



Building Evidence to Promote Financial Inclusion

Stephen Nuñez

with

**Embry Owen
James Riccio**

March 2014

Funding for this paper was provided by MetLife Foundation.

We would like to acknowledge the help of MDRC staff members who contributed to the production of the paper. Gordon Berlin, Gayle Hamilton, Richard Hendra, John Hutchins, Robert Ivry, Frieda Molina, and Caroline Schultz read earlier drafts and provided valuable feedback. Stephanie Rubino served as project manager. Joshua Malbin edited the paper and Carolyn Thomas prepared it for publication.

Dissemination of MDRC publications is supported by the following funders that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, The Harry and Jeanette Weinberg Foundation, Inc., The Kresge Foundation, Laura and John Arnold Foundation, Sandler Foundation, and The Starr Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

The findings and conclusions in this report do not necessarily represent the official positions or policies of the funders.

For information about MDRC and copies of our publications, see our Web site: www.mdrc.org.

Copyright © 2014 by MDRC®. All rights reserved.

Contents

List of Exhibits	v
How to Read This Paper	1
Introduction	1
Theoretical Advancements That Have Influenced the Design of Interventions to Improve Financial Inclusion	3
Evidence on the Effectiveness of Financial Inclusion Interventions	5
Conclusion	29
Appendix A: Program Evaluation: How to Tell Whether a Program Is Effective	31
References	39

List of Exhibits

Box

1	How to Read Impact Tables	7
2	The American Dream Demonstration: Impacts on the Ownership of Real Assets	11
3	SaveUSA: Impacts on Allocation of 2010 Federal Tax Refund	15
4	Banarjee, Duflo, Glennerster, and Kinnan's Group-Lending Microcredit Study	22

How to Read This Paper

This paper was prepared for MetLife Foundation to guide its work and investment in the field of financial inclusion. The main body of text assumes a basic familiarity with the components of formal program evaluation, and the distinctions among experimental, quasi-experimental, and nonexperimental analysis. Readers who require an introduction to these concepts or who would benefit from a refresher are encouraged to view the included appendix. There interested readers can also find additional information on the Social Innovation Fund framework used to rank the quality of evidence associated with each of the programs discussed. This includes official definitions of the three tiers of evidence quality: strong, moderate, and preliminary.

Introduction

In the last 30 years, practitioners and scholars interested in designing programs and policies that improve the lives of low-income individuals have increasingly focused their attention on issues pertaining to financial inclusion: the expansion of access to mainstream credit, banking, and financial services. This new emphasis is motivated by several interconnected observations concerning the ways in which lack of access to the mainstream financial system, lack of ability to navigate that system, or insufficient ability to effectively manage one's own finances can worsen the economic security of low-income groups or make it harder for them to become financially self-sufficient. First, "the poor pay more" as a result of their lack of access to mainstream credit and banking. Check-cashing fees and interest on payday, pawn, or "rent-to-own" loans (financial instruments that emerged to exploit the unmet banking and credit needs of low-income households) can quickly strip low-income families of whatever slack they may have between earnings and living expenses and make the development of an emergency "cushion" impossible. Second, without access to mainstream credit or a savings cushion, the poor have few effective ways to deal with shocks like loss of income or unforeseen expenditures. Third, without savings and access to mainstream credit and banking the poor can be barred from potentially important pathways out of poverty, such as business ownership or postsecondary education. And fourth, many low-income individuals lack the information or experience necessary to make good financial choices about savings, asset building, or financial products. Therefore, expanding access to mainstream financial services and helping individuals make better financial decisions are increasingly viewed as important elements in fighting poverty, reducing material hardship, and improving opportunities for low-income populations.

These observations have motivated efforts to develop interventions aimed at increasing access to and participation in mainstream credit and financial markets, and increasing individuals' ability to manage their finances effectively. Some programs offer access to accounts

provided by banks and credit unions for both aspirational and emergency savings, and give further incentives for savings with automatic deposits or matching funds. Other programs include, in various combinations, financial literacy and credit repair, drives to sign up the “unbanked” (people without bank accounts) for no-cost checking and savings accounts, and the development of easy-to-use planning and budgeting products. Microcredit/microfinance organizations offer another potentially important service by providing group or individual access to loans for a variety of uses. To the extent that such interventions are modular, practitioners have searched for ways to effectively bundle them with job-training, employment, and other antipoverty programs already operating.

Interest and investments in such strategies for improving financial inclusion have grown dramatically, and this growth has fed an increasing desire to understand how well these kinds of interventions “work.” Unfortunately, high-quality evidence in this field is in short supply, although it is slowly accumulating. This paper attempts to offer some guidance on the issue of evidence by exploring the current state of evidence in the financial inclusion field. Toward that end, the paper focuses attention on a selection of programs, each of which illustrates a certain type of approach within the field of financial inclusion. It concludes with a broad assessment of the state of evidence in this field and suggestions for how to strengthen the evidence base further.

Because it is a useful way to rank evidence quality, this paper uses the Social Innovation Fund’s three-tiered evaluation framework in the later section that reviews particular examples of financial inclusion programs. This framework categorizes evidence quality as strong, intermediate, or preliminary. Strong evidence includes results from one or more well-designed, large-sample randomized controlled trials; moderate evidence includes results from well-designed quasi-experimental analyses or experimental studies with small sample sizes or other limitations; preliminary evidence includes results from tracking studies or other non-experimental investigations of reasonable hypotheses.¹ It is important to emphasize that high-quality *evidence* does not mean that a program is effective. It may be that some programs are very effective, but lack strong evidence to demonstrate their effectiveness. Alternatively, high-quality evidence may demonstrate conclusively that some programs are *not* effective.

Before discussing particular programs, the paper briefly addresses theoretical advancements that have influenced the rationale and structure of alternative approaches being tried and evaluated, in order to provide background for understanding the nature and effectiveness of the interventions.

¹See the Appendix for a description of the Social Innovation Fund and a detailed description of its evidence categories.

Theoretical Advancements That Have Influenced the Design of Interventions to Improve Financial Inclusion

A variety of innovative theories and perspectives have influenced the design of financial inclusion interventions. Knowledge of these is useful in understanding the diverse kinds of approaches that have emerged to try to improve financial inclusion among low-income populations across the globe. The following section briefly summarizes the kinds of theories that have typically informed such interventions.

Insights from Behavioral Economics and Psychology

Behavioral economists have moved away from the unrealistic assumptions about decision making made by classical economists and have instead embraced assumptions that are more consistent with psychological research. First, and most obviously, people have limited ability to calculate the gains and losses associated with various options and can be overwhelmed by choice and complexity; they make mistakes. Behavioral economists also recognize that people do not always understand their preferences or needs well. In particular, people are quite bad at anticipating their future needs and conditions. As a result they may plan poorly for the future by failing to set aside enough money or food. In addition, emotions can distort decision making by inducing impulsive behavior or by limiting the amount and clarity of analysis done before making a decision. Examples of this include impulse purchases, “crimes of passion,” or judgment clouded by depression or desperation. Finally, people are subject to a variety of “cognitive biases” that can greatly influence decision making and lead to suboptimal outcomes. For example, research suggests that people will make different decisions when faced with the same choice depending on *how* (for example, what words are used) and *by whom* (for example, an authority figure or a friend) the choice is presented (the “framing effect”). When making a choice, people tend to avoid or ignore options for which they have limited information or which they do not understand well (rather than seeking to fill in these gaps), preferring options they understand and are familiar with (the “ambiguity effect”). People will also tend to prefer the status quo over change, even when the risks or efforts associated with change are minimal (the “status quo” or “default effect”).

These insights have influenced both the structure and presentation of programs designed to provide financial education, promote savings, and build credit. For example, the observation that individuals may not accrue sufficient savings because of poor impulse control has led practitioners to offer savings accounts with a variety of restrictions on how much and how often money can be withdrawn. Because the savings are not easily accessible, they are in theory safe from impulsive decisions that could drain savings and later be regretted. Other programs exploit the “status quo” effect by automatically enrolling the employees of a participating organization, for example, into a savings program that automatically deducts a small

amount from paychecks and deposits it into a savings account. Participation is greater than if the program were simply offered as an option to which employees would have to “opt in.”

Behavioral economic insights also factor into the design of materials to educate individuals about the potentially pernicious effects of usurious lending schemes (for example, payday lending) or the dangers of overborrowing. There is evidence that the wording of educational materials can affect their success in steering individuals away from potentially dangerous options. Describing the cost of payday loans, for example, in terms of annual percentage rates (for example, “Did you know you are paying over 465 percent interest per year?”) has proven ineffective; individuals who received these materials had difficulty understanding the concept of annual percentage rates and, furthermore, had no way to relate the number to the costs and experiences familiar to them. Describing interest as a fee and providing a currency value (for example, \$1,500 per year) both generated larger changes in later borrowing behavior, as did describing what that money could have been used for — researchers have found that people are particularly averse to activities described as creating losses or lost opportunities.

Insights from Research on Social Networks

Practitioners in the field of social inclusion have recognized the importance of social networks, including the realities of being part of a web of reciprocal obligations, in their program designs or outreach efforts. For example, microcredit organizations (described below) use social network connections to recruit borrowers. They expect borrowers to spread the word to friends and family, who will then become borrowers themselves and do the same. Some microcredit organizations use group-lending models (described in more detail below) in which each of four or five individuals with strong social connections (same village, same family, good friends, etc.) receives a loan but all are responsible for the repayment of each loan; if any one person defaults no group member can receive further assistance. Because loan group members have strong preexisting connections, they have reciprocal obligations to help struggling members, thus making default less likely. In information-poor environments where credit checks and other forms of vetting are difficult or impossible, lenders may therefore substitute a social network-based strategy for a formal underwriting process. Some researchers also note that simple, low-cost savings accounts are useful not merely to provide distance between savings and the impulsive desires of the saver, but to “hide” the money from friends and family and thus

protect it from a cycle of reciprocal obligations.² A desire to do so appears to be a motivating factor for those who sign up for such accounts.

Insights from Research on Institutional Economics

Efforts to improve financial inclusion are aided or constrained by the larger context of government and law. As economists and political scientists who focus on institutional economics point out, legal and institutional arrangements that are fundamental to many economic transactions and asset-building opportunities, and that are taken for granted in highly developed countries, often do not exist or do not function well in less developed economies. Consequently, in those latter contexts financial inclusion strategies must often work around fundamental weaknesses in the institutional environment. For example, when property rights are poorly defined or documentation is incomplete, it is difficult to secure formal credit because people cannot easily prove ownership of property for the purpose of collateral.³ Poor transportation and technological infrastructure can also make it difficult for people to gain access to banking and other financial services.

Lack of formal property rights or documentation has led to efforts to find different means of securing or underwriting loans in developing nations, including the group-lending approach discussed above. Issues with transportation and access to computers have forced practitioners to become creative in efforts to connect people in developing nations with mainstream banking services. For example, some organizations have developed banking applications for mobile phones that give people access to many of the services typically provided in person in developed nations. Even in poor rural areas with little infrastructure, many people possess mobile phones.

Evidence on the Effectiveness of Financial Inclusion Interventions

Financial inclusion is currently a somewhat ambiguous term, with no widely accepted common definition. However, construed broadly, any intervention can be thought of as promoting financial inclusion if it expands access to mainstream credit, banking, and financial services, or

²In poor communities social ties are often the primary source of aid in emergencies such as job loss, sudden illness, and necessary repairs. That support, however, often comes with an obligation for reciprocal support when others find themselves in need. (For example, “My friend helped me when I was down so I owe her.”) The constant stream of reciprocal obligations may prevent members of a community from falling into abject poverty but it also makes it difficult for them to advance economically; any savings or windfall is likely to be quickly redistributed to help others currently struggling. Sociologists term this the “leveling effect.”

³This insight was made popular and promulgated by the economist Hernando De Soto in his book, *The Mystery of Capital*. De Soto (2000).

expands the possibility of participating in economic activities that rely on such access (like small business ownership), and if it does so in a manner that is not exploitative or damaging to participants. This is consistent with the definition MetLife Foundation has adopted to guide its investment strategy:

Financial inclusion means that households and businesses have access to a full suite of quality financial services, provided at affordable prices, in a convenient manner, and with dignity for the clients. Furthermore, these households and businesses effectively use these financial services in a manner which allows them to improve management of their incomes and assets. Such services are delivered by a range of providers, most of them private, and must be provided responsibly and sustainably, in an appropriately regulated environment.

With this definition in mind, this section presents an overview of various types of interventions carried out under the rubric of financial inclusion. Together they illustrate some of the broad diversity in philosophies and strategies currently used to increase financial inclusion among low-income populations.

For each *type* of intervention discussed, the paper provides one or more examples of an actual program or organization. In practice, most programs include elements from several different types of approach. Consequently, the classification scheme used here places each program or organization into the category that represents the central component of its strategy. For example an organization that primarily provides group microloans would appear in the section describing “microcredit/microfinance” even if it also provides financial literacy training to its borrowers. The organizations discussed below are included because they illustrate various approaches. Their inclusion is not meant to signify that they are the “main” or “best” players in this field.

The discussion highlights the kinds and quality of evidence available on the effectiveness of the illustrative approaches. In doing so, it ranks that evidence by applying the Social Innovation Fund’s three-tiered framework of “strong,” “moderate,” and “preliminary” evidence (see above). For a number of studies, “impact” findings from randomized controlled trials are presented in tables, to provide concrete illustrations of what may be viewed as strong or moderate evidence. To help the nontechnical reader understand those tables, Box 1 provides a general guide.

Box 1

How to Read Impact Tables

Most impact tables in MDRC impact analyses use a similar format, illustrated below with an example from MDRC’s SaveUSA evaluation. The data show requests by SaveUSA group members and Regular Tax Filers concerning how they wished to receive their 2010 federal tax refunds. For example, the table shows that about 98 (98.2) percent of the SaveUSA group and about 76 (75.6) percent of Regular Tax Filers asked the IRS to directly deposit all or part of their refunds in bank accounts — either savings or checking.

Because individuals were assigned randomly either to the SaveUSA group or to the Regular Tax Filers group, the effect of the program can be estimated by the difference in outcomes between the two groups. The “Difference (Impact)” column in the table shows the difference between the two research groups’ rates — that is, the program’s *impact* on requesting a particular way of allocating tax refund dollars. For example, the impact on the incidence of requesting that the IRS directly deposit tax refund dollars in a bank account can be calculated by subtracting 75.6 percent from 98.2 percent, yielding 22.6 percentage points.

Differences marked with asterisks are “statistically significant,” meaning that it is quite unlikely that the differences arose by chance. The number of asterisks indicates the level of statistical significance of the impact (the lower the level, the less likely it is that the impact is due to chance). One asterisk corresponds to the 10 percent level; two asterisks, the 5 percent level; and three asterisks, the 1 percent level. The p-values show the exact levels of statistical significance of the difference to three decimal places, ranging from .000 (extremely unlikely to have occurred by chance) to .999 (extremely likely). By convention, three asterisks are used for any p-value below .001, and the difference is described as being statistically significant at the 1 percent level. For example, as shown below, the SaveUSA group had a statistically significant impact of 22.6 percentage points at the 1 percent level on the measure of asking the IRS to directly deposit tax refund dollars in a bank account.

Impacts on Allocation of 2010 Federal Tax Refund

Outcome	SaveUSA Group	Regular Tax Filers	Difference (Impact)		P-value
<u>Allocation of tax refund (%)</u>					
To any bank account	98.2	75.6	22.6	***	0.000

Promoting Savings

Some “financial inclusion” approaches focus on ways to effectively promote savings in formal accounts among participants. Savings can be used for different purposes:

1. Emergency savings can provide a buffer against “expenditure shocks” such as sudden injury or illness or a necessary repair. They can also be tapped during periods of unemployment or underemployment. With meager incomes, it is difficult for low-income households to accrue such savings. A single shock can drive a household into crippling debt by forcing reliance on credit cards, payday loans, or money obtained from loan sharks in the community.
2. Aspirational savings can be put toward the acquisition of a vehicle (expanding access to employment and services), the purchase of a home or consumer durables such as an appliance (improving quality of life), or the attainment of a certification or educational degree (expanding employment opportunities). Such savings can also be invested in children’s health and education, improving the life chances of the next generation.

Depending on the type of savings the program is attempting to facilitate, practitioners deploy a variety of strategies.

Automatic Savings Programs

An automatic savings program sets up an automated process by which money is periodically deposited into a savings account, often newly created when a person begins participation. Behavioral economists have, as noted, examined the potential for emotional states to distort optimal decision making and the difficulty many people have in anticipating their future needs. Advocates of such programs therefore argue that in addition to providing a safe place to store money (instead of, for example, under the mattress, where it can be stolen) they promote savings by sheltering a portion of earnings from impulsive spending. Some automatic savings programs impose a variety of restrictions on how and when a participant can withdraw funds. This further enforces regular savings growth. Some people may sign up for such programs as a way of imposing self-control on their spending habits; that is, they recognize that at some point in the future they may otherwise be tempted to spend unwisely.⁴ Some practitioners encourage

⁴An example from ancient Greek literature of similar “preemptive self-control” often cited in literature on automatic savings programs can be found in the *Odyssey*. Odysseus is informed that his ship will soon pass the island of the Sirens, wicked creatures with beautiful voices. Those who hear the Sirens’ call find it irresistible; they cannot help but steer their ships toward them, which inevitably leads to a wreck and certain death.

(continued)

employers to automatically enroll employees as program participants and allow them to opt out. This exploits the status quo bias uncovered by behavioral scientists and discussed above. Of course such actions are controversial because they assume the creators of the program “know best,” that is, that they have a better understanding of participants’ needs than the participants themselves.

Example: AutoSave

Evidence quality: Preliminary

AutoSave is an exploratory pilot program implemented through a partnership between the New America Foundation and MDRC. The program was designed to be easily “inserted” into a preexisting payment architecture in order to increase low-income workers’ savings and connection to mainstream banking. MDRC worked with interested employers to integrate AutoSave into their payroll systems. Through a payroll deduction, the program automatically diverts a small amount of the wages of low-income and moderate-income workers into savings accounts. Unlike most existing workplace saving programs, which focus on building retirement assets, AutoSave savings are intended to be fully liquid and available to cover short-term needs. They also may potentially increase workers’ attachment to mainstream financial services or serve as building blocks to longer-term asset accumulation.

Ideally, employees of participating employers would be automatically enrolled and allowed to opt out (applying the status quo bias that behavioral economists have identified). However, a number of legal and practical obstacles — such as employers’ inability to open bank accounts for employees without their explicit consent — have made it difficult to construct this program on an opt-out basis. Thus, the attempts to date have had to employ opt-in mechanisms. In an initial pilot test, several private, public, and nonprofit employers implemented an opt-in version of AutoSave using traditional savings accounts; about 350 employees signed up. MDRC and some employers have continued to explore opportunities for opt-out strategies; one such idea would make use of payroll cards (prepaid cards used by employers to deliver pay) as a savings vehicle. However, there remain gray areas about what uses of payroll cards are permissible under current (though still evolving) federal and state regulations, and what would be considered best practice from a consumer perspective. Given these difficulties, there are no plans at this time for a formal randomized controlled trial and impact analysis.

Odysseus instructs his crew to plug their ears. He, however, wants to experience the beautiful song of the Sirens without meeting his doom. Therefore he has his men tie him to the mast of the ship.

Individual Development Accounts (IDAs)

Individual development accounts are financial instruments designed to help low-income households meet savings goals in pursuit of a particular end, such as purchasing a car or home.⁵ Typically such accounts include a matched savings component: for every dollar the participant saves or for every milestone the participant reaches, the program contributes a certain amount of money toward the savings goal (this can be but need not be dollar for dollar).⁶ Participants may forfeit matching funds if they do not meet milestones by a certain deadline, or if they withdraw funds early. Thus, the offer of a match is intended to discourage the impulsive use of savings. IDAs can logically focus on any desired end, but typically they have been designed to promote home ownership, postsecondary education, or small business ownership.

Example: American Dream Demonstration

Evidence quality: Strong

The American Dream Demonstration was Professor Michael Sherraden's test of the effectiveness of IDAs.⁷ It ran from 1998 to 2002. The study included 13 sites, which allowed researchers to vary the components of the program to determine the most effective and cost-effective manner to market the accounts and achieve the desired result of increased savings. The IDAs offered were bundled with financial counseling and educational services of varying intensity. Sherraden also experimented with the savings match rate, offering between 1-to-1 and 6-to-1 matches for meeting savings targets.

Because of the strong design of the study, its findings were conclusive: IDAs significantly increased the long-term savings of the program group over the control group and did so without inadvertently causing material hardship (some had worried that participants would cut back significantly on consumption to meet savings goals). (See Box 2; also refer to Box 1 for general guidance on how to read an impact table.) Increasing the match rate boosted savings contributions from participants. However, the effect rapidly diminished when the match rate exceeded a 3-to-1 ratio. Interestingly, the higher match rates caused unanticipated problems with recruitment; potential study participants were suspicious of the "free money" and worried that the program was a scam.

⁵IDAs were first proposed by Michael Sherraden, a professor of social work at Washington University in St. Louis and a leading scholar in the fields of financial empowerment and inclusion. See Sherraden (1991).

⁶Note that automatic savings account programs may also include a matched savings component.

⁷Evaluators often use the term "demonstration" to describe a study commissioned to both design and test a new program. The evaluator will take part in developing the program's theory of change and components, and then design and implement a test of the program. Because the program is developed as part of the study, it is set up with the requirements of a robust evaluation in mind.

Box 2

**The American Dream Demonstration:
Impacts on the Ownership of Real Assets**

The table highlighted in this box shows the effects, or “impacts,” on real asset ownership of the American Dream Demonstration’s IDA and financial counseling program. The top half of the table presents impacts at about 18 months after random assignment; the bottom half presents impacts on the same measures at 48 months after random assignment. At 18 months, researchers found no evidence of program impacts on home, business, vehicle, or other property ownership. Had they stopped there, they might have concluded that the program failed to have any of these intended effects. However, follow-up at 48 months revealed a statistically significant 6.2 percentage-point increase in home ownership among the program group compared with the control group. While both program and control group members gained in home ownership during the intervening period (potentially reflecting the motivation and savings affinity of study participants as a whole), the program group’s gain was greater. Given the time it takes to save for a home, for financial counseling to produce results, and for IDA matching funds to accumulate, it might not be reasonable to expect to see impacts only 18 months after random assignment. Of course, the table cannot speak to what happened next. It is possible that future follow-up will reveal that impacts have faded, either because program-group home ownership rates have dropped, because control-group members’ home ownership rates have grown (so they eventually caught up with the program group), or some mixture of the two. Or the effects of the program might grow even larger over time.

(continued)

Financial counseling also improved the savings rate, although it added significantly to the program’s expense. In his cost-benefit analysis, Sherraden argued that intensive counseling and a high match rate probably made the program too expensive to expand to a large scale, given the benefit provided. He also noted that IDA programs would have relatively high fixed start-up costs and more expensive maintenance than other account types. However, economies of scale meant that “light-touch” versions of the program — those involving less direct counseling — could be cost-effective and offered more widely.

Because the American Dream Demonstration was carried out at 13 sites, it allowed Sherraden to analyze the effects of changes to individual components of the program. Component analysis is often employed once a program taken as a whole has proven to be effective. A randomized controlled trial of a program treats it like a “black box,” in the sense that any impacts uncovered are the product of the program as a whole. Evaluators can use component analysis to determine the relative contribution of each major component of the program. Evaluators may also try adding new components or varying the intensity of a component to see whether this can improve the program model in terms of overall impact or cost-effectiveness.

Box 2 (continued)				
Outcome	Program Group	Control Group	Difference (Impact)	P-value
<u>Ownership of real assets at month 18 (%)</u>				
Home ownership	35.3	34.9	0.4	0.87
Business ownership	9.4	10	-0.6	0.74
Other property ownership	3.2	3.6	-0.4	0.76
Vehicle ownership	89.9	90.1	0.2	0.93
Sample size (total = 764)	NR	NR		
<u>Ownership of real assets at month 48 (%)</u>				
Home ownership	49.1	42.9	6.2	* * 0.04
Business ownership	10.3	10.5	-0.2	0.92
Other property ownership	5.7	4.7	1	0.58
Vehicle ownership	89.9	90.3	-0.4	0.87
Sample size (total = 840)	412	428		
<p>SOURCE: MDRC presentation of data analysis results reported in the final evaluation report of the American Dream Demonstration. See Mills, Patterson, Orr, and DeMarco (2004).</p> <p>NOTES: Estimates were regression-adjusted using probit models, controlling for pre-random assignment characteristics of sample members. The sample is weighted to adjust for a change in the random assignment ratio early in the demonstration.</p> <p>Rounding may cause slight discrepancies in calculating sums and differences.</p> <p>A two-tailed t-test was applied to differences between the outcomes of the program group and the control group. The p-value indicates the likelihood that the difference between the program group and control arose by chance. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; and *** = 1 percent.</p> <p>“NR” means that these values were not reported.</p>				

For example, as noted, the American Dream Demonstration found that intensive financial counseling improved savings outcomes, but only modestly and at great expense. Therefore, Sherraden recommended that large-scale implementation of IDA programs avoid “high-touch” counseling services.

Low-Cost/High-Liquidity Savings Account Programs

Special types of savings accounts can offer low-income households easier access to their money than IDAs, without many of the fees and sometimes confusing requirements (for example, those related to minimum balances) associated with savings accounts at for-profit banking institutions. However, they may include incentives to discourage quick withdrawal of funds. Scholars and practitioners who advocate for such accounts argue that:

1. While impulse control is a factor, savings accounts also provide a secure location to store money and one that can be hidden from friends and family members looking for aid. These accounts provide a way to circumvent the cycle of reciprocal obligations that emerges from participation in social networks of support (discussed above) and that can prevent economic advancement.
2. Low-income households also need emergency savings they can draw on to deal with an expenditure shock or loss of income. If an account holder cannot easily gain access to funds because of limits on frequency and amount of withdrawals, she cannot use her savings to meet her immediate needs. Therefore it is not sufficient to provide access only to the low-liquidity accounts described above.

Example: SaveUSA

Evidence quality: Strong

SaveUSA is a demonstration project currently underway in four cities. It is modeled after an earlier prototype called \$aveNYC.⁸ It aims to encourage individuals to have more savings on hand to pay for financial emergencies, to allow them to make necessary purchases and reduce debt, and to help them develop a habit of saving. SaveUSA builds on the free tax preparation services provided by participating Volunteer Income Tax Assistance organizations. Starting in 2011 (or the 2010 tax filing season), SaveUSA offered both single filers and couples who filed jointly the opportunity to open SaveUSA accounts at local financial institutions by directly depositing a portion of their tax refunds into them, and to earn matching incentives by leaving their savings untouched for about one year.

The SaveUSA account has special features that facilitate small savings by account holders, such as no ATM card, no minimum deposit requirement, and no dormancy fees. When preparing their tax returns, participants instructed the Internal Revenue Service or state taxing agency to directly deposit at least \$200 from their tax refunds into special savings accounts. In

⁸Azurdia, Freedman, Hamilton, and Schultz (2013).

each of the three years the program has been offered, participants could pledge to keep a certain amount of their initial deposits, from \$200 to \$1,000, in their accounts for approximately one year. A participant who fulfilled this pledge would receive a 50 percent savings match, up to \$500, about a year later. Account holders whose balances dropped below their pledge amounts at any time during the follow-up year would lose their eligibility for a match, even if they subsequently replaced the funds.

MDRC is conducting a randomized controlled trial to test the effects of SaveUSA in New York City and Tulsa, Oklahoma. An interim report is currently in production. It will show how much the program increased both the proportion of tax filers who have short-term (nonretirement) savings and the total amount of such savings, relative to the corresponding outcomes among the control group. It will also report on impacts on attitudes toward savings revealed through survey responses, and on the impacts of the program on participants' debt, net worth, financial hardship, and other aspects of financial or material well-being. Given the possibility that such impacts could take time to manifest, MDRC will continue to track program participants and release a report on longer-term impacts in 2015.

SaveUSA represents an example of the importance of the counterfactual. The program was designed to be voluntary and was marketed to low-income households that had the desire to build savings. An analysis of the characteristics of those who expressed interest revealed that most already had some savings and used checking and savings accounts. Without access to a counterfactual provided by an experimental analysis (a randomized controlled trial) or quasi-experimental analysis it would be impossible to determine the value added by this program: How much did SaveUSA increase savings and connection to mainstream banking above and beyond what would have occurred *without* this intervention among this relatively motivated and already largely connected group?

Box 3 illustrates the preliminary impacts of the program on one early stage in the designers' theory about how the model should produce change: the likelihood that participants will allocate tax refunds to savings vehicles. Planned evaluation reports will present similar types of impact results regarding a wide variety of financial inclusion and economic security outcomes.

Example: Fondo Esperanza's emergency savings account program

Evidence quality: Strong

Recently, Fondo Esperanza, a Chilean microfinance institution, and Banco Credichile, a large commercial bank, collaborated to provide customers of Fondo Esperanza access to a savings account with no associated minimum balance requirements or maintenance fees. The organizations worked with Felipe Kast and Dina Pomeranz, of Harvard University and the

Box 3

SaveUSA: Impacts on Allocation of 2010 Federal Tax Refund

This table presents some preliminary findings for the SaveUSA demonstration first described in the April 2013 policy brief *Encouraging Savings for Low- and Moderate-Income Individuals*.^{*} Evaluators look for two types of impacts: implementation and participant outcomes. Significant implementation impacts demonstrate that members of the program group received the intended services. SaveUSA, as noted, offered program group members the chance to deposit their tax refunds into special savings accounts. The table shows that 98.2 percent of program group members deposited all or part of their tax refunds into bank accounts compared with 75.6 percent of control group members. The impact size is 22.6 percentage points and is statistically significant at the 1 percent level. This allows researchers to be quite confident that offering these special accounts leads to a large change in behavior. However, the end goal of the program is to decrease material hardship and reduce debt; increasing the number of people who save money in bank accounts and the amount that they save in these accounts is an intermediate step, a means to an end. This table does not speak to the SaveUSA program's impacts on these end goals. It remains to be seen whether successfully increasing deposits of tax refunds into bank accounts such as the special SaveUSA account will lead to the desired outcomes, as anticipated by the program's theory of change.

Researchers are often also interested in whether the program's success in implementation or in producing impacts on participant well-being varies according to participant characteristics. For example, a program might have stronger impacts on women than men given their particular needs or goals. In this case, the study was conducted at two sites: Tulsa, Oklahoma, and New York City. It is natural to ask whether the program had different effects at these different sites because the characteristics of the sites themselves and the populations served there may vary considerably. The bottom section of the table repeats the analysis in the top section but breaks down the results by site, providing separate impact estimates for Tulsa and New York. This is known as "subgroup analysis." Just as researchers conduct statistical analysis to determine with confidence that the differences uncovered between the control and program group are not simply the product of chance, they conduct statistical analysis to determine with confidence that the differences in impacts measured across subgroups are not the product of chance either. Here researchers estimated the impact of the program in New York City on the percentage of people who deposit their tax refunds into any account at 29.2 percentage points (compared with the overall impact of 22.6 described above); the impact estimate for Tulsa is 13.1 percentage points.

Comparison of the two impact numbers through statistical analysis shows this difference is statistically significant at the 1 percent level (indicated by the three daggers presented in the site row: †††). One can be fairly confident that the program did indeed have different impacts at different sites. There are many reasons why this might be so, exploration of which requires thorough analysis of program implementation and sample characteristics.

(continued)

^{*} Azurdia, Freedman, Hamilton, and Schultz (2013).

Box 3 (continued)					
Outcome	SaveUSA Group	Regular Tax Filers	Difference (Impact)		P-value
<u>Allocation of tax refund (%)</u>					
<u>To any bank account</u>	98.2	75.6	22.6	***	0
Savings account	93.1	14.6	78.5	***	0
<u>By site:</u>					
<u>Received a tax refund deposit into any bank account (%)</u>					
Site				†††	
New York City	98.3	69.1	29.2	***	0
Tulsa	98.0	84.9	13.1	***	0
<u>Deposited money into a savings account (%)</u>					
Site				†††	
New York City	93.5	8.2	85.3	***	0
Tulsa	92.1	23.9	68.2	***	0
Sample size (total = 1,578)					
	794	784			
<p>SOURCE: MDRC calculations from random assignment module data and 2010 tax return records.</p> <p>NOTES: The sample includes New York City and Tulsa sample members who were 18 to 64 years old at their time of random assignment.</p> <p>Sample sizes for subgroups are as follows: New York City = 922; Tulsa = 656.</p> <p>Sample sizes for specific outcomes may vary because of missing values.</p> <p>Estimates were regression-adjusted using ordinary least squares, controlling for pre-random assignment characteristics of sample members. No special weights were applied to responses to adjust for differences in sample size by site.</p> <p>Rounding may cause slight discrepancies in calculating sums and differences.</p> <p>A two-tailed t-test was applied to the differences in outcomes between the SaveUSA group and Regular Tax Filers group. The p-value indicates the likelihood that the difference between the SaveUSA group and Regular Tax Filers arose by chance. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; and *** = 1 percent.</p> <p>The H-statistic is used to assess whether the difference in impacts between sites or subgroups is statistically significant. Significance levels are indicated as follows: † = 10 percent; †† = 5 percent; ††† = 1 percent.</p>					

National Bureau of Economic Research respectively, to roll out the product in conjunction with a randomized controlled trial; two-thirds of Fondo Esperanza's clients were randomly assigned to receive an offer to set up a savings account.⁹

The program increased savings (by an average of 52,300 Chilean pesos, about \$105) and improved consumption smoothing for program group members (for example, when program group members suffered a loss of income, they had to cut back their consumption by 44 percent less than control group members who suffered a loss of income).¹⁰ The program also improved measures of subjective economic well-being. Take-up patterns and survey responses suggest that the accounts were used, in part, to shelter savings from friends and family who came to participants looking for aid. Half of the program group received the additional treatment of access to self-help peer groups. This was designed to encourage self-control. Those who received this additional treatment accumulated significantly higher savings than those who received only the base treatment of access to the savings accounts.

The study of Fondo Esperanza's savings account program is a good example of how organizations can work with evaluators to pilot test a new service while simultaneously enhancing their business plans and contributing to the field's understanding of effective strategies to combat poverty. When an organization is interested in exploring a new program or service before offering it to all customers, random assignment is a fair and informative method of determining who gains access to the new product. In this case, Kast and Pomeranz were able both to demonstrate the value of the accounts offered and to identify the motivations of those who expressed interest in them, thus providing Fondo Esperanza with a better understanding of the needs of the people it serves.

Microlending/Microcredit

Microlending involves offering small loans to low-income households that are not served by mainstream banking institutions. Banks typically cannot or will not provide loans to such households for two reasons:

1. Bank loans typically have a high fixed cost for paperwork and "due diligence" efforts, such as underwriting. A \$10,000 loan has the same costs associated with it as a \$100,000 loan. Therefore a loan below a particular dollar amount will create a net loss for the bank unless the loan is offered at an extremely high (typically illegally

⁹Kast and Pomeranz (2013).

¹⁰"Consumption smoothing" refers to putting aside income or paying off debt during good times and drawing down savings or incurring debt during periods of limited income, keeping consumption more or less stable over time.

high) interest rate. It is often not feasible for many institutions to lend amounts appropriate for low-income households (for example, \$500 or \$1,500).

2. Low-income households typically have minimal connection to mainstream banking and credit. As a result they may have no credit history for banks to refer to when underwriting the loan. When low-income households do have a credit history it is often quite poor. In either instance such households constitute a substantial credit risk for a bank, which therefore may be unwilling to extend credit. This was noted as an important challenge in the section above on “Insights from Research on Institutional Economics.”

Microlending organizations “solve” these problems in a variety of ways. As noted above, some microlenders offer group rather than individual loans in order to tap into the information and peer pressure that social networks can provide. This may mitigate the risk of default by excluding those who cannot find others willing to join a loan group with them (perhaps because they have a poor reputation or are not considered trustworthy in the community) and by encouraging group members to step in when a member begins to fall behind on payments. Alternatively, organizations like Accion USA will offer very small (around \$500) “credit-building” loans to borrowers and, by doing so, position them to qualify later for Accion’s larger small business loans (see below).¹¹

Microloans are most commonly given for business development and micro-entrepreneurship and are offered by numerous for-profit and nonprofit organizations internationally.¹² Borrowers are required to use the funds to start or expand small businesses. For example, the funds may be used to purchase a food cart or other capital goods necessary to sell wares. A borrower who successfully repays such loans may become eligible for loans of greater size. Eventually borrowers may “graduate” into borrowing from a bank rather than a microlender. Microlenders often operate in areas with poorly developed or nonexistent labor markets. In such areas receiving a loan to open a small business may be a more feasible means of economic advancement than finding a job.¹³

Microlenders can largely be divided according to two loan strategies already noted above: those that provide individual loans and those that provide group loans.

¹¹If applicable, the microlender will also report repayment of such loans to credit rating agencies to help borrowers build credit history and eventually gain access to credit cards and other financial services.

¹²A small number of microlenders offer loans for emergency expenditures and debt relief. Others may offer “starter” loans to build credit history.

¹³Recently microlenders have expanded into countries with well-developed formal and informal labor markets. It is an open question as to whether small business ownership, with its associated risks, should be promoted in such places.

Individual Microlending

Example: Accion USA

Evidence quality: Moderate

Accion USA provides business microloans and financial counseling to individual clients throughout the United States. The organization provides support for both new and preexisting small businesses. Unlike many other microfinance organizations, Accion USA does not target specific groups of entrepreneurs, such as women or immigrants. Furthermore, the organization provides financial education through in-person classes and online modules, but does not require its clients to participate. In order to assist clients in building credit history, Accion USA reports all loans and payments to major credit bureaus. The organization is connected to Accion, an international umbrella organization that provides investment and technical assistance to microfinance organizations around the world.

Although Accion USA is the largest and best-known individual microlender in the United States, it in particular has not been subject to a randomized controlled trial or quasi-experimental analysis. Several other organizations employing individual-level microlending models have been evaluated with such methods, however. For example, some have been evaluated with a strong quasi-experimental evaluation in the form of “regression-discontinuity” analysis. Studies employing this approach take advantage of the fact that the microlender sets a strict upper bound on income to determine a person’s eligibility for a loan. Those who earn more than the upper income limit are not usually able to receive a loan. In reality, individuals whose incomes are a small amount higher than the eligibility threshold are probably not much different in background characteristics or financial situation than those whose incomes fall just under the upper income limit. Thus, researchers may consider the group just above the eligibility threshold to be a good comparison group (counterfactual) for those just below the threshold. The impact of the loan could be determined by comparing the outcomes of those who were near the income limit but still qualified with the outcomes of those who were above that threshold.

In other studies, the researchers used random assignment in conjunction with upper or lower income limits. Half of those outside the usual income limit were randomized to receive loan eligibility anyway, even though they would not normally qualify. The researchers then compared outcomes for those in this category (that is, those outside the usual income limit) who were offered loans with the outcomes for those in this category who did not receive loan offers.

Overall, the above studies have found modest impacts on earnings, material hardship, and business ownership. (The lenders evaluated offered loans for business ownership.)

These studies have the drawback that they investigated the effects of the programs only on the upper range of qualified participants (the regression discontinuity studies) or on a

population for which the program was not designed (the randomized controlled trials). A well-implemented randomized controlled trial or quasi-experimental analysis can only provide information about the impacts of a program on those in the study. If study participants differ in important ways from another population, evaluators cannot extend their findings to that other group.¹⁴ In this example, lenders may impose lower and upper limits on income for their customers/clients because they recognize that the very poor may not have the necessary skills or resources to utilize the program to the fullest, and because they believe that their program would provide little added value for those who are already relatively financially stable. The programs may have much stronger impacts on the “middle” population between the income limits, given that population’s needs and the nature of the service provided, but these studies cannot speak to that possibility. The regression discontinuity studies described above can only speak to the impacts of the programs on those close to the income maximum. The randomized controlled trials can only speak to the impacts on those who fall below the programs’ income requirements or who exceed their maximums. Thus, although the current literature suggests individual microloans can lead to modest positive impacts for the populations studied, the magnitude of the impacts for the intended populations remain unknown.

Group Microlending

Example: Grameen Bank

Evidence quality: Moderate

Grameen Bank is a nonprofit microcredit/microfinance organization that provides small loans to women to help them create or expand small businesses.¹⁵ The organization views entrepreneurship as an important and viable pathway to economic advancement. Loans are distributed to groups of five; each person within a group must repay her loan before anyone else in the group can receive additional funds. The group structure ensures that members monitor each other’s repayments and step in if a group member falls behind, and allows members to provide business advice and other support to each other. Grameen’s clients receive financial literacy training, and are required to open and regularly contribute to savings accounts. In the United States (see below), Grameen reports clients’ loan repayments to major credit bureaus. The Grameen Bank opened the first branch of “Grameen America” in 2008 in New York City. It now operates three branches in the outer boroughs and has expanded operations to other cities across the United States.

¹⁴This is known as the problem of “external validity.”

¹⁵Some academics and practitioners argue that women are more likely to repay microloans and are thus a safer investment. There is little empirical evidence that this is the case.

Most existing research on the effects of group-lending microcredit models on participants' poverty and economic well-being is based on nonexperimental methods that cannot effectively rule out concerns about selection bias and other threats to the validity of impact estimates. A few other studies have tried to produce better evidence using quasi-experimental methods, but certain other problems in the research designs have raised questions about the validity of their conclusions.

Only one reliable randomized controlled trial has been conducted of a group-lending model.¹⁶ This study randomly assigned over 50 branches of a microlender, Spandana, to impoverished neighborhoods in Hyderabad, India. That is, in program neighborhoods, Spandana opened a branch and offered microloans. In control sites, it did not. See Box 4 for an illustration of some results from this study.

While the basic research design and analyses were rigorous, the study had important limitations. The Spandana bank branches were introduced randomly into a set of neighborhoods that already had several competing microlenders. Therefore even in the “control” neighborhoods, people had access to a similar service. The survey uncovered only a 9 percentage-point net treatment differential — in other words, the percentage of people in the study who were living in the “intervention” neighborhoods and who actually received microloans was only 9 percentage points higher than the proportion of people in the control neighborhoods who received microloans (albeit from sources other than Spandana). Thus, when examining impacts on other economic outcomes for participants, the study can only show the effect that a 9 percentage-point increase in receipt of microloans can have on participants' consumption levels. Whether a larger increase (that is, a bigger treatment differential) would have a larger impact on consumption outcomes cannot be determined from this study.

In general in impact studies, the smaller the treatment differential, the larger the sample size must be to determine whether a given-size difference in outcomes between the program and control groups is really just a statistical fluke (that is, a product of chance). Alternatively, a program being studied with a small sample must produce a much larger impact in order to conclude with confidence that the observed difference is a real one.

MDRC believes that quasi-experimental methods are not a good option to evaluate a program like Grameen America because they probably could not be implemented in a way that would rule out selection bias. On the other hand, a group-randomized controlled trial (assigning groups of five women to either a program or a control category) seems feasible, with the

¹⁶Duflo, Banarjee, Glennerster, and Kinnan (2013).

Box 4

**Banarjee, Duflo, Glennerster, and Kinnan's
Group-Lending Microcredit Study**

In general, a larger sample size allows for more sensitive measurement of program effects. When evaluators design a study they create a sample large enough to detect impacts of the expected size. For example, if an intervention is anticipated to increase the rate of bank account ownership by about 5 percentage points, a study to test its effectiveness would need to enroll more sample members than if the anticipated effect was 10 or 15 percentage points. Every study has a “minimum detectable effect size” that is determined by sample size and a variety of other factors. If the minimum detectable effect size is larger than the anticipated effect of the program, the study cannot uncover impacts of that size or speak to the program’s expected effectiveness. Such studies are often termed “underpowered.” This table presents results from a survey conducted as part of Banarjee, Duflo, Glennerster, and Kinnan’s randomized controlled trial of a group microlender’s program. At 6,827, the sample size is quite large, in fact much larger than is typically required to uncover small or medium-sized impacts (those of under 10 percentage points) in a randomized controlled trial. However, as noted in the text, the treatment differential between the program and control group is only about 9 percentage points: many in the program group unfortunately did not receive program services and many in the control group received similar services from alternative sources. Minimum detectable effect sizes are extremely sensitive to drops in the treatment differential; with only a 9 percentage-point differential a study cannot detect small or medium-sized impacts even with a sample size of over 6,800 people across participating neighborhoods and villages.

The table shows few program impacts on participants’ consumption levels. This could be because the program had no effects on certain measures, or it could be because the study did not have enough statistical power to determine whether the small estimated differences in consumption outcomes between the program and control groups were true effects. Those estimates had too much statistical uncertainty associated with them to conclude that the differences reflect true program effects. Nonetheless, it is possible to say with confidence that the program’s treatment differential of 9 percentage points did not have large impacts on the outcome measures presented here. Of course, if more of the program group had received the service and had been compared with a control group in which a much smaller proportion of the sample received a similar service, perhaps the impacts would have been larger. However, it is not possible to know that from this study.

When administrative records are available evaluators collect these and compare them with survey response data. Certain types of survey response data are unreliable because respondents may have trouble recalling required information, may have difficulty understanding the question, or may be ashamed or embarrassed to answer truthfully (or at all). Previous research comparing survey responses with administrative records and other data has shown that self-reported income can be wildly incorrect for these reasons. However, people generally respond truthfully and accurately to questions about consumption/expenditure. Therefore, when

(continued)

Box 4 (continued)

administrative records are not available and survey responses are the only source of outcome data, researchers may prefer to focus on consumption rather than income measures when assessing impacts on material well-being. This study focused on impacts on residents of impoverished neighborhoods in India. The authors did not have access to good administrative data from the Indian government, and, given that the work activities typical of this population are unreported and informal, such data would be inadequate in any case. Furthermore, study participants might earn from numerous small and unstable sources, making it likely that failures of recall and calculation would render self-reported income figures unreliable. Therefore the researchers chose to focus survey questions on consumption rather than income and present impact estimates for these measures.

**Impacts on Consumption Measures at 15 to 18 Months After Random Assignment
(Survey Wave 1)**

Outcome	Program Group	Control Group	Difference (Impact)	P-value
<u>Monthly expenditure per capita (2007 Indian Rupees)</u>				
Total (sample size = 6,827)	1,429.3	1,419.2	10.1	0.79
Nondurable (sample size = 6,781)	1,298.2	1,304.8	-6.6	0.83
Temptation goods (sample size = 6,863)	75.2	83.9	-8.7 *	0.07
<u>Yearly household expenditure (2007 Indian Rupees)</u>				
Durable (sample size = 6,781)	7,763	6,609	1,154 *	0.09
Festivals (sample size = 6,827)	2,969	3,732	-763 *	0.09
Sample size (total = 6,864)	N/R	N/R		

SOURCE: MDRC presentation of data analysis results reported in Banarjee, Duflo, Glennerster, and Kinnan (2013).

NOTES: Sample sizes for specific outcomes may vary because of missing values. Breakdown of responses by study group is not reported in the original document.

Estimates were regression-adjusted using ordinary least squares, controlling for pre-random assignment characteristics of sample members. Results are weighted to account for oversampling of borrowers from Spandana (the microlender being studied).

Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to differences between the outcomes of the program group and control group. The p-value indicates the likelihood that the difference between the program group and control group arose by chance. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; and *** = 1 percent.

potential to recruit a sample large enough that even relatively modest effects would be likely to show up as statistically significant. The control group in such a randomized controlled trial would also be unlikely to receive an alternative group-based microlending service; few such opportunities exist where a study like this would be fielded.

Microinsurance

Microinsurance allows low-income individuals to manage risk by providing them protection against specific perils in exchange for regular payments. Adverse events — such as a sudden illness or severe weather that ruins crops — can be especially catastrophic for the poor. Traditionally, low-income individuals have had minimal access to mainstream insurance products due to cultural and economic barriers; instead, many have relied on informal, community-based insurance schemes. Microinsurance, whether provided by a community organization, a larger organization with an on-the-ground partner, or a for-profit insurance company, can protect low-income individuals against drought and flooding, and can protect their assets in health, life, and property.

Like other financial products, large numbers of potential clients will take up microinsurance only if they are properly educated about the product, and if the product is significantly better than the informal insurance mechanisms they have previously relied upon. Scholars have found that, where existing informal insurance schemes are strong, microinsurance is crowded out of the market, especially because products may be offered by companies with no previous history in the community, and because products may be difficult to understand.

Example: RedSol

Evidence quality: Preliminary

The effective provision of microinsurance is often contingent on the collaboration of large insurance agencies and local intermediaries. In southern Mexico, AMUCSS, a network of local credit unions, and Zurich Financial Services, an international insurance firm, have joined together to deliver life, crop, and other insurance products to low-income rural individuals. Insurance products are sold through AMUCSS' RedSol network of local microfinance banks and are backed by Zurich. AMUCSS's knowledge of the rural poor's financial needs, as well as the organization's existing presence on the ground, made it an ideal partner for a large firm with

the financial size needed to provide microinsurance. RedSol has now expanded to provide credit life insurance (for clients who have existing microloans) and remittance insurance.¹⁷

In contrast to scholarship on microlending, research on microinsurance is in its infancy. Although numerous studies of microinsurance have been conducted, the vast majority have been nonexperimental, and most have focused on health insurance, which is quite different in purpose and usage than insurance for crops or livestock. However there are a number of randomized controlled trials underway for which results are not yet available. The quality of available evidence on microinsurance should improve significantly over the next several years.

Connecting Individuals to Mainstream Financial Services

Some programs promote financial inclusion by providing a combination of financial education, credit counseling, and access to mainstream financial services. Successfully navigating the world of credit, debt, and banking requires familiarity with concepts, like APR (annual percentage rate), that may be difficult to understand. It also requires knowledge of the particulars of managing one's finances. For example, banks may require a minimum balance in a checking account, charge fees to speak to a teller, or automatically enroll customers in overdraft protection or other programs whose cost implications are not immediately apparent.¹⁸ Furthermore, individuals with low financial literacy may have difficulty filling out the paperwork to apply for loans or to set up direct deposit of paychecks to their bank accounts, a service employers may offer and one that can save an individual hundreds of dollars in check-cashing fees yearly. Behavioral economists have shown that people tend to ignore choices they do not fully understand rather than take efforts to learn more and thus make informed decisions, as discussed above.

Credit counseling includes efforts to help households struggling with debt (often at high interest rates) and to build or repair credit history, which will allow participants access to larger and lower-interest mainstream loans. Credit counselors help clients develop repayment plans and may work with creditors to forgive a portion of their debts (for example, lateness or delinquency fees). Some work to consolidate debt from multiple sources into a single loan with a lower interest rate. Decreasing monthly fees in this manner can obviously increase the

¹⁷Credit life insurance pays off debts in the case of incapacitation or death. It is mostly useful for individuals who have group-liability loans (or loan cosigners) and who want to protect those other people from further obligation to repay. "Remittances" are the sums immigrants send to their families in their home countries each month. Remittance insurance provides a lump sum payment to beneficiaries if an individual sending remittances dies or can no longer work.

¹⁸Automatic enrollment in overdraft protection was outlawed in the United States by the Dodd-Frank financial reform bill. Dodd-Frank Wall Street Reform and Consumer Protection Act (2010).

disposable income remaining to individuals after their bills have been paid, reducing the need for them to cut back on consumption.

Paying down arrears and regaining good standing with current creditors can also improve credit scores. However, credit counselors may inadvertently damage clients' credit scores by encouraging them to become debt-free and to avoid *any* borrowing in the future. Although bad debt certainly damages credit scores, individuals who are debt-free for extended periods have little "footprint" in the documentation used by credit rating agencies to generate credit scores. Thus, counterintuitively, someone who avoids debt may become categorized as a poorer credit risk than someone who, for example, maintains a running credit card balance of moderate size. More recently some credit counseling programs have begun to emphasize that just as it is possible to have too much of the wrong kind of debt, it is also possible to have too little of the right kind. Such programs generally work with clients toward access to standard credit cards and as steps toward this goal they may first provide microloans (as described above) and help obtaining secure credit cards.

Example: LISC's Credit-Building Approach

Evidence quality: Preliminary

LISC Financial Opportunity Centers provide low-income individuals with financial coaching in over 25 cities across the United States. The centers bundle coaching with employment services and assistance in applying for public benefits. They provide group financial education as well as one-on-one coaching to help clients resolve credit card debt, budget, and plan for the future. To help clients with no credit history, the centers work with them to obtain secure credit cards as a first step toward gaining access to mainstream credit. Finally, the centers provide free tax preparation services each spring. The financial coaching services offered at the centers are designed to be modular, such that they can be integrated into service visits at a variety of other program offices. For example, the program might be integrated into the itinerary for a visit to an unemployment, parole, or TANF office. The idea is, again, to make exposure the default with an option to opt out, as is recommended by behavioral economists.

There have not been any experimental studies of the LISC Credit-Building Approach, although the structure itself does not appear to pose peculiar challenges to experimental design.¹⁹ If the financial coaching services are indeed modular and easily integrated into the architecture of preexisting programs, researchers would then have to focus mainly on identifying interested service providers and working with them find a stage in their intake processes or

¹⁹As part of the Social Innovation Fund, LISC's Financial Opportunity Centers are currently being evaluated using matched comparisons, a quasi-experimental approach. No impact results are yet publicly available.

appointment routines for random assignment, baseline information collection, and informed consent procedures. For example, New York City’s Office of Financial Empowerment worked with the organizations providing Jobs-Plus to include a “financial inclusion” module in that program. Jobs-Plus is a program originally designed by MDRC, the U.S. Department of Housing and Urban Development, and the Rockefeller Foundation. It is intended to increase employment and earnings among residents of public housing developments. The program includes personalized job coaching, on-site and referral job services, rent-based work incentives, and a social capital component that involves neighbor-to-neighbor outreach and support. Based on evidence from an earlier, rigorous impact study of the program’s success in increasing residents’ earnings, New York City and San Antonio, Texas have replicated the model. In the New York City replication, the program now includes a financial inclusion component in which financial specialists provide one-on-one financial counseling to help participants make good financial decisions about banking, savings, credit, and debt, as they simultaneously work with employment specialists to try to improve their success in the labor market.

The LISC Credit-Building Approach and Jobs-Plus examples illustrate how a financial inclusion program can be embedded into existing employment programs for low-income populations. However, no impact data are available (and no impact study is planned at this time) on whether the addition of that component will improve participants’ financial outcomes, or even help the program have larger effects on participants’ employment outcomes.

Example: Community Trust Prospera

Evidence quality: Preliminary

Community Trust Prospera is a California credit union that also functions as a check casher. The organization seeks to meet low-income individuals reliant on alternative financial service providers where they are, providing them financial services in a comfortable and familiar environment while helping them transition into the financial mainstream. Community Trust Prospera branches are designed to look and feel like check cashers, but they also offer a full complement of credit-union services, from savings and checking accounts to home loans. Customers can cash checks, purchase money orders, and send money to relatives for less than they would pay at for-profit check cashers. Over time, customers can transition into opening checking or savings accounts, and eventually taking out loans. Community Trust Prospera’s model is an innovative and transparent alternative to check cashing that seeks to address the fundamental barriers that keep many low-income individuals reliant on alternative financial service providers and out of the financial mainstream.

Although this approach is innovative and may hold promise, no data currently exist on the performance of this program relative to any kind of counterfactual.

Budgeting Tools to Help People Understand and Manage Their Finances

Mobile technology has led to the creation of Web sites and apps that are designed primarily to help customers create monthly budgets to track their expenses and efficiently allocate funds, and to make such thinking habitual. While technologies that make budgeting simpler can be useful to people of any socioeconomic background, they may be particularly helpful for low-income individuals who cannot afford to be imprecise in estimating costs; a single dollar error can lead to a bounced check (and its associated fees) or even the loss of heat or electricity for the month due to an unpaid utility bill. Furthermore, research suggests that the stresses of poverty can diminish people's ability to make decisions and render mistakes more likely. Web sites and apps that collect an individual's financial information in one user-friendly interface, and that are available on the go, can help low-income individuals take control of their finances.

Example: Mint.com

Evidence quality: Preliminary

Mint.com is a free, secure Web site and app that allows users to see all of their financial information in one place, create budgets, and save for goals. Financial information is pulled directly from a user's banks and lenders and is presented in an easy-to-understand, visually pleasing way. Mint allows a user to establish a budget and then categorizes all of the user's purchases made with debit and credit cards, tracking these against that budget. Mint applies various behavioral economics principles by providing an alert system for low account balances, bills due, etc.; visually tracking progress toward the user's savings goals; and visually representing how today's spending choices will influence account balances down the road.

Mint.com is an example of a program that is very difficult to study experimentally. Because it is a free Web page that is open to all, it would be impossible to prevent control group members from using it. Therefore, there might be little treatment differential between program and control groups. As noted above, if too few members of the program group receive a service and too many members of the control group gain access to it (or similar services), it can become difficult to detect the effects of the program on outcomes.

At the same time, one could imagine rigorously testing the effectiveness of individual pieces of a program such as this one. For example, it would be possible to use randomized controlled trials to test how different page layouts, instructions, and options presented on the Web site affect the experiences and behaviors of the Web site's users. Indeed, companies like Google conduct thousands of small randomized controlled trials every year by randomizing the IP addresses of those loading their sites to see different presentations of information, and then tracking usage. This can lead companies to alter their sites to promote ease of use and to change their home page "pitches" to encourage new visitors to sign up.

Conclusion

Overall, the evidence on the effectiveness of the various programs promoting financial inclusion is quite limited. Only a relatively small number of interventions have been subjected to the most rigorous types of impact evaluation, and many models or approaches have not been evaluated at all. Given the contexts in which these programs are operating, it may of course be challenging to design and carry out credible studies. However, some of the work mentioned above shows that it is feasible to conduct rigorous experimental and quasi-experimental evaluations in many circumstances, and the field would benefit from more such studies.

This review has focused primarily on understanding the *effectiveness*, or impacts, of financial inclusion programs and policies. However, as discussed in the Appendix, a comprehensive evaluation would also include an analysis to understand the implementation and operation of the program and a cost-benefit analysis. Efforts to build the scope and quality of evidence of financial inclusion interventions would benefit from stronger analyses of those types as well. Benefit-cost analyses are especially rare in this field.

For many organizations that deliver services to improve the lives of low-income populations, participation in rigorous evaluations, especially randomized controlled trials, would impose extra administrative burdens. In addition, individual-level randomized controlled trials (as opposed to randomized controlled trials of groups or even whole villages) often put organizations in the uncomfortable position of having to deny services to some applicants in order to form a control group. Yet this reticence can often be overcome, especially when it is explained to a staff that the organization may already turn away, or not recruit, other potential participants because of limited capacity. In other words, the organization may not be able to serve everyone who qualifies for its services anyway, and random assignment can offer a fair way of allocating scarce slots or resources.

For decades, countless nonprofits and other organizations have been willing to take part in rigorous evaluation studies, and hundreds of thousands of individuals and families have been willing to participate as well. (Institutional review boards help ensure that such studies are done ethically, and that human subjects are protected.) Understandably, securing the cooperation of service organizations typically requires financial support to cover the added staff time usually needed for evaluations. And special efforts must be taken to minimize the extent to which research activities, such as the process of enrolling participants into a research sample, disrupt an organization's normal flow of work. But these are usually manageable problems, and many service organizations welcome the opportunity to be part of important studies that aim to test innovative ways of trying to make a difference in people's lives.

Appendix A

**Program Evaluation: How to Tell Whether
a Program Is Effective**

An ideal evaluation has three major research elements:

1. An **implementation and process analysis** examines how the program model functions in the “real world,” what factors make it easier or harder to operate it well, how it is viewed and experienced by the people it intends to serve, and the extent to which those participants actually receive the type and amount of services called for by the model.
2. An **impact analysis** determines whether access to the program did, in fact, improve the outcomes of clients (for example, their financial hardship, monthly incomes, and debt levels) in the expected ways. This analysis attempts to determine whether any *change* in outcomes was *caused* by the program and not by other possible factors. It thus addresses what is typically meant by a program’s “effectiveness.”
3. A **benefit-cost analysis** compares the benefits of the program with its costs. It measures, separately, the overall economic gain or loss attributable to the program for those who participate in it, and for the government or other funders that pay for it. The results for funders can also be represented in terms of a “return on investment.”

The remainder of this section focuses on impact analysis and the challenge of identifying an appropriate “counterfactual” (or what is more generally referred to as a “control group”) with which the outcomes of the intervention group should be compared.

Perhaps the most crucial and difficult challenge in evaluating a program’s effectiveness is finding an appropriate counterfactual. The counterfactual indicates what *would have happened* to the program participants *in the absence of the program*. It is important to establish this in order to determine whether the program has actually had its intended effect. It is possible that any gains witnessed over time among program participants are actually the result of the characteristics of the participants or other factors rather than the result of exposure to the program itself. For example, would the person who borrowed from a microcredit organization and increased her income have done just as well (or even better) if she had *not* had the opportunity to obtain that loan? Would she have achieved the same improvement in income by taking a loan from her family, or would she have found another way to increase her earnings, such as through alternative employment? How can we know?

Researchers have developed a variety of techniques to construct a counterfactual. The most effective — in fact, the standard against which all others are judged — is random assignment. In its simplest form, a random assignment study divides a group of potential participants in a program into two groups on the basis of chance:

1. A **program group** that will go on to become participants in the program being studied. Its members will receive all the services outlined in the program's theory of change.
2. A **control group** that will not participate in the program or receive any program services.

Who within an eligible target group ends up in the program or control group is determined by the equivalent of a coin flip. In other words, people with particular traits have no higher or lower chance of being selected for the program than people without those characteristics. With a large enough sample, random assignment ensures that the distribution of characteristics of people in the program group will be the same as the distribution of those characteristics in the control group — just as with enough flips of a coin, the proportion of flips resulting in “heads” will be the same as the proportion resulting in “tails.” Thus, both the program and control groups should have the same proportion of people with traits that may be related to the outcomes of interest, such as the same proportion of men and women, the same average age, the same range of educational backgrounds, and even traits that are difficult to measure, such as motivation, an entrepreneurial personality, intelligence, and access to a supportive social network. Consequently, if the program group has better outcomes, it could not be because it had certain advantages to begin with (for example, better traits). Instead those better outcomes can be confidently attributed to the intervention itself.

A study that employs random assignment to test the effectiveness of a program is referred to as a “randomized controlled trial.” Because a properly executed randomized controlled trial leaves little doubt as to the effectiveness of a program for the participants, researchers engaged in program evaluation strongly prefer this option when designing a study.

Unfortunately, there are times when a randomized controlled trial is not feasible or ethical given the nature of the program being studied. In these instances researchers look to a variety of second-best options that belong to a class of analysis referred to as “quasi-experimental.” If certain conditions are met, a quasi-experimental analysis can produce “impact estimates” (measures of the program's outcome success) that are almost as good as those produced by a random assignment design. The two most reliable are “regression discontinuity” and “interrupted time series.” Common, but less reliable means of quasi-experimental evaluation include comparison group designs, including those that try to improve the baseline comparability of groups through a method called “propensity score matching.” The benefit of these approaches is that they do not require denial of services to those who would otherwise be eligible. The drawback is that the results are not always as definitive as those that can be derived from a randomized controlled trial. The conditions necessary for quasi-experimental methods to

produce reliable estimates are often not present, and whether a particular study has met these conditions is commonly subject to debate.

If a study cannot be designed to meet the conditions necessary for quasi-experimental analysis, any statistical investigation of the program is considered “nonexperimental.” Most such studies involve the use of statistical techniques based on multiple regression to try to control for the potential influence of preexisting differences in traits between participants and nonparticipants. While certain traits, such as gender and age, are easy to measure and control for statistically, other important traits (such as motivation, intelligence, or access to supportive social networks) are not. Suppose, for example, people who chose to enroll in a special savings program came out with more savings than others who chose not to enroll in that program. Furthermore, suppose that a statistical analysis of the program controlled for differences between participants and nonparticipants in terms of their demographic characteristics. It would still be possible that the people who enrolled in the program group were more motivated to save, or were in circumstances that made it easier for them to save — conditions that would not be easy to measure and control for statistically. Thus, if the program group saved more, it would be impossible to know whether this was truly the result of the program. Researchers refer to this problem as “selection bias.” Evaluations relying on nonexperimental methods may, because of this bias, produce misleading results.

Finally, some evaluations assess programs solely on the basis of the outcomes of those who participated. For example, a study might show that in one program, a high proportion of people — say, 75 percent — saved money, whereas in another program, a low proportion — say 30 percent — saved. Although it is tempting to conclude that the first program was more successful than the second, it is impossible to know for sure based solely on these outcomes. In fact, exactly the opposite may be true. It could be that the first program enrolled people who, for various reasons, were likely to save anyway. The second program might have enrolled people who faced more obstacles to saving. Perhaps without the program even fewer of them would have saved (for example, only 15 percent). If that were true, the second program would really be the more effective one, because it produced outcomes that were higher than what those outcomes would have been otherwise. Without a good “counterfactual” for each program, it is impossible to judge whether either program was “successful” in the sense of improving the rate of savings. In other words, it is impossible to know whether the program “added value.”¹

¹Although simple outcome tracking studies are inadequate as evidence of program impact, they may be useful as measures of compliance, uptake, and interest, and as evidence that programs have been successfully launched and implemented.

The SIF Evaluation Framework: Ranking Evidence Quality

In a time when it has become fashionable to talk about “evidence-based policy,” it can be very difficult for funders and policymakers to assess the quality of the evidence being presented. It is therefore helpful to have a framework for classifying different types of evidence by quality. One helpful example of such a framework comes from the federal Social Innovation Fund (SIF). The SIF is a strategy by which the Obama administration is attempting to use evidence to promote investment in and expansion of *effective* social programs operated by nonprofits. It offers funding to replicate and expand programs that have strong evidence behind them, but also offers funding to build stronger evidence for promising strategies that have not yet been proven to be effective. To guide decisions, the designers of the SIF constructed a three-tiered evaluation framework that could be used to classify programs according to the quality of evidence provided by studies that have examined them (if any). The findings of well-implemented randomized controlled trials and quasi-experimental designs are ranked highest.

This framework can also be useful for classifying the quality of evidence regarding financial inclusion programs. It distinguishes among three levels of evidence:²

1. **Strong evidence** means evidence from studies whose designs can support causal conclusions, and studies that in total include enough of the range of participants and settings to support scaling up to the state, regional, or national level. The following are examples of strong evidence: (1) More than one well-designed and well-implemented experimental study or well-designed and well-implemented quasi-experimental study that supports the effectiveness of the practice, strategy, or program; or (2) one large, well-designed and well-implemented randomized controlled multisite trial that supports the effectiveness of the practice, strategy, or program.
2. **Moderate evidence** means evidence from studies whose designs can support causal conclusions, but have limited generalizability. The following are examples of studies that could produce moderate evidence: (1) At least one well-designed and well-implemented experimental or quasi-experimental study supporting the effectiveness of the practice, strategy, or program, with a small sample size or other conditions of implementation or analysis that limit generalizability; (2) at least one well-designed and well-implemented experimental or quasi-experimental study that does not demonstrate equivalence between the intervention and compari-

²Corporation for National and Community Service (2014).

son groups at program entry, but that has no other major flaws; or (3) correlational research with strong statistical controls for selection bias and for discerning the influence of internal factors.

3. **Preliminary evidence** means evidence from studies that is based on a reasonable hypothesis supported by research findings. Thus, research that has yielded promising results for either the program or a similar program will constitute preliminary evidence and will meet CNCS's criteria. Examples of research that meet the standards include: (1) outcome studies that track program participants through a service pipeline and measure participants' responses at the end of the program; and (2) pre- and post-test research that determines whether participants have improved on an outcome of interest.

References

- Azurdia, Gilda, Stephen Freedman, Gayle Hamilton, and Caroline Schultz. 2013. *Encouraging Savings for Low- and Moderate-Income Individuals: Preliminary Implementation Findings from the SaveUSA Evaluation*. New York: MDRC.
- Corporation for National and Community Service. "Evidence & Evaluation." Web site: www.nationalservice.gov/programs/social-innovation-fund. Accessed March 6, 2014.
- De Soto, Hernando. 2000. *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. New York: Basic Books.
- Dodd-Frank Wall Street Reform and Consumer Protection Act. 2010. Pub.L. 111-203, 124 Stat. 1376-2223, H.R. 4173.
- Duflo, Esther, Abhijit Banerjee, Rachel Glennerster, and Cynthia G. Kinnan. 2013. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." NBER Working Paper 18950. Cambridge, MA: National Bureau of Economic Research.
- Kast, Felipe, and Dina Pomeranz. 2013. "Do Savings Constraints Lead to Indebtedness? Experimental Evidence from Access to Formal Savings Accounts in Chile." Harvard Business School Working Paper 14-001. Cambridge, MA: Harvard Business School.
- Mills, Gregory, Rhiannon Patterson, Larry Orr, and Donna DeMarco. 2004. *Evaluation of the American Dream Demonstration: Final Evaluation Report*. Cambridge, MA: Abt Associates, Inc.
- Sherraden, Michael. 1991. *Assets and the Poor: A New American Welfare Policy*. Armonk, NY: M.E. Sharpe Inc.

