

## MIT Open Access Articles

### *Field Experiments and the Practice of Policy*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**As Published:** 10.1257/AER.110.7.1952

**Publisher:** American Economic Association

**Persistent URL:** <https://hdl.handle.net/1721.1/135989>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



## Field Experiments and the Practice of Policy<sup>†</sup>

By ESTHER DUFLO\*

I was not destined to be an economist. As the daughter of a mathematician, I was quite sure I would become an academic. My heroes were Gauss, the mathematical genius, and Emmanuel Le Roy Ladurie, the quantitative historian who found peasants interesting, rather than kings. But as the daughter of a physician who spent time trying to be helpful in countries where children were victims of war, I also aspired to be a change maker. I felt that the only repayment for the incredible luck I had in my life was to do whatever I could to try to improve the lives of the many people who were not that lucky. My heroes were Mother Teresa and Albert Schweitzer. Of course, I had no idea how to combine those two aspirations, but I hoped that one day I would find a way.

Until quite late in my college career, economics did not occur to me as a plausible path for accomplishing these goals. I had studied some economics as an undergraduate, but, like most people, I trusted neither economics nor economists. Indeed, a YouGov Poll from 2017 in the United Kingdom shows economists as being among the least trusted professionals regarding their own field of expertise: only 25 percent of the poll's respondents trust economists about economics (Smith 2017). This is half of the trust enjoyed by professional weather forecasters. Only politicians are perceived with more distrust.

My 20-year-old self very much shared this distrust. Armed with just a few introductory classes, I thought of economics as an elaborate hoax (or at best a Panglossian illusion) aimed at justifying the world and keeping it exactly as it was; using simple mathematics to describe some very rudimentary version of it, and “proving” that any attempt to intervene against the smooth functioning of the market would wreak havoc. Economics certainly did not appear to be a field for an aspiring change maker.

And yet, here I am, an economist. I chose this field because, ultimately, I came to believe that economic science could be leveraged to make a positive change in the world.

A year spent working in Russia as a research assistant for teams of academic economists in 1993–1994 led me to discover—with a mix of horror and fascination—the

\*Department of Economics, Massachusetts Institute of Technology (email: [eduflo@mit.edu](mailto:eduflo@mit.edu)). Abhijit Banerjee and I prepared two lectures with parallel titles. They are companion papers and should probably be read together. I would not have been able to get to the point of even giving this lecture without the help and influence of a great many people in my life. There would be too many to list in a single footnote, but everyone involved in the projects cited here is named and thanked at the end of this article. Here, I would like to thank the economics committee of the Royal Swedish Academy of Science for selecting me for this incredible honor. I would also like to thank my co-laureates Abhijit Banerjee and Michael Kremer for many years of an incredible collaboration. In addition, Abhijit Banerjee and I discussed this lecture in great detail. Garima Sharma and Gabriella Fleischman provided excellent editing and research assistance.

<sup>†</sup>This article is a revised version of the lecture Esther Duflo delivered in Stockholm, Sweden, on December 8, 2019 when she received the Bank of Sweden Prize in Economic Sciences in memory of Alfred Nobel. This article is copyright © The Nobel Foundation 2019 and is published here with permission of the Nobel Foundation. Go to <https://doi.org/10.1257/aer.110.7.1952> to visit the article page.

enormous influence that some academic economists have on the world. It seemed at the time as if several large-scale experiments were being conducted on the Russian economy, without much control. By 1994, these experiments were already running into serious difficulties. Yet, it was amazing to witness policymakers' willingness to listen to economists' sweeping pronouncements and recommendations for wholesale change, which had seemingly little empirical backing. I realized then that economics was the path to combining an academic career with the chance to have an influence on the world, and I also learned to be wary of this influence. I resolved to learn economics to realize my goal of making the world a little bit better, but also to move gingerly and with some humility.

Almost three decades later, working with many others (researchers, NGO practitioners, government officials, donors), I have indeed become something of a change maker. The Abdul Latif Jameel Poverty Action Lab (or J-PAL), the network that Abhijit Banerjee, Sendhil Mullainathan, and I started in 2003, and that was at first led by Rachel Glennerster, and now by Iqbal Dhaliwal, has affected policies in multiple ways and on every continent. By our count (which we try to keep conservative), over 400 million people have been reached by programs that were scaled up after being evaluated (and found effective) by J-PAL affiliated researchers. There are also many ways, less easily quantified, in which J-PAL has influenced policy. People have been indirectly affected as a result of ineffective policies being scaled down. Entire states have decided to adopt different policies because of a body of evidence. These effects are so diffuse that we do not attempt to count people affected through such channels.

The process by which J-PAL (and its affiliates) has influenced policy is quite different from the process I witnessed in Russia, with professors flying back and forth between Moscow and the United States, providing pieces of advice for the macroeconomy consistent with economic theory (or their intuition). It is also quite different from the influence of the "Chicago boys" who advised on macroeconomic policy in Chile (whose office J-PAL Latin America is, ironically, currently occupying).

J-PAL's approach is less about big ideas and more about specific suggestions. We take seriously both guiding principles and the less glamorous, but still crucial, realities of day-to-day policy implementation. For when economists get the opportunity to help governments around the world design new policies and regulations, they must shoulder the responsibility of getting the big picture and broad design right. In addition, as these designs are implemented in the world, they are also responsible for the many details about which their models and theories give little guidance. This is a role that RCT researchers have embraced in collaboration with government.

In this lecture I would like to discuss how this policy work happens in practice for researchers who do randomized controlled trials. I hope to illuminate how we can leverage good science to improve the effectiveness of policies that serve the poor worldwide, and also how we might use challenges posed by the world as sources of inspiration for our science.

## I. The Strawman

It is useful to start with the strawman: what the process of policy influence does *not* look like for researchers conducting randomized controlled trials.

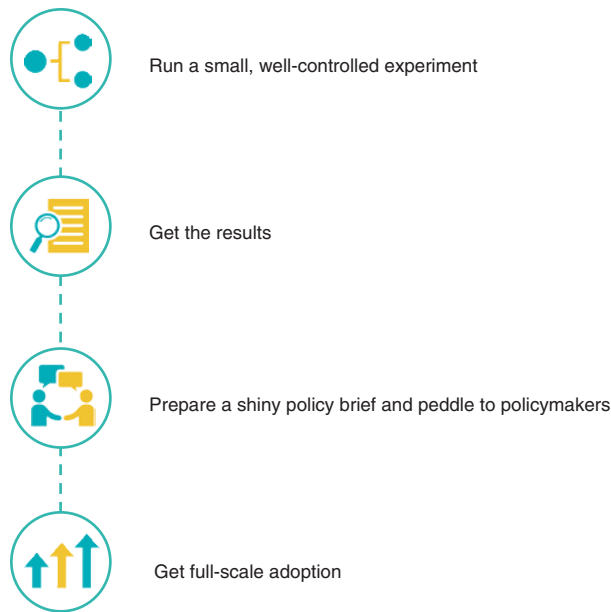


FIGURE 1. THE STRAWMAN

The strawman (illustrated in Figure 1) views the researcher as running a small, well-designed, and tightly controlled experiment (say, with 100 treatment schools, 100 control schools), implemented by excellent partners. She uncovers some results. If they are negative, she shelves the paper. If they are positive, she prepares a shiny policy brief and peddles it to policymakers, who adopt and scale up the policy.

Some version of this strawman is the basis of numerous critiques of the RCT movement, or at least of the hope of using RCTs to influence policy (e.g., Deaton and Cartwright 2018, Pritchett and Sandefur 2013). These critiques argue that the results of small, “gold-plated” experiments might not apply when programs are run at scale by less-than-perfect people. First, the argument goes, results may be highly context dependent. Second, the process of shelving what does not work may lead to selection bias, reflecting researcher luck more than reality. Third, even the most carefully controlled experiments can experience issues that prevent drawing robust insights: the sample might be too small to draw precise conclusions, compliance with treatment assignment might be imperfect, some people might be lost during follow up. Fourth, implementing a program at scale might affect outcomes that are not altered by a smaller-scale RCT intervention (as discussed in detail in our own work, see Banerjee et al. 2017, Muralidharan and Niehaus 2017): for example, prices might respond, spillover effects might affect non-participants, political economy reactions might alter program effects, and so on and so forth. Finally, policymakers might anyway be unlikely to pay attention to researcher recommendations, unless these recommendations match their politics. Even if they did, the reasons outlined above would prevent them from generalizing insights of one experimental context to another. The idea that you can go from a small experiment to widespread adoption would, under this strawman, therefore prove to be a myth. And we are

wasting valuable money in a slew of experiments that never lead to any meaningful policy influence.

These criticisms would have some bite if they accurately reflected the path of policy influence pursued by J-PAL and its affiliated researchers (and other “randomistas”). However, reality is quite different: one does not simply run an experiment, write one’s policy brief, and disappear while the policy is being scaled up. Actual policy dialogue in the RCT movement has followed quite a different path.

## II. How Lessons Are Drawn: Microcredit

The first flaw in the strawman is its misunderstanding of how RCT advances science. RCT researchers do not come to sweeping conclusions about the potential impact of a program based on any single experiment. Instead, each experiment is like a dot on a pointillist painting: on its own it does not mean much, but the accumulation of experimental results eventually paints a picture that helps make sense of the world, and guide policy. It is the accretion of results that makes sense and justifies the whole enterprise.

Perhaps the closest to the idealized example the strawman presented in Figure 1 is offered by how RCT research on microcredit came to influence perceptions among policymakers and the general public. What makes it relatively close to the process shown in Figure 1 is that this is a relatively rare instance of the results of a research program directly impacting policy, without any subsequent follow up. But it is evident that it is in fact very different ...

In the 2000s, microcredit was all the rage. As many seemingly “win-win” propositions, it gained popularity among both policymakers and the media: you could help people without spending any money, by simply lending to them and being reimbursed! You could even make money. Microcredit was expanding extremely rapidly, bolstered by successes in both public opinion and the commercial domain. Muhammad Yunus received the Nobel Prize for Peace in 2006, and some well-publicized IPOs of microfinance institutions (MFIs) made their funders very rich.

Few interventions that benefit the poor have such vast reach. Microcredit had almost 100 million clients in 2009, and 139.9 million in 2018. It would indeed have the potential to change the world were its impact actually positive on that many people. And indeed, microfinance’s more enthusiastic backers believed it as having the potential to transform people’s lives. The Consultative Group to Assist the Poor (CGAP), an organization housed at the World Bank, and originally dedicated to promoting microcredit, at one point declared on its web site: “there is mounting evidence to show that the availability of financial services for poor households can help achieve the MDGs” (including universal primary education, child mortality, and maternal health).

Unfortunately, little empirical evidence either supported or countered such propositions. The little evidence that existed was largely based on case studies, often self-produced by MFIs. For many supporters of microcredit, anecdotes seemed sufficient, at least at the time. In the late 2000s, however, the tone of the conversation on microcredit seemed to shift (shortly after Yunus’ Nobel Prize, which is perhaps enough to make some of us concerned...). Waves of farmer suicides were linked to high microcredit indebtedness; negative stories of farmers trapped in debt made

their way in the media. This shift in narrative impacted policy. In October 2010, only two months after the successful IPO of SKS, a prominent for-profit microlender in India, the Andhra Pradesh government blamed it for the suicide of 57 farmers, claiming that loan officers' coercive recovery practices had put clients under unbearable pressure. The government arrested a few loan officers and passed an ordinance forbidding the weekly collection of loans. By November, all credit officers of all the major MFI were sitting idle and losses were mounting. Anecdotes describing successful borrowers did little to help SKS at this time.

When these events unfolded, many of us had been looking for a while for a partner to evaluate the impact of this very important program. But when we approached MFIs (starting around 2002) with the proposition to rigorously evaluate their product, their usual response was, "Why do we need to be evaluated any more than an apple seller?" By which they meant that microcredit had to be beneficial as long as clients kept coming back for more. Of course, this ignored the fact that microfinance is often implicitly subsidized, as well as that irrational borrowers may borrow more than is good for them.

The real reason for the initial resistance was probably that MFIs did not see any reason to rock the boat: they were being hailed for their successes, and did not wish to run the risk of potentially refuting a positive narrative with data. But under mounting pressure from the critics of microfinance, and especially from policymakers, some MFIs decided that evaluation was worthwhile.

We conducted one of the first evaluations of microfinance with Spandana in Hyderabad. Believed to be one of the most profitable organizations in the microfinance industry, Spandana had been a chief target of government activism (in fact, it was eventually shut down during another policy-induced massive microfinance crisis of 2010). Our evaluation encompassed Spandana's expansion into some areas of the city of Hyderabad. Out of 104 neighborhoods, 52 were chosen at random for the organization to enter. The rest were left as a comparison group.

When we compared households in these two sets of neighborhoods 20 months after Spandana started lending, we found clear evidence that microfinance was doing, in many respects, what one might expect. Households in Spandana-covered neighborhoods had more businesses, and were more likely to have purchased large durable goods such as bicycles, refrigerators, and televisions. There was no clear evidence of the reckless spending that some observers feared would result. In fact, we saw exactly the opposite; households started spending *less* money on what they viewed as small "wasteful" expenditures, such as tea and snacks.

At the same time, there were no signs of radical transformation in the lives of microfinance borrowers. We found no evidence that women were feeling more empowered, at least along measurable dimensions. Nor did we see any differences in spending on education or health. There was no effect on household consumption. Even the business effects came from household starting more businesses if they were already business owners, and not from new households becoming entrepreneurs. These new businesses were small. Three years later, the effects were very much the same and most new businesses had shut down.

The study ruffled some feathers. Its results were mainly quoted for the negative findings, and as proof that microfinance was not the panacea it was made out to be. Although some MFIs accepted the results for what they were, the big international



players in microfinance decided to go on the offensive. They were even more concerned because a contemporaneous study in the Philippines (Karlan and Zinman 2011) had found equally lukewarm results.

As Abhijit Banerjee and I report in our book, *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, representatives of the “big six” MFIs in the world (Unitus, ACCION International, Foundation for International Community Assistance [FINCA], Grameen Foundation, Opportunity International, and Women’s World Banking), held a meeting in Washington, DC shortly after the microfinance studies were made public. They put together a SWAT team charged with responding to any new study (apparently convinced that all studies would be negative). A few weeks later, this SWAT team produced its first attempt at damage control, releasing six anecdotes on successful borrowers and an op-ed attacking the studies written by the CEO of Unitus, Brigit Helms, in the *Seattle Times*.

This strong reaction was surprising to us because we had been very careful not to take an extreme position. First, although the studies did not show microfinance as being a miracle, they also did not show for the disaster described by some (if anything, the Helms editorial exaggerated how negative the findings were). Second, our evaluation was in Hyderabad, the hotbed of microfinance in India, which was saturated with other microfinance agencies. High baseline access to microfinance could well underlie the lack of transformational impact in this context. We did not think we had enough evidence to draw emphatic conclusions, and we waited for results from other studies.

By 2015, evaluations of microfinance had concluded in seven countries: the Philippines, Morocco, Mongolia, Mexico, India, Ethiopia, and Bosnia. They were all published together in a single issue of the *American Economic Journal: Applied Economics* using a common reporting format (Angelucci, Karlan, and Zinman 2015; Attanasio et al. 2015; Augsburg et al. 2015; Banerjee et al. 2015; Crépon et al. 2015; Karlan and Zinman 2011; Tarozzi, Desai, and Johnson 2015) (for full disclosure, I was then the editor of *AEJ: Applied*). Each study was run by a different team. Some were rural and others were urban.

The common reporting template for these studies conducted in very different settings allowed us to tackle the challenge of “external validity,” frequently cited in the strawman as a drawback of RCTs. In particular, Rachael Meager set out to determine the differences (or similarities) in results across contexts (Meager 2018, 2019). The difficulty with this exercise is that the observed variation in effects across studies conflates the true variation in treatment effects with variation in the estimated effect that stems from having randomly sampled individuals from a population. To get around this problem, Meager used Bayesian hierarchical analysis. The basic idea is to first assume that the real treatment effect in each site is drawn from a standard normal distribution. We then add some noise to each real treatment effect to account for sampling variation. Even this minimal amount of structure on the problem allows a statistician to determine the extent to which effects “pool” across studies, that is, the extent to which the “real effects” in each site are close to those in others. It also enables computing the overall average effect, as well as country-specific results that can incorporate results from other places.

Meager examined the effect of access to microcredit on household business profits, expenditures, revenues, total consumption, spending on consumer durables, and

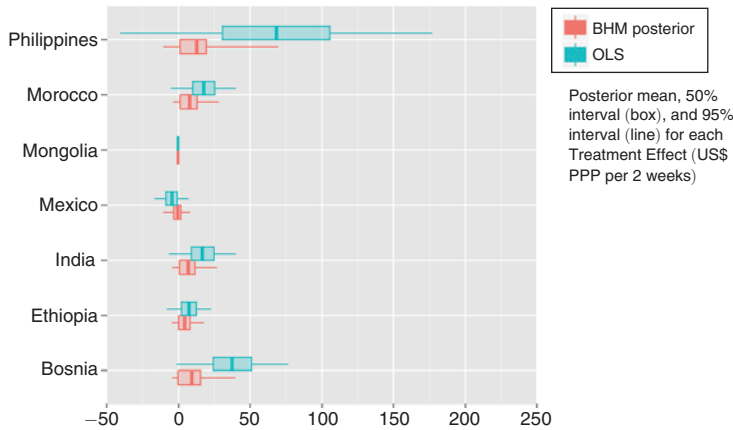


FIGURE 2. BAYESIAN HIERARCHICAL MODELING MF RESULTS: PROFITS

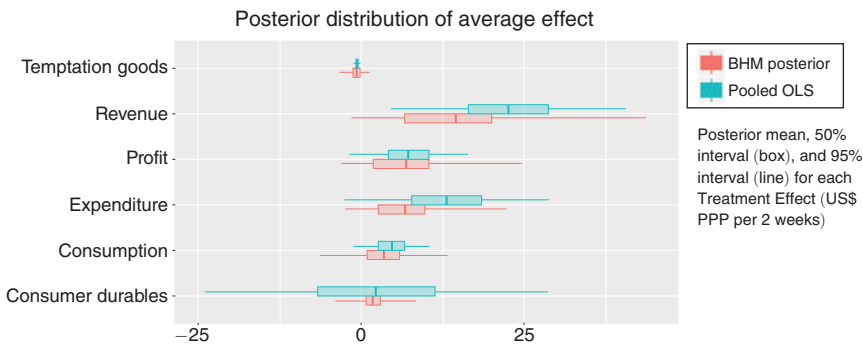


FIGURE 3. AVERAGE ESTIMATED EFFECT AND RANGE, 6 COUNTRIES

spending on “temptation” goods such as cigarettes. Overall, as illustrated in Figures 2 and 3, she finds generally small and very uncertain effects (about 7 percent of the mean outcome, with zero a very likely impact for all variables). This analysis largely confirmed our initial underwhelming findings from the Spandana study, and also (sadly) showed that the one positive result we had found—a decline in spending on temptation goods—was not in general robust across contexts.

One finding that is robust, however, is that households who were business owners prior to microcredit entry (and who had therefore proven their enterprising nature) actually did benefit from microfinance. In fact, we continued to follow them in Hyderabad and find them as experiencing large increases in business revenues, profits, and average consumption 10 years following the introduction of microcredit (Banerjee et al. 2019).

The overall conclusion from the above body of evidence was, therefore, not that microcredit is harmful or even that nobody benefits from its introduction. Rather,



across a variety of contexts, it does not enable the average person to exit poverty, or to experience impressive transformation in their lifestyle. Even so, some (existing) entrepreneurs benefit greatly from microfinance loans, and many others use it as consumption finance.

The reaction of the MFI community to the accumulated body of evidence was quite different than to the first couple of studies. We organized a joint conference with CGAP in Washington DC, followed by another one at Harvard Business School. Both were well attended by microfinance practitioners. Participants focused on redesigning microfinance using insights from the studies, as opposed to trying to kill their results. Remarkably, even the media was measured in its coverage of the event and the underlying research, with *The Economist* for example describing the results in a piece titled “A Partial Marvel.”

Results from microfinance RCTs had successfully shifted the policy debate away from shouting matches between “disaster” or “miracle,” changed the view of many promoters of microfinance,<sup>1</sup> and eventually changed microfinance itself. The objective of the researchers was of course never to undermine microfinance: in fact, much of modern development economics is predicated on the fact that financial markets work very badly for the poor, and that this constrains their occupational choice and leads to poverty traps (e.g., Banerjee and Newman 1993). What these results suggested, however, is that the “one-size-fits-all” approach that had been the hallmark of the microfinance movement since Muhammad Yunus (one loan, given once a year, and repaid in weekly, equally sized installments) was perhaps not ideal, given the extreme heterogeneity in borrowers’ needs and types. While some people needed consumption finance or even just a good savings products, a minority of real entrepreneurs needed business lending with larger and more flexible loans.

The second wave of microfinance studies was very much focused on these topics. They sought to ask not whether microfinance worked, but how to modify it to make it better. For example, some researchers asked whether the group structure, which is quite constraining, is really necessary (Giné and Karlan 2014), some experimented with a one-month grace period prior to the start of repayment (Field et al. 2013), and with changes to the frequency of repayment (Field et al. 2012). Recent research focuses on how best to identify the most entrepreneurial clients using community information (Hussam, Rigol, and Roth 2016).

The main contribution of the overall research agenda has been not to prescribe the scale up or scale down of microfinance (in fact, the number of microfinance clients has continued to grow following the first studies), but to help the sector and policymakers think about microfinance in a richer, more subtle way. One can see that the path from RCTs to policy influence was not straightforward. It involved several studies and careful analyses. It did not culminate in a “thumbs up” or “thumbs down” recommendation, but in an invitation to rethink financial services and the financing of entrepreneurs. This rethinking is very much ongoing, and combines exciting research with innovative product design.

<sup>1</sup>Most notably CGAP, which broadened its mandate to be a promoter of financial inclusion more generally and put its energy behind a program of assets transfers to the ultra-poor, described in Abhijit Banerjee’s Nobel lecture (Banerjee 2019).

### III. From Proof of Concept to Impact at Scale: Teaching at the Right Level

The microcredit example is unusual in the sense that results from RCTs were sufficient to change perceptions and policy. A more typical case of policy influence follows a long chain from the first experiments to the final adoption of policy, on the way tackling the many difficulties involved with scale-up. One prominent example where the entire chain can be traced is the “Teaching at the Right Level” program.

#### *A. Teaching at the Right Level*

In many developing countries, children are in school but are not learning very much. This is also very much the case in India, where less than one-half of all children in grade 5 can read a simple paragraph at the grade 2 level. The performance is even worse in mathematics, and, sadly, the situation is not improving over time (ASER 2015). The current state of affairs of course represents a huge waste of resources. Many experiments have attempted to examine reasons for and solutions to this problem of children not learning in school—including Michael Kremer’s very first experiment on textbooks (Glewwe, Kermer, and Moulin 2009).

The key issue appears not to be just a lack of inputs, a lack of incentives for teachers to exercise effort, or even the inability of children to learn. Rather, the pedagogy employed in schools is completely inappropriate. In particular, teachers are required to teach and complete very demanding curricula, and nothing is really done to help students catch up when they get lost. Most developing countries still have elite-biased school systems stemming to some extent from their colonial history. These education systems were originally set up to educate a small elite that was going to support the colonial power. They were expanded as is at the time of independence, in part because scaling back the ambitious curriculum might have appeared to shortchange children, which is difficult to justify politically. As a result, children in these countries are taught not at the level at which they can learn, but at some aspirational level far above what most normally constituted children can digest (Banerjee and Duflo 2011).

One might think that the solution would be to reform the curriculum, but this has not been feasible for the political reasons discussed above. The second best, deceptively simple, strategy is to teach children what they are capable of learning, whenever possible, despite the curriculum. In our first RCT, Abhijit Banerjee, Shawn Cole, Leigh Linden, and I worked with the wonderful organization Pratham to evaluate exactly this approach to solving the problem: teach children at the right level, using whatever margin can be pried open within or outside the school system.

This approach has now come to be called “Teaching at the Right Level” (or TARL). The core principle behind TARL is to frequently assess children and offer activities that correspond to their current level of knowledge. Children are assessed, grouped, taught at the level that is right for them at this exact moment, and frequently reassessed and regrouped.

B. *From Mumbai Slums to over 20 Million Children:  
15 Years and Many Experiments*

The partnership between Pratham and J-PAL is J-PAL's longest (and Pratham's), and certainly among the most influential in terms of policy impact. Revisiting the history of this partnership is instructive in understanding how one goes from a good idea to a policy that affects millions of children. It will make very clear that this process does not follow the strawman's template. This section is largely adapted from Banerjee et al. (2017) and also owes a lot to Rukmini Banerji's wonderful recollection of this journey (Banerji 2019, Banerji and Chavan 2020).

The partnership between researchers and Pratham started with a "proof of concept" randomized controlled trial of Pratham's "Balsakhi Program" (the ancestor of the Teaching at the Right Level program) in the cities of Vadodara and Mumbai, conducted in 2001–2004 (Banerjee et al. 2007). In this program, third- and fourth-grade students identified as "lagging behind" by their teachers were removed from class for two hours per day, during which they were taught remedial language and math skills by community members (*balsakhis*) hired and trained by Pratham. This RCT would have looked like the first "well-controlled" experiment in Figure 1, except that it was everything but: one year, we had to discard all the tests because it was evident that children had copied from each other; another year, test papers were given back to children before they could be doubled-entered; one year, a massive earthquake shook Baroda; another year, communal riots disrupted the city, shutting down schools and the program. Despite these setbacks, the results were clear; children's learning levels (measured by second-grade-level tests of basic math and literacy) increased by 0.28 standard deviations on average. These gains were entirely accounted for by children at the bottom of the test score distribution, who were the ones who in fact received the remedial help.

The second randomized controlled trial of what would become TARL was conducted in Jaunpur district of Uttar Pradesh in 2005–2006: this was a test of a volunteer-led, camp-based Learning-to-Read model, set in a rural area. The results were once again very positive: attending the classes made children 22.3 percentage points more likely to read letters and 23.2 percentage points more likely to read words. Nearly all children who attended the camp advanced one level (for example, from reading nothing to reading letters, or from reading words to reading a paragraph) over the course of that academic year (Banerjee et al. 2010).

This second study established that the pedagogical idea behind the *Balsakhi* program could survive the change in context (from urban to rural) and program design (from paid assistants in schools to volunteers outside schools), but it also revealed new challenges. There was substantial attrition among the volunteers, and many classes ended prematurely. Also, because the program targeted children outside of school, take-up was far from universal. Only 17 percent of eligible students were treated, and they were not even the ones who needed it the most.

In order to reach all children who needed remedial education and to more effectively use school time, Pratham started collaborating with state governments in running the Read India Programs. But since the program was now going to be implemented by public school teachers, it was not obvious that it would work as well as it had with volunteers. This change required a new wave of evaluation.

### *C. A First Attempt to Scale Up with Government*

Starting in 2008, Pratham and J-PAL embarked on a series of new evaluations to test Pratham's approach when integrated with the government school system. Two randomized controlled trials were conducted in the Indian states of Bihar and Uttarakhand over the two school years of 2008–2009 and 2009–2010. Although the evaluations covered only a few hundred schools, they were embedded in a full scale-up effort: as of June 2009, the Read India program in Bihar was being run in 28 of the 38 districts in Bihar, reaching 2 million children in approximately 40,000 schools. In Uttarakhand, before the evaluations were launched, Pratham was working in all of 12,150 schools in 95 "blocks." For the experiments in Bihar and Uttarakhand, we "carved out" a district where some schools were kept as the control group, allowing us to evaluate the effectiveness of a program run at scale.

This approach of evaluating at scale is diametrically opposed to the one described in the strawman we discussed before. Here the program is run at scale, and the control sample is kept small. This helps ensure that all issues associated with scaling up a program are addressed. Indeed, this design voids much of the concerns voiced in the strawman (the gold plating, the external validity, the political economy concerns). Much can be learnt from this kind of experimentation.

In the first intervention (evaluated only in Bihar during June 2008), remedial instruction was provided during a one-month summer camp, run in school buildings by government school teachers, who were paid extra by the government. This evaluation (which, to be perfectly honest, was a last-minute addition to the research project, made possible by Rukmini Banerji's and Michael Walton's keen attention to how the program unfolded on the ground and quick action to preserve the possibility for an experiment (Banerji 2019)) showed significant gains in language and math. In just a few weeks of summer camp, the treatment on the treated effects were of the order of 0.4 standard deviations.

The other three interventions were conducted during the school year. The first model distributed Pratham materials with no additional training or support. The second included materials, training of teachers in Pratham methodology, and monitoring by Pratham staff. Teachers were trained to improve teaching at all levels through better targeting and more engaging instruction. The third and most intensive intervention included materials, training, and volunteer support. The volunteer part was a replication of the successful model evaluated in Jaunpur, wherein volunteers conducted evening learning camps that focused on remedial instruction for students directed to them by teachers.

The results were striking and mostly disappointing. The materials-alone and materials-plus-teachers interventions had no effect in either Bihar or Uttarakhand. The materials-teachers-volunteer treatment in Uttarakhand also had no discernible impact. Only the materials-teachers-volunteer intervention in Bihar found significant impacts on reading and math scores, comparable to the earlier results from Jaunpur. So, the standard Pratham model worked, but the transfer to government teachers was unsuccessful.

At this point one might have been tempted to assume that teachers were just unable or unwilling to implement an intervention that really focused on children's learning. But the positive impact of the summer camps, which were teacher-led,

suggested otherwise (and as Rukmini Banerji recalls, this summer camp experiment was essential in reinstating my trust in teachers (Banerji 2019)). We drew on qualitative and process data we had collected throughout the project to ascertain why the school year intervention did not work. These data contained information on the relationship between Pratham and the government (Kapur and Icaza 2010, Sharma and Deshpande 2010), as well as perceptions of children, parents, and teachers.

Process monitoring revealed considerable support at the top of the hierarchy for the program in Bihar (less so in Uttarakhand), as well as effective delivery of basic inputs: two-thirds of the teachers were trained, they received the material, and they used the material over one-half of the time. Despite these successes, the key component of Pratham's approach, its focus on teaching at the children's level, was generally not implemented by schools in either state. When regular teachers were in charge, they almost never grouped students by levels.

Teachers told us they found the activities valuable, but had no time to implement them given the requirement to still complete the prescribed curriculum. Paraphrasing teachers interviewed in Bihar, Sharma and Deshpande (2010) write: "[T]he materials are good in terms of language and content. The language is simple and the content is relevant (...) However, teaching with these materials requires patience and time. So they do not use them regularly as they also have to complete the syllabus." Incidentally, completing the syllabus is required of teachers by law, so they cannot be blamed for their focus. Of course, implementing teaching at the right level was now *also* part of their job, since the program had been scaled within the government. But this had not been clearly conveyed. In the presence of potential tension between the new and old objectives, teachers decided to stay safe by focusing on the status quo.

Armed with the results of this study, the Pratham team and us attempted to find a solution to the problem of no TARL implementation. The answer proved two-pronged. First, we recommended carving out a time during the year or a time during the day to focus on teaching at the right level, so as to avoid direct competition between TARL and the completion of the curriculum. Second, we recommended either convincing teachers to take teaching at the right level more seriously by working with their superiors to build it into their mission; or cutting out the teachers altogether and implementing a volunteer-style intervention in schools. These ideas guided the design of the next two interventions.

### *Getting Teachers to Take Teaching at the Right Level Seriously*

In 2012–2013, Pratham in partnership with the Haryana State Department of Education adopted new strategies to embed Teaching at the Right Level as a "core responsibility" for teachers. To promote teacher buy-in, Pratham emphasized that the program was fully supported and implemented by the Government of Haryana, rather than by an external entity. Pratham first gave four days of training and field practice to teacher supervisors, or "Associate Block Resources Coordinators." Upon the completion of the practice period, these coordinators in turn trained and monitored teachers in their jurisdiction.

In addition, the program was implemented during a specific hour of the day. During this TARL hour children were grouped by level, not by grade. The time delineation made clear that TARL was part of a teacher's job, and that she did not

have the discretion to convert it back to regular class time. This new version of the program was evaluated in 400 schools during the 2012–2013 school year; 200 of these schools were in the treated group and received the program. The results this time were positive. Hindi test scores increased by 0.15 standard deviations (significant at the 1 percent level) (the program did not cover math).

#### *D. Using the Schools, but Not the Teachers: In-School Learning Camps*

An alternative model was to sidestep teachers altogether, and instead use volunteers to teach in schools in a “learning camp” model. Learning Camps are intensive bursts of teaching-learning activity using Pratham’s methodology. Pratham volunteers and staff administer them during school hours when regular teaching is temporarily suspended. These camps were held for 50 days per year. On “camp days” children from grades 3–5 were grouped according to their ability level and taught Hindi and math by volunteers for about 1.5 hours.

The model was tested in a randomized evaluation in Uttar Pradesh in the year 2013–2014. A sample of schools was selected and randomly divided into two camp treatment groups, a control group, and a materials-only intervention, with approximately 120 schools per arm. The learning camp intervention groups varied the length of the camp, with one receiving four 10-day rounds of camp, and the second receiving two 20-day rounds. The two interventions had similar impact, with test score gains of 0.6 to 0.7 standard deviations.

#### *E. Scaling Up*

It took five randomized controlled trials in India and several years to traverse the distance from concept to a policy that could succeed at scale. But it has been effective: since 2013–2014 when the Haryana RCT was concluded, formal partnerships with government to scale up a “Haryana style” model in schools have reached 21.3 million children across the country. And the scale up did not stop there. Paralleling the India experiments, researchers evaluated similar or identical approaches in Africa (in Kenya, children were tracked for two years according to their first semester grades; in Ghana, teams from the ministry of education visited Pratham and the Pratham model was tested in schools). TARL became one of the few projects selected by Co-Impact (“a global collaborative for systems change, focused on improving the lives of millions of people around the world”) for massive scale up through government. Figure 4 shows locations across Africa where TARL Africa (a joint venture of J-PAL and Pratham) is working with the government to scale TARL up.

### **IV. Improving Programs that Run at Scale, by Helping Government Address “Plumbing Problems”**

The sections above give us a sense of what it takes to go from proof-of-concept to a scalable policy. One lesson is that it takes many experiments. Another clear lesson is that the researcher’s role is not restricted to giving advice from some sort of a pedestal. Along with the government, researchers jointly try and err. They co-create.





FIGURE 4. TEACHING AT THE RIGHT LEVEL TODAY

The researcher has not been particularly useful if she only provides a general idea without engaging with the muddled process of implementation. Co-creation is now happening with governments as well. Much of the work done by affiliates of J-PAL, IPA, or other organizations that run RCTs, now helps governments better design and implement their own programs.

I have called this approach the “plumbing” approach (Duflo 2017). In plumbing problems, the government is not asking itself whether it should invest in health or education, or even in any particular intervention. Rather, it is asking a question of the form: “We are running this particular program and there are issues with it. What can we do to address these issues and achieve our objectives?”

Trying to answer this question is not the sweet spot for most economists. Banerjee (2007) writes that economists tend to think in “machine mode”: they want to find the button to start the machine, and identify the root cause of what makes the world go round. He writes:

*The reason we like these buttons so much, it seems to me, is that they save us the trouble of stepping into the machine. By assuming that the machine either runs on its own or does not run at all, we avoid having to go look for where the wheels are getting caught and figuring out what small adjustments it would take to get the machine to run properly. To say that we need to move to a voucher system does not oblige us to figure out how to make it work—how to make sure that parents do not trade in the vouchers for cash (because they do not attach enough value to their children’s education) and that schools do not take parents for a ride (because parents may not know what a good education looks like). And how to get the private schools to be more effective? After all, at least in India, even children who go to private schools are nowhere near grade level. And many other messy details that every real program has to contend with.*

In contrast, an economist who cares about the details of policy implementation must heed complications that may appear far below her pay grade (e.g., the font size on posters) or far beyond her expertise level (e.g., the intricacy of government budgeting in a federal system). She must apply her economist mind to tackle



incentives, information, imperfect rationality, etc. She must keep a close eye on the impact of any recommended change. What makes this process of implementation akin to plumbing is that the economist will typically not even have the safety net of a bounded set of assumptions. She knows she will not know for certain the determinants of success. Nonetheless, she will put her best foot forward: using her knowledge of the science, the contextual knowledge of her partner organization, and prior experience. There will remain genuine uncertainty about the best way to proceed on many details, because the solution depends on a host of factors that are not easy to quantify, or sometimes even to identify, in the abstract. (These are the “unknown unknowns”: all the issues we cannot predict but know will arise). In the pursuit of good implementation of public policy, the economist is willing to tinker and try again. And in the presence of uncertainty, field experimentation becomes her tool of choice: the best way to determine what works, and to adjust. Policymakers are also often willing to experiment on questions of implementation, because they recognize that they do not have a clear path forward.

One example of a project to improve the quality of implementation is offered by the rice distribution program in Indonesia (Banerjee et al. 2018). This program (then called Raskin) is massive, reaching over 17.5 million households. It is funded centrally but administered locally. As with many programs of this scale, it experiences many issues with implementation. For example, many eligible households do not receive the program, many who receive it end up paying more and getting less than they should, and a substantial part of the program’s budget “leaks” into the pockets of government officials responsible for implementation. As a result, potential beneficiaries only receive about 30 percent of the benefits to which they are entitled. The government at the time believed poor information about eligibility as one main reason for these problems. Rema Hanna and Ben Olken, the co-leaders of the J-PAL South East Asia office in Indonesia, have established a long-running collaboration with the Indonesian government, which leads the government to frequently bring up these types of concerns with J-PAL and to collaborate on policy-oriented research projects. In this particular instance, the government originally wanted to distribute cards to increase awareness about program eligibility. The researchers (Ben and Rema, joined by Abhijit Banerjee, Jordan Kyle, and Sudarno Sumarto, the leader of an Indonesian Think Tank) were keen to explore this idea in a large-scale experiment that exploited the reach of the program, given that distributing cards is inexpensive. They proposed an evaluation that enabled them and the government to not only learn *whether* the cards made a difference, but also *how* to structure the card’s content and its distribution.

They asked a series of pertinent questions. Should the card inform recipients of the correct price? Should everybody in village get a physical card, or is it sufficient to deliver it to a subset of individuals but publicly post the entire list of beneficiaries? Should a village be plastered with posters, so that, in addition to beneficiaries knowing of their eligibility, officials also know that beneficiaries know, and the villagers in turn know that the officials know that they know, and so on and so forth, creating “common knowledge” (potentially changing how people bargain)? Should the card have clip-off coupons that officials are required to send to their supervisors to enhance perceived accountability?

When the research team implemented the experiment in over 550 villages, they evaluated not only the impact of giving a card, but also answered the questions above. Multiple treatment groups enabled them to provide insight on which version of the card and distribution mechanism was the most effective and cost-effective. The best strategy, it turns out, is to include price information, distribute the card to everyone, and create common knowledge. The accountability piece was not particularly important. This best strategy increased take up of the program and reduced the price paid, leading to an overall 26 percent increase in the value of the subsidy received by eligible households.

Because this intervention was evaluated in response to the Indonesian government's interest and demand, it was almost immediately scaled up to over 60 million participating families. This scale-up immediately following research was possible due to the close collaboration between J-PAL and the government, as well as the fact that it involved ramping up operations already occurring at significant scale in the same context. By going "inside the machine" the researchers found an immediately relevant way to make it work.

Of course, the project also yielded insights that can prove helpful in other contexts. In particular, it demonstrated the key role of specific and verifiable information on the bargaining process between beneficiaries and government officials. Similar "plumbing" projects often yield more general lessons that can be applied in other settings or in other types of programs (with more fine-tuning, and perhaps a new experiment).

Today, this kind of direct collaboration with government comprises a very important way in which RCT researchers play a role in the policy process. I could cite several similar examples, but will mention just one more. It highlights a scenario in which close collaboration between researchers and the government, good knowledge of economics, and excellent knowledge of local institutions eventually led to statewide reform of policy.

In this project, Michael Greenstone, Rohini Pande, Nick Ryan, and I collaborated with the Gujarat Pollution Control board (GPCB) to help them reform and revive a third-party environmental audit system. Gujarat is the Indian state with the fastest industrial growth, and, partly as a consequence, is also the state with the fastest growth in pollution. Some of the most polluted places on earth are in Gujarat. A few years ago, the Supreme Court ordered the Government to set up a third-party audit system, wherein each plant in highly polluting sectors would have to obtain (and pay for) an annual audit administered via a private firm. The audit report would be shared with the GPCB, which could impose sanctions. This is a great idea in principle, since it forces the polluter to pay and allows the government to harness private competencies it does not possess. Unfortunately, however, the structure of the program created a natural conflict of interest between the auditor and the firm: since the firm chooses to hire and pays the auditor, the latter has every reason to give them a clean bill of health. This dysfunction was common knowledge at the start of our collaboration with the GPCB. Business associates were even suing the government to remove the scheme, arguing that the information collected was so useless that the audits just ended up serving as an extra tax.

A GPCB lawyer initiated contact with one of us (Rohini Pande) during a visit to the Harvard Kennedy School. They were interested in reforming the system to

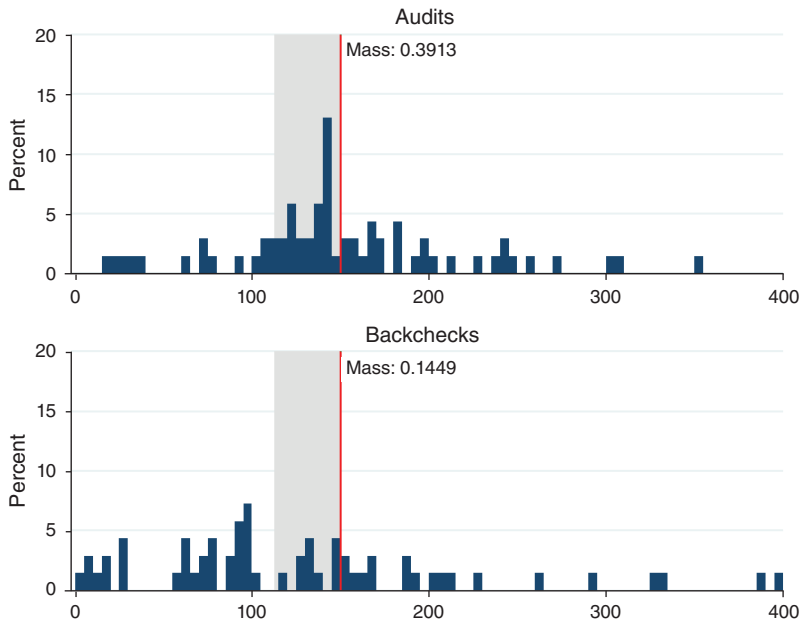


FIGURE 5. POLLUTION REPORTS VERSUS REALITY

give it more bite. To verify that the system was indeed not working, we began by collecting “back-check data” on the audited firms. As part of these “back-checks,” we sent a second audit team (comprised of students and faculty from a local engineering college) to collect information on the same pollutant examined in the original (private-firm-administered) audit. As illustrated in Figure 5, there was a stark contrast between the audit report and the back-check. Whereas most audit reports showed pollution levels just below the acceptable threshold, true levels of pollution were very different. Many firms were found in the back-check as polluting much more than in the original audit, while others were polluting much less. It was apparent that the auditors did not even bother visiting the plants to collect samples: they were just making up a plausible-sounding number. This had the extra advantage of making the audit very inexpensive... the going rate for an audit report was not even sufficient to cover the cost of testing the samples.

Following extensive conversations with GPCB over many months (which turned into a fruitful collaboration over several years), we proposed a three-part solution to alleviate the apparent conflict of interest and make the auditor loyal to society as opposed to the audited firm. First, we proposed breaking the financial link between the audited company and the auditor, by creating a central pool from which auditors would be paid. Second, we proposed making the monitor feel responsible for accuracy. In the first year, this was achieved by threatening to discontinue their participation in the scheme for low accuracy, and in the second year by rewarding them with higher payment for high accuracy. Third, we began measuring accuracy through back-checks. We designed a randomized controlled trial to test this new system: audit-eligible firms were randomly assigned either to the status quo system or to the new system. We found audit reports as being

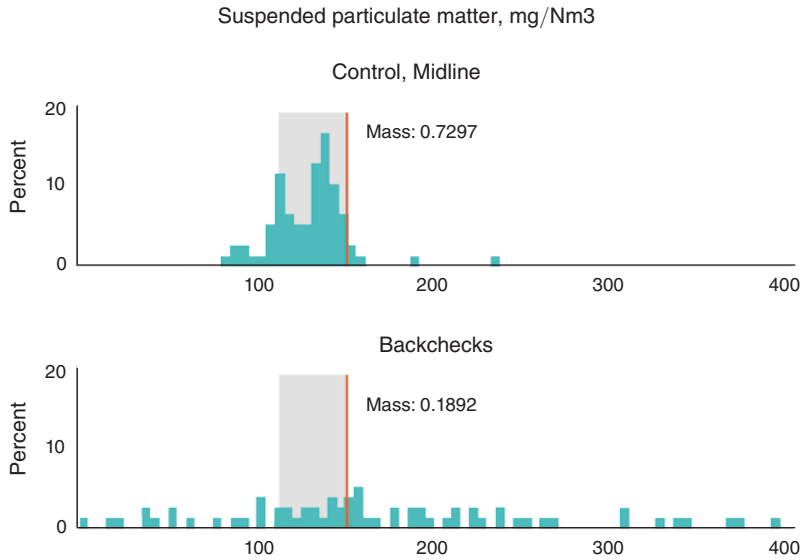


FIGURE 6. IMPACT OF THE REFORM

much more accurate under the new system. This is illustrated for one particular pollutant in Figure 6, where we show that the new system causes the excess mass for firms reportedly polluting right below the acceptable level to disappear. Moreover, perhaps because of greater scrutiny, pollution (measured in an independent endline survey) also declined, particularly for the worst offenders (Duflo et al. 2013).

Based on these results, GPCB successfully convinced the court and state administration to change the rules governing the scheme's implementation. These changes came into effect in 2015, with new guidelines requiring the random assignment of environmental auditors to firms, instituting back-checks, and imposing a fee schedule.

In this example, we combined basic knowledge of fundamental economic principles with a deep knowledge of ground realities (gained from extensive qualitative interviews with the GPCB staff) to help the government redesign rules to solve a very specific plumbing problem: ensuring that the audit system achieved its stated objectives.

The collaboration between GPCB and the research team did not stop at this one policy. In another project, we studied both the impact of inspections and the optimal way to assign them (randomly or using discretion) (Duflo et al. 2018). Contrary to our own instinct (and that of many economists and policymakers), we found that the staff at GPCB has and uses significantly relevant information to “fish out” the worst offenders for pollution. It would therefore be inefficient to require them to randomize the first inspection. Today, Rohini Pande, Michael Greenstone, and Nick Ryan continue to work with GPCB; they are piloting a novel Emission Trading Scheme that could be a template for India and beyond.

The scale and the ambition of the researchers' collaboration with GPCB goes far beyond a set of recommendations. By establishing a long-run relationship built on

trust and collaboration, they are able to pilot ideas which would be impossible to implement and test in any other setting. Having research collaborators also allows the government to try policy changes that they may otherwise not have the bandwidth to implement.

The research-government relationship also helps bureaucrats and politicians create a space for innovation in the policymaking process: with explicit experimentation, they have the license to try out new things or to do things differently. They no longer need to inflate the benefits of a recommended project, because it can be shut down if it does not produce expected gains. Failure is no longer stigmatized. This culture of learning will perhaps be the deepest and most lasting policy influence of J-PAL. The ultimate success, of course, would be for this culture of innovation and trial and error to become so deeply ingrained that it occurs even in the absence of J-PAL.

Of course, the ultimate objective of this kind of policy work is to reach a point where, as an organization, we would largely be irrelevant, because the culture of learning and the capacity for doing this work would be so widespread that governments would take over the whole project themselves. We are working on it: this is why the third pillar of J-PAL is training. Training takes many forms, from short executive courses to semester-long online courses, and even a blended masters program at MIT, called “Data and Economics for Development Policy,” where students take one semester worth of online classes, on the basis of which they are admitted to MIT. To be completely honest, we are far from the point where we can declare our work done and shut down our offices. But there are signs of progress. In Peru, the ministry of education created a unit called the “Minedu Lab” that is devoted to policy innovation and evaluation and is actively engaged in RCTs. In Tamil Nadu, India, the government has a long-standard memorandum of understanding with J-PAL, whereby departments or researchers can propose policy innovation to test and innovate. Each of these partnerships takes us closer to a world where our ultimate policy influence will be that we are not needed any more.

## V. Conclusion: A Prize for a Movement

It should be clear from this lecture that I did not become either the kind of academic or the kind of change maker that I dreamt of being. I did not make a difference through the solitary pursuit of science. And I am not a savior or a hero. The only reason we managed to change the practice of economics, as Abhijit Banerjee (2019) describes in his lecture, or the practice of policy, as I describe here, is because we were part of a movement. This movement is not one that is constituted only of academics: while the academic plays a key role, they could not even do their work without their partners, and their staff, who are often much more experienced than them about ground realities.

Each project described here involved numerous people: researchers, research assistants and field staff, J-PAL leadership and staff, and the leaders and staff of NGOs. These individuals are sometimes, but not always, coauthors on a final paper, but their role never starts or ends with the paper. They are essential at every step, to prepare the project, implement it, and ensure follow through.

This lecture would not be complete if I did not attempt to list the people who participated in these projects. When I delivered the lecture in Stockholm, I asked those

present and associated with our movement to stand up. The written version gives me an opportunity to include many others who were not with us in person.

This list is necessarily partial: it would be impossible to give a complete list of the field staff. And, of course, this is only a small set of projects from a much larger body of work. But even this partial effort should give a good sense of the extent to which the essence of these projects is collective.

**Researchers:** Manuela Angelucci, Orazio Attanasio, Britta Augsburg, Rukmini Banerji, James Berry, Emily Breza, Shawn Cole, Ralph De Haas, Pascaline Dupas, Rachel Glennerster, Michael Greenstone, Rema Hanna, Heike Harmgart, Harini Kannan, Dean Karlan, Stuti Khemani, Cynthia Kinnan, Michael Kremer, Jordan Kyle, Leigh Lindon, Costas Meghir, Shobhini Mukerji, Andrew Newman, Benjamin Olken, Rohini Pande, Nicholas Ryan, Marc Shotland, Sudarno Sumarto, Michael Walton, and Jonathan Zinman.

**Leadership of J-PAL and IPA:** Iqbal Dhaliwal, Rachel Glennerster, Annie Duflo, Shobhini Mukherjee, Tithee Mukhopadhyay, Ruben Menon, Shagun Sabarwal.

**Leadership of Pratham, Spandana and AI Mama:** Fouad Abdelmoumni, Madhav Chavan, Rukmini Banerji, Pratima Bandekar, Lekha Bhatt, Shekhar Hardikar, Rajashree Kabare, Aditya Natraj, Padmaja Reddy, and many others.

**Policymakers and Government Officials:** Santhosh Matthew, Mitra Samya, the Indonesian National Team for the Acceleration of Poverty Reduction, Bambang Widianto, Suahasil Nazara, Sri Kusumastuti Rahayu, and Fiona Howell.

**Funding Organizations (Including the Key Staff Who Interacted with Us):** Amrita Ahuja (Marshall Family Foundation); Dana Schmidt (Hewlett Foundation), Smita Singh (Hewlett Foundation), Lynn Murphy (Hewlett Foundation), Ward Heneveld (Hewlett Foundation); International Initiative of Impact Evaluation; Institut Veolia Environment, DFID, The AFD, The Australian Government, the National Science Foundation, the Government of Haryana, the Regional Centers for Learning on Evaluation and Results, the ICICI corporation, the World Bank, the Alfred P. Sloan Foundation, and the John D. and Catherine T. MacArthur Foundation, the Sustainability Science Program (SSP), the Harvard Environmental Economics Program, the Center for Energy and Environmental Policy Research (CEEPR), the International Initiative for Impact Evaluation (3ie), the International Growth Centre (IGC), The Vanguard Charitable Endowment Program, Spandana, J-PAL, Agence Francaise de Developpement, Trust Fund for Environmentally and Socially Sustainable Development (TFESSD) and the DIME initiative at the World Bank.

**Key Staff Members and Research Assistants for Those Projects:** Parul Agarwal, Angela Ambroz, Adie Angrist, Vipin Awatramani, Sugat Bajracharya, Tamayata Bansal, Bruno Barsanetti, Susanna Berkouwer, Jim Berry, Shaher Bhanu Vagh, Nandit Bhatt, Ozgur Bozcaga, Janjala Chirakijja, Logan Clark, Ofer Cohen, Aparna Dasika, Anupama Deshpande, Diva Dhar, Eric Dodge, Madeline Duhon, Leonardo Elias, Harris Eppsteiner, John Firth, Blaise Gonda, Nick Hagerty, Jonathan Hawkins,



Zoe Hitzig, Shehla Imran, Seema Kacker, Dan Keniston, Nurzanty Khadijah, Chaerudin Kodir, Dhruva Kothari, Gabriel Kreindler, Sanjib Kundu, Zakia Lalaoui, Christian Larroulet, Alyssa Lawther, Eric Lewis, Taylor Lewis, Tracy Li, Yuanjian Li, Adrien Lorenceau, Lina Marliani, Jonathan Mazumdar, Richard McDowell, Jacqueline Merriam, Aditi Nagaraj, Sam Norris, Purwanto Nugroho, Aurélie Ouss, Cecilia Peluffo, Mukesh Prajapati, Manaswini Rao, Kevin Rowe, Hector Salazar Salame, Mitra Samya, Wayne Sandholtz, Paribhasha Sharma, Kartini Shastry, Joseph Shields, Marc Shotland, Zakaria Siddiqui, Bondan Sikoki, Freida Siregar, Stefanie Stantcheva, Sneha Stephen, Laura Stilwell, Cecep Sumantri, Yuta Toyama, Yashas Vaidya, Pankaj Verma, Melanie Wasserman, He Yang, Fatim-Zahra Zaim, and Gabriel Zucker.

## REFERENCES

- Angelucci, Manuela, Dean Karlan, and Jonathan Zinman.** 2015. "Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco." *American Economic Journal: Applied Economics* 7 (1): 151–82.
- ASER Centre.** 2015. *Annual Status of Education Report (Rural) 2014*. New Delhi: ASER Centre.
- Atanasio, Orazio, Britta Augsburg, Ralph De Haas, Emla Fitzsimons, and Heike Harngart.** 2015. "The Impacts of Microfinance: Evidence from Joint-Liability Lending in Mongolia." *American Economic Journal: Applied Economics* 7 (1): 90–122.
- Augsburg, Britta, Ralph De Haas, Heike Harngart, and Costas Meghir.** 2015. "The Impacts of Microcredit: Evidence from Bosnia and Herzegovina." *American Economic Journal: Applied Economics* 7 (1): 183–203.
- Banerjee, Abhijit.** 2007. *Making Aid Work*. Cambridge, MA: MIT Press.
- Banerjee, Abhijit.** 2019. "Field Experiments and the Practice of Economics." Nobel Prize Lecture. <https://www.nobelprize.org/prizes/economic-sciences/2019/banerjee/lecture>.
- Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton.** 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives* 31 (4): 73–102.
- Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani.** 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy* 2 (1): 1–30.
- Banerjee, Abhijit, Emily Breza, Esther Duflo, and Cynthia Kinnan.** 2019. "Can Microfinance Unlock a Poverty Trap for Some Entrepreneurs?" NBER Working Paper 26346.
- Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden.** 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122 (3): 1235–64.
- Banerjee, Abhijit, and Esther Duflo.** 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: PublicAffairs.
- Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan.** 2015. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." *American Economic Journal: Applied Economics* 7 (1): 22–53.
- Banerjee, Abhijit, Rema Hanna, Jordan Kyle, Benjamin A. Olken, and Sudarno Sumarto.** 2018. "Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia." *Journal of Political Economy* 126 (2): 451–91.
- Banerjee, Abhijit, and Andrew F. Newman.** 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy* 101 (2): 274–98.
- Banerji, Rukmini.** 2019. "Banerjee and Duflo's Journey with Pratham." *Ideas for India*. <https://www.ideasforindia.in/topics/human-development/banerjee-and-duflo-s-journey-with-pratham.html>. November 13, 2019.
- Banerji, Rukmini, and Madhav Chavan.** 2020. "A Twenty-Year Partnership of Practice and Research: The Nobel Laureates and Pratham in India." *World Development* 127.
- Co-Impact.** <https://www.co-impact.org/> (accessed January 27, 2020).
- Crépon, Bruno, Florencia DeVoto, Esther Duflo, and William Parienté.** 2015. "Estimating the Impact of Microcredit on Those Who Take it Up: Evidence from a Randomized Experiment in Morocco." *American Economic Journal: Applied Economics* 7 (1): 123–50.



- Deaton, Angus, and Nancy Cartwright.** 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine* 210: 2–21.
- Duflo, Esther.** 2017. "Richard T. Ely Lecture: The Economist as Plumber." *American Economic Review* 107 (5): 1–26.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer.** 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (5): 1739–74.
- Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan.** 2013. "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128 (4): 1499–1545.
- Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan.** 2018. "The Value of Regulatory Discretion: Estimates from Environmental Inspections in India." *Econometrica* 86 (6): 2123–60.
- The Economist.* 2009. "A Partial Marvel." <https://www.economist.com/finance-and-economics/2009/07/16/a-partial-marvel>, July 16, 2009.
- Field, Erica, Rohini Pande, John Papp, and Y. Jeanette Park.** 2012. "Repayment Flexibility Can Reduce Financial Stress: A Randomized Control Trial with Microfinance Clients in India." *PLoS ONE* 7 (9): e45679.
- Field, Erica, Rohini Pande, John Papp, and Natalia Rigol.** 2013. "Does the Classic Microfinance Model Discourage Entrepreneurship among the Poor? Experimental Evidence from India." *American Economic Review* 103 (6): 2196–2226.
- Glewwe, Paul, Michael Kremer, and Sylvie Moulin.** 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* 1 (1): 112–135.
- Giné, Xavier, and Dean S. Karlan.** 2014. "Group versus Individual Liability: Short and Long Term Evidence from Philippine Microcredit Lending Groups." *Journal of Development Economics* 107: 65–83.
- Hussam, Reshmaan, Natalia Rigol, and Benjamin Roth.** 2016. "Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design in the Field." Private Enterprise Development in Low-Income Countries Research Paper.
- Innovations for Poverty Action.** 2018. "Evaluating the Teacher Community Assistant Initiative in Ghana." <https://www.poverty-action.org/study/evaluating-teacher-community-assistant-initiative-ghana> (accessed January 27, 2020).
- Kapur, Avani, and Lorenza Icaza.** 2010. "An Institutional Study of Read India in Bihar and Uttarakhand." J-PAL Working Paper.
- Karlan, Dean, and Jonathan Zinman.** 2011. "Microcredit in Theory and Practice: Using Randomized Credit Scoring for Impact Evaluation." *Science* 332 (6035): 1278–84.
- Meager, Rachael.** 2018. "Aggregating Distributional Treatment Effects: A Bayesian Analysis of the Microcredit Literature." Unpublished.
- Meager, Rachael.** 2019. "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments." *American Economic Journal: Applied Economics* 11 (1): 57–91.
- Muralidharan, Karthik, and Paul Niehaus.** 2017. "Experimentation at Scale." *Journal of Economic Perspectives* 31 (4): 103–24.
- Pritchett, Lant, and Justin Sandefur.** 2013. "Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix." Center for Global Development Working Paper 336.
- Sharma, P. and A. Deshpande.** 2010. "Teachers' Perception of Primary Education and Mothers' Aspirations for Their Children: A Qualitative Study in Bihar and Uttarakhand." J-PAL Working Paper.
- Smith, Matthew.** 2017. "Leave Voters Are Less Likely to Trust Any Experts: Even Weather Forecasters." YouGov. February 16, 2017. <https://yougov.co.uk/topics/politics/articles-reports/2017/02/17/leave-voters-are-less-likely-trust-any-experts-eve> (accessed January 27, 2020)
- Tarozzi, Alessandro, Jaikishan Desai, and Kristin Johnson.** 2015. "The Impacts of Microcredit: Evidence from Ethiopia." *American Economic Journal: Applied Economics* 7 (1): 54–89.