

# What We Don't Know That We Ought To

Remarks by Chris Dunford, Freedom from Hunger

Microfinance Impact and Innovation Conference

New York City October 22, 2010

Let's remind ourselves that evaluation is a larger field to which impact research makes distinctive contributions. In other words, impact research by itself is not sufficient for evaluation. Just consider the major questions about microfinance that have been in the news lately and continue to spread into public consciousness:

- Does microfinance have positive impact on the poor?

To answer this, we need to know definitions. What do you mean by microfinance? What do you consider to be positive impact? Who are these poor people you refer to? These definitional questions have to be answered by the people who will use the results of our impact research, and we have to design our research around their definitions.

Then there are value questions:

- Are interest rates too high? Are the poor getting over-indebted? Are IPOs good or bad?

They beg the Goldilocks judgments: too cold, too hot, or just right? Value judgments are made in the light of philosophical and practical considerations—again, by consumers of impact research results, not by researchers themselves.

So the larger field of evaluation is much more than impact research, it involves making a lot of definitions and value judgments. I suppose that's why it's called "evaluation." One of the main things we researchers ought to know, but usually don't, is exactly how the results of impact research relate to this larger evaluation activity. Put this another way: How do our results get used by decision-makers? Just as we study the psychology of economic decisions by the poor, we should be putting a good deal more effort into understanding the decision-making of policymakers, funders (both donors and investors), program designers and program managers, remembering that each is a distinctly different audience and user of impact research results.

You know that impact research results, especially from RCTs, don't figure into the decisions of these audiences much at all. We had better figure out why not.

Our narrative has been that high-quality research, mainly RCT designs, should be carried to test the merits of an innovation before it is rolled out to scale. That's the ideal that drives us. However, the reality is that we're chasing the innovation and rollout process as it storms forward with or without us. If you think the investors in rollout of mobile banking will wait patiently while we do a number of RCTs around the world to see if it's okay to go ahead, I suggest you wake up and answer your cell phone.

Why are we running to catch up?

An obvious answer is a combination of two facts on the ground: first, RCTs take a long time, are very expensive, and they produce results that are too often too difficult to interpret in practical terms AND second, decisionmakers at all levels are in a hurry and use whatever evidence seems to point in a clear direction—not necessarily due to bias, except a bias to action; they want a clear direction to go in—blissfully unaware of the flaws in the evidence or their interpretation of it.

So that's my first recommendation for more research—the psychology of decision-making by the intended users of impact research.

My second recommendation relates to my hypothesis that too often RCTs produce false negatives. Granted that's better than producing false positives, which affirm innovations that are actually worthless. However, because RCTs are increasingly carried out in the context of live program operations by independent researchers who are not part of the day-to-day operations, we have many situations where the RCT is testing quality of delivery rather than efficacy of design. Yesterday, Nathanael said that the benefit of learning from RCTs is that they cut through the noise of day-to-day operations. That's very true, but that is the ideal.

The problem is that the innovation to be tested, the “treatment” in an RCT, is often just that, an innovation. It's new to the MFI managers and staff and they are still struggling to get the delivery right, even as the RCT rolls forward—especially when we are looking at the impacts of a complex, human-mediated innovation, like an education intervention or building a new lending/borrowing relationship. For the researchers responding to the policymakers whose global question is “Does the innovation work?” the important difference is between the people who get the designed innovation and the people who don't. For the managers and staff, the lived reality is more complex.

Let me illustrate this problem with a hypothetical scenario rather than use a real example and embarrass anyone here in particular.

Imagine if health science had tested the efficacy of a measles vaccine only by comparing populations that had access to a vaccination campaign with populations that did not. And

imagine that the operational mechanics of vaccination campaigns were still in a process of trial-and-error development by the campaign managers during the comparison period. There is a good chance that the research results would not tell us whether the vaccine (the new design) actually works, because that test got mixed up with a test of a separate question, whether the quality of delivery of the innovation could be maintained during expansion to reach large numbers of people in a short period of time. It is a confusion of design and delivery in the same test.

So my second research recommendation is that we look at quality of delivery and how the heterogeneity of the treatment affects the heterogeneity of outcomes or impacts.

Because of the operational difficulties of doing impact research with an MFI for whom the treatment is an operational innovation, it seems that the main utility of RCTs, in reality, is when they are used to confirm that a well-established innovation, like microfinance (however defined!), does or does not merit all the investment of time and effort and careers put into its development and massive dissemination. And of course, such confirmation or refutation of this massive prior commitment to microfinance comes only from multiple RCTs in a variety of places with a variety of microfinance designs. My third recommendation then is that we push ahead with this tedious work, because it is so critical to getting a nuanced but satisfactory answer to the question “Does microfinance have positive impact for the poor?” We’ve still got years to go yet!

However, the other and perhaps more important value of the RCT approach is the opportunity to generate counter-intuitive results, the unexpected, the paradigm busters, the “Wow, I didn’t see that coming!” moments we all cherish and dread at the same time. For these purposes, I’ll venture to say that there are much less expensive and more timely ways to generate counter-intuitive results than the large samples and highly structured interviews of the classic full-blown RCT. If you think about it, wouldn’t you say that we of the microfinance world got a great many more “Wow, I didn’t see that coming!” moments when we read *Portfolios of the Poor* than by reading the recent RCT study reports? In citing financial diaries, I’m not minimizing the crucial importance of randomization between treatment and control to really attribute effects to causes, far from it, but maybe our questions of the treatment and control groups need to be more open, to allow for learning that was totally unexpected. A little less resistance to the value of open-ended questions, please! That’s my fourth and last recommendation.

Perhaps because of the frustrations of doing RCTs for policymakers and funders long after the innovation horses are out of the barn, there has arisen a new fashion for RCTs that focus on much smaller innovations (like lowering interest rates) within the bigger innovations (like microfinance), oriented to improving rather than just proving microfinance. We can see this new fashion coming into its own in this conference. I don’t see as much interest in testing the

value of a whole sector as in testing the value of tweaking the microfinance design this way and that. This is certainly much more helpful to practitioners than the more global research questions.

With that in mind, my research colleagues at Freedom from Hunger, Megan Gash and Bobbi Gray (who will moderate the next session), have compiled a relatively short list of practitioner-oriented research questions that beg for RCTs. However, I won't go through this list with you now. Instead, I ask those who are interested in such things to give me or Megan or Bobbi your card or email address and we will share the list with you and be happy to discuss it with you.

In closing, I'll recap the four recommendations for research I've made here:

1. We should understand the psychology of decision-making by the intended users of impact research and shape our conduct and communication of impact research accordingly.
2. We should pay closer attention to the quality of delivery and how the heterogeneity of the treatment affects the heterogeneity of outcomes or impacts.
3. We should push ahead the tedious work of multiple RCTs in a variety of places with a variety of microfinance designs to answer the "Does it work?" questions in terms of where and when and for whom it works.
4. And last, and perhaps the most controversial, we should master quantitative analysis of qualitative information, so that our questions of the treatment and control groups can be more open, to allow for learning that was totally unexpected.