

Stanford SOCIAL INNOVATION^{Review}

Features

Caught in a Fake Debate

By Annie Duflo & Heidi McAnnally-Linz

Stanford Social Innovation Review
Summer 2017

Copyright © 2017 by Leland Stanford Jr. University
All Rights Reserved



Finding solutions for the global poor need not require choosing between theory and decision making. Good evaluations focus on both.

BY ANNIE DUFLO & HEIDI MCANNALLY-LINZ

CAUGHT in a FAKE DEBATE

At Innovations for Poverty Action (IPA), a research and policy non-profit founded in 2002 dedicated to discovering and promoting effective solutions to the problems of global poverty, we have worked with more than 400 academics to carry out more than 650 rigorous evaluations around the world in countries from Ghana to the Philippines.

In pursuing these efforts, researchers and practitioners have worked closely together to identify, design, and rigorously evaluate solutions that are guided by theory and on-the-ground experience. These evaluations help inform both shorter-term, practical policymaking and longer-term understanding and decisions. Consequently, we find recent debates among evaluation experts about decision-based versus theory-based evaluations confounding.

According to one side of this debate, rigorous evaluations should focus more on helping decision makers with their time-sensitive decisions—such as how best to roll out a cash-transfer program. According to the other side, they should be designed with theory in mind, helping us understand how and why, for instance, cash transfers work, and if they can succeed in other contexts. We think that the two sides of this debate present a false dichotomy.

This distinction implies that theory-based evaluations are not decision-based, when in fact they often are. We have learned that good evaluations can, and often should, aim to inform both. If evaluators are interested only in advancing a theory, it does disservice to their partner. Focusing only on helping the particular program under evaluation means missing an opportunity to help improve the hundreds of other programs and organizations working on the same problem.

Certain types of decisions (especially the more immediate, practical ones about delivery) do not necessarily need to be answered with a theory-based evaluation. But it is always much more powerful when they are.

Take our 2010–2014 community health workers study in Zambia, led by Nava Ashraf, Oriana Bandiera, and Scott Lee. The evaluation compared two different recruitment strategies for community health workers: one that emphasized the opportunity to grow one's career and one that highlighted helping communities. The career-oriented messaging was found to attract workers who were more qualified and performed better on the job. This study is very much decision-based; the ministry of health used the findings to make a decision about how to recruit more effective workers. Yet the evaluation was also theory-based and helped us learn about the motivations of these

kinds of workers and the mechanisms through which one recruitment strategy might lead to better health outcomes than the other.

Not every evaluation will allow for immediate decision making. But if the primary purpose of evaluations is to help improve the lives of the poor, decision makers need to be able to decide not only what to do with existing programs and policies now, but also what to do over the long term. They need to know what new or innovative policies to adopt, what has worked elsewhere and in what circumstances, and how a certain type of intervention will work in another context.

THREE-TIERED PROCESS

We have built IPA's strategy both to help answer these questions and to guide practitioners in these kinds of decisions. But rather than pigeonholing evaluations into theory-based and decision-based categories, we prefer to think of our approach as a three-tiered process that reflects the kinds of decisions that need to be taken at different stages of a fully developed programmatic cycle.

1. Proof of concept. The majority of our evaluations are what we call “proof of concept” studies, meaning that they evaluate whether an idea is effective or not for the first time. These could involve experimenting with a completely new idea (such as an innovative savings product), evaluating an already well-established intervention for the first time (such as microcredit), or comparing different ways to deliver a program (such as price subsidies or recruitment strategies).

Often these studies are designed in partnership with researchers and practitioners, and help the partners make a decision about the program or policy they run or might run. For example, the community health worker study in Zambia is helping the government ministry more effectively recruit 5,000 community health workers.

Some of these studies were not initially conducted with an implementing partner and might be deemed only theory-based, but the concept was subsequently adopted by an implementing organization. For example, our 2009–2012 study led by economists Ernest Aryeetey, Isaac Osei-akoto, Dean Karlan, and Chris Udry on how to encourage farmers in Ghana to invest more in better seeds and tools compared two solutions: one that offered farmers cash directly and one that subsidized rainfall-index insurance to help them manage farming risks. We found that it was primarily risk, rather than lack of capital, that constrained their investment. Although we acted as the insurance agent for the study, the results persuaded the Ghanaian insurance industry to adopt the model.

In many cases, at the proof-of-concept stage, we have helped an organization make programmatic decisions. And in almost all

cases, we have also learned something about the effectiveness of the mechanism at work and about human behavior, and helped fill gaps in our knowledge—which, in turn, informs future decisions.

2. Adaptation of concept. This stage involves testing whether particular aspects of a program matter, such as where a particular mechanism is implemented, who runs it, what the different programmatic models are, what parts of the program are most cost-effective, and so on. When we adapt the concept to a different context, we learn more about the mechanism at work and are thereby able to generalize more.

This is where field replications become critical. These can range from simple adaptations that help us refine our understanding of the mechanism at work to fully coordinated multicontext trials, such as our 2015 six-country study of the ultra-poor graduation model. This large project showed that a “big push” program that addressed the many challenges of poverty simultaneously boosted livelihoods, income, and health among the ultra-poor. Such field replications may test whether something will work in a different region or country (like the graduation model), through a different type of institution (such as a nonprofit or government), or at a larger scale (for example, when a program scales from a couple of districts or regions to a nationwide program).

At this stage, theory combined with field replications enables us to understand why something may or may not work beyond the context in which it was initially evaluated, and how to adapt a concept from one context to another. When these adaptations show that a particular theory holds across contexts, the policy impact can be powerful. The successful replication of the ultra-poor graduation model, for example, has spurred governments and development agencies to expand the model to millions of people.

3. Advocacy, institutionalization, and scale. This is the stage where we help get the successful mechanism embedded into existing systems (for example, the aforementioned ultra-poor graduation models in government social-protection schemes), where we advocate for donors or governments to fund particular mechanisms at scale (for example, school-based deworming), and where we inform implementers about the most effective way to run programs (give away bed nets for free instead of charging).

While we present this as a third stage for simplicity, we have learned that laying the groundwork for these efforts in the first and second stages by proactively engaging the right decision makers to address their questions, understand the context, and regularly update them is crucial for securing their buy-in.

The goal is not just to scale successful ideas. Rather, it is about building an evidence-based decision-making culture. We can vastly expand the use of evidence by supporting governments in institutionalizing its use. To take one example, we joined our sister organization, The Abdul Latif Jameel Poverty Action Lab (J-PAL), to partner with the Ministry of Education in Peru to help them set up MineduLAB, a lab within the ministry to test innovative education solutions and apply the effective ones to policy.

INCREASING EVIDENCE-BASED DECISION MAKING

We have done much more of the first stage than we have of the second and third, but in order to facilitate evidence-based decision making, we believe that there needs to be more field replications, more advocacy, and more institutionalization. We have been less focused on these priorities for a number of reasons. For one, the field

ANNIE DUFLO is the executive director of Innovations for Poverty Action.

HEIDI MCANNALLY-LINZ is associate director of policy and communications at Innovations for Poverty Action.

is new: There was a genuine lack of evidence. Also, the pressures of funding cycles and the reality of academic career tracks incentivize proof-of-concept studies, which lend themselves to shorter timelines and testing new innovations. But we can relax these constraints and help further the evaluation revolution beyond proof of concept, if each partner in the process takes a few key steps.

Donors should consider evaluations as forward-looking, R&D tools rather than accountability tools. Development will become a field of knowledge acquisition only when it prioritizes the funding of this kind of investigation. To pursue this kind of learning, certain programs simply need good monitoring data, other programs need to apply evidence, while still others can benefit from a strong evaluation. Donors should not give incentives to over-evaluate—yes, you read that right. This can lead to poor evaluations, which wastes money and adds unnecessary confusion to debates about solutions. Donors should also be tolerant of failure and encourage organizations to be transparent about it and about what they will do to adapt.

Funders should commit to field replications and testing multiple variations of a program to determine why it works and to disentangle which components are most cost-effective. This commitment can cost more than your average “program versus no program” randomized evaluation, but it can also lead to clearer policy wins. This, in turn, will help organizations working in different contexts to make decisions about whether and how to adapt effective ideas tested elsewhere.

Finally, donors should also consider funding the third stage: advocating for the use of evidence and supporting its institutionalization. While it might seem less tangible than funding a specific study or intervention, it is a crucial stage for evidence to be used in decision making.

Academics and evaluation organizations should treat decision makers as their evaluation clients, too. Yes, adding to the body of knowledge is the researcher’s job, but good partner relationships can also make for better and more impactful evaluations. We have an internal campaign called “Impact: One Project at a Time” that encourages our staff to cultivate good relationships with evaluation partners and other organizations that might take an interest in the study, which can help improve research quality and ultimately enable the researcher to make a real impact.

Practitioners should not push to do impact evaluations just to please a donor or stakeholder when they do not need to evaluate. But they should strive to use data to inform their decisions, by using both existing rigorous evidence to design (or modify) their programs whenever possible and simple monitoring data to track whether their program is implemented as designed. When the time is right for an evaluation, practitioners should evaluate only if they are willing to abandon their prior assumptions and change what they do based on the results.

Rigorous impact evaluations should always help decision making, whether immediately or in the longer term, whether for a new idea or for adapting an idea from one context to another. Evaluations should also add to the body of evidence and help us arrive at generalizability. There is no contradiction between these two goals, and at IPA we strive to do both. ■