

## Randomized Trials for Strategic Innovation in Retail Finance

Nathanael Goldberg, Dean Karlan & Jonathan Zinman

January 2008

---

*Contributions to this research made by a member of The Financial Access Initiative and Innovations for Poverty Action.*

The Financial Access Initiative is a consortium of researchers at New York University, Harvard, Yale and Innovations for Poverty Action.

NYU Wagner Graduate School  
295 Lafayette Street, 2nd Floor  
New York, NY 10012-9604

T: 212.998.7523  
F: 212.995.4162  
E: [contact@financialaccess.org](mailto:contact@financialaccess.org)

[www.financialaccess.org](http://www.financialaccess.org)

Innovations for Poverty Action applies rigorous research techniques to develop and test solutions to real-world problems faced by the poor in developing countries.

Innovations for Poverty Action  
85 Willow St, Building B, 2nd Floor  
New Haven, CT 06511

T: 203.772.2216  
F: 203.772.2428  
E: [contact@poverty-action.org](mailto:contact@poverty-action.org)

[www.poverty-action.org](http://www.poverty-action.org)

## Randomized Trials for Strategic Innovation in Retail Finance

by

Nathanael Goldberg  
Innovations for Poverty Action

Dean Karlan  
Yale University

Jonathan Zinman  
Dartmouth College

January 2008

The financial world is full of indications that microfinance has “arrived”. 2006 saw Muhammad Yunus and Grameen Bank win the Nobel Peace Prize. 2007 saw the \$1.5 billion IPO of the Mexican microfinance institution Compartamos. Now rumors swirl that the next watershed may be the entry of investment giants such as the Carlyle Group and Blackstone into microfinance.<sup>1</sup>

Yet 50-80% of the adult population in developing countries remains unbanked. Hundreds of millions of (working) poor households and entrepreneurs still lack access to basic financial products and services.<sup>2</sup>

How will this vast untapped demand be met? Already, retail financial institutions are reaching out to poor households and entrepreneurs. But clearly meeting this demand is not as simple as hanging a sign and ushering the poor into branches. Poor customers have particular financial needs: they need to be able to deposit and borrow small amounts of money, often at irregular intervals; many live in remote areas; many are illiterate. How are banks to overcome these obstacles? Innovations for Poverty Action (IPA) has been working with financial institutions (both for-profit and non-profit) in a variety of settings to explore techniques for expanding financial access for the poor.

Our advice to financial institutions wishing to serve the poor is to *scale down but emulate up*. Many leading firms (Capital One, H&R Block, Amazon, etc.) have adopted systematic experimentation as part of their risk management, marketing, pricing, and innovation strategies. Our partners in developing countries are finding that using scientific rigor to create a “learning organization” is a source of strategic innovation and long-run comparative advantage. IPA has worked with banks and microfinance institutions to study marketing, risk assessment, product development, and more. In this paper we will describe how retail financial institutions can use randomized controlled trials to improve both profits and broader social objectives.

---

<sup>1</sup> *MicroCapital Monitor* Vol. 2 Issue 10, October 2007

<sup>2</sup> World Bank. *Finance for all? Policies and pitfalls in expanding access*.

### *Why Randomized Trials?*

Suppose you want to know how successful a new type of savings account is at bringing in new clients and increasing savings balances. One naïve approach is simply to track account activity following the launch of the new product. Say you do this and find that 1,000 new accounts were opened since the launch, and that savings balances increased by 10% on average. Does this mean the new product was a huge success? Not necessarily. Maybe demand for saving products increased in your market for reasons unrelated to the new product? Maybe you improved sales or customer relationship management practices as part of the new launch?

The impact question simply stated applies both to measuring impact on poverty as it does to measuring impact on profits for the firm: “How did the lives of the clients and the profits of the bank change relative to how they would have changed had the policy not been implemented?” The first part of that question is typically easier to answer. One can follow clients to see how their lives changed (and measure a myriad of outcomes to one’s heart’s content), and one can measure profits of the firm. The trick is measuring how these things would have changed had the policy not been implemented.

Establishing causality is important in order to avoid drawing wrong conclusions. Wrong conclusions can harm clients, and it can also be costly for firms to change operations (or not change operations) when they should not (or should). So it is essential to verify whether innovations truly achieve the desired effect, and to learn about how and why they are (not) working well in order to maximize effectiveness. There is almost always more than one way to design a product, market a service, run a business. Even if one could be confident that the product or service is working well, is it the *best* way to do it? Is it the *most* profitable, does it have the *greatest* impact on clients? Only through accurate evaluation is it possible for firms to make such comparisons and determine their optimal operational strategy.

But, establishing causality is hard to do. A scientific approach is a must because human behavior is so complex. Customer decisions are driven by many factors, and we typically cannot measure all of these factors perfectly (or even close to perfectly).

Randomized controlled trials (RCTs) solve this problem by randomly varying the business proposition we want to evaluate—e.g., the offer of a new product, or the features or marketing of a product—and holding everything else about the offer fixed. So in our saving product example a RCT would randomly assign some existing bank clients to be offered the new product. Other clients would be randomly chosen to not receive the offer (yet). A large enough sample ensures that the “treated” clients (those who got the offer) are essentially identical to the “control group” (those who did not get the offer). After all, the only thing that determined who got the offer is a random number generator (the computer code equivalent to drawing names out of a hat). The bank could then compare client retention and balances over some period of time (say 6-12 months, or sometimes even longer) to measure whether the new product had a *causal* effect on client behavior and profitability. Then if the product was indeed a success the bank could offer it everyone.

RCTs are the “gold standard” research methodology because they allow us to establish causality. Non-experimental methods typically require strong assumptions about *why* some people use a product in order to conclude that the product itself—and not some unobservable characteristic of the individual, or some environmental influence—*caused* a change in an outcome of interest. The randomization takes care of both of these issues because individuals do not *choose* whether to be randomly offered the product or not, and because there is no need to control for environmental influences since they influence both the treatment and the control groups. So whether we are studying the marketability of a new product or the effect of a new human resources policy, an RCT evaluates the *causal impact* of that innovation. By impact we mean the effect of the innovation *compared to it not existing*. Only this approach identifies the true value-added of an innovation, whether the innovation is in the product, marketing, or product space, and whether the metric of value is profits or social impacts or both.

We will devote the remainder of this paper to demonstrating through several examples that RCTs can provide a source of strategic innovation for financial institutions. These studies are part of a larger trend in scientific experimentation among innovative organizations.

#### *Direct Mailing: Pricing and Marketing*

Some settings are more natural than others for conducting RCTs. There is perhaps no more opportune setting than a direct mailing (or mass emailing, phone banking operations, or text messaging) to potential customers. Direct communications can provide a large sample size and a carefully controlled environment in which several versions of a marketing or pricing pitch can be evaluated simultaneously. Even better, where an institution is already conducting direct individualized communications, the only additional cost of an RCT is in managing the permutations and carefully tracking their response rates.

We conducted an RCT with a South African consumer finance company in which we used a direct mailing of more than 50,000 credit offers to former clients to evaluate the lender’s pricing, marketing, and risk strategies. Each of these presents problems for financial institutions, especially when expanding into new markets, because there is surprisingly little evidence available to guide them in making key decisions in these areas: how to price products, who to lend to, and how to sell it to them. Sure, it is simple enough to mimic competitors’ products and pricing, and to lend only to obviously qualified customers, but a profit- or socially minded firm should aim for a higher bar: what is the *optimal* strategy in each of these cases? How do we maximize returns?

To explore these questions the lender mailed “pre-qualified,” two- to six-week limited-time offers to all former clients of in 86 urban and rural branches who had borrowed from the Lender within the past 24 months. The client base was largely working poor. The mailing was limited to former clients who were in good standing with the lender, and did not have a loan outstanding. The offers contained randomized interest rates ranging from 3.25% per month to 14.75% per month,<sup>3</sup> 96% of which were lower than the

---

<sup>3</sup> Readers unfamiliar with the South African consumer credit industry may be surprised at these rates. They are high by global standards but, at the time of the experiment, normal for South Africa.

lender's standard rate.<sup>4</sup> Former clients eligible for maturities longer than four months also received a randomized example of either a four-, six-, or 12-month loan.

In our sample frame, 4,540 clients out of 53,810 applied for a loan at the offered interest rate, for an 8.4% take-up rate. Among the 99% of the sample who received offers at or below the standard rate for their risk category we found a 100-basis-point decrease in the monthly interest rate increased take-up by 3/10 of a percentage point. This implies that a price decrease from the maximum rate offered to this group (11.75%) all the way to the minimum (3.25%) would only increase take-up by 31%. Raising rates, however, had a much larger effect, with a price sensitivity six times as great. One potential explanation is that when customers know your standard rates they perceive an increase as “unfair” and react accordingly.

While borrowers turned out to be rather insensitive to price cuts, we found that they were much more sensitive to the maturity of the loan. Though they would ultimately pay much more interest with longer loans, poorer clients increased the amount borrowed when offered a longer maturity loan, a finding which suggests the poor are severely liquidity constrained. The wealthier half of our sample did not increase their loan size when offered a longer-maturity loan, and were more sensitive to the interest rate offered.

We used the same direct mailing to test out the effectiveness of several marketing approaches drawn from the psychology literature.<sup>5</sup> By simultaneously testing marketing and price within the same experiment we were able to calibrate each marketing feature to the change in interest rate one would have to offer to achieve the same effect in take-up. The results were surprisingly large. We found that some marketing treatments—at no cost to the lender—were as powerful as dropping the interest rate substantially. One in particular, the simple inclusion of a female photo on the mailer, was equivalent (for male clients) to dropping the interest rate by 4.5 percentage points. For some clients the lender would have to cut the interest rate in half to achieve the same effect!

In another test we compared the presentation of a single example loan offer to a table with several choices of loan sizes and monthly payments. While economic logic might suggest that more information is better for the consumer, our results corroborated a psychology theory which suggests consumers can be overwhelmed by “choice overload”. We found that employing a simple description of the offer has roughly the same effect on take-up as decreasing the interest rate by 2.3 percentage points (a 25% reduction in the interest rate for some!).

---

<sup>4</sup> Slightly more than 1% of the offers were higher than the normal rate and 3% were at the normal interest rate.

<sup>5</sup> We did this by independently randomizing price and marketing features to the same sample. Each household in the sample received a single mailer with a randomized loan offer and a separately randomized marketing pitch “selling” the offer. The pricing experiment does not affect the marketing experiment because in the aggregate clients received each pitch across the full range of prices. In effect the high and low prices cancel each other out. The same logic holds in the other direction: the marketing experiment does not affect the pricing experiment because those who received low rates would get *all* the marketing pitches, as would those who received high rates.

## *Risk Assessment*

How do you identify a creditworthy client? How should one balance “soft” information (human judgment) with “hard” information (quantitative models), and is there a way to get the best of both worlds? In this section, first we discuss how to use credit scoring to both make better credit assessments, and to measure impact on clients’ welfare and poverty. Then we discuss how one can set up randomized trials to determine the most effective decision-making process.

Lenders use information available to them such as borrowers’ credit history, income, and assets to estimate the risk of default for each credit applicant. But where is the optimal place to draw the line between creditworthy and not creditworthy such that lenders are not leaving money on the table? In other words, how do banks know for certain that the next client below the approval cutoff would not be likely to bring in more revenue than he or she would cost in collections and unpaid arrears? Since credit scores are imperfect the best way to find out is to take some calculated, controlled risks by actually lending to riskier clients in an RCT that allows you to measure where to draw the line between creditworthy and uncreditworthy.

We did a second RCT with the South African consumer lender described above to optimize its risk assessment criteria. Here the senior management felt its branch staff were being overly conservative. By randomizing some approvals for riskier clients who were rejected under the lender’s previously standard procedures (but not deemed obviously uncreditworthy by loan officers), we were able to limit the lender’s risk exposure while measuring the profitability of the riskier loans (both in an absolute sense, and compared to loans that were just-above-the-bar).

We found that the riskier loans were profitable, although substantially less profitable than the loans made under the standard, stricter criteria. The loans extended to applicants just below the bar were less likely to have been paid back in full than loans that were just *above* the bar (71.5% vs. 76.4%). With costing input from the lender we were able to calculate the profitability of the below- and above-the-bar loans as well. The marginal loans were profitable in an absolute sense (yielding R201 (\$32) per loan), but less profitable than the standard above-the-bar loans (which yielded R284 (\$45) per loan).

Risk assessment RCTs can also shed light on the *best* way to expand. Take the simple case where profitability is the only metric of success. Then a mid- or large-size MFI could compare the profits obtained from systematically liberalizing risk assessment within branches as in the RCT above, to the profits obtained from opening new branches. The key here would be to randomize whether branches or other types of field operations (including village banks) “dig deeper” (i.e., make riskier loans), or “spread farther” (by expanding into new areas or segments where there are potential clients with more typical risk profiles).

Risk assessment RCTs are also a natural place to evaluate the piece of a double bottom line that pertains to impacts on the borrowers. In the risk assessment RCT described above IPA hired a survey firm (and used follow-up credit reports obtained by the lender) to measure a range of economic and subjective

well-being outcomes in both the treated (randomly assigned to get a loan they otherwise would not have gotten) and control (randomly assigned to remain rejected) groups. Despite concerns about the welfare effects of consumer credit we found that treated applicants were better off than their control group counterparts 6-27 months after initially applying for the 200% APR loan. Treatment group members were significantly more likely to retain their job over the study period, and their incomes were significantly higher. They were less likely to report hunger, and were even less likely to be below the poverty line.

In all the findings from this study suggest expanding financial access can be good for the “double bottom line” of profitability and social welfare. As such the findings from this RCT provided important evidence to the policy community and valuable operational feedback to the lender.

Credit scoring can include hard and soft information, and in particular one can set up randomized trials in order to learn the most effective decision-making process. What training is effective in guiding decision-makers to emphasize the most important borrower characteristics? Who should get the final decision, the computer or the credit officer? If the computer, what are the “human” inputs (e.g., perception of trustworthiness, perception of financial sophistication and planning capabilities) that the credit officer can provide that can guide the quantitative model the best?

### *Product Development*

Rolling out a new product is a risky proposition for an FI as well, and RCTs provide a methodology for taking carefully controlled and measured risks in order to produce effective product innovations. One example of this approach was a savings product IPA developed in cooperation with a rural bank in the Philippines. The product design was built on cutting-edge research at the intersection of psychology and economics. This work suggested that two of the main obstacles individuals face in meeting their savings goals are self-control and spousal control. We designed a product that provided clients with more control: a “commitment savings” product called SEED. SEED clients established a goal (i.e., a commitment to reach a target balance within a pre-specified time period) upfront and voluntarily restricted their right to withdraw any funds in their own accounts until they reached a self-specified goal. Clients could opt to restrict withdrawals until a specified date (e.g., in a month when school fees were due), or until a specified savings amount was reached (e.g., a certain amount of money for a new roof). The clients had complete flexibility to choose which of these restrictions they would like on their account. However, once the decision was made it could not be changed, and SEED clients could not withdraw funds from the account until they met their chosen goal amount or date.

To evaluate the impact of this new product the rural bank implemented an RCT where it randomly assigned about 1800 individuals to either receive an offer to open the SEED account or not. The bank’s primary metric of success was savings balances, and here we found a strong effect: after twelve months, average balances increased by 80 percent in the treatment group (i.e., those who got the SEED offer) compared to the control group. We were also interested measuring whether the spousal control

channel was important, and found a significant increase in women's decision-making power within the household, which we attribute to the account helping women to gain control over household assets.

### *Conclusion*

These examples represent just a few of the randomized trials we have conducted with retail financial institutions. We have used RCTs to evaluate door-to-door deposit collection, emergency savings accounts, health insurance, and more. Our current work includes projects testing loan collection strategies, deposit reminders via SMS, and crop price insurance for poor farmers. We hope we have demonstrated both the flexibility of RCTs and their utility in helping financial institutions to improve their operations, profitability, and impact.

Of course, RCTs will be easier (and cheaper) to implement in some settings than others. The simplest RCTs (such as direct mail, text messaging and phone banking) can be implemented by banks themselves using a spreadsheet to randomize and their own accounting software to track clients. Where randomizations are more complicated, outcomes are harder to measure, and there is less direct control of the process (e.g., where marketing is performed outside the bank branch) RCTs will be more expensive and better left to experienced researchers. Still, for key decisions the benefits of better information can easily outweigh the costs. The trend towards commercialization of microfinance will provide firms with incentives to get these decisions right. We predict firms will follow the lead of consumer finance companies in America and begin integrating randomized trials into everyday operations. These RCTs can be used to develop product and process innovations that are both profitable to financial institutions, and beneficial to their clients.