

# Randomized Controlled Trials and Financial Capability

## Why, When, and How

Brett Theodos, Margaret Simms, Rachel Brash, Claudia Sharygin, and Dina Emam

Metropolitan Housing and Communities Policy Center Brief

June 2014



In the current economic climate, it is more important than ever that low-income families have the ability to make sound, informed financial decisions. Poor financial decisions can significantly affect long-term stability for individuals and families. Therefore, helping consumers develop healthy financial habits is critical. While methods to help consumers make better financial decisions are clearly needed, little formal evidence demonstrates which educational approaches, supportive services, or technological strategies best improve financial outcomes for these households. The evidence base in this area is growing, but rigorous empirical support remains scant relative to other fields. Service providers, financial institutions, policymakers, and funders are unsure if the programs and services they offer give consumers the skills they need.

Policy researchers agree that randomized controlled trials (RCTs) often produce the highest standard of evidence about an intervention's effectiveness. Given the promise that RCTs hold, but also the difficulty of successfully implementing these studies, the Urban Institute and the Consumer Financial Protection Bureau (CFPB) convened a roundtable on the benefits and challenges of RCTs in April 2013. In addition to Urban Institute and CFPB staff, 26 evaluators, funders, and program staff—all of whom were involved in ongoing or recent evaluations of financial literacy, service, or education programs—discussed their experiences and shared practical insights into successful strategies and pitfalls of rigorous evaluations of financial capability interventions. This brief shares those insights with other researchers, practitioners, and funders undertaking or contemplating rigorous effectiveness research in the financial capability realm.

## Why Conduct Financial Capability Research?

Financial capability programs have proliferated in recent years, but research into which programs or methods are actually effective has not kept pace. Researchers have conducted very few rigorous studies in this field, and practitioners, policymakers, and public and private funders are increasingly calling for research to help identify programs and methods that improve consumers' financial decisions.

The CFPB is just one federal agency intent on determining “what works” for improving financial capability and decisionmaking. The Government Accountability Office prioritized research that evaluates the relative effectiveness of different initiatives and approaches in its survey of federal efforts to promote financial literacy (Cackley 2012). The Treasury Department's Financial Literacy Education Commission also puts effectiveness at the top of its research agenda and has supplemented its [mymoney.gov](http://mymoney.gov) website with an online clearinghouse of evidence-based research and evaluation studies, innovative approaches to financial literacy, and general best practices.<sup>1</sup>

The Urban Institute is itself engaged in such research. CFPB has contracted with the Urban Institute to rigorously and quantitatively evaluate two financial coaching programs in order to better understand which interventions can improve consumers' financial decisionmaking skills. The study focuses on how the two interventions—one offered by Miami's Branches, the other by New York City's Financial Clinic—improve financial behavior and knowledge and reduce financial stress among program participants.

## What Is a Randomized Controlled Trial?

Randomized controlled trials select units (individuals, schools, neighborhoods, etc.) from the same population and randomly assign them to one of at least two groups: treatment (sometimes called experimental) or control. Such a process helps make the treatment and control groups equivalent—in motivation, ability, knowledge, socioeconomic and demographic characteristics, and so on—at the start of the study. If all goes well, any differences in outcomes between the treatment and control groups observed after the intervention can be attributed to it. In other words, the control group is the counterfactual that helps observers understand what would have happened to the treatment group were it not for the intervention.<sup>2</sup>

RCTs first took hold as a research strategy in medical research, with the experimentation typically occurring in controlled laboratories. The realities of social science often deviate from these ideal conditions. For example, many RCTs have difficulty with recruitment or attrition of study subjects. In fact, the only threat that RCTs control for is selection bias. Further, RCTs are inappropriate in certain scenarios, such as where it is infeasible or unethical to deny treatment to some individuals, as is the case in many place-based interventions.<sup>3</sup> Accurately capturing the impact of a particular intervention (“internal validity”) does not ensure that the findings are generalizable to other people, programs, geographies, or time spans (“external validity”). Researchers can consider less-expensive and less-challenging research methods, such as a pre-

post or quasi-experimental study design. But while RCTs often fall short of their unfortunate “gold standard” moniker, they are often the best, and sometimes the only, way to draw conclusions about an intervention’s causal impact.

## When Is a Randomized Controlled Trial Suitable?

While the number of financial capability programs may be on the rise, not all programs are appropriate subjects for rigorous evaluation. Program evaluation—and RCT evaluations in particular—can be time-intensive and expensive, and many programs may not have the operational capacity or client volume to justify participation in such a study. Even well-established programs with large client bases may find it difficult to participate in an evaluation without external support. Funding supporting these evaluations is limited, which puts more pressure on the studies that are conducted to be planned and executed well.

Participants in the convening agreed that, where feasible, the RCT is the preferred way to measure a program’s effectiveness. The participants identified key criteria for determining whether to go ahead with a potential evaluation. We view the first four criteria as factors that cannot easily be addressed in the short run. The final two criteria are likely more amenable to change.

- **Applicability, scalability, and replicability.** Because RCT studies have high opportunity and actual costs, research funding should be directed toward studies whose findings will be widely relevant to the field or applicable to other sites. Ideally the interventions that are the subject of the studies should be scalable—that is, made available on a wider basis—and replicable.
- **Organizational capacity and scale.** The program should be well implemented and managed and have the operational capacity to help implement the study. Further, the program should be large enough (in number of individuals enrolled in the study and treated)—or be capable of growing large enough—to support an evaluation.
- **Program stability.** Program design can shift in response to staffing changes or funding mandates. It is usually important that programs demonstrate a track record of stable service delivery, preferably for multiple years. An established track record is less critical for interventions that vary a discrete component of a program, or if the intervention is less complex or less reliant on people for the delivery of services, such as an RCT involving a new application of technology.
- **Adequate funding for program operations and evaluation.** Programs should have sufficient resources to implement the model being tested in the research design. This may require additional funding to increase the number of individuals served or gathering and entering data.
- **Buy-in and research planning participation from frontline staff.** Frontline staff who will need to be involved in study implementation day to day need to be fully invested in the study and involved in designing implementation and data-collection strategies. This includes the staff responsible for recruiting study participants, assigning participants to treatment and control groups, providing the financial intervention, and managing data. As one researcher said, “I won’t do a research project with a program without the IT person signing off.”
- **Close working relationship between evaluators and providers.** Both parties should see the evaluation as a partnership, have an equal commitment to the fidelity and success of the study, and be willing to listen and contribute to making it work. Both sides should expect a heavy dose of up-front talks during research design and early implementation, as well as frequent ongoing communications once the study is up and running.

## What Hinders a Strong Study? How Can These Challenges Be Overcome?

Researchers face challenges in designing and executing RCT evaluations of financial capability programs. The convening participants identified common obstacles in research design, research implementation, and data collection and analysis. They also offered ideas for overcoming these challenges.

### Research Design

The first set of obstacles researchers and programs face relate to research design. Can the study support an RCT design? What should target enrollment be? How will study participants be randomized? Do new partners need to come aboard for the study to be viable? For questions such as these, the attendees of the convening identified nine areas that require special attention from researchers and program partners during the study design phase.

- **Scope of evaluation.** One major challenge in research design is determining the appropriate scope of the evaluation—for example, how many outcomes to measure and how long to follow study participants. Certain outcomes, such as building up savings or improving credit scores, may take months or years to develop or may be costly and difficult to measure. Researchers, practitioners, and funders should realistically approach the tradeoff between the importance of obtaining data on numerous financial behavioral outcomes versus the feasibility of observing change in these outcomes within the scope of the study.
- **Limited sample size.** The biggest challenge facing most RCTs is generating a large enough sample of study participants to measure program effects with statistical precision. Limited resources, difficulty recruiting participants, and participants dropping out before they attend sessions or complete all surveys all hamper evaluators' efforts to reach target sample sizes. A limited sample size may require a research design focused on one or two central outcomes. Other outcomes might be considered or explored but not rigorously tested owing to limited capacity to examine impacts on different subgroups. Many researchers hope that the financial capability field will identify one or two “priority” outcomes that measure basic financial capability across a broad range of populations, contexts, and interventions—analogue to measuring patients' blood pressure and body mass index in health research.
- **Small, difficult-to-detect changes in behavior.** Often, the expected changes in behavioral outcomes are small, making them even more difficult to measure. For this reason, it may be prudent to focus on behaviors where single decisions or changes at the margin can have substantial long-term effects. Examples of such behaviors are yes-no decisions, such as whether a study participant established a savings account or an automatic bill payment, or threshold outcomes, such as whether a study participant was able to qualify for a standard credit card (to get away from predatory lending) or a mortgage refinance (out of a subprime loan).
- **Discomfort with denial or delay in service due to randomization.** Funders and service providers sometimes balk at the idea of delaying or denying service to control participants. To allay this discomfort, the RCT might be conducted as a study of a pilot program or a new service. Another option is to perform random assignment before individuals are even offered the service. This approach may work best in a setting where individuals show up for other services—such as a workplace program, tax-filing assistance, or applying for a loan—and the treatment is framed as an additional benefit.
- **Program and study enrollment selection effects.** Certain kinds of interventions rely on individuals to opt in to the treatment. In these cases, people who choose to participate in an intervention are likely more concerned about and motivated to address their financial behavior than the general population. Researchers must be cautious in drawing conclusions about the general population from these results.
- **Point of randomization.** Researchers acknowledged that in some circumstances, randomly assigning individuals to treatment and control groups as they enroll, the “flip a coin” ideal, may be infeasible. Some alternatives to this design that preserve the RCT approach include randomly assigning similar groups of individuals, or enrolling participants to the treatment group on a first-come, first-served basis and using the waiting list as a control group. Of course, an RCT need not compare an all-or-nothing dichotomy; it is possible to compare interventions with different dosages, designs, or durations and not fully exclude any individuals seeking service.
- **Fidelity to model.** Many agreed that once a suitable program is chosen, “the less you can mess with standard business practice, the better,” as voiced by one convening attendee. A study will be more useful if it evaluates what a program does during its normal course of business. It is also more likely to be faithfully implemented if it places fewer new requirements on program staff.
- **Input from program partners.** Researchers need to seek staff input on all areas of study design, including determining the study timeline, engaging with potential study participants, measuring changes in financial decisionmaking among their target population, survey design, and data analysis strategies. Staff working directly with clients often provide much needed reality checks for the research design: will these plans actually work as intended, or will some idiosyncratic factor unexpected to researchers but well known among staff on the ground suggest a change in plans? For example, Target gift cards will not be useful in an area with no Target stores; proposed survey question wording may not be familiar to study participants. Frontline staff knows the study participants and program operations best, and it is important to use this knowledge.
- **Process or implementation studies.** Convening participants agreed that impact evaluations can be strengthened by including an implementation study, a descriptive account of the organization's goals, strategies, methods, and activities combined with the researchers' account of the rollout of the RCT evaluation. Process studies provide continuous reporting on the evaluation's progress and help put the eventual study results in context.

## Research Implementation

Even the most elegantly designed RCT can encounter difficulties when study implementation begins. Practitioners need the most support when transitioning from planning the study to starting the study. Participants had the following insights for moving smoothly from design to implementation.

- **Prepare, prepare, prepare.** The move to implementation should include some preparatory steps if possible. These include field-testing surveys and doing practice runs of recruitment pitches, study consent procedures, randomization tools, and follow-up approaches. Some organizations need to hire additional service providers to handle the increased number of clients resulting from the study. Staff training and oversight, specifically for the study, are important at the outset and for continuing buy-in as the study gets under way. At every point along the way, trainers and evaluators should make sure that staff's questions are answered quickly and easily.
- **Enlist a research coordinator.** RCT studies can greatly benefit from a dedicated research coordinator whose primary responsibility is to manage the evaluation on site at the evaluated organization. The additional work associated with a research study may be too much for program staff to handle alone, and having a research coordinator who can devote significant time to this effort may be the difference between a well-executed study and a poorly executed one. The coordinator might be employed by either the service provider or the evaluators. Making the coordinator a part of the organization's staff (or choosing an organization that already has a coordinator on staff) has some advantages. A permanent research coordinator gains the experience and authority within the organization to build the capacity to participate in rigorous evaluations in the long term. Alternatively, a research coordinator directly employed by the research organization—potentially, a coordinator with greater experience in RCT design and implementation—may be beneficial. Regardless of who employs the research coordinator, having someone in this capacity means he or she is better able to assist the practitioners as they navigate everyday challenges, particularly difficult tasks like denying or delaying service to needy participants or asking sensitive and potentially upsetting survey questions.

### Research coordinators can...

- understand both the program's need to keep its primary focus on providing services, and the research staff's requirements to push the evaluation forward and maintain fidelity to the research design;
  - translate the relative importance of each side's priorities;
  - mediate between the program staff and researchers when the requirements of each may conflict;
  - minimize the evaluation's disruptive effects on program operations;
  - provide oversight and quality control for research implementation and fidelity to evaluation design; and
  - improve the program's ability to internally track its performance.
- **Align data collection efforts.** In most studies, evaluators are responsible for collecting the data that will be used in the analysis. Evaluators should coordinate with organization staff to minimize duplicative data collection and participant contact. Automated randomization and data-collection tools that can synch easily with the organization's existing tracking system reduce data-entry burden, ensure fidelity to the research design, and alleviate the stress of randomization.
  - **Expect bumps along the way.** Even with ample preparation, it is entirely normal for things to go wrong. Both evaluators and program staff need to be prepared for pitfalls like lower-than-expected rates of consent into the study, difficulty tracking down participants to receive services, important survey questions being misunderstood or skipped entirely, and technological problems. Good communication between researchers and practitioners is key to moving past these problems. Problems at start-up are less likely to derail an evaluation as long as they are documented and dealt with as soon as they arise.

## Data Collection and Analysis

Convening participants recommended that data analysis—the capstone of the research—be sufficiently incorporated into the planning and research design stages and expand as study data become available. Participants offered the following advice on the use of data, from the early stages of a study to the final analysis:

- **Look at data early and often.** Researchers first use program data to determine whether the program is a good candidate for evaluation, and to design their own data-collection tools to complement the programs' existing intake forms and tracking instruments. Study coordinators can also use participants' data to troubleshoot randomization, data collection, and data entry. Initial responses to financial knowledge and behavior questions can also highlight areas where the follow-up survey may need to be revised.

- **Don't undervalue intermediate results.** Program and research funders are also interested in short-term and intermediate study results to keep the conversation active around ongoing evaluation projects and to help fulfill responsibilities to their stakeholders. Research designs that incorporate theories of change for improving financial decisionmaking and logic models for the impact of program interventions allow evaluators to connect early data on potential long-term outcomes.
- **Choose data sources deliberately.** Administrative data collected by third-party organizations, and other forms of what one funder referred to as “naturally renewable data sources,” are increasingly useful to evaluations and to financial capability training more generally. For example, it is becoming standard practice for financial capability programs to review individuals’ credit reports as a part of understanding their financial history and how their behaviors affect their ability to obtain credit. Programs or interventions operated by financial institutions or third-party personal financial management firms frequently download information from participants’ financial accounts. While some financial outcomes from administrative data—such as credit scores—are potential candidates for the “priority outcomes” tracked in all financial capability evaluations, researchers suggested caution in prioritizing measures that are convenient to collect but that might not reflect appropriate goals or likely outcomes for all participants.
- **Expect realistic effects.** Researchers, practitioners, policymakers, and funders stressed the importance of setting expectations for the evaluation’s results, both at the outset and during data analysis. In many cases, the results of an intervention may be statistically significant but small. To avoid mischaracterization of the study’s conclusions, all parties should give significant thought to how the results will be framed.
- **Consult practitioners on interpretation of findings.** Researchers should gather input from practitioners and policymakers, including ones not directly involved with the study, to help the researchers understand the meaning and implication of the findings. In some cases, a “negative” result might mask a positive outcome. In one example cited at the convening, researchers found that debt levels increased among program participants, which initially seemed like a failure to improve financial behavior. When the researchers discussed the results with the program staff, they realized that the result was driven by low-income participants who now had access to credit, which allowed them to smooth financial disruptions by borrowing. Looking at these individuals’ credit reports, researchers saw that they were using credit responsibly. Flexibility, thoughtfulness, and communication are key to interpreting findings.

## How Does This Discussion Shape Practice, Policy, and Funding?

Researchers, practitioners, policymakers, and funders can take several concrete steps to expand rigorous, data-driven, experimental research on financial capability programs. A number of key themes emerged from conversations during the convening about how these different groups can prime the field for future evaluations in this space.

Two key steps for practitioners:

- Encourage a “culture of data” in their operations, and prioritize tracking participants and collecting as much information as possible on their activities and outcomes.
- Use data to isolate key impact metrics, those indicators that programs feel they are able to shift the needle on.

Programs that build this internal data collection and analysis capacity will be able to document their efforts, refine their approaches, and communicate the value of their work to external stakeholders without a formal, external evaluation.

Three key steps for evaluators, together with funders and policymakers:

- Develop strategies and systems to make data collection easier, faster, less expensive, and more efficient. Both researchers and practitioners will benefit from standardized, affordable, and easy-to-use data collection and management programs.
- Work with programs to develop “priority outcomes,” accepted financial capability outcomes shared across evaluation studies.
- Design standardized data-collection efforts around the priority outcomes, once established, and ensure that these outcomes are included in subsequent evaluations.

In addition to concrete actions described above, these groups need to be supported in critical ways. Practitioners need to be supported financially in their efforts to expand their data collection and analysis capacity. Something of a chicken-and-egg problem exists: Programs need funding to set up data-collection systems but have trouble fundraising without data-driven evidence of their program’s impact. Reducing the expense and complexity of installing these systems will help address this issue. Access to a set of priority outcomes, accepted by researchers, practitioners, and funders, can motivate funders to help programs adopt these practices. CFPB and others are currently engaged in this area.

Policymakers are interested in the potential benefits of integrating financial capability into other programs and services. Some attendees suggested supporting return-on-investment studies into the costs and benefits of such integration. Others were interested in seeing how the programs' impact varies by individuals' initial behavior and attitude, in addition to their demographic characteristics.

Finally, all participants emphasized that external, rigorous evaluations are expensive and time-consuming. It is important, therefore, to choose evaluations strategically, focusing on those with the most potential to further knowledge in the field.

## Notes

1. See <http://www.mymoney.gov/> and "Financial Literacy and Education Commission," US Department of the Treasury, <http://www.treasury.gov/resource-center/financial-education/Pages/commission-index.aspx>.
2. For a more thorough explanation of RCTs see Wholey, Hatry, and Newcomer (2010) or Shadish, Cook, and Campbell (2002).
3. See Barnow (2010).

## References

- Barnow, Burt S. 2010. "Setting Up Social Experiments: The Good, the Bad, and the Ugly." *Journal for Labour Market Research* 43 (2): 91–105.
- Cackley, Alice Puente. 2012. "Financial Literacy: Enhancing the Effectiveness of the Federal Government's Role." GAO-12-636T. Washington, DC: Government Accountability Office.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. 2nd ed. Stamford, CT: Cengage Learning.
- Wholey, Joseph S., Harry P. Hatry, and Kathryn E. Newcomer, eds. 2010. *Handbook of Practical Program Evaluation*. 3rd ed. San Francisco: Jossey-Bass.

## Roundtable Participants

Xiaoling Ang, Consumer Financial Protection Bureau  
Sarah Bainton, Consumer Financial Protection Bureau  
Rachel Brash, Urban Institute  
Camille Busette, Consumer Financial Protection Bureau  
Haidee Cabusora, The Financial Clinic  
Anna Cash, Innovations for Poverty Action  
Dina Emam, Urban Institute  
Lars Gilberts, Branches  
Nathalie Gons, New York City Office of Financial Empowerment  
Sarah Gordon, Center for Financial Services Innovation  
Mae Watson Grote, The Financial Clinic  
Cindy Guy, Annie E. Casey Foundation  
Leena Im, Consumer Financial Protection Bureau  
Julian Jamison, Consumer Financial Protection Bureau  
Michelle Jones, CredAbility

Scott Karol, Clarifi  
Darren Liddell, Branches  
Jose Loya, Branches  
Genevieve Melford, Consumer Financial Protection Bureau  
Marina Myhre, US Department of Housing and Urban Development  
Gautam Nair, Innovations for Poverty Action  
Louisa Quittman, US Department of the Treasury  
Regina Salliey, Annie E. Casey Foundation  
Irene Skricki, Consumer Financial Protection Bureau  
Claudia Sharygin, Urban Institute  
Rebecca Smith, The Financial Clinic  
Margaret Simms, Urban Institute  
Jonathan Spader, Abt Associates  
Brett Theodos, Urban Institute

Copyright © June 2014. Urban Institute. Permission is granted for reproduction of this document, with attribution to the Urban Institute.

The roundtable was convened by the Urban Institute under contract number CFP-12-Z-00006 for the Consumer Financial Protection Bureau. The views expressed are those of the authors and do not necessarily reflect those of the Consumer Financial Protection Bureau or of the Urban Institute, its trustees, or its funders.