Conducting influential impact evaluations in China: The experience of the Rural Education Action Project (REAP)

Matthew Boswell, Scott Rozelle, Linxiu Zhang, Chengfang Liu, Renfu Luo, and Yaojiang Shi

Abstract

Impact evaluation has become an increasingly integral part of development project design and execution in recent years. Many questions remain, however, about what methods yield the most compelling evaluations, and how best to implement them. The Rural Education Action Project (REAP) is among the most successful impact evaluation groups currently operating in China. The goal of this paper is to share five practical strategies that REAP has employed to maximize the effectiveness of our impact evaluations. These strategies include the use of randomization and other experimental and quasi experimental research designs; pursuit of local and international collaboration; strict attention to policy relevance; a modular, incremental research approach; and robust outreach.

Working Paper 212

October 2010

reapchina.org/reap.stanford.edu

Conducting influential impact evaluations in China:
the experience of the Rural Education Action Project (REAP)

An experience-sharing paper for 3ie

Mathew Boswell and Scott Rozelle
Stanford University

Linxiu Zhang, Chengfang Liu and Renfu Luo
Institute of Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences

Yaojiang Shi
Northwest University of Xi'an

Conducting influential impact evaluations in China: the experience of the Rural Education Action Project (REAP)

Impact evaluation has become an increasingly integral part of development project design and execution in recent years. Many questions remain, however, about what methods yield the most compelling evaluations, and how best to implement them. The Rural Education Action Project (REAP) is among the most successful impact evaluation groups currently operating in China. The goal of this paper is to share five practical strategies that REAP has employed to maximize the effectiveness of our impact evaluations. These strategies include the use of randomization and other experimental and quasi experimental research designs; pursuit of local and international collaboration; strict attention to policy relevance; a modular, incremental research approach; and robust outreach.

The first section of this paper provides a brief background on REAP's work to date. The following sections will offer specific details on each of these strategies. A concluding segment will present a summary of REAP's lessons learned and recommendations for good practice.

This paper draws on the experience of team members involved in all aspects of our impact evaluation work, including principal investigators, project managers, volunteers, implementation teams, and government partners. In the past five years REAP has carried out more than 25 impact evaluations. For this reason we are also able to synthesize lessons across several studies. It is hoped that a review of REAP's experiences in this regard will offer guidance to other organizations and encourage the further integration of empirical impact evaluation methodologies into development initiatives worldwide.

REAP: organizational background

China is currently transforming into a modern, knowledge-based economy. Building a skilled labor force is a vital component of this transition. The enhanced human capital that underpins such a labor force requires the equitable delivery of quality education throughout the country. However, today millions of rural and underprivileged Chinese are unable to access adequate schooling. For this reason, serious questions remain about whether China's labor can rise to the challenge of establishing an innovation-oriented economy.

The Rural Education Action Project (REAP) is an impact evaluation organization that aims to inform sound educational policy in China. A collaboration of the Center for Chinese Agricultural Policy (CCAP) in the Chinese Academy of Sciences in Beijing and the Freeman Spogli Institute for International Studies at Stanford University, REAP's goal is to help students from vulnerable communities in China enhance their human capital and overcome obstacles in education so that they can escape poverty and better contribute to China's developing economy. REAP research focuses on three key platforms: a.) nutrition, health and education; b.) technology and human capital; and c.) access to quality secondary and tertiary education.

REAP's core competency is impact evaluation through experimental design. We strive to identify attribution between a program's inputs and outcomes by collecting panel data, employing experimental and quasi-experimental research methods, and conducting large scale randomized

control trials (RCTs). Subsequent data driven analysis seeks to produce compelling evidence about which programs work and which do not.

REAP has a track record of success. Our evaluation services are in demand from many national ministries and commissions. We have collaborations with scores of partners. Our annual training events are oversubscribed. Formal policy briefs based on our findings have been recognized by a host of regional and national leaders, including the prime minister, and incorporated into new policy initiatives. To date, we have advised provincial and county departments of education, finance and health in four provinces. We have published 20 papers in refereed, SSCI-indexed international journals since 2006.

Strategies for Making Impact Evaluation More Effective

Strategy 1: Use of Randomization and Other Experimental and Quasi-experimental Research Designs

REAP is interested in discovering which programs work and which do not. In seeking to solve problems and inform sound policy, many ideas and approaches can appear sensible. However, it is often difficult—and in some cases impossible—to use traditional monitoring and evaluation techniques to ascertain whether a given program generates the desired outcomes. In order to more reliably establish causation, REAP employs randomized interventions to help China's authorities determine whether policies deliver desirable outcomes. Like a clinical trial in a medical setting, randomized trials involve treatment and control groups that are statistically identical. When evaluating an intervention, the control group serves as a basis of comparison for the treated group. Randomizing ensures that we can account for key variables in a way that is impossible in a corresponding observational study. For this reason, most researchers agree that randomized experiments are superior to all other designs in terms of statistical reliability.

REAP's empirical approach is a key aspect of our outreach to policymakers as well. Randomized experimental designs are capable of producing simple and compelling evidence. Numbers, after all, do not lie, and policymakers rightfully trust them. For this reason our quantitative results have been instrumental in convincing authorities of a given policy's utility. By sharing our results with relevant authorities, REAP has successfully persuaded policy makers to scale up effective pilots, modify or stop ineffective ones, and otherwise answer difficult questions about the impact of policies, programs, and public investments.

A good example of this can be found in REAP's work in high school education. The provincial government in Shaanxi has been considering an increase in subsidies to students in a bid to raise high school enrollment among the province's poor. This plan could pose a costly investment and the authorities are keen to grasp its potential effects. Through consultation with the Shaanxi Province Policy Advice Desk, REAP is rolling out a randomized control trial in which we offer vouchers to poor junior high students. The vouchers could help defray the costs of attending high school that many believe are preventing poor young people from staying in school and performing according to their potential. The grades of poor students that receive the vouchers will then be compared to poor students who do not, thereby illuminating whether or not the

promise of cheaper high school can motivate children to work harder to pass high school admissions tests. REAP's quantitative analysis of the data will then serve to inform future subsidy measures rolled out by Shaanxi authorities.

Strategy 2: Pursuit of collaboration

Collaborations lie at the heart of REAP's work. About a dozen researchers and program staff at Stanford University and their counterparts at CCAP comprise the core team at REAP. This group cooperates to oversea all of REAP's initiatives and make all strategic decisions. Beyond this nucleus, however, REAP actively pursues links to outside entities, and to date has collaborated with over 50 academic institutions, foundations, non-governmental organizations, and government bodies. These collaborations can be roughly divided into two categories: those that leverage subject matter expertise and those that leverage implementation expertise.

Leveraging subject matter expertise. In designing our projects, REAP leverages the intellectual resources of interdisciplinary partners. For example, on the Stanford campus, we collaborate with researchers from the School of Education, School of Medicine, School of Engineering, and key social science departments, such as Economics and Sociology. In China we have collaborations with researchers in departments of medicine in several major universities, and with the Center for Disease Control at the national, provincial and local levels. Frequent collaborations with other economists in China and Stanford have also proven an enormous help in establishing content for survey forms and other experimental tools. Incorporating the knowledge and experience of such diverse parties throughout academia greatly magnifies the rigor and precision of our interventions. What is more, these collaborations link REAP with top scholars that have their own contacts in the government and institutions of education in China.

Take for example REAP's work on anemia among primary school children. REAP is composed largely of economists, not doctors. In order to conduct interventions that target anemia rates, we consult regularly with doctors, health economists, and nutritionists to ensure we adhere to best practices. For our anemia relevant studies, partnerships with a nutritionist at Emory University, a health economist and a pediatrician at Stanford, and a nutritionist Xian Jiaotong University have been instrumental in our choosing what blood tests to use, what vitamins to administer and in what amounts, what ethical considerations to account for, and how to incorporate age and altitude adjustments into our anemia rate assessments, among other things. Sometimes these experts also co write our papers and help design future anemia related experiments. Without the input of such partners, we could not design our health related interventions or analyze their data effectively.

Leveraging implementation expertise. Large-scale randomized control trials require an enormous amount of human resource capacity to carry out. Just one such trial can involve multiple physical visits to tens, if not hundreds, of villages, clinics, schools, or households across sprawling, remote areas. These visits are necessary to collect background information, conduct baseline and evaluation surveys, and other relevant tasks. Dozens of support staff are needed to carry out these functions. REAP achieves this capacity by collaborating with implementation teams sourced close to project sites. A signature such collaboration has developed between REAP

and the Northwest Social Science Development Research Center (NSDRC). This network of seven universities in China's western region is based at the School of Economic Management at Northwest University in Xian City, Shaanxi Province. Affiliates from NSDRC have been instrumental in gaining official endorsements and drafting the professional and volunteer staff necessary to implement several of REAP's evaluations. REAP has forged implementation partnerships with many other entities as well, including eight provincial-level government bureaus, the China Center for Disease Control, and NGOs with staff stationed throughout poor areas of the country. Without these partnerships REAP could not reach the large number of subjects required by rigorous experimental designs.

One of the several projects in which such implementation expertise have been critical is a computer assisted learning (CAL) project REAP is carrying out in Shaanxi boarding schools. During the early stages of planning REAP's primary contact at the NSDRC contacted the relevant education officials in Shaanxi to finesse their support for project implementation. The same individual arranged for a team of student volunteers from his university in Xian to call each school in three target counties to collect data on class sizes, school assets, and other relevant information for our intervention. From these data our CAL project manager in Beijing established an appropriate sample frame for a randomized control trial, selecting some schools to receive a computer tutoring intervention and some not to. The initial school data is also instrumental in conducting the baseline and evaluation tests as the project moves forward. Without the help of our implementation partner in Xian, it would have been very difficult to establish credibility with schools and undertake the legwork required to collect initial school data.

Establishing such broad networks of collaborations has costs as well as benefits. For one thing, it is time consuming to involve others. Credit often has to be shared. Sometimes senior authorship and speaking rights at conferences have to be given to the staff of the collaborators. However, as noted above, there are also many benefits. More collaborators mean more capacity to do more work. When collaborators are chosen carefully, there can be synergies that make the quality of impact evaluation higher and the subsequent efforts to upscale projects more effective. The key is choosing the right number of the best collaborators so the benefits outweigh the costs.

Strategy 3: Pursuing Policy Relevance

Policy relevance is a cardinal tenet of REAP's research. We seek to only design interventions that policymakers are likely to take an interest in. If an intervention cannot be adopted or expanded as a broader policy, it is of no practical use outside of an economics journal. REAP takes several steps to ensure that our interventions maintain policy relevance.

Emphasizing the policy angle. When choosing and designing an intervention, REAP is careful to address policy relevant problems (ways to improve student nutrition, for example). Equally important, in seeking solutions to these problems we only assess approaches that are feasible from a policy making stand point. We ask ourselves: Would policy makers care about what we are trying to figure out? Would they understand the intervention? If a given intervention is successful, would the relevant authorities have the resources to carry it out on a larger scale? Keeping these questions in mind helps make sure our results resonate within officialdom.

For example, several of REAP's experiments have focused on the link between rampant anemia and other nutritional deficiencies to poor academic performance in poor areas of China. To see if improved nutrition can raise test scores, it would seem simple enough to purchase nutrition supplements (multivitamins or quality cafeteria food, for example), give them to a randomly selected group of treatment schools, make sure that the students ate the supplements, and measure the effect on test scores (against a control group) after a certain period. However, it is unclear if this is the intervention that China's government would choose to pursue on a large scale, no matter how effective such supplements might be in raising test scores. As a result of this concern, we designed an experiment to see if we could convince local actors, in this case school principals, to take on the burden of improving nutrition themselves, by using subsidies, training, or the promise of personal cash inducements based on their success in reducing anemia. By designing the intervention in this way, it becomes simple from the point of view of a policy maker to not only learn the most effective treatment model, but also transform that model into a useful policy.

Making policy makers stakeholders. REAP endeavors to invest policy makers in the outcome of a project. We are careful make sure they understand the methods and rationale behind randomization. In many projects that REAP evaluates, we ask policy makers to serve on a temporary "board of advisors" for the project. We want them to be involved from the very start as we discuss details of the intervention and the design of the randomization. Ideally, policymakers even consider the policy or program to be "their project." In this way we can secure the endorsements necessary for large-scale implementation, and also attain the support required to scale up those interventions that we find to be effective.

Policy makers in a poor rural county in Shaanxi have become firm stakeholders in a REAP intervention that measures the effect of eliminating high school tuition fees on enrollment. In consultation with REAP, the county in question, Ningshan County, has used its own funds to eliminate tuition fees across the county, and is eager to see evidence of the effects. REAP is comparing performance levels in the treatment county versus performance in two adjacent counties where the policy does not exist. County officials in Ningshan have been on hand at all stages of the intervention. They trust our evidence-based results, and will use them when assessing whether or not it is worth their while to continue their program.

Keeping the message simple. For the most part, REAP is not interested in highly complex interventions. These are difficult to carry out and tend to be geared more toward proving someone's grand theory than addressing a policy challenge. Instead, we focus on interventions that reveal simple yet forceful lessons that officials can understand and act on. When we complete an intervention we need only to show a bar graph that indicates the change in the treatment group versus the change in the control group and the message of the intervention is transmitted.

By emphasizing policy relevance in these ways, REAP has been successful in attracting the attention of policy makers. We have also been able to keep their attention (and not get bogged down in complicated statistical analysis). In this way REAP strives to operate beyond the narrow confines of academia and bring the fruit of rigorous impact evaluation—i.e. effective policy—to the people that need it the most.

Strategy 4: Incrementalism

Patience and a willingness to take a modular and incremental approach have also served REAP well. After exploring opportunities and settling on an area where we can add value, we start small with a relatively limited pilot and an evaluation. After analyzing these results we will move to a larger test of a concept, often incorporating several different approaches. In testing the larger pilot, we are careful to evaluate not only the efficacy but also the cost effectiveness of different approaches. Lessons from the pilot and follow on studies allow the organization to achieve a firm understanding of the dynamics of a given problem as well as fruitful solutions. With this understanding we move forward to work with upper level governments to roll out projects on a larger scale in a way that affects the lives of tens of thousands or even millions. An example can illustrate this incremental approach:

As recently as two years ago, anemia and other micronutrient deficiencies among students were not on anyone's radar screen in China—in either the Ministry of Education or the Center for Disease Control's child health division. In the past two years, REAP has used large randomized sampling exercises to test more than 30,000 children from ages 3 to 12 in five provinces and in more than 40 counties. We have learned that more than 30 percent of school-aged children in poor rural areas are iron deficient.

The first pilot program in this area assessed the effectiveness of curing anemia on educational performance, and was done in performed in eight poor counties in Shaanxi Province. The objective: to learn whether improving nutrition could reduce anemia and raise standardized test scores. Even though the study site was circumscribed, it was the earliest RCT that that had ever been undertaken on the issue. The intervention also incorporated a sample size just large enough to achieve the statistical power necessary to reliably assess the impact of the intervention. In carrying out the trial, we showed that when students take a simple over-the-counter vitamin, at a cost of US $4 per year, per student, iron deficiencies fall and educational performance rises.

The Shaanxi government saw these results and took action. They ordered that every primary school student in the province should receive an egg a day. This was a breakthrough for a school system a poor province that had never before taken on responsibilities for the health and nutrition of students. Educators were trying to see if nutrition was a productive input to quality rural education. There were skeptics, however, and debate arose whether or not this program would raise the educational performance of Shaanxi children. In response, the Shaanxi government is having REAP conduct an independent evaluation. The project is called "Is One Egg Enough?" In a set of randomly selected "check and intervention schools," we are providing vitamins, nutritious supplemented meals and nutritional training for parents, teachers and educators to see if these additional inputs can enhance the effectiveness of the one egg per day program.

Local governments have been impressed enough to try the vitamin project in their schools. In one poor county, the government is testing a vitamin initiative, using its own resources in half of its schools, and working with REAP to conduct an independent evaluation. If vitamins are shown to increase educational performance and lower rates of malnutrition, the county bureau of education is planning to give vitamins to all of their students for at least the next five years.

After completing these two upscaling projects, using nutrition as an input is gather momentum and spreading. The Ningxia Provincial Departments of Education and Finance will roll out a REAP-designed; REAP-advised nutritional training program in the 2010-11 academic year in four counties. More than 50,000 nutritional guides and pamphlets will be passed out during intensive training sessions. In addition, Qinghai and Ningxia provinces are experimenting with nutritional inputs in schools. They are using a "paying for performance" approach to see if principals and teachers are able to use resources effectively and solve anemia in their own way and with their own resources.

In these ways, a project that started as a small, randomized trial has grown into several major policy initiatives that affect large swaths of northwest China.

Strategy 5: Robust Outreach

Economics journals are important, but not very many people read them. Part of REAP's strategy for success is to disseminate our findings to as wide an array of audiences as possible in order to generate interest in our work. To ensure "cross border" accessibility, the majority of our outreach materials are available in both English and Chinese. REAP publicity targets three main groups: general audiences, policy makers, and academia.

General audiences. Quantitative econometric data does not excite the casual reader. To connect our research to the lay observer, REAP works hard to strip the numbers from our findings and relate them in reader friendly avenues. To this end, we maintain a comprehensive website that contains summary descriptions of all of our projects, their motivations, and findings. The REAP website also contains interviews, videos, and photo journals that document the lives and stories of the poor people in China whom we target in our research. We also publish a *REAP Brief* series of pamphlets that documents each of our projects in a glossy, reader friendly format. Other outreach efforts in this area include writing press releases about our findings, pursuing interviews with major media outlets in China and the US, and sending out quarterly e-newsletters about our ongoing projects. In addition to keeping lay readers informed, these outreach efforts are critical to generating interest in our work from potential collaborators and funders.

The REAP brief series can be accessed on our website, http://reap.stanford.edu.

Policy makers. Whenever REAP's research generates particularly compelling findings, we send a formal policy brief to China's national authorities. This requires distilling volumes of data into a four page document with a cogent, resonant message. Core staff at REAP work hard to craft this message. The resulting brief is then forwarded through the Chinese Academy of Sciences to the policy advice desk at the Office of the Premier.

A recent REAP policy brief was based on the findings of our intestinal worms work in Guizhou Province. A yearlong project there found that about 30 percent of children in the province carried intestinal worms. Such worms have been shown to adversely affect health in a variety of ways, not least of which by causing anemia. The policy brief was read by Liu Yandong, the vice premier and politburo member in charge of education and health. In her notes she expressed shock and her office soon authorized the allocation of 200 million yuan to eradicate

worms in the province by 2012. REAP will be on hand to conduct a follow up canvas survey at that time.

REAP has submitted seven such briefs over the past three years, and they have played important roles in informing subsequent policy measures. REAP also takes great care to share our findings with government partners at the provincial and county levels as well. REAP's success in informing policy in China is due in large account to these outreach efforts to policy makers.

Academia. Despite our focus on informing policy, REAP is still functions in large part as a research organization. Nearly all of our projects generate at least one paper in an SCII-indexed journal. In fact, we see little conflict—and lots of complementarities—between publishing in academic journals and focusing on policy. Publishing our work ensures that we are carrying out our work at the highest quality standards of the field. Our policy results are also being scrutinized by other scholars in the field—those in the China field and those in the evaluation field. This published work is linked to our website alongside full text versions of our numerous working papers. Our survey instruments and data are also available on our website (http://reap.stanford.edu/docs/reap_survey_instruments) to anyone interested in examining or using them.

In addition, we are focused on training and sharing our skills in Impact Evaluation. Each year REAP holds at least two workshops to share our results with academics, as well as interested officials, educators, NGOs and journalists. We work hard to make all of our data and survey instruments available on-line to serve as a reference to interested parties in academia. We do not consider these our intellectual property – if others are interested in using our tools or copying our methods, we welcome them – the more rigorous impact evaluation there is in the world, the better.

Conclusion: Lessons Learned and Recommendations for Good Practice

Four years and twenty-five impact evaluations after its inception in 2006, REAP's experience in impact evaluation can be distilled into four key lessons and recommendations.

Emphasize Empiricism. There are thousands of government entities, private organizations, and research institutions around the world—and especially in China today—that are dedicated to solving problems for vulnerable populations. Often they are awash with money and good intentions. Yet the problems they are committed to solving persist. REAP believes this is partly because very few organizations are able to convincingly answer a fundamental question about their efforts: Do they work? Quantitative, experimental design is the best means to reliably measure success and effectively channel ideas and investments that target the world's most pressing problems.

Leverage Local Expertise. Rigorous randomized impact evaluation requires the involvement of large numbers of people and intimate knowledge of local circumstances. REAP is a group of economists, some of whom are from non-Chinese institutions. We have not tried to come in from the outside, do all of the evaluation work, and lecture policy makers about our findings. Instead whenever possible we try to leverage local partnerships so that stakeholders on

the ground can understand our methods and become invested in our work. This increases the quality of our interventions and extent to which our findings are heard.

Focus on the Practical. There is a tendency among research organizations to focus excessively on theoretical abstractions rather than delivering practical solutions. By incorporating local partners, focusing on policy relevant problems, shedding light on workable solutions, and keeping our messages simple, REAP has been able to not only produce cutting edge research results, but also deliver useful policy solutions to people that need them.

Communicate Effectively. An excellent policy prescription that has hard data to back it up should be accessible to people other than just econometricians. REAP's capacity to perform depends on sharing our results and reaching out to groups whose interests align with our capabilities. The only way to discover and leverage these synergies is to disseminate news of our work and results to as wide an audience as possible. An effective policy intervention will do no one any good if nobody knows about it.

Rigorous impact evaluation is a vital component of effective development work. By adhering to the above lessons, REAP has made important strides in informing sound education and health policy in China. We hope that sharing these lessons can help other groups deliver on the abundant promise of strong impact evaluations.