

Field Testing in Development Economics

Education, microcredit, health policy.... How can we really measure the effectiveness of a public policy? Esther Duflo talks about the principles of the experimental method that she has developed and perfected in several situations around the world.

Esther Duflo is a development economist at the Massachusetts Institute of Technology (MIT). In 2008-2009 she held the *Savoirs contre pauvreté* (Knowledge against Poverty) chair at the Collège de France in Paris. She is a co-founder of the Abdul Latif Jameel Poverty Action Lab. She defends the use of *in vivo* experiments in economics. Modelled on the clinical trials that are carried out in the “hard sciences”, these randomized experiments have certain advantages, notably that they can control the parameters, to avoid skewed results. However, they raise some ethical and epistemological questions. In this interview (in Paris, January 2009), Professor Duflo reflects on the randomized approach in economics, on the kinds of projects that are pursued using this approach, and on managing such projects. She also responds to some of the questions raised by this approach.

Assessing Development Policies

Books & Ideas: As the 2008-2009 *Savoirs contre pauvreté* Professor at the Collège de France, you have been presenting your work on development economics. How would you describe the method of randomized experiments?

Esther Duflo: It is a method used to try to assess the impact of a program or a project. My work focuses especially on the combat against poverty in developing countries, in such areas as education, health, corruption, credit, and so on. But these methods can be and have been used in other countries, too.

The basic idea is to approximate as closely as possible the clinical trial method. You compare people who have had a treatment – for a clinical trial this would be a new drug – with people who have not had it. To do this we make every effort to ensure that the two groups of people are in other respects as comparable as they can be. In reality, what one discovers when comparisons are made between beneficiaries and non-beneficiaries of a program – for example, the construction of a school – is that the way programs are allocated generally means that the two groups are not at all comparable. For example, you can put schools in places where the people there most want them, in which case the level of education will rise, or you can put them in places where the people there most need them, and in that case the result will be less impressive.

The aim of a randomized experiment is to work with field partners – for example, NGO’s, local governments, or private companies who want to implement a program – to build

into the program conditions that will initially make beneficiaries and non-beneficiaries completely comparable. First you define a sample, for example 200 villages where some schools are going to be built, and you randomly choose the places in which to put these schools. For example, if an NGO has enough money to build 100 schools, you select 200 villages instead of 100 (which the NGO would have selected in any case). And afterwards, you collect data on all 200, so that you can compare for example school attendance in the two categories of village. Then, in general, after the experiment is over, schools are built everywhere.

Books & Ideas: For the most part, you carry out these development aid programs and these experiments through the structure of the Poverty Action Lab (J-PAL), which currently has about a hundred projects on the go. What kinds of projects do you do?

Esther Duflo: There is quite a range of projects, from education projects, in which we are interested in both access to and quality of schools, to health projects, where we are less interested than doctors are in the effectiveness of various treatments than in understanding people's health practices and the features of available health facilities, and how these facilities could be improved. There are programs concerned with the credit market, or more broadly with access to financial services such as savings, credit and insurance. Others concern the governance of corruption, for example the impact of particular rules and decisions in decentralized governments; for instance, is it better to have votes or meetings? And there are other programs that can't easily be categorized, for example trying to measure the level of discrimination in a society.

Books & Ideas: How are things initiated, at the government level? Are you contacted by states and local governments? Or are you proactive? And in the definition and implementation of a project, how are the different levels of decision making related, especially involvement with the local public authorities?

Esther Duflo: For a project of this kind, three things are always necessary. First, a field partner who wants to implement the project, and who wants an assessment, in spite of the pressures that this implies. This field partner can be a government, an NGO, or a private company. Next, you need a researcher or a research team whose interest in the project is strong enough to make them want to do it. And finally, you need someone to pay for it.

You need to have all three of these things united, and it's not always the same one that comes first. There have been projects on which a researcher or research team really wanted to work, for example on the impact of microcredit, because there had not yet been any assessments of this kind. In other cases, it is the field partner who really wants a project to be assessed; that might be an NGO, for example an Indian NGO contacting us to see if we can work with them on ways to improve their health facilities. Or take another example. You were asking about governments. The state government of Rajasthan (the Police and the Home Department) contacted us to find out if we would like to work with them to find ways to improve the effectiveness of the police, to reduce corruption, and to improve the public image of the Rajasthan Police. We replied, saying "here's what we can do in working with you." We began by looking at some ideas in the literature, and talking with police officers, judges, user associations, and so on, to find things to try; then we tested them, so we could afterwards examine what worked and what didn't. And then there is a third scenario, in which the initiative comes from the person with the money, who is interested in the impact of some particular thing. For example, the French Development Agency (*L'Agence française du*

développement) finances a lot of microcredit projects, to find out more about their various activities in that sector.

The starting point can be any of the three I just mentioned. But after that, there's going to be such close collaboration among the three parties – field partner, finance provider, and researcher – that you can never talk about a project organized by just one of them.

Books & Ideas: Does this collaboration occur smoothly, or are there conflicting interests at some point in the process?

Esther Duflo: There are always clashes – small clashes – but they are never about fundamentals. I've never found myself in a situation where the results were so unsatisfactory that the initiator of the project wanted to bury it. That's because we never get involved without being extremely open-minded about the result; above all, we insist on a procedure with a real result. If you just want an assessment to show that a project is working, you can always find someone to do that; that's always going to be a much easier and less expensive option, and it will give you exactly the result you want. When a partner wants to participate in our procedure, they know what they want and what they're getting into.

The clashes and conflicts of interest that might arise in following the procedure are much more localized. For example, in the normal course of their activities, field partners habitually adapt themselves to situations; say they'd decided to work at a certain place, but there is a problem, in that the local population turns out to be not very cooperative, so they change to another place. We ask them not to do that. So there can be conflicts between the need to do the project in the most effective and least expensive way from the organization's point of view, and the need to respect the proper procedure; but we always manage to adapt. Correct procedures must be respected in the main; and afterwards, in analysing the data, we can take into account differences from the ideal laboratory situation. On the other side, field partners are well aware that the situation is a bit special, so each side adapts.

Books & Ideas: How much do these projects cost, and who pays for them?

Esther Duflo: The cost varies enormously. If you do a project in urban schools, all you need to do is go into the schools to organize testing the children, which is not very expensive. That can be done for €50,000 per year. If you do a project in which you measure the impact of AIDS prevention programs on the rate of infection, you have to have enormous samples, because the rate of HIV infection fortunately is not very high, and because all these people are very isolated in the countryside. That could very quickly mount up to two million dollars.

As to who pays, that's a good question, because it is important to remember that the assessment is a prospective tool, not a retrospective one: the lessons we draw from one program – whether it's a failure or a success – are general. The knowledge gained from these kinds of assessment is a public good, and therefore should be financed not just by the people with a specific interest in the project, but by institutions that finance public goods. For example, the World Bank finances a lot of assessments, by us and others. Research institutions, the same ones that finance clinical trials (for example the National Institute of Health), and foundations like the Hewlett, Gates, and MacArthur Foundations finance programs of both research and action. Occasionally it is small foundations, or even private donors, interested in something specific, for example Veolia's interest in the importance of

water in Morocco. There are also some bilateral donors (the French Development Agency among others).

The Ethics of Doing Experiments

Books & Ideas: This new experimental method in development economics has prompted a number of questions and debates. First of all, doesn't the implementation of some of these projects raise ethical problems? Isn't there a problem in the fact that a potentially beneficial project is assessed by arbitrarily deciding that some people but not others will receive treatment? Is it possible that as a result of a project some people in the treatment group will end up in a worse situation than before, and if so, are there any compensation procedures?

Esther Duflo: Since it's to do with research, in which all the researchers are academics, ethical issues are looked after by university ethics committees (which unfortunately do not exist in France, but do in the United States). For example, all of the projects that I participate in are reviewed by the MIT ethics committee, which is the committee that handles ethical problems in clinical trials, and actually in all research – not just experiments – that involve human subjects. All of these ethical issues are considered in the framework of protocols that have been set internationally, protocols that appear for example in the Belmont Report in the United States, which are perhaps not followed worldwide, but which the ethics committees apply, at least in all of the countries that I know about. In general, a project receives advice from the ethics committee of all of the universities with which the researchers are involved, as well as from the ethics committee of the country where the research is done. For example, in Kenya the ethics committee that deals with all of the experiments and clinical trials also handles ethics questions for the kind of projects that we do.

There are also some rules of common sense, one of which is to limit as much as possible the risks to people participating in the experiment. In the kind of projects that we do, there are actually very few negative effects and risks, because we don't ever give out medicines (or if we do they are medicines that are already well known). In any case, part of the planning to protect subjects is always to ask the question "What harm might come to them?" In our case, the main problem is preserving confidentiality, and avoiding situations in which information held by researchers could potentially harm subjects if someone else got hold of it. We take great pains to respect personal confidentiality and freedom: people are not obliged to participate in an experiment, and can be beneficiaries of a program without taking part in the experiment. For example, there might be people in the treatment group who refuse to respond to research questions, but since they live in the village where the program is being implemented by the NGO, they nevertheless have the same right of access to it. We are very careful to respect this freedom.

Then there is the question whether we can deny people the benefits of a program. In general, this does not happen, because we are working in situations in which there are already so many budgetary and implementation capacity restrictions in the programs. For example, when we work with NGOs, they in any case have very limited budgets. If they haphazardly select 100 villages, that's 100 villages full stop; but if their program is part of an experiment, we select the 100 at random from 200 villages, which makes it possible to know whether or not the program has worked, and if it does, that creates a powerful argument for its extension. Having said that, we often try to work where there is a gradual extension of the program, so the people who constitute the control group are in fact just later beneficiaries than the treatment group. For example, in an AIDS training program, the training takes place in waves,

and we just organize the wave that gets the training done in a certain number of schools, teachers first and others later on. In general, that speeds up the process, in comparison to what would otherwise happen, because there could well be funding that is specifically devoted to the research, which can make things go more quickly than if things had been left as they were. So in practice, we do not experience any difficult ethical issues.

Or, if there are any, we do not do the project. For example, an Indian NGO asked us to work with them on their support program for severely malnourished children. They identify such children, then they give to the mothers both advice and food supplements. We were asked to assess this program. I refused, because the moment such children are identified, ethically one cannot not nourish them; so it was ethically impossible to participate in this program. Another example is organizations that deal with post-conflict refugees, for example in Liberia, in Congo-Brazzaville, and so on. At the moment, these organizations are asking themselves a lot of questions about the effectiveness of what they do, and are wondering how to do these things better. They are very attracted to the experimental approach, so they invited me to talk to them about it. I told them that often they are in no position to use the experimental approach because if they are handling a refugee camp, they are taking care of their refugees and cannot treat people in the camp differently; it's neither ethically nor practically possible. So there are some barriers. We don't get involved with situations like that. (That said, I think the significance of such barriers is exaggerated.) There are some very clear limit cases, to be avoided. Another limit, also very clear, is that we want to do no harm to people.

How Can the Conclusions of Local Experiments Be Generalized?

Books & Ideas: What is being measured in these randomized experiments? These programs are always implemented in a particular country, in particular villages, in particular conditions; how much generality is there in the results thus obtained? Do we learn things very particular, very relative to the country in question, or is it possible to generalize from the results of this research?

Esther Duflo: That is a crucial question, because if there were no generalization, i.e. if what we learned from a hundred villages in Kenya could not be generalized to apply to a hundred villages nearby, there would be no reason to undertake these projects. They are useful only for the general lessons that can be drawn from them; it is those lessons that justify the expense and efforts. The question of context is always relevant. As Montesquieu says, there are no good or bad laws, there are only laws that are good or bad in their context. This is equally true for these programs. On the other hand, we shouldn't push that too far. Remember Hume, who wondered if it is possible to learn anything general. We see that the sun rises every day, but that does not in the least imply that the sun will rise tomorrow, without a theory about the relationships between stars and planets. It's the same for these programs: you can't generalize a particular experiment – or a particular observation (experimental or not) – without a theoretical framework, implicit or explicit. Still, we all have such a theoretical framework. In fact, it's the very basis of political economics: political economics would not be possible if we thought that a particular experiment regarding someone had absolutely no relevance to that someone's neighbour. It is simply not thinkable that there be no possibility of learning something general from a particular experiment. Yes, context counts, but a theory about what explains the impact of a program is still what lets us decide if it will work in one place or another. For example, there can be very simple theories such as physiological theories.

Michael Kremer assessed the impact of deworming children: intestinal parasites make them ill, so if you give them a pill to cure them, they will be less ill and therefore can go to school more often. Presumably that finding can be generalized. In other cases, for example getting people to monitor their teachers and nurses, presumably the possibility of organizing a large number of people depends on having a democratic context. That is what makes it possible to predict that something is going to work for example in Kenya but probably not in India.

Books & Ideas: Do you raise this question about generalization and context before, after, or during the experiments?

Esther Duflo: A bit of each. We really prefer to raise the questions as soon as possible. To be quite honest, what often happens is that beforehand, on the basis of the theory that we have about the impact of a program on a series of variables, we make predictions. And then, on all of the intermediate variables that are anticipated, we make the assessment. Take for example the impact that women being in power has on political decisions, which I first studied in West Bengal. I started with a rather simple idea: a woman in power will make political decisions that resemble – that better reflect – the preferences of women more than those of men. Therefore, first of all we had to measure the political preferences of women and men, which we did by studying the complaints filed by them. Then we looked at the effects, the decisions that women and men politicians make. And we saw in this West Bengal example that women put more emphasis on drinkable water and roads than they did on education. So our prediction was that there will be more water and roads and fewer schools. Afterwards we tried to replicate that somewhere, to make sure that it was not just valid for Bengal, where women have a lot of power. In another and different state, Rajasthan, where there are a lot of young girls murdered, etc., the prediction was the same: women politicians will do more things that agree with what women want. So we measured what the women wanted there, which was not the same as in Bengal; they too were interested in water, but not at all in roads. So our prediction was now different: where there are more men, there will be more roads and less water. We collected the data, and found that that was the case. Thus we had different results, but we knew why, because we had that intermediate variable, the variable of preferences. So the prediction was not that with more women you'll have more water, but rather that there will be preferences better corresponding to what women want. That is a prediction at a more general level, which we can measure.

Books & Ideas: For a long time now, development economists have been interested in macroeconomic determinants – development requires wealth, wealth requires growth – and there has been an effort to identify the determinants of growth. The experimental approach is by definition conducted at the level of the behaviour of agents, i.e. persons. How does this microeconomic research applied to development economics relate to a much more macroeconomic approach?

Esther Duflo: There are two answers to that. One is to say that you need to shift up, to try to do experiments at the macroeconomic level, not necessarily at the country level, but sometimes at the market level. The trouble with a microeconomic experiment is that it gives us some parameters – of elasticity, productivity capital, etc. – but they are obviously in partial equilibrium. If there are general equilibrium effects, by definition you cannot spot them while doing an experiment within a particular market. Take school privatization as an example. In many countries, including the United States but also in many poor countries, the question arises whether the state should just finance schools, without trying to build them, leaving that to the market, and give vouchers to people so they can buy private education. Here, the first

question is a partial equilibrium question: if you get a voucher and I don't, in comparing us it could be seen that you obviously have more chance of going to a private school and having better results. That shows us a partial equilibrium. This is an experiment that has been conducted in Columbia, for example, where the private schools are better than public schools, but that does not show us that we should stop funding public schools and give vouchers to everyone, because once everyone receives a voucher you've changed the basic characteristics of the market. Private schools will change. So will the remaining public schools (which will continue to exist, since people can use their vouchers in them). Perhaps they'll have much worse pupils, perhaps better ones (perhaps there will be more competition). So we get no predictions about moving from the current system to a comprehensive voucher system, which really could go either way. But what we can do is to say, let's do an experiment at the market level. There is one on this voucher issue currently underway in Andhra Pradesh, where an Indian foundation (which belongs to the founder of Wipro, one of the big computer companies) gives vouchers to people right across the market, so general equilibrium effects are captured because of the level of the experiment. That's the first answer.

That answer has some limitations. It works for the school market because people won't travel very far: schoolchildren have short legs! But for example the job market is different; in it people will be much more prepared to move in order to find a better job. That's why we can learn nothing about a political program from job market experiments that are done locally. So it becomes important to combine an experimental approach that gives us some useful parameters with macro models. And here, we are really in the early days of this art – and it is indeed more an art than a science. This is something that for example Robert Townsend does on questions about growth, when he uses microeconomic models – usually models of career choice (are you going to be an entrepreneur rather than an employee, etc.) – that take into account the pressures of the credit and insurance markets. From these models he gets microeconomic estimates and various parameters that don't yet figure in the experiment but that could in the future, and then he tries to make his model function and to calibrate it to an economy, in this case the economy of Thailand. For fifteen years now, he has been collecting data on thousands of Thai households; he's watched them grow – in this period there's been a lot of growth in Thailand – and he's tried to see if this model, with the right parameters, replicates the performance of the economy, both in terms of growth and in terms of inequalities, career choices, geographical inequalities, migrations, and so on. I think this is very promising research.

Books & Ideas: Are these models good at making predictions?

Esther Duflo: Yes, they do predict pretty well. It's possible to do much better, but on the basis of what we know how to do, we already go much farther than if we were using a model on its own, without any of these guidelines. There's been very clear progress, which has come from combining model calibrations with microeconomic estimates. Better, more precise estimates will produce a more useful macroeconomic calibration. And that seems to me more promising than trying to do regressions, i.e. to assess a statistical model on the basis of macroeconomic data that is too aggregated, because we know that all theorems that allow aggregation depend on assumptions, for example that resources always flow to the best use, and so on. We know that it is not always reasonable to make such assumptions in developing countries, so it's not even worth trying to assess these models without taking that into account.

Expertise and Politics

Books & Ideas: Along the same lines, suppose I'm a head of a government with limited resources, in charge of the country's policies for economic development. Will the results that you get from the experimental method enable you to advise me in choosing among pursuing policies in, say, education, health, access to credit, and agrarian reform?

Esther Duflo: At that level of generality, I'd prefer not to respond to that question. It's a political issue, i.e. it's more to do with a society's choice, than with what can be provided by a technical adviser (which is what I see myself as). Where I could help you as the head of a developing country is after you have decided that you are going to prioritize, say, education.

Books & Ideas: Does that mean that we don't really have a clear idea about the returns on the various different policies?

Esther Duflo: We do have very clear ideas – this is what we really work on – regarding the respective returns on different policies that try to achieve the same result. For example, we are beginning to know extremely well the costs and benefits of a dozen different approaches to schooling so that children will have the maximum number of school days. So a choice can be made among these approaches. It's the same for health: if you want to vaccinate children, or even improve the access to preventive care, which is very low in developing countries, there are lots of ideas and approaches that can be subsidized – e.g. having information policies, or having more Primary Healthcare Centres. We are certainly able to compare all of these approaches. And the more projects that get done, the more things there will be to compare in order to choose. There are also cost-benefit ratios, dollars per dollars; for example one could say that a better-educated child will be earning for a longer time, so if I invest x dollars in school books, that will ultimately bring in $x+y$ dollars in extra revenues. But to me such approaches seem tricky, insofar as there are other benefits than the precise dollar result of each policy. For example, education can be seen as a good in itself, as can health; and it is rather a society's choice to say what intrinsic value should be attributed to education or health, or to access to credit, or to insurance. So I think it would be illusory to say that there is a technical answer to this question. The question demands a political response, in the noble sense of the term political.

Books & Ideas: That brings us to the question of the role of the researcher in using these new *in vivo* methods in experimental economics. Does the role of the researcher who adopts the experimental method change completely?

Esther Duflo: In research in development economics, we have moved from very simple assessments and conceptual questions, to more ambitious, more demanding questions, which entail more sophisticated experimental designs. This means the creativity lies less in the statistics than in the economics. And the more that we know of economic theory, the better we are able to think up interesting experiments that get beyond the issues of context and so forth. Thus, finally the role of the researcher as economist does not change all that much: the researcher is no longer a pure statistician, but an economist, who reflects on economic theory and its implications. The statistics becomes more a matter of thinking up experiments, experiments that are not lacking in statistical subtlety. There is a branch of biology called biostatistics, which studies the ideals of experimental design: how to process the data, how to manage problems and assessments in the implementation of the design, etc. Here, the research mode does change a little, when compared with empirical economists sitting at their desks

gathering data to analyse. In other words, the physical organization of projects is different; these are big projects with big teams, from a fresh university graduate who will be spending a year in the field to set things up, to a postdoctoral research assistant, to a young professor, to an experienced professor who will be finding the money and making some initial contacts. All of these people have a place in every project. So these are projects that are perhaps more like biologists' projects, with one portion for researchers.

Books & Ideas: Now, regarding the researcher's role in the world and in public discussions, a moment ago you said that what you do is provide technical answers. At the same time, it appears that all of these projects are carried out with governments. Do political scientists have a different position?

Esther Duflo: In economics there is a tradition of refusing to comment on politics, and there are claims to describe the world as it is; but there is also a normative tradition, which our work belongs to, which thinks that economists in their particular field have things to say about the best way to tackle for example a market failure, or the reorganization of the operation of a public service. So we are not forbidden to have ideas. Either we ourselves have some ideas, in which case we make suggestions directly, or we use the knowledge held by the whole field to assess the ideas of others.

Therefore, since we work very closely with field partners, either NGOs or governments, we don't just act as an assessor when everything is all over, we get involved from the outset of the introduction of a program, and in general even before that, during the reflections about the program to be implemented. I always try to spell out very clearly that I want to be informed of the objectives that I am supposed to be helping to achieve, and it is here that there is a difference between the science and the politics: the politics has to supply the objectives. Once these have been formulated, you have to attain them as well as you can, and here the scientist might well have something to say. That is the first role of the scientist in the world. The secondary role is assessment. Which is not all that easy to implement: though it is not very difficult conceptually, in practice it is a bit difficult.

Books & Ideas: And for the last question: does your work finish with the completion of the assessment, or do you, out of concern for political advice, look after promoting the results, in a kind of lobbying, since you know that there are some things that work in attempting to spread these best practices?

Esther Duflo: That is exactly why we created the Poverty Action Lab (now the Jameel Poverty Action Lab), because while all researchers think their job is to write papers, it constitutes a great loss to always be assessing without getting involved in the follow-up, in order to make information available to decision makers. Our motto is "translating research into action", i.e. taking care of promoting the results. I don't know if one can call this lobbying, but it does mean making the results part of the political discourse. Of course, it is not just the results that decide the adoption of a policy, but nevertheless they do play a role. There is specific funding for that – there are donors specifically interested in it – for moving from a result to the adoption of a policy. For example, that explains the success of deworming. We have made tremendous efforts in this field, and once you know that its cost is very low and its effects on health and education very high, you can try to discuss it with people who make the pills, and with governments that can set up programs, and so on. There are several tens of millions of children who have benefited from deworming.

Another example is mosquito nets. There is a great debate about whether or not they should be distributed free of charge. A study by Pascaline Dupas and Jessica Cohen has shown that it is better to give them out for free, because many more people get them, and they are just as likely to use them. Effective coverage is much more important; with that, the mosquitoes go somewhere else. So if you give them out for free, the cost-benefit ratio is paradoxically better. This result persuaded Population Services International, a big social marketing NGO that distributes mosquito nets, to distribute them free instead of charging for them.

There are even simpler cases. For example we assessed the impact of a literacy program for children already in school who have not learned to read, by a big Indian NGO called Pratham. We found that this was really remarkably effective. We supported their application to the Hewlett Foundation and the Gates Foundation, and they eventually received enough money to set up the program in a third of India. This was an easy case, because there was someone ready to fund the program as soon as information about the results was available.

Further reading:

The experimental approach to development economics, A. Banerjee and E. Duflo
<http://econ-www.mit.edu/files/3158>

Publications by J-PAL
<http://www.povertyactionlab.org/papers/>

The new development economics. We shall experiment, but how shall we learn? , D. Rodrik,
<http://ksghome.harvard.edu/~drodrik/papers.html>

La révolution des randomistas
<http://www.ecopublix.eu/2009/01/la-rvolution-des-randomistas.html>

Chaire annuelle Savoirs contre pauvreté, Collège de France
http://www.college-de-france.fr/default/EN/all/cha_int/

Web page of Esther Duflo (on the MIT site)
<http://econ-www.mit.edu/faculty/eduflo/>

The Abdul Latif Jameel Poverty Action Lab
<http://www.povertyactionlab.com/>

Interview by Florian Mayneris, transcribed by Florence Brigant.

First published in French on laviedesidees.fr, 5 May 2009. Translated by John Zvesper with the help of the Institut Français. Published on booksandideas.net, May 30th 2013.

©booksandideas.net