrowers. Or in the FINCA example, recipients of credit with education may choose to borrow because they value their children’s education more than non-recipients do.

Because of endogenous program placement and endogenous program participation, those who are not borrowing are often not a good comparison group for those borrowing. As a result, the observed difference in the 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 use 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

\[ \text{It can be shown statistically that with a sufficiently large sample, on average, the two groups are identical, as follows. Suppose } X \text{ is any observable (e.g., education, age) or unobservable (e.g., capability, taste) characteristic of the population of potential clients. Suppose further that } X \text{ follows any distribution (normal or non-normal) with mean } \mu \text{ and variance } \sigma^2. \text{ From this population we randomly draw a treatment group } X_T \text{ and a control group } X_C \text{ of size } N_T \text{ and } N_C, \text{ respectively. Then, with a sufficiently large sample, by the law of large numbers sample averages } \bar{X}_T \text{ and } \bar{X}_C \text{ follow a normal distribution with mean } \mu \text{ and variance } \sigma^2/N_T \text{ and } \sigma^2/N_C, \text{ respectively. That is, on average, } X_T \text{ and } X_C \text{ are identical, for any } X, \text{ the condition we need for a credible control group. This statistical result can occur only if } X_T \text{ and } X_C \text{ are randomly drawn from the population } X. \]
through the same external events throughout the same period of time, and thus encounter the same other 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 new product feature. 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 control 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 using some service in the program.

The flowchart in Figure 2 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. Identifying the problem, potential solution, and conduct 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.
An operating plan should include specifics on all necessary operations components to effect 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. One needs a control group as the previous section has discussed.

b. Identify treatment assignments

Often a proposed solution has a main change, but many minor issues that need to be decided. For instance, when testing Credit with Education, we had to select which type

---

2 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.
of education modules to offer, and when testing individual liability we needed to
determine what loan size is optimal. A careful experimental design can include tests such
of 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 it
was 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 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. See the following
section for a description of 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. We recommend the free software Optimal
Design for helping to determine sample sizes.3

4. Critiques and myths of experimental designs

Internal Validity

The internal validity of experimental methodology rests on integrity of the data from
treatment and control groups. The experimental protocol must remain intact throughout
the study. In some experimental designs, it is acceptable for a control group to get
treated. These are called encouragement designs, in which a treatment group is given an
encouragement to participate in the treatment, but is not required to participate.
Similarly, the control group is not given the encouragement, but is allowed to participate
if they so choose.4 What matters most is that assignment to treatment is used, not
treatment, to differentiate the groups when analyzing the results. In other words, an
individual assigned to treatment but who does not participate is still considered part of the

3 This can be downloaded from http://www.ssccentral.com/otherproducts/othersoftware.html.

4 To evaluate the impact of the change, it is critical that the proportion in the treatment group who participate is
significantly bigger than the proportion who participate in the control group. The bigger the difference between the
proportions, the smaller the sample size one needs to have confidence statistically in the results.
treatment group. In this sense they were “treated” with encouragement to participate. Similarly, a control group individual who does participate cannot be moved to the treatment group, and cannot be dropped from the study, but rather must be included in the control group. The more often this occurs, the larger the sample size that will be necessary to measure any observed changes from the program.

**Spillovers**

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 we have found 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 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.

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, 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 amongst 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 quality for the consumers in the community).

Ethical considerations

Stakeholders sometimes have ethical arguments about randomization, as some perceive them as arbitrarily and “unethically” depriving the controls from positive benefits. This argument 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 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 organization 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 through the careful test not 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 typically 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.

Cost of Randomized Experiments

Experimental methodologies are often perceived as more difficult or costly than non-experimental methodologies. Relative to no evaluation at all, then 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 optimize the long-term decisions and planning for the institution. 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. The analysis for an experimental evaluation is, if designed correctly, quite simple: one obtains the answer
simply by comparing mean outcomes between treatment and control groups. 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 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.

5. Evaluating Group- versus Individual-Liability Loans, the Philippines

This section introduces the example of a randomized control experiment in the Philippines, which evaluates 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 control trials to rigorously evaluate 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 for those in the microfinance community to know what the real advantages and disadvantages are of each—and therefore to formulate policies on this topic. 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 peer pressure, not just legal pressure, 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.
- The group process may help build social and business relationships.

However, 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, as described above, 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 control 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

---

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

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

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 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.

**Pilot Phase**

Since this change to individual liability is significant, careful testing is required before it can be fully 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 in fact perhaps is only good for the best groups.

This first pilot phase began in August 2004. In early November 2004, after the program was in place for 2-4 months, 24 more centers were randomly converted. The full pilot phase as of May, 2005 includes 93 converted centers and 93 original (group liability) centers.

**New Area Plan**

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 control 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 discard 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 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 individuals 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 true impact of group- versus individual-liability programs. However, we cannot draw a generalized 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.

The methodology presented here is relatively easy to replicate and requires less risk than the typical process lenders use to innovate new product designs. The replication could help identify the specific environments, such as clients’ characteristics, institutional framework, and loan characteristics, in which individual-liability is most effective. The issue of generalizability is further discussed in Section 5.

**5. Generalizability**

As already mentioned, the replication of the study is crucial to be able to draw a more general conclusion about the impact of a particular program. In the case of the experiment with Green Bank, many factors may make the results of the evaluation unique to Green Bank and its context.
• **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.

6. **Pilot Experimental Approach for Other Lending Product Innovations:**

In the previous section we used an example from a randomized control 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. **Credit with Education**

Credit with Education is one of the most popular add-in features of lending programs; yet, the impact of education programs is not well known. Educational programs such as business training may help clients with financial management and lead to more efficient allocation of credit, a service which clients may value. However, if these educational programs are offered mandatory, clients might find it time-consuming and leave the program all together. See McKernan (2002).

b. **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.
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 et al. 2003; Ashraf, Karlan et al. 2004).

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, 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. Little empirical evidence exists on these issues. Experimental studies can be done to study the relationship between interest rates, demand for credit and repayment rates. See Karlan and Zinman (2005) for an example of such a study.

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 control trials. Randomized evaluations can be performed ethically and cost-effectively, and the accuracy of their results makes them valuable both to institution implementing the evaluation and to the microfinance community at large. Through the example of the evaluation of the introduction of individual-liability loans at a microfinance program in the Philippines we showed the steps involved in performing an experimental evaluation. A great many questions remain untested, 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. To stimulate the experimental evaluations of more program innovations we have provided a list of several modifications which could be tested using similar methodology.

8. Bibliography


