

# Banerjee, Duflo, Kremer, and the Rise of Modern Development Economics\*

Benjamin A. Olken

MIT

May 2020

## **Abstract**

In 2019, Abhijit Banerjee, Esther Duflo and Michael Kremer received the Sveriges Riksbank Prize in Economic Sciences in memory of Alfred Nobel. These three scholars were recognized “for their experimental approach to alleviating global poverty.” This paper reviews the contributions of these three scholars in the field of development economics, to put this contribution in perspective. I highlight how the experimental approach helped to break down the challenges of understanding economic development into a number of component pieces, and contrast this to understanding development using macroeconomic aggregates. I discuss pioneering contributions in understanding challenges of education, service delivery, and credit markets in developing countries, as well as how the experimental approach has spread to virtually all aspects of development economics.

## ***Introduction***

Development economics, broadly speaking, seeks to answer two questions. The first question development economics poses is: why are some countries poor and some countries rich, and relatedly, what can poor countries do to grow and become richer in the future? The second question of development economics is: how do the many and varied phenomena studied by economists differ systematically in environments characterized by low levels of development?

---

\* I thank the three Laureates -- Abhijit Banerjee, Esther Duflo, and Michael Kremer – for countless conversations over the many years, including comments on this article. Special thanks also to Rema Hanna and Seema Jayachandran for helpful comments. All errors are my own. Contact information: bolken@mit.edu.

The field of development economics has changed dramatically over the past 25 years. Today, the field of development economics is largely a micro-empirical field, with countless studies on all aspects of economic development. While many micro-empirical approaches are used, development economics has led the way in using randomized field experiments to tackle a wide variety of development challenges. By comparison, 25 years ago, there was far less research in development economics – and what there was much more heavily dominated by theory, and to the extent it was empirical, it was much more heavily oriented towards macro aggregates.

The explosion in the field has been driven by the combination of two key, connected ideas: 1) the idea that one should focus on understanding the development challenges through the use of empirical micro studies to break down the development challenges into smaller pieces; and 2) the idea that one can achieve this in particular through running randomized controlled trials. The combination of these two ideas – the first a substantive approach to understanding development, and the second a methodological approach, has driven a complete revolution in the field.

While of course many scholars have been involved in this dramatic shift, the 2019 Nobel Memorial Prize was awarded to three individuals who have led this transformation, Abhijit Banerjee, Esther Duflo, and Michael Kremer, “for their experimental approach to alleviating global poverty.”

As a measure of the speed with which the field experiment based approach has swept the field of development economics, today, less than 25 years after Kremer’s first field experiment on child sponsorship was fielded in Kenya, the Abdul Latif Jameel Poverty Action Lab – a center

at MIT cofounded by Banerjee, Duflo, and Sendhil Mullainathan (and which I codirect) which seeks to advance the use of randomized evaluations – counts over 1,000 randomized evaluations either completed or ongoing by its member academics worldwide – including many by senior economists who were not originally trained in this type of work but who have begun to incorporate it into their toolkit. This is surely a substantial underestimate of the total number of such studies. The amount of high-quality empirical evidence used in development economics is almost unrecognizable compared to what it was before this revolution took off. And while randomized field evaluations appear throughout microeconomics, this change happened first and foremost in development economics: for example, a review of top economics journals from 2009-2013 found that there were more RCTs published in development economics than in all other fields of economics combined; development also had the highest percentage of published papers in top journals that were randomized trials (46 percent) (Finkelstein and Taubman 2014). I did a similar review of the *American Economic Review*, *Quarterly Journal of Economics*, *Journal of Political Economy*, *Review of Economic Studies*, and *Econometrica* twenty years earlier - from 1989-1994 – and there were zero – zero! – development RCTs published in those journals during that period. It's been a tsunami.

Of course, it is important to note that the idea of using randomized controlled trials in science is not new, nor do the works cited here represent the first use of randomized field experiments in economics. For example, prior to these three individuals' work there had been some high-profile randomized trials in the US, such as the Negative Income Tax Experiments carried out in the late 1960s and early 1970s, the RAND health insurance experiment (Newhouse 1993), and the Moving to Opportunity project (Katz, Kling et al. 2001), as well as other program evaluations sponsored by the US federal or state governments (see, e.g., Gueron 2017). There

were also a few high-profile programs randomly evaluated in developing countries at the same time these authors randomized trials were being started, most notably Mexico's Progresa (Gertler 2004, Schultz 2004, Skoufias 2005), which was started in 1997 under the leadership of Santiago Levy. And, of course, randomized trials have been ubiquitous in the medical community for some time. But together the contribution of these three individuals has been to reshape development economics as a micro-empirical field, using randomized trials as a critical tool in this endeavor. This has, in turn, led to an explosion of empirical research on virtually all aspects of economic development.

While most of the coverage of this year's prize has focused on the methodological contribution – the use of randomized field experiments to shed light on both the challenges of economic development and various approaches to improving them – my goal in this article is to step back somewhat, and situate this approach and this shift as part of the broader development economics endeavor. I therefore begin with some intellectual background, setting the stage by discussing briefly some of the contributions to understanding the problem of development in the 1990s. I focus in particular on how development was approached from a macroeconomic perspective, which in some sense teed up the micro-experimental revolution. I then review some of Banerjee, Duflo, and Kremer's early pioneering work using field experiments, focusing in particular on the following three thematic areas: education, governance of local service providers, and market failures. I close with some discussion of recent innovations and directions in the field, showing how the basic tool pioneered by Banerjee, Duflo, and Kremer – intentionally designed randomized experiments to understand development questions – has been adapted in countless new directions.

Before beginning, it is also worth saying a bit personally about these three Laureates. I was fortunate to get to know all three of them when I was a PhD student at Harvard in the early 2000s. Kremer was the chair of my PhD committee, and both Banerjee and Duflo provided outstanding advice on my dissertation from MIT; for the past decade or so I have also been fortunate enough to have Banerjee and Duflo as colleagues at MIT. Part of the reason their work has been so influential is that all three of them are deeply dedicated to training and supporting an entire generation of students and scholars who have run with these ideas in all kinds of ways. I certainly benefited in innumerable ways from all three of them, and am deeply grateful for their mentorship and support – as are so many of our colleagues throughout the profession.

### ***I. Background: Macro Development Research in the 1980s and 1990s***

I begin by stepping back to review some other important contributions from the 1980s and 1990s to set the stage for the contributions I discuss in more depth. These contributions are important to understand since, in some sense, the extent of the modern development economics revolution can be best understood by first understanding where the field was beforehand.

Throughout much of the 1990s, many of the questions about development approached the topic from the perspective of economic growth. This, after all, seems intuitive: if one is interested in why some countries are rich, and some are poor, it makes sense to start by examining this question at the broadest level. This of course builds on a long- and well-established tradition of understanding development from this perspective going back many years.

One important contribution in the literature on economic growth and development is the idea of endogenous growth theory, developed in the late 1980s, for which Paul Romer was

awarded the 2018 Nobel Memorial Prize. The point of this line of research is that, as was shown in growth models going back at least to Solow (1956), the long-run rate of economic growth is driven by the rate of growth of total factor productivity. Solow viewed this as exogenously driven technical change, but the point of these new models was that technical change is instead endogenously determined by what happens in the R&D sector. Key contributions include Romer (1986), Romer (1990), Grossman and Helpman (1991), Aghion and Howitt (1992), as well as many others. These contributions focused on what happens when ideas are nonrival and endogenize the production of ideas.

These contributions were critical to help understand the long-run evolution of growth at the technological frontier, but by themselves the benchmark models are not particularly helpful for understanding the process of economic development for countries that are far from the frontier. That is, for more developing countries, growth is determined primarily by catch-up to the frontier rather than technical innovation at the frontier. To the extent that these endogenous growth issues are relevant for development, what is important is that they predict that technical change will be biased systematically in favor of rich countries, where the returns to R&D activities will be higher. For example, if technology is specific to a particular capital-labor ratio (as the model of Basu and Weil (1998) and others) or specific to a particular skill mix (as in the models of Acemoglu (1998) and Acemoglu (2002)), then since the markets for high capital-labor ratio or high-skill intensity innovation are largest, the endogenous R&D mechanisms described in Romer (1990) and others will lead to more innovation of the types of technologies relevant for developed economies, and hence worse outcomes for developing countries.

A second area which received substantial attention in the 1990s was the empirical determinants of economic growth from a macro perspective. This literature had two somewhat

separate, though clearly related, components. The first component was the “development accounting” (and related “growth accounting”) debate, i.e., the debate about whether the macro facts about development are largely about capital and human capital accumulation (e.g. Mankiw, Romer et al. (1992), Young (1995)), or about total factor productivity (TFP) (e.g. Klenow and Rodriguez-Clare (1997), Hsieh (2002)); see Caselli (2005) for an excellent summary of this literature. Mankiw, Romer et al. (1992) in some sense touched off this debate by claiming that observable factors (in particular physical and human capital accumulation, plus population growth) explain most of economic growth. Subsequent refinements of this approach, however, argued that capital, including human capital, explain much less of cross-country growth differences than one might have thought, and hence there is more of a role for productivity in explaining human capital differences. This substantive idea – that actually an important part of the development problem is that poorer countries are less efficient at using the resources they have than richer countries – motivates much of the modern empirical work on development economics, as I will return to in more detail below.

The second major component of the empirical macro-growth literature was the work on the empirical determinants of growth. This was set off to a large extent by Barro (1991). (One way to think of the first debate touched off by Mankiw, Romer et al. (1992) as being essentially about the  $R^2$  in a growth regression, whereas the second debate touched off by Barro (1991) is about the coefficients on the regressors in a growth regression.) This paper looked, empirically, at the determinants of economic growth from a regression perspective. Substantively, the key point of the Barro paper was to test the idea that falls out of the neoclassical growth model that, conditional on the long-run determinants of the steady-state, countries that are further behind will grow faster as they converge to their steady state. This is an important idea: it suggests that if

you get the long-run fundamentals right, countries grow rapidly to reach the frontier. This approach, for example, explains both Japan's rapid growth in the post-war period and its slowdown once it reached US levels of income (though it does not explain why Japan has grown more slowly than the US since the 1980s, which must be a different phenomenon); applied to China, it predicts rapid growth for a long time to come as long as the basic policies are right, but an eventual growth slowdown once it approaches the frontier.

While there is strong evidence of this type of "conditional convergence" (see, in particular, Barro and Sala-i-Martin (1992) on convergence among US states), as a development phenomenon, the problem is that this approach doesn't have as much explanatory power as one might like. Two key critiques were Easterly, Kremer et al. (1993) and Pritchett (1997). Pritchett's provocative paper, "Divergence: Big Time", makes the simple point that while "conditional" convergence may be present in the data, it is swamped at the macro level by other forces, so that, in general, the development experience is explained by divergence between rich and poor countries, not convergence. Thus, over the long run, the "conditional" in conditional convergence ends up doing a lot of the work, and in fact the development experience has not been characterized by convergence. Easterly, Kremer et al. (1993) look over shorter periods (e.g. decades) with two main findings. First, they show that growth is not particularly persistent across decades. In the Barro framework, this would mean that the long-run determinants of the steady-state must have changed. But, second, they show that policy variables they can measure do not appear to change much across decades. So, the decade-to-decade variance in growth they observe isn't driven by Barro-type policies, either. Combined, these two papers argue that while conditional convergence may exist in and be important in some well-controlled settings (US states, European countries in the postwar period; perhaps China since the reforms of Deng

Xiaoping), it doesn't seem to have a lot of explanatory power for growth experiences of most developing countries over the medium run (e.g. decades) or the long run (e.g. a century).

In addition to the work on conditional convergence, the Barro (1991) paper spurred a series of papers looking for other empirical causes of growth. These "growth regressions" were all the rage in the 1990s (e.g. Mauro (1995) on corruption in growth, Easterly and Levine (1997) on ethnic heterogeneity and growth, etc.), but they quickly ran into two problems. First, there are more possible regressors than countries, so one quickly runs out of data. Second, it is very hard to establish causality since all these regressors are jointly and endogenously determined. In some ways, the culmination of this line of the literature was the Sala-i-Martin (1997) paper "I Just Ran Two Million Regressions," which tried to deal with the issues of more regressors than countries in a serious way, but also ends up pointing out the inherent challenges in the whole exercise.

These developments in the macro-growth side of development led to a bit of a conundrum. By the late 1990s, it was pretty clear that it was going to be hard to make progress on the development question empirically from a macroeconomic perspective. In some sense, this led to the emergence of modern, empirical micro development economics in the late 1990s and throughout the 2000s.

## ***II. Modern Development Economics and the Experimental Approach***

Given this background and the challenges of understanding development from macro aggregates, the question was: how can we make progress next? The answer the field has taken is to delve into the micro structure of development. Return, for a moment, to the macro development accounting equation, with an aggregate production function  $Y=AF(K,H)$ . The

micro approach has taken apart each component of A, K, and H, and sought to tackle them from a microeconomic perspective. For example, rather than trying to understand how, say, average schooling affects output at the country level (the approach taken in the macro growth regression and growth accounting perspective, as done by e.g. Bils and Klenow (2000)), the new micro approach asks how can we make progress on understanding the development problem by understanding the details of human capital from a microeconomic perspective, with separate studies seeking to understand why people don't have more human capital, why a given year of schooling in a developing country may be worth less than a year of schooling in a developed country, and how schooling in developing countries can impact more human capital.

This approach proceeded in a few distinct strands, with many important contributors, but Kremer, Duflo, and Banerjee played a crucial role in this transformation. In particular, starting in the mid-1990s, these three scholars pushed forward the methodological revolution which has subsequently swept through the field of development economics: the use of carefully designed field experiments to isolate each individual component of the development problem. The field experiment approach solved two problems at once.

First, a key challenge in many empirical studies is understanding causality; does X cause Y, or is it merely correlated with Y. While there have been dramatic improvements throughout economics over the past 30 or so years in isolating causality using a variety of econometric techniques, in particular focusing on so-called 'natural experiments', randomized experiments provide a convincing solution to this problem. In particular, since the variation in X is randomly assigned, we know by construction that there is no omitted variable that is influencing both X and Y. Put another way, since we know that which subjects were assigned to treatment and control by chance, we know that there is no other systematic difference between treatment and

control groups except for the experimental intervention given to the treatment group. Statistical methods then allow us to compute the probability that any observed difference we see between treatment or control could have occurred by random chance, as opposed to being truly ‘caused’ by the treatment.

Second, if the goal of the new micro-empirical approach is to break down a problem into its many component parts, by *designing* randomized field experiments, one can design a study in which the variation one is interested in captures one particular piece of the problem. Rather than ask a broad question such as ‘are people poor because they don’t have access to capital,’ one can start to disentangle the puzzle, asking about the returns to a particular type of program – say, microfinance – or even about the particular *features* of microfinance programs, such as group lending, regular meetings with loan officers, and so on.

The use of this approach has spread from initial simple trials on different inputs into the education production function into almost all areas of the field: agriculture, insurance, credit markets, tax, poverty policies, governance, and so on. In so doing, it has allowed scholars to investigate *both* of the core development economics questions I outlined above: why are some countries poorer than others (and what to do about that), and how do various economic phenomena we study elsewhere differ in developing countries.

Each of this year’s Laureates played a critical role in this transition. As I will discuss below, a number of the earliest of this new wave of trials, which sought to break down the development challenge into component parts that each could be rigorously evaluated and tested, were done by Michael Kremer and his coauthors, starting in the mid-to-late 1990s in western Kenya. These include, for example, trials evaluating the impact of various educational policies

such as deworming children (Miguel and Kremer 2004), providing educational inputs such as flip charts (Glewwe, Kremer et al. 2004), textbooks (Glewwe, Kremer et al. 2009), school meals (Vermeersch and Kremer 2005), child sponsorships (Kremer, Moulin et al. 2003), and changing the incentives for teachers (Glewwe, Ilias et al. 2010).

Banerjee and Duflo also began working on randomized trials during the late 1990s and early 2000s, in various combinations. For example, Banerjee, Jacob et al. (2004) launched an RCT in early 1997 in Udaipur, India that randomly added an additional teacher to half of sampled schools and investigated the impact on test scores. Banerjee and Duflo also started a large number of other early randomized trials trying to unpack different aspects of development, in various combinations (e.g., Banerjee, Cole et al. 2007, Banerjee, Duflo et al. 2008, Duflo, Kremer et al. 2008, Banerjee, Banerji et al. 2010, Banerjee, Duflo et al. 2010, Duflo, Kremer et al. 2011, Banerjee, Duflo et al. 2015), and which range in topics from education to fertilizer to immunization to microcredit.

More broadly, all three of these scholars worked together in various combinations to advance the broader idea that a series of experiments could unpack and make progress on the development phenomenon. They also did so through their writing to advocate for the new micro-founded development economics, with a particular emphasis on the use of a series of randomized trials tackling various aspects of the development problem to accomplish this (see, e.g., Kremer (2003), Duflo, Glennerster et al. (2004), Banerjee and Duflo (2005), Duflo and Kremer (2005), Duflo, Glennerster et al. (2008), Banerjee and Duflo (2009)); these are of course all published dates, but most of these were circulating as working papers for several years prior. Banerjee and Duflo were also instrumental in fostering the rapid spread of RCTs throughout development

economics with their cofounding of the Abdul Latif Jameel Poverty Action Lab (together with Sendhil Mullainathan) in 2003, with an explicit mission to help support this type of research.<sup>1</sup>

These scholars have also led a wide range of methodological innovations in broadening the types of questions that can be answered by these types of trials, moving beyond simple “treatment vs. control” analysis to more nuanced designs that shed light on more subtle questions. For example, Miguel and Kremer’s (2004) study of deworming in Kenya shows how to use the fact that different numbers of children live nearby treated vs. control schools to measure the ‘externality’ from deworming treatment. To take another example, the study of job placement assistance by Crépon, Duflo et al. (2013) showed how, by intentionally varying the percentage of people in a labor market randomly offered assistance, one can estimate not just the direct effects of the program on treatment groups, but also what the total effect on the market is, i.e. whether this leads to more job placement in aggregate. Both of these design innovations have become the pattern for scores of new studies on a huge range of topics. And these innovations continue. To name just a few recent examples, Duflo has been working with econometricians on how to use innovations in machine learning to predict heterogeneous treatment effects from randomized trials (Chernozhukov, Demirer et al. 2018), and Banerjee has been working with other theorists to articulate a theory that explains why randomization is so convincing, and exploring implications for experimental design (Banerjee, Chassang et al. forthcoming).

While I have framed the experimental revolution in modern development economics in the discussion above as, in some sense, a reaction to the challenges of approaching the

---

<sup>1</sup> They are not alone of course in founding centers to promote this type of research. For example, Dean Karlan co-founded Innovations for Poverty Action at roughly the same time as the Poverty Action Lab was founded by Banerjee, Duflo, and Mullainathan, and Ted Miguel, founded the Center for Effective Global Action (CEGA) at Berkeley several years later.

development economics problem from a macroeconomic perspective, it is also crucial to note that it builds substantially on other work that was happening at the same time in the microeconomics of development. This was happening in several important respects. First, many scholars, such as Pranab Bardhan, Dilip Mookherjee, Andrew Newman, and Debraj Ray (as well as Banerjee, in his pre-experimental days) were developing theoretical models to help understand a wide range of development phenomena, including public goods (and underprovision thereof), labor markets, credit constraints, insurance arrangements, elite capture, and more. These theories (and more) posed a wide range of questions and suggested avenues that the experimental approach was then able to investigate.

Second, scholars such as (for example) Andrew Foster, Mark Rosenzweig, Christina Paxson, John Strauss, Duncan Thomas, Robert Townsend, and Chris Udry, were already seeking to understand the micro development challenge using data in various (non-experimental) ways. The work of Angus Deaton, who was awarded the 2015 Nobel prize for his groundbreaking analysis of consumption, also was a crucial component, as it showed how to use micro data obtained by household surveys to shed light on welfare, poverty, and other questions. Finally, throughout the same period, the natural experimental revolution was also occurring outside of development economics, pushed forward by scholars such as Joshua Angrist, David Card, Alan Krueger, and many others. This approach sought to find ways to more carefully identify cause and effect using careful natural experiments. One can therefore also see the experimental revolution in development economics led by the three Laureates as a synthesis of these forces: theoretical work highlighting how the microstructure of markets in developing countries can lead to inefficiencies, empirical work seeking to test these ideas using non-experimental methods and

micro data, often collected through household surveys, and general increased attention that was happening at the time to more rigorous causal identification in econometric studies.

### ***III. Some Early Examples of the Experimental Approach***

In what follows, I illustrate this transition by in three thematic areas: education, governance of local service providers, and market failures, highlighting the substantive contributions of the three Laureates as well as early and important randomized trials they conducted. I intentionally focus here on some of the early innovations here to illustrate the start of this movement; I touch on some more recent new directions below.

#### *III.I Education*

The macro development literature asserted an important role for human capital.<sup>2</sup> Yet this literature had several challenges. First, in principle, the cross-sectional returns to education are biased, due to selection effects – for example, people with lower discount rates get more education, and the low discount rate itself may be correlated with higher returns in the job market. Second, the returns were measured all from years of schooling – and not all years of schooling are the same in terms of human capital acquisition. The latter suggested that understanding the educational production function was important to understanding the acquisition of human capital.

Duflo (2001) provided the first convincing micro evidence in developing countries on the first point, the returns to schooling. Her idea was to use a natural experiment, not a field

---

<sup>2</sup> For example, human capital issues figured prominently in both the analysis of Mankiw, Romer and Weil (who noted that one needs to augment the Solow regression with human capital to get a high  $R^2$ ) and Barro described above. And, growth accounting required understanding the returns to human capital, which was done largely by running cross-sectional Mincer regressions in as many countries as possible (Psacharopoulos 1973, Psacharopoulos 1985, Psacharopoulos 1994)

experiment – in the form of Indonesia’s massive school construction boom in the 1970s – to estimate the returns to schooling. The crucial difference between Duflo’s paper and previous work is that she uses plausibly exogenous variation in schooling – for a given district, those cohorts born prior to school construction tended to have low schooling, since schools were difficult to access, while those cohorts born after school construction tended to have high schooling, since schools had now been built in all villages. The fact that Indonesia determined school construction centrally, based on a formula, allowed her to argue that the change in school construction was exogenous to individual decision making or other trends, thus avoiding the identification problems that had plagued the previous literature. This paper swept the development world because it was among the first papers to apply rigorous micro-empirical techniques (of the sort that had been developed in the US labor economics literature, by Angrist, Card, Krueger, and others) to development problems, and as such, it really marked the beginning of a new wave of empirical research in development.

It was on the second front – understanding the heterogeneity in how years of schooling translate into human capital due to differences in the education production function – that randomized trials were introduced. Specifically, in the mid-1990s, Michael Kremer, working with an NGO, launched a series of randomized, controlled trials in Busia, Kenya that sought to test the impact of different school inputs. In particular, the early rounds of these experiments featured several such trials: a trial of the impact of child sponsorships, including textbooks and school uniforms (begun in 1994; see Kremer, Moulin et al. (2003)); two trials testing school inputs such as textbooks (begun in 1996; Glewwe, Kremer et al. (2009)) and flip charts used as teaching aids (begun in 1997; Glewwe, Kremer et al. (2004)), and two trials testing health

interventions, such as deworming children (begun in 1998; Miguel and Kremer (2004)) and school meals (begun in 2000; Vermeersch and Kremer (2005)).

These studies had mixed findings. To the surprise of many, the school inputs studies basically found zero effects. Textbooks did not lead to average improvements in test scores, but did improve test scores for those students who scored highest on a pre-test. The authors hypothesized that the reason may be that the textbooks were too hard: they were aimed at the top students, so students who were already behind found them useless. The flip charts had no effect. On the other hand, the deworming intervention reduced school absenteeism by children by 25 percent, though it did not improve test scores. School meals also made a difference: participation also increased by about 30 percent.

The positive findings, particularly the deworming study, have had a large impact, because they conclusively demonstrated the health – education linkage. (Results were also largely replicated by Bobonis, Miguel et al. (2006)). Deworming in particular has had a large policy impact since it is so cost-effective (it is orders of magnitude cheaper than school meals): in 2013/14, the Deworm the World Initiative – founded by Duflo and Kremer in 2007 to use the results of the randomized trials on deworming and transform them into policy and now part of the NGO Evidence Action – dewormed 37 million children in India and Kenya, as well as supported a deworming initiative in Bihar and a national deworming campaign in Ethiopia.

Another early randomized trial on education – begun in 2001 just after the Kremer et al studies in Kenya – was a study done by Banerjee, Cole, Duflo, and Linden in India schools (Banerjee, Cole et al. 2007). This project sought to examine two approaches to changing the education production function: a remedial education program that hired para-teachers to assist

lagging children, and a computer-assisted learning program – both of which were effective. In some sense, these results suggested that the early results from Kremer’s studies should not be interpreted to mean that school inputs don’t matter, but rather that choosing exactly the right inputs is important: textbooks kids can’t read don’t help, but computer programs that can adapt to where children are can be useful.

The negative findings from some of the early studies also had an important effect in terms of informing future research. For example, a key finding of the textbooks study was that they did not help lower-achieving students because they were too hard for poorly performing students. One implication of this is that introducing heterogeneity in classes – so that teachers can focus separately on high- and low-performing students – could be beneficial to all students. A follow-on study explored exactly this idea, building on the textbooks study and the para-teacher study. In 2005, Duflo, Dupas, and Kremer conducted a trial where schools with one teacher were awarded an extra teacher (Duflo, Dupas et al. 2011). The schools were randomly divided into two groups, one in which students were unsorted, and one in which they were sorted by initial ability. Consistent with the hypothesis from the textbook study, the schools where students were sorted based on initial ability did better, suggesting that customizing education to the needs of the students is important and – at least in this context – outweighs potential benefits to lower students from learning from their higher-performing peers.

Beyond the specifics of the study, I want to highlight the methodological approach here, which typifies the core of the field experimental approach. Each trial typically focuses in on a very specific hypothesis, which it is designed to test. In the process of analyzing the results, the trials then typically generate a new set of hypotheses that form the basis for future work. Given this iterative approach, in most cases a breakthrough will not come from a single particular

paper, but rather knowledge is accumulated through a series of experiments on related topics. It is bringing this methodological approach to development – breaking down the social science challenge into more precise and specific hypotheses, and then *designing* randomized trials to investigate each aspect of this challenge – that has spurred such a wide-ranging revolution in the field.

The authors here started this process in the field of understanding the microeconomics of schools in developing countries, but of course they are not the only ones to continue it. Some prominent examples of important randomized trials that seek to unpack different aspects of the education production function include studies on school uniforms (Evans, Kremer et al. 2008), incentives for teachers (Glewwe, Ilias et al. 2010, Muralidharan and Sundararaman 2011), incentives for students (Kremer, Miguel et al. 2009), student-teacher ratios (Duflo, Dupas et al. forthcoming), school choice (Angrist, Bettinger et al. 2002, Muralidharan and Sundararaman 2013), school block grants (Das, Dercon et al. 2013), information to parents about school quality (Andrabi, Das et al. 2000), and others. This is by no means an exhaustive list, but what should be apparent is that these initial trials in Kenya and India led to a substantial literature using randomized trials to understand various aspects of the education production function.

### *III.II Local Governance and Service Delivery*

These early studies focused on school inputs (i.e. things that could be budgeted for), but in the process, various combinations of the same set of authors realized that there was another problem in service delivery in developing countries: conditional on the inputs provided, the service providers did not necessarily deliver. A key problem was absence of service providers. (I group this separately from the education studies since studies are roughly equally in education

and health contexts, and the issues do not seem to be sector specific). An early paper here is a project carried out in 2002 and 2003 by Kremer and coauthors, which documented shockingly high rates of absence of teachers and health providers across Indian states and across several countries (Chaudhury, Hammer et al. 2006), and which was prepared as part of the work for the 2004 World Development Report on service provision. Banerjee, Deaton et al. (2004), documented similar findings in their 2002-2003 survey of health facilities in Rajasthan. These studies were not necessarily the first to note that absenteeism was an issue, but they made the point systematically that absenteeism was a first-order problem.<sup>3</sup>

This same set of authors then began a systematic set of experiments that sought to determine how one could address this issue. Let me highlight several notable early papers from this literature. In one case, Duflo, Hanna et al. (2012) (in an experiment begun in 2003) examined the impact of changing teacher salaries to compensate them explicitly for attendance. Specifically, working with an NGO that operated single-teacher schools in rural India, they randomly selected some schools in which teachers' pay made was a linear function of the number of days above 10 in a month in which they could provide a time-stamped photograph showing from the beginning and end of the day of the teacher with the student, proving that they attended. The results document not only that incentives improved attendance, but that this translated into higher test scores for the children – documenting the importance of the phenomenon. This paper was also methodologically innovative, combining an experiment with a

---

<sup>3</sup> The WDR sites a few earlier studies, such as Schleicher, Siniscalco et al. (1995) and Thomas, Lavy et al. (1996), that also have statistics on absenteeism, though it was not the main point of these studies and I was not aware of them until I went back just now and checked the footnotes in the WDR. I also found references to a 1995 survey done by Banerjee, Kremer, and coauthors that documented high absenteeism rates (Banerjee, Jacob et al. 2005), but I cannot find the original paper.

structurally estimated model, where the model helped shed light on the mechanisms in the experiment.

Interestingly, a similar study with nurses in India found surprisingly different results. Banerjee, Duflo et al. (2008) examined similar incentives for nurses in India. In that experiment, the incentives for attendance were broadly similar to those in Duflo, Hanna, and Ryan: nurses who were recorded absent more than 50 percent of the days in a month would have their pay reduced by the number of days they were recorded absent, and those nurses who were absent more than 50 percent of the days in two consecutive months would be suspended from government service. Nurses used a protected time/date stamp machine to verify attendance. In their study, while there was initially a substantial treatment effect, the effect diminished over time, and was zero at the end of their study. Although they do not have the data to confirm this, anecdotal evidence suggests that the decline may be due to nurses learning how to exploit loopholes in the systems, and recording more exempt absences over time. A related finding is reported by Kremer and Chen (2001), who found that in Kenya, an incentive program for teacher attendance that rewarded highly-attending teachers with a bicycle resulted in higher attendance reported to those awarding attendance, but not higher attendance when verified through an independent survey.<sup>4</sup>

An alternative approach is community monitoring. In theory, for both education and health, many developing countries have mechanisms through which community members can hold service providers accountable, yet there is the view that many of these mechanisms do not

---

<sup>4</sup> Note that I can't find the original version of this paper, but it is discussed at some length in a survey article on absenteeism by Banerjee and Duflo (2006).

work effectively. Banerjee, Banerji et al. (2010) report an experiment they conducted starting in 2005 in which they tried to strengthen these mechanisms. Interventions included training members of the school committee (at baseline, only 13 percent of school committee members had ever received training on what it meant to be in a school committee), and doing a survey where they investigated the baseline status of kids' knowledge and informed the school committee of this. Neither intervention made a difference. On the other hand, an intervention where they taught local volunteers how to teach kids to read made an enormous difference in kids' reading ability, showing that it was possible to improve education in this context.

Substantively, the impact of community monitoring has proven to be heterogeneous, with different randomized trials finding different effects: a contemporaneous field experiment study by Bjorkman and Svensson (2009) found dramatic improvements from an intervention that provided communities in Uganda with information about local health centers and local health and helped them formulate an action plan to solve them; a contemporaneous field experiment I conducted (Olken 2007) found that enhanced community monitoring in Indonesia reduced corruption but only for certain types of expenditures; a contemporaneous trial by Duflo, Dupas et al. (forthcoming) found that enhanced training of school committees reduced capture of a block grant by school staff. A more recent field experiment by Pradhan, Suryadarma et al. (2014) on enhancing school committees in Indonesia found mixed effects, as did a follow-up study by (Björkman Nyqvist, De Walque et al. 2017) that studied a less comprehensive form of a similar intervention.

As is evident here, understanding how to improve service provision in education and health has become a huge focus of research. These authors were instrumental in documenting that a) poor service provision, and in particular absence of service providers, is a key issue that

leads to low quality health and education; b) addressing the absence issue can lead to improvements in educational outcomes, but c) reducing absence in a sustained way over time is challenging, as community interventions often don't work and incentives can be undermined.

### *III.III Credit constraints and other types of market failures*

A third important area of micro research in the economics of developing countries has been understanding market failures, particularly credit constraints. These market failures can lead to inefficient allocation of capital (i.e. high productivity firms). To the extent there is lumpiness in investment, it can also generate poverty traps.

Several early papers by Banerjee, Duflo, and coauthors helped make these points about credit constraints forcefully. One early example is the micro study by Banerjee and Duflo (2014) (this paper in fact was in circulation by 2002) and the related paper with Munshi (Banerjee, Duflo et al. 2003). These papers provide direct evidence of credit constraints by examining a directed lending program in India, which essentially instructed banks to lend to a certain class of firms. Using a differences-in-differences identification strategy, they find that the expansion of credit led to substantial increases in sales and profits for targeted firms. They conclude that this is evidence of credit constraints for large firms: had these firms been unconstrained, a targeted lending program might have affected the financial portfolio of the firms, allowing them to pay down more expensive debt, but would not have affected real behavior. Banerjee and Munshi (2004) also use micro data – in this case, differences between the production processes of two different ethnic groups – to argue that the different ethnic group's access to capital was an important determinant of the differences in their production outcomes.

Banerjee and Duflo nicely articulated the view of how credit constraints and misallocation of capital could be a first-order issue for economic development in their 2005 handbook of economic growth piece (Banerjee and Duflo 2005). This piece clearly articulated the link between the development problems documented in the micro-development literature and the macro-economic phenomenon of low TFP in developing countries, suggesting that the credit constraints and the resulting misallocation of capital among firms could be a first-order cause of the low TFP documented in the macro literature. A subsequent piece by Hsieh and Klenow (2009) is well known for documenting systematically the differences in allocation of capital matter for understanding the development problem, building in part on the Banerjee and Duflo (2005) piece and others.

These papers collectively make clear that understanding the microstructure of capital allocation is a critical part of the development puzzle. In some sense, this is the next step from the development accounting literature I reviewed above; as Jones (2016) nicely puts it, misallocation of capital is a plausible ‘theory of TFP.’ Once we have this, the next step is to turn the same set of micro experimental tools used in health, governance, and so on to understand the allocation (and misallocation) of capital in developing countries.

This problem has then been tackled in various ways. For example, a commonly advanced solution to the problems of credit constraints in developing countries was microfinance. Indeed, Muhammad Yunus and the Grameen Bank were awarded the 2006 Nobel Peace Prize for their work in pioneering micro loans which would allow the very poor to obtain credit. Yet, until relatively recently, there was virtually no rigorous evidence of whether such loans actually made a difference. Several studies by Banerjee, Duflo, and their coauthors were among the first to experimentally evaluate these programs (e.g., Banerjee, Duflo et al. 2015, Crepon, Devoto et al.

2015). The key point of these papers was to show, experimentally, that microfinance has relatively modest impacts. Of course, these studies are not the only field experiments on microfinance: other authors, such as Erica Field, Rohini Pande, and their coauthors, and Dean Karlan and his coauthors have run a series of very nice field experiments trying to unpack which features of microfinance matters, and to test various theories about why (see, for example, Karlan and Zinman 2009, Karlan and Valdivia 2011, Feigenberg, Field et al. 2013, Field, Pande et al. 2013). In some sense, this literature parallels the approach taken to the educational production function I outlined above: taking a broad phenomenon – microfinance – and unpacking all of its pieces, testing each one experimentally to develop a better understanding of whether and how it works and why. Townsend and coauthors have a number of non-experimental papers that also seek to unpack the impact of better finance, most notably Kaboski and Townsend (2011), which combines a plausibly exogenous reduced form analysis of a credit expansion in Thailand with a structural model that helps explain heterogeneity in how households respond to the credit expansion.

Duflo and Kremer also helped contribute to understanding the puzzle of high average (as well as highly variable) returns in agriculture markets in developing countries. They give two answers, in the context of fertilizer. Duflo, Kremer et al. (2008) give one answer, which is that while there are high returns available, it is hard to get it right, and you can easily end up with negative returns. In particular, they show experimentally, by varying the amount of fertilizer used by farmers on different plots, that the returns to fertilizer use are highly concave – so that if one uses, for example, a half-teaspoon amount of fertilizer per plant, the expected rate of return is 70 percent, but using one teaspoon the expected rate of return is -17.8 percent. Even worse, following the government's recommended use yields a return of -48 percent. Moreover, this is

hard for farmers to figure out—returns are increasing in fertilizer use (so more is good), so one would need to do careful calculation and measurement to figure out the average returns.

A second reason, also from the same series of experiments (which began with a series of pilots in 2000), has to do with behavioral distortions in farmers' purchasing decisions. In particular, Duflo, Kremer et al. (2011) show that offering small, time-limited discounts at harvest time – when farmers are flush with cash – leads to substantial adoption of fertilizer. The impact is much larger than offering even larger discounts at the time when fertilizer is applied, suggesting that there is something important about farmers' inability to save that is driving the decision. Of course, there are many important contributions to agriculture in developing countries, such as Foster and Rosenzweig (1995) and Conley and Udry's (2010) work on social learning and land issues in particular stands out, but see Jack (2013) for a more comprehensive review – but these projects are an important contribution in understanding again why there are high returns on the table that are not taken up.

The three substantive areas I have focused on above – education, improving service delivery, and credit constraints – are just three early examples where the micro-experimental approach of Banerjee, Duflo, and Kremer, has substantially changed both the methodological direction of the field and our understanding of it. But it is no means an exhaustive list. To take just a few more examples, on health, Duflo and coauthors have run field experiments on the adoption and consequences of clean cookstoves (Hanna, Duflo et al. 2012) and piped water (Devoto, Duflo et al. 2011), and Banerjee, Duflo and coauthors have investigated the impact of using incentives to improve immunizations (Banerjee, Duflo et al. 2008). Kremer and coauthors have a series of papers exploring the market failures through which communities do not provide

sufficient clean water (Kremer, Leino et al. 2011). One could continue on a breathtakingly wide range of topics.

#### **IV. Moving forward and future directions**

The experimental revolution described in the three examples above has spread widely to virtually all aspects of the development field. Indeed, in 2017 Banerjee and Duflo edited a “Handbook of Field Experiments” which documents (among other topics) the experimental revolution in the study of health and health care, social protection programs, agriculture, governance, labor market issues, voting, discrimination, and marketing, in addition to many of the issues discussed in more depth above.

In many of these areas, there are now a large number of well-identified, microeconomic studies (often but not exclusively randomized trials) by a large number of authors, moving well beyond the three Laureates whose work I have focused on here. Indeed, what has really transformed the field of development economics is not simply the work done by these three authors, but rather the experimental approach that they pioneered. That work inspired a generation of scholars to take this approach to tackle an enormous variety of problems in development. Together, these represent over 1,000 studies that have, collectively, transformed the field, helping understand both why developing countries remain poor – i.e. why they have lower levels of labor, capital, human capital, and TFP (i.e. L, K, H, and A) in the development accounting sense – and also how the wide variety of economic forces studied throughout economics behave differently in developing country contexts.

In addition to a blossoming breadth of topics, in the years since these three authors began the field experimental movement, the types of studies have also evolved in myriad ways. For

example, one relatively recent approach has been to take the idea of designing an experiment to the next level by designing the entire workplace. For example, to test theories about labor markets, Emily Breza, Supreet Kaur, and Yogita Shamdasani created their own factory workshops in Odisha, India, in which people were hired to produce various products (e.g. rope, plates, etc.). This goes beyond typical lab experiments or lab-in-the-field models – in these cases, people worked for a month, producing real products for a real wage – but at the same time they were able to design virtually all aspects of the environment, testing the effects of pay inequity (Breza, Kaur et al. 2018). This type of approach is being applied to test a wide range of topics. In just the labor space alone, this approach has recently been used to test ideas from the effects of sleep (Bessone, Rao et al. 2019) to the effects of noise (Dean 2019) on labor market outcomes.

Other scholars have designed experiments that move in what is some sense the opposite direction: experimenting “at scale” (Muralidharan and Niehaus 2017). One challenge for answering some questions with experiments is that experiments do not necessarily capture the effect a policy or intervention would have if it was scaled up. This could be because of equilibrium market effects (e.g. effects on prices; see for example Cunha, De Giorgi et al. (2019)). It could also be because implementation quality at scale is different (Bold, Kimenyi et al. 2013). Or it could be because the places chosen for experimentation are not typical (Allcott 2015). Given this, some scholars are working on randomized trials with large, representative samples and large units grouped into treatment and control– not just randomizing individuals or schools but randomizing entire subdistricts (Olken, Onishi et al. 2008, Muralidharan, Niehaus et al. 2016), labor markets (Crépon, Duflo et al. 2013), or even provinces (Banerjee, Hanna et al. 2018).

A third direction a number of scholars have taken is to studying the longer-run implications of their work. A number of the early randomized trials I discussed above were conducted 10 or 15 years ago or more. This has led, in recent years, to a number of remarkable studies tracing their impacts over the longer run. For example, Baird, Hicks et al. (2016) return 10 years later to the original set of children who were dewormed in Miguel and Kremer (2004) to measure the impact on labor market outcomes; similarly, Duflo, Dupas et al. (2017) follow children who received scholarships in 2008 for 11 years to trace out their impacts. Long-run social science trials with outcomes in 5-6 year (e.g., De Mel, McKenzie et al. 2012, Cahyadi, Hanna et al. forthcoming) or even decade-long range (e.g., Blattman, Fiala et al. forthcoming, Bloom, Mahajan et al. forthcoming) are becoming increasingly common, shedding light on how these types of interventions affect people and firms over much longer time horizons.

Scholars are also finding a wide range of approaches to more closely-knit economic theory with the design of experiments. This can happen in a multitude of ways. Some papers develop an explicit theory and find a reduced form way to test it, such as Hanna et al's (2014) study of limited attention with multiple dimensions to observe, which they examine by studying seaweed farmers, Casaburi and Willis's (2018) explanation of low insurance demand, in which they posit that regular insurance conflates moving money across time and states and then design an experimental test of this idea, and Aggarwal et al's (2019) study of the impact of behavioral biases on incentive contract design, which they study in the context of a program incentivizing exercise for diabetics. Other papers used theory to explicitly design an intervention's design (e.g., Khan, Khwaja et al. 2019), structurally estimate parameters using experimental moments (e.g., Bai 2016, Kreindler 2020), or use theory as a more general guide to help interpret the results and provide guidance for further tests (Banerjee, Chandrasekhar et al. 2019). The

increasing and wide-ranging interplay between economic theory and experimental design is a hallmark of many of the most recent wave of field experiments.

Finally, the results from this type of work are increasingly finding their ways into policy debates. This is not a coincidence; indeed, I want to close by noting that all three of these scholars share a dedication to helping use the results of their research – and of rigorously evaluated research more generally – to inform development policy. As described above, Banerjee and Duflo, along with Sendhil Mullainathan, founded the Abdul Latif Jameel Poverty Action Lab which seeks to promote both the use of randomized field experiments among experiments, but also the uptake of the results of these experiments in the policy world. Kremer helped found the Development Innovation Ventures program at USAID, and the new Global Innovation Fund. These programs seek to produce reliable evidence on effective development policies and then provide the funds to scale them up if found successful. As mentioned before, Kremer and Duflo were cofounders of Deworm the World (which became Evidence Action), to take the evidence from the deworming trial and bring it. In work not highlighted here, Kremer also helped turn an academic idea of his – that pre-invention purchasing commitments can solve the global market failure by which pharmaceutical companies do not develop drugs for developing countries because markets are too small – into reality, with a \$1.5 billion pre-market commitment for the development of a pneumococcal vaccine made by a consortium of donors in 2007. They are not alone, of course, among academic development economists in trying to bring development policy into action – but it is worth noting that these authors have gone the extra step to move beyond research and to translate the results of the research into development policy.

## References

- Acemoglu, D. (1998). "Why do new technologies complement skills? Directed technical change and wage inequality." Quarterly Journal of Economics: 1055-1089.
- Acemoglu, D. (2002). "Directed technical change." The Review of Economic Studies **69**(4): 781-809.
- Aggarwal, S., R. Dizon-Ross and A. Zucker (2019). "Incentivizing Behavioral Change: The Role of Time Preferences."
- Aghion, P. and P. Howitt (1992). "A Model of Growth Through Creative Destruction." Econometrica: 323-351.
- Allcott, H. (2015). "Site selection bias in program evaluation." The Quarterly Journal of Economics **130**(3): 1117-1165.
- Andrabi, T., J. Das and A. I. Khwaja (2000). "Report cards: The impact of providing school and child test scores on educational markets."
- Angrist, J., E. Bettinger, E. Bloom, E. King and M. Kremer (2002). "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." American Economic Review **92**(5): 1535-1558.
- Bai, J. (2016). *Melons as Lemons: Asymmetric Information, Consumer Learning and Seller Reputation*, The Field Experiments Website.
- Baird, S., J. H. Hicks, M. Kremer and E. Miguel (2016). "Worms at work: Long-run impacts of a child health investment." The quarterly journal of economics **131**(4): 1637-1680.
- Banerjee, A., A. G. Chandrasekhar, E. Duflo and M. O. Jackson (2019). "Using gossips to spread information: Theory and evidence from two randomized controlled trials." The Review of Economic Studies **86**(6): 2453-2490.
- Banerjee, A., S. Chassang, S. Montero and E. Snowberg (forthcoming). "A Theory of Experimenters: Robustness, Randomization, and Balance." American Economic Review.
- Banerjee, A., A. Deaton and E. Duflo (2004). "Wealth, Health, and Health Services in Rural Rajasthan." American Economic Review **94**(2): 326-330.
- Banerjee, A. and E. Duflo (2006). "Addressing Absence." Journal of Economic Perspectives **20**(1): 117-132.
- Banerjee, A., R. Hanna, B. A. Olken and S. Sumarto (2018). *The (lack of) distortionary effects of proxy-means tests: results from a nationwide experiment in Indonesia*, National Bureau of Economic Research.
- Banerjee, A., S. Jacob and M. Kremer (2005). "Moving to Universal Education: Costs and Tradeoffs" Mimeo. J. Lanjouw and P. Lanjouw, MIT.
- Banerjee, A. and K. Munshi (2004). "How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur." Review of Economic Studies **71**(1): 19-42.
- Banerjee, A. V., R. Banerji, E. Duflo, R. Glennerster and S. Khemani (2010). "Pitfalls of Participatory Programs: Evidence from a randomized evaluation in education in India." American Economic Journal: Economic Policy: 1-30.
- Banerjee, A. V., S. Cole, E. Duflo and L. Linden (2007). "Remedying Education: Evidence from Two Randomized Experiments in India." Quarterly Journal of Economics **122**(3): 1235-1264.
- Banerjee, A. V. and E. Duflo (2005). "Growth theory through the lens of development economics." Handbook of economic growth **1**: 473-552.
- Banerjee, A. V. and E. Duflo (2009). "The experimental approach to development economics." Annu. Rev. Econ. **1**(1): 151-178.
- Banerjee, A. V. and E. Duflo (2014). "Do firms want to borrow more? Testing credit constraints using a directed lending program." The Review of Economic Studies **81**(2): 572-607.

Banerjee, A. V., E. Duflo and R. Glennerster (2008). "Putting a band-aid on a corpse: Incentives for nurses in the Indian public health care system." Journal of the European Economic Association **6**(2-3): 487-500.

Banerjee, A. V., E. Duflo, R. Glennerster and C. Kinnan (2015). "The miracle of microfinance? Evidence from a randomized evaluation." American Economic Journal: Applied Economics **7**(1): 22-53.

Banerjee, A. V., E. Duflo, R. Glennerster and D. Kothari (2008). Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of Immunization Campaigns with and without Incentives, MIT.

Banerjee, A. V., E. Duflo, R. Glennerster and D. Kothari (2010). "Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives." Bmj **340**: c2220.

Banerjee, A. V., E. Duflo and K. Munshi (2003). "The (mis) allocation of capital." Journal of the European Economic Association **1**(2-3): 484-494.

Banerjee, A. V., S. Jacob, M. Kremer and with J. Lanjouw and P. Lanjouw (2004). How Much Would Universal Primary Education Cost? , MIT.

Barro, R. J. (1991). "Economic Growth in a Cross Section of Countries." Quarterly Journal of Economics **106**(2): 407-443.

Barro, R. J. and X. Sala-i-Martin (1992). "Convergence." The Journal of Political Economy **100**(2): 223-251.

Basu, S. and D. N. Weil (1998). "Appropriate Technology and Growth." The Quarterly Journal of Economics **113**(4): 1025-1054.

Bessone, P., G. Rao, F. Schilbach, H. Schofield and M. Toma (2019). Sleepless in Chennai: The Consequences of Increasing Sleep among the Urban Poor, mimeo.

Bils, M. and P. J. Klenow (2000). "Does Schooling Cause Growth?" American Economic Review **90**(5): 1160-1183.

Bjorkman, M. and J. Svensson (2009). "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." Quarterly Journal of Economics **124**(2): 735-769.

Björkman Nyqvist, M., D. De Walque and J. Svensson (2017). "Experimental evidence on the long-run impact of community-based monitoring." American Economic Journal: Applied Economics **9**(1): 33-69.

Blattman, C., N. Fiala and S. Martinez (forthcoming). "The long term impacts of grants on poverty: 9-year evidence from Uganda's Youth Opportunities Program." American Economic Review: Insights.

Bloom, N., A. Mahajan, D. McKenzie and J. Roberts (forthcoming). "Do management interventions last? Evidence from India." American Economic Journal: Applied Economics.

Bobonis, G. J., E. Miguel and C. Puri-Sharma (2006). "Anemia and School Participation." Journal of Human Resources **41**(4): 692.

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a and J. Sandefur (2013). "Scaling up what works: Experimental evidence on external validity in Kenyan education." Center for Global Development Working Paper(321).

Breza, E., S. Kaur and Y. Shamdasani (2018). "The morale effects of pay inequality." The Quarterly Journal of Economics **133**(2): 611-663.

Cahyadi, N., R. Hanna, B. A. Olken, R. A. Prima, E. Satriawan and E. Syamsulhakim (forthcoming). "Cumulative impacts of conditional cash transfer programs: Experimental evidence from Indonesia." American Economic Journal: Economic Policy.

Casaburi, L. and J. Willis (2018). "Time versus state in insurance: Experimental evidence from contract farming in Kenya." American Economic Review **108**(12): 3778-3813.

Caselli, F. (2005). "Accounting for cross-country income differences." Handbook of economic growth **1**: 679-741.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan and F. H. Rogers (2006). "Missing in Action: Teacher and Health Worker Absence in Developing Countries." Journal of Economic Perspectives **20**(1).

Chernozhukov, V., M. Demirer, E. Duflo and I. Fernandez-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, National Bureau of Economic Research.

Conley, T. G. and C. R. Udry (2010). "Learning about a new technology: Pineapple in Ghana." American economic review **100**(1): 35-69.

Crepon, B., F. Devoto, E. Duflo and W. Pariente (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): 123-150.

Crépon, B., E. Duflo, M. Gurgand, R. Rathelot and P. Zamora (2013). "Do labor market policies have displacement effects? Evidence from a clustered randomized experiment." The quarterly journal of economics **128**(2): 531-580.

Cunha, J. M., G. De Giorgi and S. Jayachandran (2019). "The price effects of cash versus in-kind transfers." The Review of Economic Studies **86**(1): 240-281.

Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan and V. Sundararaman (2013). "School Inputs, Household Substitution, and Test Scores." American Economic Journal: Applied Economics **5**(2): 29-57.

De Mel, S., D. McKenzie and C. Woodruff (2012). "One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka." Science **335**(6071): 962-966.

Dean, J. T. (2019). Noise, cognitive function, and worker productivity, mimeo.

Devoto, F., E. Duflo, P. Dupas, W. Pariente and V. Pons (2011). Happiness on Tap: Piped Water Adoption in Urban Morocco, National Bureau of Economic Research.

Duflo, E. (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." American Economic Review **91**(4): 795-813.

Duflo, E., P. Dupas and M. Kremer (2011). "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." The American Economic Review: 1739-1774.

Duflo, E., P. Dupas and M. Kremer (2017). "The impact of free secondary education: Experimental evidence from Ghana." Massachusetts Institute of Technology Working Paper Cambridge, MA.

Duflo, E., P. Dupas and M. Kremer (forthcoming). "School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools." Journal of Public Economics.

Duflo, E., R. Glennerster and M. Kremer (2004). "Randomized evaluations of interventions in social service delivery." Development Outreach **6**(1): 26-29.

Duflo, E., R. Glennerster and M. Kremer (2008). Using Randomization in Development Economics Research: A Toolkit. Handbook of Development Economics. T. P. Schultz and J. A. Strauss, Elsevier. **4**.

Duflo, E., R. Hanna and S. P. Ryan (2012). "Incentives work: Getting teachers to come to school." The American Economic Review **102**(4): 1241-1278.

Duflo, E. and M. Kremer (2005). "Use of randomization in the evaluation of development effectiveness." Evaluating development effectiveness **7**: 205-231.

Duflo, E., M. Kremer and J. Robinson (2008). "How high are rates of return to fertilizer? Evidence from field experiments in Kenya." The American Economic Review: 482-488.

Duflo, E., M. Kremer and J. Robinson (2011). "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." American Economic Review **101**(6): 2350-2390.

Easterly, W., M. Kremer, L. Pritchett and L. Summers (1993). "Good Policy or Good Luck: Country Growth Performance and Temporary Shocks." Journal of Monetary Economics **32**(3): 459-483.

Easterly, W. and R. Levine (1997). "Africa's growth tragedy: policies and ethnic divisions." The Quarterly Journal of Economics: 1203-1250.

Evans, D., M. Kremer and M. Ngatia (2008). The Impact of Distributing School Uniforms on Children's Education in Kenya, Harvard University.

Feigenberg, B., E. Field and R. Pande (2013). "The economic returns to social interaction: Experimental evidence from microfinance." Review of Economic Studies **80**(4): 1459-1483.

Field, E., R. Pande, J. Papp and N. Rigol (2013). "Does the classic microfinance model discourage entrepreneurship among the poor? Experimental evidence from India." American Economic Review **103**(6): 2196-2226.

Finkelstein, A. and S. Taubman (2014). Using Randomized Evaluations to Improve the Efficiency of US Healthcare Delivery, MIT.

Foster, A. D. and M. R. Rosenzweig (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." Journal of Political Economy **103**(6): 1176-1209.

Gertler, P. (2004). "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment." American Economic Review **94**(2): 336-341.

Glewwe, P., N. Ilias and M. Kremer (2010). "Teacher Incentives." American Economic Journal: Applied Economics: 205-227.

Glewwe, P., M. Kremer and S. Moulin (2009). "Many children left behind? Textbooks and test scores in Kenya." American Economic Journal: Applied Economics **1**(1): 112-135.

Glewwe, P., M. Kremer, S. Moulin and E. Zitzewitz (2004). "Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya." Journal of Development Economics **74**(1): 251-268.

Grossman, G. M. and E. Helpman (1991). "Quality ladders in the theory of growth." The review of economic studies **58**(1): 43-61.

Gueron, J. M. (2017). The Politics and Practice of Social Experiments: Seeds of a Revolution. Handbook of Field Experiments. A. Banerjee and E. Duflo, Elsevier. **1**.

Hanna, R., E. Duflo and M. Greenstone (2012). Up in smoke: the influence of household behavior on the long-run impact of improved cooking stoves, National Bureau of Economic Research.

Hanna, R., S. Mullainathan and J. Schwartzstein (2014). "Learning through noticing: Theory and evidence from a field experiment." The Quarterly Journal of Economics **129**(3): 1311-1353.

Hsieh, C.-T. (2002). "What explains the industrial revolution in East Asia? Evidence from the factor markets." American Economic Review: 502-526.

Hsieh, C.-T. and P. J. Klenow (2009). "Misallocation and Manufacturing TFP in China and India." The Quarterly Journal of Economics **124**(4): 1403-1448.

Jack, B. K. (2013). "Constraints on the adoption of agricultural technologies in developing countries." Agricultural Technology Adoption Initiative, J-PAL (MIT) and CEGA (UC Berkeley).

Jones, C. I. (2016). The facts of economic growth. Handbook of macroeconomics, Elsevier. **2**: 3-69.

Kaboski, J. P. and R. M. Townsend (2011). "A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative." Econometrica **79**(5): 1357-1406.

Karlan, D. and M. Valdivia (2011). "Teaching entrepreneurship: Impact of business training on microfinance clients and institutions." Review of Economics and Statistics **93**(2): 510-527.

Karlan, D. and J. Zinman (2009). "Observing unobservables: Identifying information asymmetries with a consumer credit field experiment." Econometrica **77**(6): 1993-2008.

Katz, L. F., J. R. Kling and J. B. Liebman (2001). "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." Quarterly Journal of Economics **116**(2): 607-654.

Khan, A. Q., A. I. Khwaja and B. A. Olken (2019). "Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings." American Economic Review **109**(1): 237-270.

Klenow, P. and A. Rodriguez-Clare (1997). The neoclassical revival in growth economics: Has it gone too far? NBER Macroeconomics Annual 1997, Volume 12, MIT Press: 73-114.

Kreindler, G. (2020). Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications, Harvard University.

Kremer, M. (2003). "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons." American Economic Review **93**(2): 102-106.

Kremer, M. and D. Chen (2001). "An interim report on a teacher attendance incentive program in Kenya." Development Economics Department, Harvard University Cambridge, MA.

Kremer, M., J. Leino, E. Miguel and A. Zwane (2011). "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." The Quarterly Journal of Economics **126**: 145-205.

Kremer, M., E. Miguel and R. Thornton (2009). "Incentives to learn." The Review of Economics and Statistics **91**(3): 437-456.

Kremer, M., S. Moulin and R. Namunyu (2003). "Decentralization: A cautionary tale." Poverty Action Lab Paper **10**.

Mankiw, G., D. Romer and D. Weil (1992). "A Contribution to the Empirics of Economic Growth." Quarterly Journal of Economics **107**(2): 407-437.

Mauro, P. (1995). "Corruption and Growth." Quarterly Journal of Economics **110**(3): 681-712.

Miguel, E. and M. Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." Econometrica **72**(1): 159-217.

Muralidharan, K. and P. Niehaus (2017). "Experimentation at scale." Journal of Economic Perspectives **31**(4): 103-124.

Muralidharan, K., P. Niehaus and S. Sukhtankar (2016). "Building state capacity: Evidence from biometric smartcards in India." American Economic Review **106**(10): 2895-2929.

Muralidharan, K. and V. Sundararaman (2011). "Teacher Performance Pay: Experimental Evidence from India." Journal of Political Economy **119**(1): 39-77.

Muralidharan, K. and V. Sundararaman (2013). The Aggregate Effect of School Choice: Evidence from a two-stage experiment in India, National Bureau of Economic Research.

Newhouse, J. P. (1993). Free for all? : lessons from the Rand Health Insurance Experiment. Cambridge, Mass., Harvard University Press.

Olken, B. A. (2007). "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." Journal of Political Economy **115**(2): 200-249.

Olken, B. A., J. Onishi and S. Wong (2008). Should Aid Be Conditional on Performance? Experimental Design and Research Questions, Harvard University.

Pradhan, M., D. Suryadarma, A. Beatty, M. Wong, A. Alishjabana, A. Gaduh and R. P. Artha (2014). "Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia." American Economic Journal: Applied Economics **6**(2): 105-126.

Pritchett, L. (1997). "Divergence, big time." The Journal of Economic Perspectives: 3-17.

Psacharopoulos, G. (1973). Returns to Education: An International Comparison. San Francisco, Jossy Bass-Elsevier.

Psacharopoulos, G. (1985). "Returns to education: a further international update and implications." Journal of Human Resources: 583-604.

Psacharopoulos, G. (1994). "Returns to investment in education: A global update." World Development **22**(9): 1325-1343.

Romer, P. M. (1986). "Increasing returns and long-run growth." Journal of political economy **94**(5): 1002-1037.

Romer, P. M. (1990). "Endogenous technological change." Journal of political Economy **98**(5, Part 2): S71-S102.

Sala-i-Martin, X. X. (1997). "I just ran two million regressions." The American Economic Review **87**(2): 178-183.

Schleicher, A., M. T. Siniscalco and T. Postlethwaite (1995). "The conditions of primary schools: A pilot study in the least developed countries." A mimeographed report to UNESCO and UNICEF. Hamburg.

Schultz, T. P. (2004). "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." Journal of Development Economics **74**(1): 199-250.

Skoufias, E. (2005). PROGRESA and Its Impacts on the Welfare of Rural Households in Mexico. IFPRI Research Report.

Solow, R. M. (1956). "A contribution to the theory of economic growth." The quarterly journal of economics **70**(1): 65-94.

Thomas, D., V. Lavy and J. Strauss (1996). "Public policy and anthropometric outcomes in the Cote d'Ivoire." Journal of Public Economics **61**(2): 155-192.

Vermeersch, C. and M. Kremer (2005). School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation. World Bank Policy Research Working Paper Series.

Young, A. (1995). "The tyranny of numbers: confronting the statistical realities of the East Asian growth experience." The Quarterly Journal of Economics **110**(3): 641-680.