



Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2019

UNDERSTANDING DEVELOPMENT AND POVERTY ALLEVIATION

The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel

THE ROYAL SWEDISH ACADEMY OF SCIENCES, founded in 1739, is an independent organisation whose overall objective is to promote the sciences and strengthen their influence in society. The Academy takes special responsibility for the natural sciences and mathematics, but endeavours to promote the exchange of ideas between various disciplines.

Scientific Background on the Sveriges Riksbank Prize in Economic Sciences
in Memory of Alfred Nobel 2019

Understanding Development and Poverty Alleviation

The Committee for the Prize in Economic Sciences
in Memory of Alfred Nobel

October 14, 2019

Despite massive progress in the past few decades, global poverty — in all its different dimensions — remains a broad and entrenched problem. For example, today, more than 700 million people subsist on extremely low incomes. Every year, five million children under five die of diseases that often could have been prevented or treated by a handful of proven interventions. Today, a large majority of children in low- and middle-income countries attend primary school, but many of them leave school lacking proficiency in reading, writing and mathematics. How to effectively reduce global poverty remains one of humankind's most pressing questions. It is also one of the biggest questions facing the discipline of economics since its very inception.

So how best to identify strategies to help the least well-off? This year's Prize in Economic Sciences rewards the experimental approach that has transformed development economics, a field that studies the causes of global poverty and how best to combat it. In just two decades, the pioneering work by this year's Laureates has turned development economics — the field that studies what causes global poverty and how best to combat it — into a blossoming, largely experimental field.

Innovations both inside and outside of this field helped sow the seeds of the transformation. Inside the field, 2015 Laureate Angus Deaton pushed the research in development economics towards microeconomic analysis. He also championed the idea that the measurement of well-being, especially the well-being of the poor, must be closely integrated into the fight against poverty. Outside the field, the so-called credibility revolution, which first took off within labor economics in the early 1990s, pushed economic research in several areas towards a stronger focus on estimating causal effects. In addition, a well-articulated microeconomic theory appeared on how incentives and information, together with behavioral constraints, shape human behavior. This theory — rewarded with several Economics Prizes — gave researchers a powerful analytical tool kit to analyze the determinants of poverty and channels of poverty alleviation. These methodological gains were prerequisites for the transformation to follow, but a core piece of the puzzle was still missing.

Specifically, a well-articulated theory may be crucial to discovering possible mechanisms behind poverty and to guiding the search for effective ways to combat it. But it is not sufficient to guide policy. While theory can pinpoint certain incentives, it

does not tell us how powerful these are in practice. To give just a few examples, theory cannot tell us whether temporarily employing additional contract teachers with a possibility of re-employment is a more cost-effective way to raise the quality of education than reducing class sizes. Neither can it tell us whether microfinance programs effectively boost entrepreneurship among the poor. Nor does it reveal the extent to which subsidized health-care products will raise poor people's investment in their own health. Knowing the right quantitative answers to such specific questions is vital for enhancing human capital, increasing income, and improving health among the poor. Answering these questions requires an empirical approach that allows researchers to draw firm conclusions about causal effects.

By pioneering an approach to empirical research for providing such answers, the 2019 Laureates — Abhijit Banerjee, Esther Duflo and Michael Kremer — have transformed development economics. Their approach remained guided by microeconomic theory and the use of microeconomic data. But it shifted focus towards identifying workable policies, for which one can make causal claims of impact.

As a result, we now have a large number of concrete results on specific mechanisms behind poverty and specific interventions to alleviate it. For example, on schooling, strong evidence now shows that the employment of contract teachers is generally a cost-effective way to improve student learning, while the impact of reduced class size is mixed, at best. On health, poor people's investment in preventive care has been shown to be very sensitive to the prices of health products or services, giving a strong argument for generous subsidies to such investments. On credit, growing evidence indicates that microfinance programs do not have the development effects that many had thought when these programs were introduced on a large scale.

The transformation of the field involved important contributions by several scholars. Three contributions by the Laureates, however, stand out.

First, in the mid-1990s, Kremer and his co-authors launched a set of randomized controlled trials on schooling in Kenya.¹ In effect, their approach amounted to splitting

¹ A randomized controlled trial is a method to estimate the causal impact of a certain intervention, program or policy. A field experiment is a randomized controlled trial in which participants make choices in their normal day-to-day environment. Section 1 gives a more detailed description.

up the question of how to boost human capital in low-income countries into smaller and more manageable specific topics, each of which could be rigorously studied via a carefully designed field experiment. Soon thereafter, Banerjee and Duflo, often together with Kremer or others, broadened the set of educational topics and expanded the scope of the research to other areas, including health, credit and agriculture.

Second, in a series of contributions, Banerjee and Duflo articulated how pieces from such microeconomic studies can help us get closer to solving the broad development puzzle: what explains the enormous difference in per-capita income across countries? They started by documenting a striking empirical fact: low- and middle-income economies encompass enormous heterogeneities in the rates of return to the same factors of production within countries, which dwarf observed cross-country heterogeneities in economy-wide (average) returns. In other words, some firms and individuals in developing countries use the latest technology, while others in the same country and sector use outdated production methods. In high-income countries, these within-sector differences in productivity are much smaller. A deeper understanding of the development problem thus requires an explanation of why some firms and individuals do not take advantage of the best available opportunities and technologies. Banerjee and Duflo further argued that these misallocations can be traced back to various market imperfections and government failures. Hence, a core step in understanding, and ultimately alleviating, poverty is to identify sources of the observed inefficiencies as well as policies that could address them.

Finally, by designing new experimental research methods and by addressing the key challenge of generalizing results from a specific experiment — i.e., the issue of external validity — the Laureates firmly established this transformed approach to development economics. This laid a solid stepping stone for a new generation of researchers in development economics and other fields.

In sum, by bringing the theory of incentives closer to direct applicability, the research by the Laureates has profoundly altered the practice of development economics. The work by the Laureates, and by many other scholars who followed in their footsteps, has dramatically increased the practical quantitative knowledge necessary to isolate key mechanisms behind poverty and behavioral responses to various policy

interventions. This work has significantly deepened our understanding of poverty in the developing world.

We start this overview by describing the core ideas behind the new microeconomic approach to development economics and by discussing the key foundational contributions (Section 1). Then, we go on to present some of the substantive research findings by the Laureates. This presentation spans several topics that help us understand the stark welfare differences between people in high-income and low-income countries, although they constitute only a small fraction of the Laureates' empirical contributions. It also describes how the research findings were achieved (Section 2). Next, we discuss challenges to external validity and sketch out how the Laureates and others have shown ways forward to resolve these challenges (Section 3). We end by a short account of how the new experimental approach has influenced policy (Section 4) and concluding remarks (Section 5).

1. The Experimental Approach to Alleviating Global Poverty

This section describes the cornerstones on which the modern approach to development economics is built. We start by discussing the three contributions highlighted in the introduction. Then, we briefly discuss the main empirical method: randomized controlled trials. Finally, we turn to how this method has been used.

Three key contributions

The modern approach to development economics relies on two simple but powerful ideas. One idea is that empirical micro-level studies guided by economic theory can provide crucial insights into the design of policies for effective poverty alleviation. The other is that the best way to draw precise conclusions about the true path from causes to effects is often to conduct a randomized controlled field trial. The systematic application of these ideas over the past 20 years has paved the way for the transformation of development research.

This revamping of the field entailed a few distinct steps, with important contributions by several scholars. However, three contributions by the Laureates stand out.

First, starting in the mid-1990s, Kremer and various colleagues launched a series of field experiments in Kenya to disentangle various components in the educational production function (Kremer 2003). In essence, his approach amounted to breaking down the question of how to boost human-capital accumulation into smaller, more manageable topics, each of which could be rigorously studied via specifically designed randomized controlled trials. Soon thereafter, Banerjee and Duflo — in many cases jointly with Kremer and other co-authors — broadened the set of educational topics and launched field experiments on several other topics. These studies played a central role in expanding the new approach to basically all branches of development economics. Moreover, the research results pointed to important market inefficiencies in the private and public sectors and provided evidence on alternative ways to address these inefficiencies.

Second, in a series of contributions, Banerjee and Duflo articulated the intellectual case for a microeconomic approach to help understand various aspects of the broader (macroeconomic) development problem (Banerjee and Duflo 2005, 2007, 2011). Among these, the paper that Banerjee and Duflo published in 2005 is a key conceptual piece that links microeconomic development issues to low aggregate per-capita income in developing countries. The starting point for this work is an important empirical observation: low- and middle-income countries have large heterogeneities in the rates of return to the same factors of production and large variation in the extent to which profitable investment opportunities are exploited. The extent of this misallocation may be severe enough to help explain the large total-factor productivity gaps between low- and high-income countries that have been highlighted in the empirical growth literature. Intuitively, when resources are allocated optimally, the economy will operate on its production-possibilities frontier. When resources are misallocated, the economy will operate inside this frontier: output and productivity will be lower than they could be. Banerjee and Duflo further argued that market and government imperfections documented in the development literature — be they government failures, credit constraints, insurance failures, externalities, family dynamics, or behavioral issues — can help explain the misallocation. The first step to understanding why some countries are poor is thus to empirically identify important sources of inefficiency and policies to address them. The book *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty* (Banerjee and Duflo 2011)

takes this argument yet another step forward. Based on results from a large set of microstudies on the causes of poverty, Banerjee and Duflo have drawn lessons for a science-based approach to improving the health, schooling and incomes of the poor.

Third, by designing new experimental-research methods — e.g., to address the key challenge of external validity (Duflo 2004, 2006a, Duflo, Glennerster and Kremer 2006, Banerjee and Duflo 2009) — the Laureates firmly established the new approach and laid out a clear course forward for a new generation of researchers. The Abdul Latif Jameel Poverty Action Lab at MIT (J-PAL), which Banerjee and Duflo founded together with Sendhil Mullainathan, also was vital to this endeavor. J-PAL has promoted research built on randomized controlled trials in many countries and promoted the acceptance of results from such trials in the economic-policy community.²

Estimating causal effects

An important boost for the transformation of development-economics research was the creation and adoption of a (more) coherent microeconomic theoretical framework. Another important catalyst was the strong shift in other areas of economics towards empirical studies explicitly designed to credibly estimate causal relations. This so-called design-based approach started among labor economists in the early 1990s (see the overview in Angrist and Pischke 2010). But unlike the original design-based approach, which relied largely on natural experiments, the new microeconomic development research relied largely, though not exclusively, on field experiments.

A randomized controlled trial is a method for assessing the causal impact of a certain intervention or program. In essence, it is designed to answer counterfactual questions: How would individuals exposed to a program have fared in the absence of the program? Conversely, how would other individuals who were not exposed have fared, had they had the opportunity to participate? These questions are

² Duflo, Banerjee and Mullainathan are not alone among leading development economists to link experimental research to policy change and advice. For example, Dean Karlan co-founded Innovations for Poverty Action (IPA) — a nongovernmental organization (NGO) with offices in a number of developing countries — roughly at the same time as J-PAL was founded. IPA plays a crucial role in assisting researchers to run and implement experiments throughout the world. A few years later, the Center for Effective Global Action Lab (CEGA) was co-founded by Edward Miguel. Duflo and Kremer co-founded Deworm the World (now part of the NGO Evidence Action) to transform into action the evidence from the early deworming trial of Miguel and Kremer (2004).

counterfactual because at any given point in time, an individual is either exposed to the program or not. This is an instance of “the fundamental problem of causal inference” (Holland 1986): estimating the impact of a program on an individual at a given time is impossible. However, it is possible to estimate the average impact of the program on a group of individuals by comparing them to a similar group of individuals who were not exposed to the program.³ To do so requires a credible comparison group: a group of people who would have had outcomes similar to those exposed to the program, in the absence of the program. But how does one find such a valid comparison group?

If we just compare the group of individuals exposed to a program with a group consisting of nonexposed individuals, the estimated difference has two components. One component is the average causal effect of the program. But the estimated difference also would include a term reflecting selection bias: the difference in counterfactual outcomes between the two groups in the absence of the program. Without a reliable way to estimate or remove this selection bias, causal effects cannot be credibly estimated.

One method to solve the selection problem is to randomly assign individuals — or more generally the units of analysis, such as households, communities or schools — to a treatment and a control group. If a unit belongs to one of these groups just as the result of a random draw, the only systematic differences across the groups arise through their exposure to treatment. When correctly designed and implemented, a randomized controlled trial thus allows researchers to estimate the causal impact of a certain intervention in an unbiased way.

Of course, randomized controlled trials have a long history in science.⁴ A century ago, agricultural researchers pioneered the approach in crop studies. In the postwar era, randomized controlled trials became closely associated with clinical trials and later field trials in medicine. In economics, some important randomized controlled trials predate the explosion of experimental work in development economics, including the negative income-tax experiments (Hausman and Wise 1985), the RAND Health Insurance Experiment (Newhouse 1993), a series of welfare reform

³ Similarly, one can estimate the impact on other moments of the distribution than the mean.

⁴ The earliest published description of a clinical trial appears to be James Lind's test in 1747 of the value of oranges and lemons in treating scurvy in sailors. For details, see Thomas (1997).

experiments in the 1980s and 1990s (Manski and Garfinkel 1992), and educational research, such as the Perry Preschool Project and Project STAR (Schweinhart, Barnes and Weikart 1993). Thus, the main method used to estimate causal effects is not new. But, as we describe next, the application of randomized controlled trials in development economics has significantly expanded its use.⁵

Quantifying causal pathways

The approach championed by the Laureates rests on a fundamental insight of economics: most outcomes one may want to influence reflect purposeful choices made by individuals. To sustainably influence the outcomes of interest, also called endpoints, one must therefore understand the choices that drive observed outcomes, as well as the mechanisms that mediate those choices. That is, one needs to pinpoint the causal pathways through which changes in incentives, constraints and information influence outcomes of interest via human behavior.

For example, the knowledge to stop life-threatening afflictions of malaria, diarrhea, tuberculosis and other deadly diseases has largely been gathered through randomized controlled trials conducted by medical researchers. Despite this established medical knowledge, millions of children in low-income countries die from preventable diseases each year. Today's discussion about reducing child mortality in low-income countries therefore largely revolves around human behavior. Why is proven and inexpensive quality care not provided or demanded? Why do providers at various links in the service-delivery chain not show up, or not perform, at work? How can those hurdles be overcome? If so, how could workable services be delivered in cost-effective ways? Such questions are at the heart of the experimental approach adopted by development economists.

The strong emphasis on incentives and constraints is an important reason why designs of field experiments differ from designs in more controlled settings. For example, double-blind field trials are not typical, by design and for practical reasons. In the case of schools, adding additional resources can only crowd out investments or efforts by parents if they are aware of the additional resources (in other words, not

⁵ The use of randomized controlled trials, often referred to as field experiments, has also increased in other fields of economics (for overviews, see, e.g., Harrison and List 2004 and Card, DellaVigna and Malmendier 2011).

blind to them). These behavioral responses are not only central to understanding the experimental results themselves, but also to understanding broader human behaviors. In fact, the focus on so-called behavioral midpoints has made field experiments powerful tools for more general tests of human behavior. The responses, and their connections to economic theory, also may shed light on broader issues, beyond the experimental intervention at hand.

Uncovering mechanisms

The quest to uncover mechanisms also helps explain why field experiments have become the dominant empirical method in the field. Randomized controlled trials are unique in giving researchers complete control not only over the assignment mechanism, which removes the selection bias, but also over the treatment itself. As Duflo and Banerjee emphasized (Duflo 2006a, Banerjee and Duflo 2009), conventional observational studies evaluate events that have already occurred in the world. Experiments allow scholars to manipulate treatments of interest to create events that have not yet been observed.

This design flexibility has a number of advantages. Researchers can test new policies or interventions they believe might be effective, based on prior knowledge or theory, even if no policymaker is currently considering implementing them. Moreover, they can introduce random variation to test predictions from existing theory.⁶

Furthermore, experiments allow for a transparent iterative research process in the same setting, with the same outcome variables and the same measurement techniques. Whether by quasi-experimental or fully experimental methods, empirical work typically raises important follow-up questions. But follow-ups in observational studies are limited by the original policy change that generated the initial data. The ability to control the treatment itself relaxes that constraint and enables researchers to learn from sequential multistep experiments, where each new step takes

⁶ The ability to design new interventions — in many cases by revising, amending or decomposing programs run by others, e.g., NGOs — separates the development literature from the evaluation (social-experiment) literature in the US and Canada (see Gueron 2017 for a discussion). In the latter, the programs to be evaluated are typically fairly comprehensive packages chosen by an implementing agency with the researchers primarily serving as professional evaluators. Notable exceptions include the research-led Negative Income Tax (NIT) experiment and the RAND Health Insurance Experiment, which were both designed to test mechanisms (income and substitution effect in the NIT experiments, and moral hazard in the Rand Health Insurance Experiment).

advantage of the results in earlier steps. As noted by Duflo (2006a), such learning in a fixed context and with a fixed population is usually associated with laboratory experiments. Unlike most laboratory experiments in other areas of economics, however, field experiments in development economics involve real-world decision-makers who make important choices in their normal day-to-day environment.

Finally, as discussed by Kremer and Glennerster (2011), planning and implementing field experiments forces researchers to engage and spend time with realities on the ground, often working closely with government or nongovernment agencies. This field-based approach allows for an inductive discovery of new mechanisms and facts, beyond those suggested by existing models. This induction, in turn, can help refine the theory and improve our ability to design better policies.

2. Evidence on Fighting Poverty in Developing Countries

This section traces the intellectual history of the experimental approach in development economics, focusing on a set of thematic areas: education, health, behavioral biases, gender and politics, and credit. A separate subsection for each of these five areas highlights the substantive contributions by the Laureates.

The following is by no means an exhaustive presentation of the Laureates' research, and even less a literature review of the five thematic areas. However, it elucidates how the experimental approach pioneered by Banerjee, Duflo and Kremer has substantially changed our factual knowledge about economic, social and political phenomena in developing countries, as well as the methodological direction of the field. It also illustrates the broad approach to poverty alleviation taken in the field, ranging from policies to boost human and physical capital accumulation, via policies aimed at facilitating adoption of better technologies, to interventions that can influence how policies affecting the poor are chosen and implemented. Moreover, a large set of well-identified microeconomic studies by many researchers have been conducted in most areas of development. To reemphasize a point from the introduction, the Laureates' research has transformed the field through the experimental approach they pioneered.

2A. Education

The macroeconomic research on growth and development in the late 1980s and 1990s emphasized human capital, often approximated by educational attainment, in theoretical work (e.g., Lucas 1988) and empirical work (e.g., Barro 1991 and Mankiw, Romer and Weil 1992). Growth accounting attempted to decompose long-run growth across countries into a set of proximate causes, including human capital. This technique required measuring the returns to human capital, which was largely done by running cross-sectional Mincer regressions that linked wages to educational attainment.

That empirical literature, however, suffered from three problems. First, estimated cross-sectional returns to education were often biased because the underlying variation in education reflected systematic selection. Second, returns were measured from years of schooling, and these are not equal across time and place in terms of human-capital acquisition. Finally, the literature was largely silent on policy, i.e., how to most effectively increase enrollment and improve student learning. The new research pioneered by the Laureates has made important progress in all these dimensions.

Effects of better schooling

In the mid-1990s, Kremer and his co-authors initiated the transformation of development economics. To investigate how supply and demand factors interact to determine educational outcomes, they launched a series of field experiments in collaboration with a nongovernmental organization (NGO) in western Kenya. Two of the experiments estimated the impact of additional school inputs: textbooks, started in 1996 (Glewwe, Kremer and Moulin 2009), and flip charts, started in 1997 (Glewwe, Kremer, Moulin and Zitzewitz 2004). Two other experiments estimated the effects of health interventions, including deworming of children, started in 1998 (Miguel and Kremer 2004), and school meals, started in 2000 (Vermeersh and Kremer 2005). One experiment, begun in 1998, provided teachers with financial incentives tied to students' test scores (Glewwe, Kremer and Moulin 2010).

These early studies illustrated the power and feasibility of focused field experiments. But they also offered substantive lessons. Given the context, simply providing more

resources had a limited impact on school quality. More textbooks per student did not improve average test scores, but did improve test scores of the most able students. Giving flip charts to schools had no effect on student learning. The two health interventions reduced school absenteeism, but did not improve test scores. In theory, the incentive program could lead teachers either to increase effort to stimulate long-term learning or, alternatively, to teach to the test. The latter effect dominated. Teachers increased their efforts in test preparation, which raised test scores on exams linked to the incentives, but left test scores in unrelated exams unaffected.

Matching teaching better to student-learning levels

The findings from the first field experiments in Kenya provided a starting point for an early randomized controlled trial regarding education in India, which started in 2000 (Banerjee, Cole, Duflo and Linden 2007). Reviewing the findings from Kenya, Banerjee, Duflo and their co-authors concluded that students appeared to learn nothing from additional days at school. Neither did spending on textbooks seem to boost learning, even though the schools in Kenya lacked many essential inputs. Moreover, in the Indian context Banerjee and Duflo intended to study, many children appeared to learn little: in results from field tests in the city of Vadodara fewer than one in five third-grade students could correctly answer first-grade curriculum math test questions.⁷

In response to such findings, Banerjee, Duflo and co-authors argued that efforts to get more children into school must be complemented by reforms to improve school quality. Additional inputs may only work when they address specific unmet needs. In particular, they argued that neither the pedagogy nor the curriculum had been adapted to the large influx of children into primary education, as also suggested by the findings reported by Glewwe, Kremer and Moulin (2009). Many of the new students were indeed first-generation learners, whose parents may not have easily followed events in school or reacted if their children fell behind. The learning

⁷ Unfortunately, these findings did not seem specific to Vadodara. Although school enrollments have been universally increasing in the past several decades in the developing world, attending school does not ensure that children learn (Glewwe and Muralidharan 2015). The UNESCO Institute for Statistics estimates that only half of the world's children achieve minimum proficiency in reading and mathematics by the time they leave school (UIS 2017).

difficulties they observed may therefore partly reflect the success in getting more children to attend school.

Working with a large NGO in government schools in India, Banerjee and Duflo studied the impact of two interventions that targeted learning by weaker students. One was a remedial education program that hired paraprofessionals to work with poorly performing third- and fourth-grade students outside their regular classroom. The other was a computer-assisted learning program where fourth-grade children played games with math puzzles on a shared computer for two hours a week. Contrary to the earlier work on general resource additions, Banerjee, Duflo and their co-authors found substantial positive medium-term effects on student learning for both interventions, after both one and two years.

The project design allowed them to examine alternative mechanisms through which the remedial program improved average test scores. The program may have directly affected children who received instruction well-matched to their learning levels. It also could have indirectly affected them via class-size effects, where children falling behind benefit from a lower student-teacher ratio, or via peer effects, where children falling behind benefit from higher-performing classmates. However, the results suggested that the entire improvement was driven by the direct effect of remedial training with no indirect class-size or peer-group effects.

Banerjee and Duflo's study also started addressing the question about the generalizability — or external validity — of the experimental findings, an issue we discuss further in Section 3. The experiment was run on a large scale, with over 15,000 students over two years, and deliberately implemented in two different cities, Mumbai and Vadodara, with separate implementation teams.

Duflo and Kremer conducted another early multiple-treatment experiment in Kenya, starting in 2005 (Duflo, Dupas and Kremer 2011, 2015) that also was motivated by the challenges of a large influx of new students with varying academic preparation in response to the introduction of free primary education. Specifically, Duflo, Dupas and Kremer took advantage of a program that gave school committees funds to hire extra contract teachers, in order to reduce first-grade class sizes. But they added two experimental variations: tracking students by prior achievement and training school committees to monitor the extra teachers. This design allowed them to analyze a

range of important questions, including the impact of class-size reduction without changing pedagogy, the impact of contract teachers working under a dynamic incentive scheme versus tenured civil-servant teachers, the impact of empowered school committees, and the impact of tracking by achievement in primary schools.

A common response to overcrowded classrooms is to add more teachers. The idea is simple: lowering the student-teacher ratio increases the amount of time teachers can spend per individual student, which could have a direct effect on learning. If students benefit from higher-achieving peers, then sorting students into separate classes based on their preparedness or their ability could disadvantage low-achieving students while benefiting high-achieving students, thereby exacerbating inequality. For that reason, tracking is a controversial practice that many oppose. But as Duflo, Dupas and Kremer (2011) stressed, tracking also allows teachers to better target their teaching to student needs.

To shed light on tracking, the authors presented a theoretical model where peer quality affects students directly. But it also affects them indirectly through teachers' choices of effort and teaching levels. These, in turn, depend on the distribution of test scores in classes and on whether teachers' payoffs are linear, concave or convex functions of test scores. The further away a student's learning level is from her teacher's teaching level, the less the student benefits; if this distance is too great, she may not benefit at all. The model generates a rich set of predictions regarding the effect of a tracking program on the distribution of student achievement.

Duflo, Dupas and Kremer (2011) provided evidence suggesting that all the students in the study benefited from tracking.⁸ Using regression-discontinuity analysis, the authors further showed that the bottom-ranked students in the upper track and the top-ranked students in the lower track gained equally from tracking.⁹ According to their model, these empirical results imply that the teachers must have adjusted their

⁸ The effect of tracking was significant at the 5-percent level when controlling for baseline test score and age and gender of the child. In schools that originally had two or more first-grade classes (19 schools; 10 tracking, 9 not tracking), it proved difficult to consistently organize tracking. Those schools were dropped from the analysis.

⁹ The rankings refer to the test score distributions at baseline.

teaching to class composition and that their reward functions were convex in the distribution of final scores.¹⁰

Random assignment of students into classes in untracked schools also provided insights into the learning process. According to Duflo, Dupas and Kremer's model, an exogenous upward shift in the distribution of prior achievement will strongly benefit students with initial achievement at the top of the distribution, both because of a positive direct peer effect and because the students will benefit from the upward shift in their teacher's instruction level. For students further down in the distribution, however, the effects of such a shift become ambiguous, as the positive peer effect is counteracted by being further away from their optimal instruction level. These effects also are evident in the data: top students benefitted from better prior achievement of their classmates, while no effect could be detected in the middle of the distribution. Duflo, Dupas and Kremer (2011) further documented a positive effect at the bottom of the distribution, which is consistent with their model if these children were already so far below the level of the instructional target that the effect of an additional mismatch is outweighed by the positive peer effect.

Teacher effort

The low apparent rates of learning for many students in low- and middle-income countries have several roots, including the aforementioned mismatch between student academic preparation and teacher target levels. A number of studies from the early 2000s (e.g., Chaudhury, Hammer, Kremer, Muralidharan and Rogers 2006, Banerjee, Deaton and Duflo 2004) highlighted an additional possibility: many teachers in low-income countries are actually not teaching when they are supposed to. That is, absence rates of teachers in many developing countries are very high.¹¹

¹⁰ A linear payoff function in the model would imply that teachers teach to the median student in their class. Students starting in the middle of the prior distribution of achievement would therefore perform worse under tracking. But students just above the median would do better than students just below because they would gain from higher-achieving peers, which is not consistent with the empirical results. If the payoff schedule for teachers is convex, on the other hand, students just below the median would benefit relative to the students just above because teachers focus on the upper tail of their students, while peer effects would benefit students just above the median more than students just below.

¹¹ These papers were not the first to note that absenteeism was an issue, but were the first to make the point systematically within and across countries. Chaudhury, Hammer, Kremer, Muralidharan and Rogers (2006), using data from seven low- and middle-income countries, documented an average absence rate of 19 percent. Moreover, using data from government-run schools in India, they found

In a series of papers in the early 2000s, Duflo and Banerjee, along with their various co-authors, began a systematic exploration of how to address teacher absenteeism. Duflo, Hanna and Ryan initiated a field experiment in 2003 that examined high-powered incentives linked to attendance (Duflo, Hanna and Ryan 2012). Working with an NGO that operated single-teacher schools in rural India, they randomly selected some schools where teachers received an additional bonus per day attended, as verified by school cameras at the start and end of the school day. They found that teacher absence dropped by half in treatment schools relative to control schools. Moreover, student learning improved.

This work by Duflo, Hanna and Ryan (2012) is one of the first examples of how randomized evaluations can shed light not only on the impacts of specific interventions, but also help estimate behavioral parameters that are of more general interest. One way of doing so is to combine experimental evidence with structural modeling.¹² Specifically, Duflo and her co-authors estimated a structural model using treatment-group data, and they validated the model using a control sample. The study provided convincing evidence on important behavioral parameters. One example is the wage elasticity of teacher labor supply, which is required to design policies for better teacher performance.

A similar study by Banerjee, Glennerster and Duflo (2008) of financial incentives for nurse attendance in government clinics in India provided a more cautionary note. In that experiment, the incentives were broadly similar to those in the experiment reported by Duflo, Hanna and Ryan (2012). Nurses recorded absent more than 50 percent in a month would have their pay reduced by the percentage of days absent, while nurses absent more than 50 percent in two consecutive months would be suspended from government service. While the study recorded a substantial initial treatment effect, this effect diminished over time and was zero at the end of the study. Anecdotal evidence suggested that the nurses learned how to exploit loopholes in the systems and record exempt absences. Comparing the two studies, a

that only 45 percent of teachers assigned to a school are engaged in teaching activity at any given point in time. A decade later, Bold and colleagues (Bold, Filmer, Martin, Molina, Stacy, Rockmore, Svensson and Wane 2017), using data from seven sub-Saharan African countries, reported similar estimates (23-percent rate of absence from school and 44-percent rate of absence from class).

¹² Other early examples of combining experimental data and structural modelling in a developing-country context include Todd and Wolpin (2006) and Attanasio, Meghir and Santiago (2012).

possible difference is that Duflo, Hanna and Ryan (2012) studied schools run by an NGO, which may have been better able than the government to enforce the incentives. The two studies thus raise the questions of whether such incentive programs can be operated at scale and whether government can implement them as effectively as an NGO. We return to these important external-validity questions in Section 3.

Duflo, Dupas and Kremer (2015) reported results from funds randomly provided to school committees to hire extra teachers on short contracts so as to reduce first-grade class sizes in Kenya, as well as the impact of providing school based management training of the school committees. These contract teachers are hired on annual contracts, outside the normal Ministry of Education civil-service channels, and are typically paid a lower wage than civil-servant teachers. The contracts are renewable, conditional on good performance in the local institutional environment. Studying temporary contract teachers, relative to permanent civil-servant teachers, is interesting for several reasons. Theoretically, conditional employment contracts rely on dynamic incentives (Holmström 1982), with good performance rewarded by continued employment and bad performance followed by contract termination. Due to their career concerns, contract teachers ought to exert more effort than permanent teachers who do not have performance-based contracts. Policy-wise, understanding the impact of contract teachers versus civil-servant teachers on student learning is important, because many teachers in developing countries are hired on short-term contracts in order to limit costs.

Duflo, Dupas and Kremer (2015) showed that students who were randomly assigned to stay with existing classes — taught by civil-servant teachers — did not raise their test scores significantly, despite classes being cut in half from 82 to 44 pupils on average. A potential reason for this finding is that teachers responded to the program in a way that could have reduced its impact. Specifically, the authors showed that civil-servant teachers in schools that received funding to hire a contract teacher both increased their absences and adversely influenced the hiring process of contract teachers. On the other hand, students assigned to locally hired contract teachers, who had lower absence rates, raised their test scores significantly. A governance program that empowered parents within school committees raised test scores for students, whether they were taught by contract teachers or civil-servant teachers.

Meta-studies

In the past 20 years, more than 100 randomized controlled trials on education have been implemented across the developing world. The growing number of high-quality studies is also mirrored by a growing number of systematic reviews of the evidence. A clear message from these meta-studies is that some of the early interventions tested by Banerjee, Duflo and Kremer are seen as the most cost-effective interventions to improve student learning. For example, Glewwe and Muralidharan (2015) concluded that

“interventions that focus on improved pedagogy (especially supplemental instruction to children lagging behind grade level competencies) are particularly effective, and so are interventions that improve school governance and teacher accountability,”

while Kremer, Brannen and Glennerster (2013) stated that

“pedagogical reforms that match teaching to students’ learning levels are highly cost effective at increasing learning, as are reforms that improve accountability and incentives, such as local hiring of teachers on short term contracts.”

2B. Health

Modern public-health technologies, such as vaccines, antibiotics and anti-malarial drugs, and effective preventive methods, such as mosquito nets and drinking-water treatment, have improved health to historically unprecedented levels even in low-income countries. Nevertheless, the risk of a child dying before age five is still almost 15 times higher in low-income countries than in high-income countries. And coverage of a range of low-cost preventive health products remains incomplete in the developing world.

The standard human-capital model views health both as a consumption good and as an investment good (Grossman 1972). Human capital can be built up by investing in health, and rational consumers undertake such investments if they expect the private marginal benefit to exceed the marginal cost. Wedges between private and social benefits generate underinvestment in health by consumers, absent efficient public intervention. For example, health investments will be too low if treatment or prevention has positive externalities, or if they are public goods, such that their marginal costs are lower than average costs. This simple model was the starting

point for a series of influential contributions by Kremer and his co-authors in an effort to understand the reasons for the apparently suboptimal uptake of public-health measures.

Externalities

Miguel and Kremer (2004) estimated the direct effects and the externalities of deworming. They argued that with externalities, studies that randomize disease control at the individual level will underestimate effect sizes, as they do not incorporate the positive externalities. While spillovers are likely to be of first-order importance — especially in countries where infectious diseases still account for a large share of the disease burden — they have received limited empirical attention in public-health and epidemiological research (Benjamin-Chung, Arnold, Abedin, Falcao, Clark, Konagaya, Luby, Miguel and Colford 2015).¹³

In contrast, Miguel and Kremer (2004) designed their study specifically to measure these spillovers. They examined an NGO program for school-based mass treatment with deworming drugs and health education. The order of treatment phase-in to 75 primary schools was determined by a list that first grouped schools geographically and then alphabetically within locations. In Miguel and Kremer's econometric model, the effect of deworming is conditional on the total density of the local-school population within a particular geographic distance. Holding constant the total number of children attending primary school who live within a certain distance from the school, the number of these children attending schools assigned to treatment should be uncorrelated to other local observables and non-observables. The exposure to treatment spillovers is thus (quasi-) experimentally designed.

Miguel and Kremer (2004) found evidence for large external effects on worm-infection rates, as well as on subsequent school-participation rates, extending about 2 miles (at least 3 km) away from treatment schools.^{14,15} The empirical approach

¹³ Experimental trials in these fields typically do not estimate such spillovers but instead try to mitigate them through appropriate design choices (Hargreaves, Aiken, Davey and Hayes 2015).

¹⁴ The initial article also documented treatment externalities at distances 2–4 miles (3–6 km) from each school. These effects, however, were partly caused by a coding error. This and other robustness checks are discussed in subsequent publications (Aiken, Davey, Hargreaves and Hayes 2015, Hicks, Kremer and Miguel 2015, Hargreaves, Aiken, Davey and Hayes 2015).

¹⁵ The World Health Organization recommends mass treatment of children to eliminate worm infections in high-prevalence settings, although the question of under what circumstances the

proposed by Miguel and Kremer (2004) has been utilized in a large number of studies within economics — of both health and non-health issues — to estimate the magnitude and spatial scope of treatment externalities.¹⁶

Public goods

Some infrastructure is a public good in the sense that additional customers can be served at low marginal cost once the infrastructure is in place, even though the service is excludable. Such goods are natural monopolies. If households heterogeneously value the infrastructure (and the supplier cannot perfectly price discriminate), there will be static deadweight losses. Society may then be better off by regulating prices to reduce those static inefficiencies, even if this may reduce the incentives to invest in infrastructure.

Kremer and his co-authors (2011) examined these issues in the context of water-infrastructure technology. In the study area, many people collected water from naturally occurring springs, which can be contaminated by feces from humans or other animals. The authors evaluated a program that protected a random subsample of springs from fecal contamination. The intervention reduced the presence of *Escherichia coli* (bacteria used as indicators of fecal matter) by two-thirds in water at the source, and households reported that children had about 25-percent lower incidences of diarrhea in the treatment versus control groups.

In addition to estimating the direct effects of the intervention, Kremer and his colleagues (2011) used the data from the evaluation to determine the optimal governance of water protection. Based on a travel-cost model, and information on households' choices of where to collect water, the authors estimated the willingness to pay for spring protection. These estimates were then used in a structural model to assess the impact of alternative policies and property-rights systems on water infrastructure investment and welfare. The results suggested that the existing system of essentially common property rights delivered higher welfare than a private-property-rights system. At higher income levels, however, private property rights

expected benefits of mass drug administration exceeds its cost remains contested (WHO 2017, Croke, Hsu and Kremer 2017).

¹⁶ See the review by Dupas and Miguel (2017) for references.

could stimulate enough investment in spring protection to outweigh the static costs associated with giving landowners local market power over water.

Pricing

Whether and how much to charge users for health products and health services have been and continue to be hotly debated policy questions. Levying fees could reduce the uptake of highly cost-effective products and impede poor people's access to health care. On the other hand, charging a positive price could help select those who value and need the product the most. A positive price may also have a psychological effect: people may use more of a good that they have bought — a sunk-cost effect (Thaler 1980, Arkes and Blumer 1985). Higher prices may also encourage usage if they are interpreted as a signal of higher quality.

Kremer and Miguel (2007) provided the first experimental assessment of how prices affect the adoption of health products in a low-income setting. Among 50 primary schools enrolled in the free-deworming program discussed above (Miguel and Kremer 2004), they randomly selected 25 of those to participate in a cost-sharing program, where parents had to pay a fee for their children's deworming pills. They found an uptake of 75 percent in schools with free deworming pills, but only 18 percent with a fee of US\$0.40 (which is still a heavily subsidized price). While this result shows that demand is very sensitive to price, and potentially raises questions about the maintained assumptions of the rational human-capital model, the evidence is not conclusive. As Kremer and Miguel hypothesized, the perceived private value of deworming may be lower than the fee charged, simply because of the treatment externalities they documented. While their paper did not disentangle this externality effect from other effects of the positive price, subsequent experiments with alternative designs have pushed the research frontier significantly forward and helped distinguish different mechanisms.

Poor quality of care

Health systems in developing countries are often highly dysfunctional. A recent estimate suggests that a majority of deaths in low- and middle-income countries are

due to poor-quality care (Kruk, Gage, Joseph, Danaei, Garcia-Saiso and Salomon 2018).

Understanding why health-service quality is so low and what policies could improve it has long been a very active research area in development economics. The early absenteeism studies discussed above provided significant impetus for this agenda, and much of the early work focused on the effort channel.¹⁷

High rates of absenteeism, and more generally poor public-service provision, served as motives for an experimental study by Banerjee, Duflo and their co-authors of ways to improve immunization coverage in rural India (Banerjee, Duflo, Glennerster and Kothari 2010). In the study area, only 2 percent of children between one and two years old had received the recommended package of basic immunizations. In this study, the researchers discussed several reasons for the low uptake rates, including poor public-service provision. For example, in the year preceding the intervention, they documented that almost half of the health staff in charge of immunizations were absent from their health centers and could not be found anywhere in their villages.

The intervention utilized mobile vaccination clinics (“camps”), where the care staff were always on site. In a random subsample of these camps, small incentives were offered to households that brought their children to be immunized. Full immunization rates reached 39 percent in communities served by “camps with incentives,” compared to 6 percent in control communities, and 18 percent in communities with “camps but no incentives.” However, regular camps were sufficient to raise the percentage of children receiving at least one shot to levels comparable with those in the camps with incentives (78 and 74 percent, respectively). The incentives were particularly effective at encouraging families to stay the course and reach full immunization. Still, even with good access, reminders of the benefits of immunization, and small nonfinancial rewards (1 kilogram of lentils valued at about US\$1) for each immunization, 61 percent of the households did not get their children fully immunized.

¹⁷ Dupas and Miguel (2017) reviewed the literature on quantifying the quality of care, as well as experimental studies on policies to improve health-care provision.

2C. Behavioral Biases

Modern development economics systematically explores the structural constraints that poverty imposes on decision-making in an environment where information is incomplete. In this case, poverty can affect behavior even if decision-makers are “neoclassical” – i.e., rational, forward-looking and internally consistent (Duflo 2006b).

Duflo, Kremer and Robinson (2011) reached an important milestone in integrating behavioral and development economics. The researchers started a series of experiments in 2000 to shed light on a big puzzle: why do so many smallholder farmers, especially in sub-Saharan Africa, fail to take up relatively simple modern technologies, such as fertilizer, despite evidence of very high returns from agricultural trials?

To answer this question, they set up a long-term sequence of field experiments with farmers in western Kenya. Their first set of findings suggested that it is not necessarily easy to use fertilizer in a correct way. Farmers may thus not use it because it is unprofitable unless the right quantity is applied (Duflo, Kremer and Robinson 2008). But these findings also suggested substantial scope for learning. The next set of experiments looked at whether lack of information could explain low adoption rates. The results suggested it cannot.

Duflo, Kremer and Robinson (2011) instead asked whether present bias can explain farmers’ behavior. They proposed a model where some farmers are stochastically present-biased — in the sense of being hyperbolic discounters — and naive, such that they underestimate the likelihood that they will be present-biased in the future. Because purchasing fertilizer has a small fixed cost, hyperbolic discounting implies that farmers who plan to buy fertilizer will defer their purchase until close to a deadline. But at that point, they will be impatient again and choose not to buy.

Using this model, Duflo, Kremer and Robinson compared two alternative policy interventions: a relatively large subsidy and a small time-limited discount on fertilizer bought at harvest time, when farmers have some money. They implemented both interventions in a field experiment and showed that farmers purchased 50 percent more fertilizer when offered the small time-limited subsidy, which took the form of free delivery. Moreover, and consistent with the theory, this effect was greater than that of

offering free delivery plus a 50-percent subsidy on fertilizer later in the season. The results are in line with present bias being an important driver of low uptake among smallholder farmers. Evidence from additional experiments that explored alternative hypotheses further strengthened this interpretation.¹⁸

Together with Duflo and Banerjee's descriptive work on the economic lives of the poor based on household surveys conducted in 13 countries (Banerjee and Duflo 2007), the Kenyan experiments by Duflo, Kremer and Robinson (2011) greatly influenced later research on cognitive and psychological decision-making by poor people. This work was conceptually novel. The designs and findings in the study came about through a series of sequential experiments. New rounds of experiments were initiated in response to results from earlier trials, with the design of each treatment guided by theory. Such an iterative learning process, using experimental methods and holding context and population fixed, is usually associated with laboratory experiments. Unlike most lab experiments in economics, however, the field experiments involved real-world professionals, in this case Kenyan maize farmers, who made high-stakes decisions.

2D. Gender and Politics

An important issue in the political economics of development is how the identity of political leaders affects observed policy choices. Duflo tackled this question in one of her very first published studies (Chattopadhyay and Duflo 2004). The research described in this paper exploited a political reform that aimed to strengthen women's political standing in India. Specifically, in 1993, India's federal government introduced a new constitutional rule that each state had to reserve a third of all positions as chair (Pradhan) of village councils (Gram Panchayats) for women. These councils had also been given an increasing role in local decisions on infrastructure, with rules that differed by state. To investigate the effect of so-called female reservations, Duflo and Chattopadhyay surveyed a sample of villages in the two states of West Bengal and

¹⁸ See Kremer, Rao and Schilbach (2019) for a discussion and a review of the behavioral development economics literature.

Rajasthan, where the former had a longer history of village elections and also more extensive decentralized powers allocated to village councils.

In both states, a specific set of rules ensured that the chairs were reserved for a woman in a random selection of village councils. Exploiting these rules and the data from their own surveys, Duflo and Chattopadhyay could thus estimate the effects of having a randomly selected female leader. They found that female leaders seemed to make decisions that accorded better with the preferences of women. In West Bengal, village women were more concerned with drinking water and roads, while village men were more concerned with education. Female leaders in West Bengal indeed invested more than male leaders in drinking water and roads, at the expense of education. Analogously, in Rajasthan, where women were more concerned than men with water but less concerned with roads, Gram Panchayats reserved for women leaders made similar priorities in their investments, spending more money for water than for roads.

In a follow-up paper (Beaman, Chattopadhyay, Duflo, Pande and Topalova 2009), Duflo and her co-authors exploited the same underlying natural experiment, but again supplemented the electoral outcomes with their own survey and experimental data from a sample of about 500 villages. They showed that repeated “female reservation” of a leadership position in a village significantly raised the electoral prospects for female candidates in future elections. Moreover, they demonstrated that an important mechanism behind this result was decreased stereotypes among voters: specifically, less prejudice against women as effective policymakers. The analysis in this paper was not only substantively important. It was also methodologically innovative, as one of the first economics papers to use so-called implicit association tests (Greenwald, McGhee and Schwartz 1998) to gauge biases in the population.

Duflo’s research on female political leaders has greatly affected subsequent research on gender and politics, both in developing and developed countries. Of course, her research on gender, politics and policy addresses just one aspect of the broader question about female empowerment and economic development. As discussed by Duflo (2012), the causation between empowerment and development can run in both directions and along multiple channels.

2E. Credit

Misallocation of capital and other inputs into the aggregate production function can reflect credit constraints for some firms. Such a mechanism was at the core of Banerjee and Duflo's work that articulated the link between the myriad of distortions in low-income economies and the large difference in productivity and incomes identified in the empirical growth literature (Banerjee and Duflo 2005).

A number of influential contributions by Banerjee, Duflo and their co-authors have further investigated the existence, causes and consequences of credit constraints. In 2002, Banerjee and Duflo first proposed an idea that provided convincing evidence of binding credit constraints by examining a directed lending program in India (Banerjee and Duflo 2014). The intervention essentially instructed banks to lend to a certain class of firms. Using difference-in-differences analysis, Banerjee and Duflo (2014) found that the expansion of credit substantially increased sales and profits for targeted firms. They concluded that this is prima-facie evidence of credit constraints for large firms. Had these firms been unconstrained, a targeted lending program might have shifted their financial portfolios — allowing them, e.g., to pay off expensive debts — but would not have changed their real behavior.

Several field experiments in development economics have evaluated the impact of popular and heavily promoted policies that have already been implemented at a large scale, rather than testing new programs or mechanisms. A case in point is the microfinance movement, for which the pioneering work of Muhammad Yunus and the Grameen Bank was awarded the Nobel Peace Prize 2006. However, little convincing evidence exists on whether microcredit — which is intended to allow the very poor to obtain loans — actually has had a significant positive impact on development.

In joint work, Banerjee and Duflo reported on the very first randomized evaluation of a standard group-lending microcredit model in the city of Hyderabad, India, which targets women who may or may not be entrepreneurs (Banerjee, Duflo, Glennerster and Kinnan 2015). This evaluation followed households for more than three years in order to capture the medium-term impacts of the program.

At a first evaluation point, after 12 to 18 months, Banerjee, Duflo and co-authors found that households do borrow more from microcredit institutions, but that overall

uptake is not very high. Only about a quarter of eligible households borrow from microcredit institutions. Moreover, some of the new loans substitute for informal loans, and as informal borrowing declines, overall borrowing does not rise significantly.

The evidence does not suggest a particularly high demand for microcredit. Moreover, the research does not find significant differences for any key development outcomes, such as per-capita consumption expenditures, health, women's empowerment, or children's education. Women with access to microcredit are not more likely to be entrepreneurs, which is defined as having at least one business, but they do invest more in the businesses they have. The profitability of these businesses, between the 5th and the 95th percentile, does not differ with the assistance of microcredit or microloans, although the program appears to help the most profitable businesses.

The next evaluation point was three and half years after the program began to be rolled out in treatment neighborhoods. At this time, control neighborhoods had gained access to microcredit as well, but "treated" households had been able to borrow for a much longer time. Very few significant differences persist at this point. Altogether, little evidence supports the idea that microfinance serves as a primary engine of growth or of development more generally.

Banerjee, Duflo and their colleagues stressed a number of caveats to be kept in mind when interpreting and generalizing these results, including several potential benefits not properly accounted for in their evaluation (spillovers, general equilibrium effects, or effects via the expectation of being able to borrow when needed). Even with these caveats, the evidence does not suggest any large and positive effects of microfinance. Similar findings have subsequently emerged from evaluations of other microcredit programs (see Banerjee, Karlan and Zinman 2015).

3. External Validity

Establishing causal claims lies at the heart of the experimental approach pioneered by the Laureates. It is generally accepted that randomized controlled trials are particularly credible when it comes to *internal* validity (Athey and Imbens 2017). This aspect of validity applies to a trial's particular intervention, specific sample of

participants, location in calendar time, and its design. Lively debate surrounds the *external* validity of randomized trials, or how well causal inferences can be generalized for a particular population, timing and setting. Many, though far from all, field experiments in development economics are implemented at a relatively small scale by a well-managed local NGO, and their external validity has been debated.

Both scale and government implementation can affect how likely a particular finding is to be generalizable. As illustrated by Bold and colleagues (2018), politics when going to scale and state capacity when implementation is moved to the government are two channels that could significantly alter how findings from local trials generalize. Equilibrium effects, spillovers, context dependence, randomization bias, and piloting bias are other potential channels.

The Laureates have contributed to this important debate. In a number of papers and other venues, they have presented, discussed and analyzed challenges to external validity (Duflo 2004, Duflo, Glennerster and Kremer 2006, Banerjee and Duflo 2009, Banerjee, Banerji, Berry, Duflo, Kannan, Mukherji, Shotland and Walton 2017).¹⁹ More importantly, the Laureates have been at the forefront of dealing with and mitigating such challenges, often on the basis of experimental evidence and methods.

Consider first the equilibrium challenge. Unlike smaller field experiments, where results can plausibly be interpreted in a partial-equilibrium model, a scaled-up program likely affects, for example, prices and wages. This may influence the overall program effects and also lead to important and perhaps unwanted distributional consequences. If these equilibrium effects are believed to be important, they can be anticipated in the experimental design. One way to do so is to randomize at the market rather than the individual level and to focus on market-level outcomes. Another approach is to randomize in two stages, assigning treatment at the market level in addition to random assignment within treated markets.

Crépon, Duflo, Gurgand, Rathelot and Zamora (2013) used such a two-step design to evaluate the direct and indirect impacts of job-placement assistance. In the first step,

¹⁹ Apart from possible randomization bias, these external-validity concerns are not specific to experimental work but are a more general concern for drawing conclusions from micro-empirical work about policy implemented at scale.

each of the employment areas in their study were randomly assigned a percent p of job seekers, with p varying between 0 and 100. In a second step, p percent of all the eligible job seekers in each area were randomly selected to be assigned to get assistance. This design provided random variation not only in who gets treated, but also in the share of job seekers treated in the area. Of course, the equilibrium effects may occur at such a high level of aggregation that they cannot be subjected to randomization. Still, experimental methods may help to credibly estimate key behavioral parameters, which can be used to analyze equilibrium outcomes at a higher level of aggregation. Spillover effects can similarly be accounted for and quantified by randomizing at a higher level of aggregation (see Miguel and Kremer 2004, Duflo and Saez 2003).

Context dependence can be addressed through replication. The microfinance evaluation discussed in Section 2E, for example, can be seen as part of such a replication program, which carried out a total of six experimental studies of microfinance. Another example is multisite projects, such as the evaluation of a multipronged approach to support the ultra-poor (Banerjee, Duflo, Goldberg, Karlan, Osei, Parienté, Shapiro, Thuysbaert and Udry 2015), with randomized trials undertaken at the same time period in six developing countries. This particular multisite project examined a “sufficiency” claim and showed that the primary goal, to substantially raise the consumption rates of the very poor, was achieved at the end of the program and maintained one year later, at a cost below the estimated benefits, in five out of six sites.

Taking context dependence into account also requires guidance from theory. Even if specific program components may not be generalizable, underlying patterns in human behavior may. Understanding these behavioral patterns is thus crucial (Banerjee and Duflo 2009).

An instructive example of other external-validity challenges can be found in the scale-up of the Teaching at the Right Level intervention discussed in Section 2A. This scale-up illustrates the challenges from randomization bias: i.e., subjects who agree to participate in a small experiment may be different from the rest of the population. It also illustrates the challenges from piloting bias: i.e., findings from a smaller project

with high degrees of monitoring and control may not be replicable in a program run at scale.

The underlying approach, where instructors teach basic language and mathematics according to the knowledge level of the students, rather than a prescribed age-grade syllabus, was developed by Pratham, an NGO in India. Banerjee, Duflo and co-authors reported how they worked in collaboration with the NGO in an iterative process to design, experiment, and redesign and experiment again. In the end, they developed two successful, replicable models of this approach for scale-up in India's government school system (Banerjee, Banerji, Berry, Duflo, Kannan, Mukherji, Shotland and Walton 2016, 2017).

This iterative process started with the proof-of-concept trial discussed in Section 2A (Banerjee, Cole, Duflo and Linden 2007), where students lagging behind received remedial language and math training outside their regular classroom. To assess the external validity of these pilot findings, a new randomized evaluation was launched in a more challenging rural context. While the results from the new evaluation were positive (Banerjee, Banerji, Duflo, Glennerster and Khemani 2010), they raised additional concerns, including low uptake of students at the bottom of the learning distribution.

Two field experiments, embedded in the first scaled-up program within the government-school system, were conducted in 2008–2009 and 2009–2010. The results were largely disappointing. While the findings suggested that government teachers are capable of delivering remedial education, most teachers choose not to do so. A revised design was therefore developed and experimentally evaluated in the 2012–2013 school year: a so-called teacher-led model tried to ensure that teachers viewed the remedial education component as a core teaching task. This time the results were positive. A complementary program — a within-school volunteer-led model — was also developed and evaluated when running at scale. This program showed positive results as well.

In the end — after five randomized controlled trials spanning several years — two versions of the remedial education program had been designed and shown to be successful on a large scale. Today, the teacher-led model has been implemented in over 100,000 schools across 13 Indian states and has reached almost 5 million

children. The within-school volunteer-led model has been implemented in over 4,000 schools across India, reaching over 200,000 children (Banerjee, Banerji, Berry, Duflo, Kannan, Mukherji, Shotland and Walton 2017).

4. Influence on Policymaking

The approach pioneered by the Laureates yields insights on new and existing programs designed to address poverty. Combining these insights with an understanding of the local and institutional context gives powerful guidelines for development policy.

But whether and to what extent such information is incorporated into real-world policy design and implementation is largely outside of the researchers' control.

Policymakers may or may not consider robust evidence of what works, and why, when they decide which policies to implement. As studied in the field of political economics, policymakers are subject to a range of constraints that can explain a failure to adopt policies that are effective (see Persson and Tabellini 2000 for a broad overview of this field). Depending on the context, these constraints can reflect the policy preferences of various actors, the technologies used in private and public production, the information and commitment capacities of policymakers, and the institutional arrangements that allocate political power. These constraints hold true both in domestic policymaking and perhaps even more in an international-aid context.

Even so, the research approach pioneered by the Laureates has had a clear impact on policy, both directly and indirectly. Although this impact is difficult to quantify, J-PAL — the global research organization founded by Banerjee, Duflo and Mullainathan — estimates that more than 400 million people have been reached by programs that were scaled up after having been evaluated by researchers affiliated with J-PAL. Of course, this is just one metric and it does not include evaluations and field experiments implemented by development economists not affiliated with J-PAL.

Moreover, incorporating evidence into decision-making is not only about scaling up effective programs, but also about not spending valuable resources on ineffective programs. For example, Banerjee, Glennerster and Duflo's (2008) evaluation of a program to reduce health-worker absenteeism by a biometric monitoring system,

discussed in Section 2A, had limited impact on attendance. This finding contributed to the government's decision to cancel the planned scale-up of the program, saving millions of dollars and countless hours of staff time needed to run it.

Evidence generated from randomized controlled trials of social and economic programs in developing countries has also shaped actual policymaking in other ways. For example, the gradual reduction of user fees for preventive health products reflects the findings in a large number of randomized evaluations by development economists — starting with Kremer and Miguel (2007) — of how pricing affects the use of such products.²⁰

In addition, the approach pioneered by Banerjee, Duflo and Kremer has profoundly changed how some government organizations and NGOs work. A growing number — though far from all — organizations devoted to global poverty alleviation have indeed begun to systematically evaluate new policy proposals (often by field experiments) and to use the evidence from those evaluations in their decision-making.

5. Final Remarks

Over the past 20 years, we have seen major changes in development economics research. Several scholars have played a vital role in this endeavor. However, the broad contributions by this year's Laureates have been essential for bringing development research to its current standing. Kremer and his co-authors pursued a set of early experiments in western Kenya that showcased the promise of splitting up the daunting global-poverty question into smaller more manageable topics, each of which could be rigorously studied via a designated field experiment. Banerjee and Duflo, often together with Kremer or other researchers, broadened and expanded the set of topics, and articulated to the research community how pieces from such microeconomic studies can help us get closer to solving the broad development puzzle. All three Laureates expanded the experimental approach to basically all branches of the field. They were also at the forefront in addressing legitimate

²⁰ The British government (UK Government 2009) cited a study by Kremer and Miguel (2007) and a related study on the demand for long-lasting antimalarial bed nets by Cohen and Dupas (2010) in calling for the abolition of user fees for health products and service in poor countries.

challenges to this experimental approach and in presenting solutions to these challenges.

The contributions by Banerjee, Duflo and Kremer have encouraged and inspired a new generation of researchers to follow their lead. As a result, development economics has not just been decisively transformed, but continues to be a vibrant and expanding field that yields growing evidence-based guidelines for development policy. The Laureates remain among the key innovators in the field, with their recent contributions spanning climate and environmental policy (Duflo, Greenstone, Pande and Ryan 2018), social networks (Banerjee, Chandrasekhar, Duflo and Jackson 2013), and cognitive science (Dillon, Kannan, Dean, Spelke and Duflo 2017).

In less than two decades, the empirical microeconomic approach pioneered by Banerjee, Duflo and Kremer has changed how development economists conduct their research. The research carried out with their experimental approach has uncovered a large body of new substantive results and keeps improving our ability to mitigate global poverty.

References

- Aiken, Alexander M., Calum Davey, James R. Hargreaves and Richard J. Hayes. 2015. "Re-analysis of Health and Educational Impacts of a School-based Deworming Programme in Western Kenya: A Pure Replication." *International Journal of Epidemiology* 44(5):1572–1580.
- Angrist, Joshua D. and Jörg-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24(2): 3-30.
- Attanasio, Orazio, Costas Meghir and Ana Santiago. 2012. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresa." *Review of Economic Studies* 79(1): 37–66.
- Arkes, Hal R. and Catherine Blumer. 1985. "The Psychology of Sunk Cost." *Organizational Behavior and Human Decision Processes* 35: 124–140.
- Athey, Susan and Guido W. Imbens. 2017. "The Econometrics of Randomized Experiments." In A. Banerjee and E. Duflo (Eds.) *Handbook of Field Experiments*, Vol 1: 73-140. Amsterdam: North Holland, Elsevier.
- Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland and Michael Walton. 2016. "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of 'Teaching at the Right Level' in India." NBER Working Paper No. 22746.
- Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland and Michael Walton. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives* 31(4): 73-102.
- Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster and Stuti Khemani. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy* 2(1): 1-30.
- Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo and Matthew O. Jackson. 2013. "The Diffusion of Microfinance." *Science* 341(6144): 1236-1249.
- Banerjee, Abhijit, Shawn Cole, Esther Duflo and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122(3): 1235-1264.
- Banerjee, Abhijit, Angus Deaton and Esther Duflo. 2004. "Wealth, Health, and Health Services in Rural Rajasthan." *American Economic Review* 94(2): 326-330.
- Banerjee, Abhijit and Esther Duflo. 2005. "Growth Theory Through the Lens of Development Economics," In *Handbook of Economic Growth, Vol. 1A*. Durlauf, Steve and Philippe Aghion (eds.), 473–552. Amsterdam: North Holland, Elsevier.

- Banerjee, Abhijit and Esther Duflo. 2007. "The Economic Lives of the Poor." *Journal of Economic Perspectives* 21(1): 141-167.
- Banerjee, Abhijit and Esther Duflo. 2009. "The Experimental Approach to Development Economics," *Annual Review of Economics* 1: 151–178.
- Banerjee, Abhijit and Esther Duflo. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York, NY: Public Affairs.
- Banerjee, Abhijit and Esther Duflo. 2014. "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program." *Review of Economic Studies* 81(2): 572-607.
- Banerjee, Abhijit, Esther Duflo and Rachel 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, Abhijit, Esther Duflo, Rachel Glennerster and Cynthia Kinnan. 2015. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." *American Economic Journal: Applied Economics* 7(1): 22-53.
- Banerjee, Abhijit, Esther Duflo, Rachel Glennerster and Dhruva Kothari. 2010. "Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of Immunization Campaigns with and without Incentives." *British Medical Journal* 340: C2220.
- Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert and Christopher Udry. 2015. "A Multi-faceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348(6236): 1260799.
- Banerjee, Abhijit, Dean Karlan and Jonathan Zinman. 2015. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics* 7(1): 1-21.
- Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics* 106(2): 407-443.
- Beaman, Lori, Raghavendra Chattopadhyay, Esther Duflo, Rohini Pande and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics* 124(4): 1497–1540.
- Benjamin-Chung, Jade, Benjamin F. Arnold, Jaynal Abedin, Lauren Falcao, Ashley E. Clark, Eugene Konagaya, Steve P. Luby, Edward Miguel and John M. Colford. 2015. "The Identification and Measurement of Health-Related Spillovers in Impact Evaluations: A Systematic Review." 3ie Systematic Review 22. London: International Initiative for Impact Evaluation (3ie).
- Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Brian Stacy, Christophe Rockmore, Jakob Svensson and Waly Wane. 2017. "Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa." *Journal of Economic Perspectives* 31(4): 1–21.

- Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang' and Justin Sandefur. 2018. "Experimental Evidence on Scaling Up Education Reforms in Kenya." *Journal of Public Economics* 168: 1-20.
- Card, David, Stefano DellaVigna and Ulrike Malmendier. 2011. "The Role of Theory in Field Experiments." *Journal of Economic Perspectives* 25(3): 39-62.
- Chattopadhyay, Raghavendra and Esther Duflo. 2004. "Women as Policy Makers: Evidence from an India-Wide Randomized Policy Experiment." *Econometrica* 72(5): 1409-1444.
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan and F. Halsey Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives* 20(1): 91-116.
- Cohen, J. and P. Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Experiment." *Quarterly Journal of Economics* 125(1): 1-45.
- Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot and Philippe Zamora. 2013. "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment." *Quarterly Journal of Economics* 128(2): 531-580.
- Croke, Kevin, Eric Hsu and Michael Kremer. 2017. "More Evidence on the Effects of Deworming: What Lessons Can We Learn?" *American Journal of Tropical Medicine and Hygiene* 96(6): 1265-1266.
- Dillon, Moira R., Harini Kannan, Joshua T. Dean, Elizabeth S. Spelke and Esther Duflo. 2017. "Cognitive Science in the Field: A Preschool Intervention Durably Enhances Intuitive but not Formal Mathematics." *Science* 357(6346): 47-55.
- Duflo, Esther. 2004. "Scaling Up and Evaluation." In Bourguignon, Francois and Boris Pleskovic (eds.) *Accelerating Development*. New York, NY: Oxford University Press.
- Duflo, Esther. 2006a. "Field Experiments in Development Economics," In *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress, Volume 2, Blundell, Richard, Whitney Newey, and Torsten Persson (eds.), 322–348. New York, NY: Cambridge University Press.
- Duflo, Esther. 2006b. "Poor but Rational?" In Banerjee, Abhijit, Dilip Mookherjee and Roland Benabou (eds.) *Understanding Poverty*. New York, NY: Oxford University Press.
- Duflo, Esther. 2012. "Women Empowerment and Economic Development." *Journal of Economic Literature* 50(4): 1051-1079.
- Duflo, Esther, Pascaline Dupas and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101(5): 1739-1774.

- Duflo, Esther, Pascaline Dupas and Michael Kremer. 2015. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." *Journal of Public Economics* 123: 92-110.
- Duflo, Esther, Rachel Glennerster and Michael Kremer. 2006. "Using Randomization in Development Economics Research: A Toolkit," NBER Technical Working Paper 333. Reprinted in *Handbook of Development Economics*, 2007, Volume 4, Schultz, T. Paul and John Strauss (eds.), 3895-3962. Amsterdam: North Holland, Elsevier.
- Duflo, Esther, Michael Greenstone, Rohini Pande and Nicholas Ryan. 2018. "The Value of Regulatory Discretion: Estimates from Environmental Inspections in India." *Econometrica* 86(6): 2123-2160.
- Duflo, Esther, Rema Hanna and Stephen Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102(4): 1241-1278.
- Duflo, Esther, Michael Kremer and Jonathan Robinson. 2008. "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya." *American Economic Review* 98(2): 482-488.
- Duflo, Esther, Michael Kremer and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101(6): 2350-2390.
- Duflo, Esther and Emmanuel Saez. 2003. "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *Quarterly Journal of Economics* 118(3): 815-842.
- Dupas, Pascaline and Edward Miguel. 2017. "Impacts and Determinants of Health Levels in Low-Income Countries." In *Handbook of Field Experiments*, Volume 2, Duflo, Esther and Abhijit Banerjee (eds.). Amsterdam: North Holland, Elsevier.
- Glewwe, Paul, Nauman Ilias and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2(3): 205-227.
- Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied* 1(1): 112-135.
- Glewwe, Paul, Michael Kremer, Sylvie Moulin and Erik Zitzewitz. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics* 74(1): 251-268.
- Glewwe, Paul and Karthik Muralidharan. 2015. "Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." RISE Working Paper 15/001.
- Greenwald, Anthony G., Debbie E. McGhee and Jordan L.K. Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74(6): 1464-1480.

- Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80(2): 223-255.
- Gueron, Judith M. 2017. "The Politics and Practice of Social Experiments: Seeds of a Revolution." In *Handbook of Field Experiments*, Volume 1, Duflo, Esther and Abhijit Banerjee (eds.). Amsterdam: North Holland, Elsevier.
- Hargreaves, James R, Alexander M Aiken, Calum Davey and Richard J Hayes. 2015. "Authors' Response to: Deworming Externalities and School Impacts in Kenya." *International Journal of Epidemiology* 44(5): 1596-1599.
- Harrison, Glenn W. and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42(4): 1009-1055.
- Hausman, Jerry A. and David A. Wise. 1985. *Social Experimentation*. Chicago, IL: University of Chicago Press.
- Hicks, Joan Hamory, Michael Kremer and Edward Miguel. 2015. "Commentary: Deworming Externalities and Schooling Impacts in Kenya: A comment on Aiken et al (2015) and Davey et al (2015)." *International Journal of Epidemiology* 44(5): 1593-1596.
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945-960.
- Holmström, Bengt. 1982. "Managerial Incentive Problems: A Dynamic Perspective." In *Essays in Honour of Lars Wahlbeck*, Helsinki, Finland.
- Kremer, Michael. 2003. "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons." *American Economic Review* 93(2): 102-106.
- Kremer, Michael, Conner Brannen and Rachel Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340(6130): 297-300.
- Kremer, Michael and Rachel Glennerster. 2011. "Improving Health in Developing Countries." In *Handbook of Health Economics*, Volume 2, Mark V. Pauly; Thomas G McGuire; Pedro Pita Barros (eds.). Amsterdam: North Holland, Elsevier.
- Kremer, Michael and Edward Miguel. 2007. "The Illusion of Sustainability." *Quarterly Journal of Economics* 122(3): 1007-1065.
- Kremer, Michael, Edward Miguel, Jessica Leino and Alix Peterson Zwane. 2011. "Spring Cleaning: Rural Water Impacts, Valuation and Property Rights Institutions." *Quarterly Journal of Economics* 126(1): 145-205.
- Kremer, Michael, Gautam Rao and Frank Schilbach. 2019. "Behavioral Development Economics." *Handbook of Behavioral Economics - Foundations and Applications* 2, Volume 2, Bernheim, Douglas, Stefano DellaVigna and David Laibson (eds.). Amsterdam: North Holland, Elsevier.
- Kruk, Margaret E., Anna D. Gage, Naima T. Joseph, Goodarz Danaei, Sebastian Garcia-Saiso and Joshua A. Salomon. 2018. "Mortality Due to Low-quality Health

Systems in the Universal Health Coverage Era: A Systematic Analysis of Amenable Deaths in 137 Countries”. *Lancet* 392: 2203-2212.

Lucas, Robert E. 1988. “On the Mechanisms of Economic Development.” *Journal of Monetary Economics* 22(1): 3-42.

Mankiw, N. Gregory, David Romer and David N. Weil. 1992. “A Contribution to the Empirics of Economic Growth.” *Quarterly Journal of Economics* 107(2): 407-437.

Manski, Charles F. and Irwin Garfinkel. 1992. *Evaluating Welfare and Training Programs*. Cambridge, MA: Harvard University Press.

Miguel, Edward and Michael Kremer. 2004. “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica* 72(1): 159–217.

Newhouse, Joseph P. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge, MA: Harvard University Press.

Persson, Torsten and Guido Tabellini. 2000. *Political Economics*. Cambridge, MA: MIT Press.

Schweinhart, Lawrence J., Helen V. Barnes and David P. Weikart. 1993. *Significant Benefits: The High/Scope Perry Preschool Study Through Age 27*. Ypsilanti, MI: High/Scope Press.

Thaler, Richard. 1980. “Toward a Positive Theory of Consumer Choice.” *Journal of Economic Behavior and Organization* 1(1): 39–60.

Thomas, Duncan P. 1997. “Sailors, Scurvy and Science.” *Journal of the Royal Society of Medicine* 90(1): 50-54.

Todd, Petra E. and Kenneth I. Wolpin. 2006. “Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility.” *American Economic Review* 96(5): 1384-1417.

UIS. 2017. *More than One-Half of Children and Adolescents Are Not Learning Worldwide*. Fact Sheet No. 46. UNESCO/UIS. Montreal.

UK Government. 2009. “PM’s Article on Universal Healthcare”. Statements and articles 2009/09 (<https://webarchive.nationalarchives.gov.uk/20100511120725/http://www.number10.gov.uk/news/statements-and-articles/2009/09/pms-article-on-maternal-healthcare-20720>).

Vermeersch, Christel and Michael Kremer. 2005. “School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation.” Policy Research Working Paper; No. 3523. World Bank, Washington, DC.

WHO. 2017. *Guideline: Preventive Chemotherapy to Control Soil-Transmitted Helminth Infections in At-Risk Population Groups*. Geneva: World Health Organization.