



**The American
University in Cairo**

School of Business

John D. Gerhart Center for
Philanthropy, Civic Engagement
and Responsible Business



**FORD
FOUNDATION**

MEASURING SOCIAL IMPACT OF PHILANTHROPY IN THE ARAB WORLD

Abdel Rahman Nagy

Director of Learning and Strategy, Sawiris Foundation
for Social Development

Engy Saleh

Research Manager, (Abdul Latif Jameel Poverty Action Lab
(J-PAL), MENA)

MEASURING SOCIAL IMPACT OF PHILANTHROPY IN THE ARAB WORLD

Abdel Rahman Nagy

Director of Learning and Strategy, Sawiris Foundation
for Social Development

Engy Saleh

Research Manager, (Abdul Latif Jameel Poverty Action Lab
(J-PAL), MENA)

The poor are no less rational than anyone else—quite the contrary. Precisely because they have so little, we often find them putting much careful thought into their choices: They have to be sophisticated economists just to survive.”

- Abhijit V. Banerjee

Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty

WHY EVALUATE THE IMPACT OF SOCIAL PROGRAMS?

Despite the fact that the role of philanthropies in the Arab world has been progressing from its charitable and religious nature to being more development-focused, it is still possible to enlarge philanthropies' contribution to society and maximize their impact in a structured and credible way that channels the philanthropies' experiences, lessons learned and best practices. One key challenge impeding philanthropies from realizing their potential is that they lack the necessary data to advocate for their work and prove that the funds allocated to projects are effective. This is because the work is not systematically planned, monitored, or reported, making it "more difficult to have a strategic approach to third sector operations" (Hamoud as cited in Hartnell, 2018, p.43). Evaluating the programs they support addresses this challenge.

The objective behind an evaluation is not always obvious, especially to those involved in the day-to-day activities of program implementation. However, evaluations in their broader meaning can help policymakers and practitioners answer critical questions about their programs, especially when faced with a problem they cannot justify:

Put simply, a program evaluation is meant to answer the question, "how is our program or policy doing?" This can have different implications depending on who is asking the question and to whom they are talking. For example, if a donor asks the NGO director, "how is our program doing?" she may imply, "have you been wasting our money?" This can feel interrogatory. Alternatively, if a politician asks her constituents, "how is our program doing?" she could imply, "is our program meeting your needs? How can we make it better for you?" ("Introduction to Evaluations").

Evaluating programs is part of a growing global trend of evidence-based policymaking that shifts the focus from inputs to results, and there is an increased demand for using those results to inform future decisions related to accountability

and budget allocations to the most effective interventions (Gertler, Martinez, Premand, Rawlings & Vermeersch, 2011). Also, as the world is now going through a global pandemic, more and more entities and philanthropic organizations are keen on knowing how to serve their target populations better and maximize their impact using scientific evidence on what works with an aim of effective altruism or doing good better. An example of these organizations is Sawiris Foundation for Social Development, which is developing its entire strategy 100% based on scientific evidence.

A comprehensive package of evaluating programs would include needs and program theory assessments, which lay out the problem or needs being addressed by the program and how the program intends to help satisfy those needs. It also includes process evaluation, where the program activities are being closely monitored by program managers, impact evaluations to ensure that the program is achieving its objectives, and finally, cost-benefit and cost-effectiveness analyses that are critical for policy decisions related to the worthiness of the program costs (the former), and the budget allocated to a particular program as opposed to others (the latter) ("Introduction to Evaluations").

The distinct feature of philanthropies, however, is that they are private initiatives for the public good, and hence the type of programs they support is aimed specifically at improving people's lives. This adds a unique and challenging social dimension to why and how philanthropies should evaluate their programs' impact. There are multiple reasons for evaluating the social impact, and all of them tie back to the broader aim of having a more effective and more accountable third sector. With limited available resources to combat social problems, impact evaluations allow for the continuous improvement and enhancement of the quality of interventions, and thus they inform program design as well as optimum budget allocation decisions. In doing so, philanthropies have the ability to build

the knowledge and evidence base on what works and present development models that are suitable and ready for scale-up. They will, therefore, be fulfilling an important part of their role in society in terms of integrating with the public and private sectors to achieve the broader development goals and improve service delivery. A great example of how impact evaluation influenced policy and the decision to scale up is from the Mexican social protection program Oportunidades/Progresá, "its impact evaluation provided credible and strong evidence that could not be ignored in subsequent policy decisions. The program's adoption and expansion were strongly influenced by the evaluation results. Today, the Oportunidades program reaches close to one out of four Mexicans and is a centerpiece of Mexico's strategy to combat poverty" (Gertler et al., 2011).

Finally, being able to justify budget allocation decisions and promoting a culture of testing, learning, and building evidence addresses a key challenge philanthropies in the Arab world face; the general public's lack of trust in NGOs.



CAUSATION VERSUS CORRELATION

After establishing Why there is a need to evaluate, the next step is to explore How to evaluate to ensure that the changes in the well-being of individuals could truly be attributed to the program. The main challenge for any impact evaluation methodology is to establish causality. In other words, to try to disentangle the program's effect or the intervention on the outcomes of interest from other factors that could be affecting the same outcomes. This is particularly relevant to practitioners and policymakers who are keen on establishing a cause-and-effect kind of relationship that will enable them to confidently advocate for - or against - the scale-up of a specific intervention or program. For example, policymakers are keen on understanding whether vocational training programs actually improve outcomes for participants. However, it is not sufficient to know that the participants' incomes increased after the program, simply because incomes could have increased for a variety of reasons other than the training; better economic conditions, favorable labor market regulations, or even individual efforts from the participants themselves (Gertler et al., 2011).

In order to infer the change in outcomes that is caused by the intervention, the evaluation method used should enable us to "examine how the people who participated in the program fared compared to how they would have fared if they had not participated in the program," or what is called the counterfactual (Glennester & Takavarasha, 2013, p.24). However, it is practically impossible to observe this counterfactual state directly because at a given point in time, it is not possible to view the same individuals with and without the program. Therefore, what the different evaluation methodologies attempt to do is to try to mimic this counterfactual state by making a number of assumptions about the comparison group, which is the "group of individuals who, in the absence of the program, would have had outcomes similar to those who were exposed to the program" (Bourguignon, Pleskovic and Duflo, 2004, p.343). This comparison group could, under certain conditions, either be a group of non-participants or the same group of participants before the start of the program, or a combination of both.

DIFFERENT IMPACT EVALUATION METHODOLOGIES

Assessing the impact of a social program could be done in multiple ways. One evaluation method is using qualitative tools, in particular with interventions that address issues that are difficult for participants to discuss openly and comfortably like domestic violence or war memories. One of the key strengths of qualitative tools is that they can generate rich insights and a wealth of information from participants that would not have been otherwise generated had the question been asked to them directly. The insights that result from qualitative methods are usually easier for the public to relate to because they paint a picture full of details and provide justifications for foreign phenomena to some contexts. That is why qualitative tools are also used in many cases to support quantitative evaluations in order to define the problem being addressed more tangibly and provide explanations and justifications for the findings.

And so in using tools like open-ended interviews, direct observation or focus group discussions with program participants, the comparison group becomes implicitly the participants themselves when they talk about how their lives had changed after the interventions. Of course, this requires a number of assumptions to hold true for the comparison group to be valid and inform us about the true effect of the program. First, evaluators should be able to clearly convey to participants that they need to differentiate between the changes that took place in their lives purely due to the program and the ones that were going on in the society regardless. Also, we need to assume that participants' answers and actions are not biased by what they expect the researcher wants to hear or see or what is considered socially desirable. Asking males for example whether providing information about the harmful consequences of wife-beating has changed their perceptions and attitudes towards their wives is expected to be susceptible to much bias which should be taken into consideration while designing the data collection tools. Finally, we have to assume that the summaries and

reporting of the findings are also not influenced by the preferences and biases of the researchers themselves (Glennester and Takavarasha 2013).

On the other hand, quantitative methods of program evaluation can take many forms which are best demonstrated using an actual program as an example. The ultra-poor graduation program, which was initiated in Bangladesh by BRAC, the largest NGO in the world, provides a holistic solution for the ultra-poor through a number of sequenced and carefully designed interventions to help them graduate out of poverty within the course of two years. The participants receive consumption support, an asset as a livelihood option, technical and financial literacy training as well as coaching and mentorship. This program was piloted and evaluated in seven countries, reaching thousands of poor households, and is also currently being piloted in the poorest two governorates in Upper Egypt, Assiut and Sohag, aiming to target 2,400 of the poorest households. The program is also expanded in several states in India and scaled up in Ethiopia and incorporated into its National Productive Safety Net Program estimated to serve around 675,000 households (Fahey, 2015).

The ultra-poor households are evaluated on a variety of outcomes that capture the improvement in their socio-economic well-being and that eventually determines their graduation from poverty. If we take for example the livelihood dimension of the program, one of the most commonly used methods of evaluating the impact is to do a pre and post comparison of the performance of the participating households on this dimension. The problem with this method, however, is that if we witness an improvement in the outcomes of the participating households in terms of income and diversification of assets, then we are assuming that there were no concurrent changes that were taking place during the two years of the program, like other development programs providing income-generating activities or technical training or even that the general

prices of livestock decreased during that time leading households to acquire more assets, which we would wrongly capture as the effect of our intervention.

Another common method is to do a simple comparison between the outcomes of the participating households and those of non-participating households. We are assuming in this case that the households that ended up actually taking the program are identical to those that did not, and that is hardly the case. Most development programs have one of two ways of enrolling participants: they either encourage them to take part and participate, or the program puts in place a number of criteria that determine whether a household qualifies to be enrolled or not. Most, if not all, of the graduation approaches worldwide use the latter method in enrolling or excluding households. In Egypt, for example, households are excluded if they satisfy one of the following criteria: owning more than six qirates of land, owning large livestock like cows or buffaloes, occupying a house that is three floors or higher, owning a vehicle...etc. This, by default, results in differences that are pre-existing between the households that are enrolled in the program and those that are not. And therefore, comparing their outcomes retrospectively will result in biased results because we can never be sure that the changes in outcomes are not attributed to their pre-existing differences. The other enrollment method of encouragement also poses the same challenge because households that made an effort and applied to the program could be inherently different, in terms of motivation, for example, than those that did not. This challenge is referred to as selection bias.

A number of evaluation tools could be used to tackle the problem of selection bias, among which is Statistical Matching, Difference-in-Difference, Regression Discontinuity Design

(RDD) and Randomized Controlled Trials (RCTs). The four methods aim at creating a comparison group that is as similar as possible to the group of participants. The only difference will be that one group received the interventions while the other did not. In the Statistical Matching technique, every participant ultra-poor household is matched to at least one non-participant household that is similar on several observable characteristics like household structure, literacy status of the head and spouse, owned land, occupation of the members...etc. Another way of performing Statistical Matching is to use those characteristics to predict the probability (or propensity) of a certain household to participate in the program. Afterward, each participant household is matched to a non-participant that has the same probability of participating, with the assumption that it is entirely due to chance - and not due to unobservable characteristics that were not taken into account - that one household received the program and another did not (Glennerster and Takavarasha 2013).

The Difference-in-Difference approach attempts to solve the selection bias problem by combining the pre and post method with the participant-non participant comparison of outcomes. In other words, we measure the changes in outcomes of the two groups across time and compare those changes to each other. The whole idea that this approach rests on is that if we are assuming that there is the unobservable characteristic of risk-taking which is driving some of the increase in income for participants constantly over time, then its effect will be factored out when we calculate the change in income for participants between the beginning and end of the program. At the same time, in order to factor out any changes in the external environment that happened and that might have affected income like the decrease in the price of livestock, we obtain the difference between the change in income of participants and the change in income of non-participants. This will eliminate the change that happened at a certain point in time but affected both groups similarly. So in this approach, we are assuming that had it not been for the program, the participant households would have had the same change in income as non-participants. And therefore, this method will be invalid if the effect of risk-taking on income actually changes

over time (leading for example to a decrease in income for engaging in risky business), or if an external event affected participants only during the two years of the program (the government decides to give an additional asset to the program participants only).

The Regression Discontinuity Design (RDD) takes advantage of a policy or rule that determine the enrollment of some households into the program and not others. Imagine if households are enrolled in the graduation program if their total monthly income does not exceed 2000 EGP. This rule will automatically place households on either side of the eligibility cut-off based on their income. The RDD design rests on the assumption that the households who just made it to the program and those who just missed the program (those just above and just below the cut-off) cannot be that different from each other. Of course in many contexts where there is a high probability of manipulating the cut-off and misreporting income, applying this method could be challenging since the manipulation introduces elements of difference between the participant and the comparison groups.

Finally, Randomized Controlled Trials (RCTs), also known as field evaluations and social experiments, could be used as an experimental approach to evaluating social programs. The main idea behind this approach is that people who benefit from the program are randomly selected from a pool of targeted population. "This ensures that there are no systematic differences between those who receive the program and those who serve as the comparison group" (Glennerster and Takavarasha, 2013, p.44). And so, the internal validity, or the unbiasedness of the estimated impact, comes from the fact that we are able to vary one factor at a time and hence estimate the causal effect (Banerjee and Duflo, 2008).

Practically, RCTs look like the following: the method starts as a collaboration between researchers and policymakers to define the specific pool or criteria for the population eligible for the program to select from. The units in this pool are then randomly assigned, whether using an actual lottery or a random number generator, to a treatment group (i.e. the group that receives the program intervention) and a comparison

group, which does not receive anything:

A key advantage of the randomized approach is that in addition to observable characteristics (such as income and test scores), unobservable characteristics that we don't have a measure of (such as motivation or talent) will also be balanced between groups. This means that, on average, the treatment and comparison groups will be comparable at the onset of the program, and their trajectory will be the same in the absence of the program (Glennerster and Takavarasha, 2013, p.44, p.47).

Starting from the premise that social programs should have a transparent way of program assignment, especially with limited and scarce resources, random participant selection, where everyone who is eligible has the same probability of getting selected, is the fairest and most straightforward way to do so (Gertler et al., 2011). And thus, unlike the non-experimental approaches discussed above, which are applied to programs retrospectively, randomized evaluations are prospective in nature, meaning that program evaluations, in this case, need to be planned in advance in order to ensure that we are creating a comparison group in a random manner (Bourguignon, Pleskovic and Duflo, 2004).

Because it starts early on in the project, the process of an RCT also allows for the accommodation of the different elements of a comprehensive evaluation mentioned above. To start with, needs assessments are usually conducted in collaboration with implementers in order to study the nature of the population and identify the problem. Also, RCTs allow researchers to conduct process evaluations that not only ensure that the program interventions are carefully implemented, but also that there is no contamination between treatment and control groups, which is usually a challenging task, especially in situations where both groups exist in proximity to each other. Finally, because RCTs are in many cases costly since they require primary data collection, specifically in developing country settings where administrative data are either inaccessible or unorganized, doing a cost-benefit analysis is a key step in order to ensure

that benefits of the program justify the costs. Therefore,

At a very basic level, randomized evaluation can answer the question: Was the program effective? But if thoughtfully designed and implemented, it can also answer the questions, how effective was it? Were there unintended side-effects? Who benefited most? Who was harmed? Why did it work or not work? What lessons can be applied to other contexts, or if the program was scaled up? How cost-effective was the program? How does it compare to other programs designed to accomplish similar goals? (“Introduction to Evaluations”)

The Ultra-Poor Graduation Program was and is being evaluated through RCTs across different countries and contexts. The most famous was able to study the graduation approach in seven countries through seven randomized evaluations in Bangladesh, Ethiopia, Ghana, Honduras, India, Pakistan, and Peru. Results from this study were very important in shedding light on the external validity of this program, meaning that the impact estimated in the evaluation sample can be generalized to the population of all eligible units, which is a very important concept to consider when evaluating social programs that are expected to be either expanded or scaled up. The results were also very informative to the policy, given that the evaluation was able to address many aspects of fighting poverty. Results showed that the graduation approach caused a long-lasting increase in household consumption of around 5.4% compared to the comparison group as well as increases in food security, assets, and savings. Also, providing productive assets as well as livelihood training led to an increase in the well-being of the participants, particularly through self-employment and entrepreneurship skills. Improvements were also witnessed in the participants’ psychosocial well-being (Fahey, 2015 and Banerjee et al., 2015).

Another large-scale evaluation took place in Bangladesh, covering over 21,000 households in 1,309 villages. What distinguishes this particular evaluation is that participant households were surveyed four times over seven years to study the long-term impacts of the graduation approach. Results showed that household earnings increased by around 37% over the seven years

with significant increases in consumption and savings (Bandiera et al., 2017 cited in “Graduating the Poor: CGAP Completion Report”). The Sawiris Foundation for Social Development (SFSD) is currently piloting the graduation approach in Upper Egypt in Assiut and Sohag villages and evaluating the program by the Abdul Latif Jameel Poverty Action Lab (J-PAL)’s affiliated professors. SFSD aims to build new projects based on scientific evidence with the hope of scaling it up on a national level in collaboration with the government and other philanthropic and international organizations.

Even though RCTs provide a rigorous and accurate way of measuring the true impact of an intervention, in some cases they may not be feasible or may not be the best methodology to use. RCTs are not suitable for evaluating macroeconomic policies such as the adoption of a fixed exchange rate. Randomized evaluations need many units to be randomized in order to give us confidence in the impacts we find. So in the case of macroeconomic policies, especially if they are adopted on a national level, it is almost impossible to conduct an RCT because nation-states cannot be randomized to different policies. In addition to this, randomized evaluations cannot accurately capture what we call the General Equilibrium Effects. The general equilibrium effects are a form of spill-overs. For example, an intervention that introduces a shock to the price of a certain commodity like gas is likely to generate effects that go well beyond what researchers are able to capture through an RCT. That is because prices are usually a result of demand and supply factors that are interlinked, in addition to the impact this might cause to the prices in the markets of complements or substitutes (Glennerster and Takavarasha, 2013). This is especially relevant in the case where an intervention is being evaluated for future scale-up, where there is the concern that effects found in a pilot RCT may not hold when applied nationwide (Banerjee and Duflo, 2008, P.21).

PROGRAM DESIGN: GOOD INTENTIONS VERSUS EVIDENCE-BASED

In his book, *More than Good Intentions*, Dean Karlan mentioned that it is essential to find individual programs that work and support them and find programs that do not work and stop doing them. This is a fundamental concept of how we understand program design in the world of development. We can see many programs developed and designed only based on good intentions. Even though they may work in some contexts, we need to test and measure the impact of these programs. We should keep in mind that ignoring conducting an impact evaluation for these programs can lead to poor program design and waste resources in a world suffering from limited resources and high needs.

In case we have an effective program and we do not measure its impact, we can miss an opportunity to learn about the magnitude of its effectiveness and how it can alleviate poverty. As part of this, we will miss the chance to understand better what exactly works in this program and why. This is in addition to learning what positively changed in the target group and determining the cost-effectiveness of this program. In the opposite case, we can keep working on a program and replicating it when it is not effective. This is the case for programs that are only based on good intentions without scientific or rigorous evidence on their effectiveness.

In 2006, for his contribution to reducing world poverty, Mohammad Yunus and the Grameen Bank were awarded the Nobel Prize for Peace. As a result, spending on microfinance programs increased rapidly. According to the Micro-Finance Barometer 2019, 139.9 million borrowers benefited from MFI programs in 2018 compared to just 98 million in 2009. Of those 139.9 million

borrowers, 80% are women, and 65% are rural borrowers; this is a very high number for one year. MFIs reported another year of growth with a total credit portfolio of \$124.1 billion (an additional 8.5% compared to 2017). But do we have evidence on the impact of microfinance? In a report published in 2010 called *What Is the Impact of Microfinance on Poor People?*, the authors conclude that microfinance made some people poorer, in particular, microfinance clients. This might have resulted from the high-interest rates they need to pay for their loans and that their business is not able to produce enough money to cover their expenses and the interest rate.

Another study conducted by Esther Duflo could not find an impact on women's empowerment or human development outcomes after 18-36 months of getting the MFI loan. Moreover, approximately 70 percent of the eligible and qualified households end up not having an MFI loan but choose to borrow from other sources, if they borrow (and most do) (Duflo et al., 2013). A systematic review of six different impact evaluations on the effect of microfinance in six different countries concludes that microfinance has a modestly favorable impact, but not transformative effects. (Banerjee, Karlan & Zinman, 2015). Although the evidence did not show a significant positive impact on all poor people who are getting these loans, a lot of funding is still allocated to micro-finance. This is an example of programs that are based on good intentions.

On the other hand, there are many programs that are evidence-based and have a significant impact, one of which is the Teaching at the Right Level (TaRL) model, an approach developed by the Indian NGO Pratham aiming to build basic math and reading skills for children in primary schools. The model starts by determining the children's learning levels at the instructional level using a simple assessment tool.

This stage helps in grouping children based on their learning levels instead of age or grade

by using a variety of interactive teaching and learning activities, and by concentrating on basic reading, writing, and mathematical skills rather than depending on the curriculum alone, as well as measuring children's progress in the typical exams. This is a fundamental change in the concept of children's classification, which used to be based only on age. Teaching at the Right Level started as a collaboration between Pratham NGO and the Indian government. Pratham put on top of their priorities: providing strong academic and technical support for teachers as well as providing continuous monitoring and assistance on-site. This is in addition to the valuable feedback at various levels of the school system. This holistic approach positively contributes to the program's success. J-PAL-affiliated researchers conducted several RCTs for over a decade, which proved that TaRL significantly improves learning outcomes. The TaRL program showed the most prominent learning gains among rigorously tested educational programs. The innovative concept and the assessment process helped Pratham develop and adapt TaRL in several contexts in collaboration with J-PAL. Up until now, TaRL has benefited millions of children in India and in African countries (Teaching at the Right Level to improve learning, 2020). SFSD, a philanthropic organization driven by social impact, is replicating this program in Egypt in collaboration with Educate-me foundation. This model can be the solution for improving reading, writing, and mathematical skills for children in primary schools in Egypt and improving their learning outcomes.

This is not the only program that produced rigorous and robust evidence. In Egypt, researchers tested the impact of providing exporting opportunities to handcraft producers in Fowa, a peri-urban town with a population of 65,000. Fowa is famous for producing handmade rugs. "These firms typically consisted of a single owner operating out of rented space or his house and typically employing one to four people using handlooms. At the time of the study, the average per capita income in the area was US\$ 3,600 (PPP-adjusted), less than half of the national average for Egypt of US\$ 6,500. Researchers focused their assessment on small and medium-sized enterprises with fewer than five employees, the majority of whom had

not exported knowingly in the past" (Atkin, Khandelwal & Osman, 2014 as cited in Varma, 2019).

Researchers found that operating profits increased by between 15 and 26 percent among SMEs offered export opportunities compared to other companies in the control group. The higher quality rugs that international retailers requested were costlier and time-consuming to produce, but the resulting higher purchase price was more than sufficient to offset more expensive production. This increase in profits, therefore, came in tandem with quality increases, as well as declines in output per hour (Atkin, Khandelwal & Osman, 2014 as cited in Varma, 2019).

In another ongoing study currently evaluated by J-PAL MENA, researchers found a significant positive impact for providing vocational training to women in Upper Egypt after the first round of data collection. However, they did not find a significant impact on providing extra counseling sessions for beneficiaries. The results of this study are similar to another study conducted by J-PAL MENA researchers in collaboration with SFSD, implemented in four governorates in upper Egypt. In this study, researchers found a positive impact on providing vocational training for women. However, adding on-the-job training and job counseling did not matter much, and the impact on men was not promising. Many other studies are ongoing now in Egypt and the MENA region that would improve our ability to find solutions for poverty challenges based on scientific evidence, not only on good intentions.



HOW IMPACT EVALUATIONS CAN HELP ORGANIZATIONS BETTER DEVELOP THEIR STRATEGIES

Impact evaluation is a great tool for informing management staff in organizations and policymakers about best practices, as well as learning more about what worked and what did not work in specific programs. As part of the impact evaluation process, management will get access to a wealth of data and information that could help them better utilize their resources, understand expected outcomes, as well as design better-informed policies. There are concrete examples from J-PAL's work within the MENA region. The first example is the Sawiris Foundation for Social Development (SFSD). It was founded with an endowment from the Sawiris family with the objective of supporting development programs to improve lives across Egypt. SFSD collaborated with J-PAL affiliated professors Bruno Crepon and Adam Osman as well as Egypt-based professors Mohammed El Komi and Reham Rizk to evaluate the impact of their development programs in 2015. SFSD's management team started by sharing their thoughts about the challenging questions they had for their programs. This process ended with agreeing on two main impact evaluations. The first was to measure the impact of providing different types of training to equip job seekers with the needed skills to join the labor market. SFSD's team was wondering if they should provide vocational training only, vocational training with on the job training, or vocational training with on the job training and job counseling. To test these different interventions, the research team developed an RCT in collaboration with four NGOs in Beni Suief, Minya, Assuit, and Luxor. On a random basis, the NGOs provided these different types of technical support. With a sample of more than 2000 job seekers, researchers started to test these interventions. Job seekers started by applying to these training opportunities, then got assigned to one out of the three treatment options or the control group. The research team, in collaboration with SFSD, were able to conduct three rounds of data collection, two of which were in person and one over a phone survey, in addition to conducting in-depth interviews and focus groups. As previously

mentioned, researchers found an exciting impact of providing vocational training to women. Currently, SFSD is working with the research on collecting long-term data (three to four years after implementation) to better understand the sustainability of the impact. During that time, the SFSD management team had several meetings with J-PAL professors and staff that enriched their ability to fully understand the process and get lots of data that helped them interpret the outcomes of their interventions.

At the same time, the SFSD team was also interested in understanding the impact of providing different types of capital assistance. They started with questions on how they can support entrepreneurs in starting their own business and whether it is better to provide loans, cash grants, or in-kind grants. Evidence shows that marginal returns to capital grants are high in many developing country contexts (McKenzie, 2017). In contrast, the microcredit literature shows that the returns to microloans are often modest (Banerjee, Duflo & Hornbeck, 2018). SFSD and three Egyptian NGOs along with J-PAL affiliates compared the impact of cash grants, in-kind grants, and loans for low-income entrepreneurs in some of the poorest parts of Upper Egypt. This is the first study to directly compare the three approaches. Implementation of the project has ended, and now the data collection and analysis phases are completed. The results are already promising. Overall, the research found that all three treatment groups increase business ownership, overall employment, business profits, and total income. In addition, researchers found substantial differences by gender, with much stronger positive impacts on women, and null effects on men, unlike most earlier studies. Next year, we will be able to see the long-term effects of these interventions as follow up data collection has begun.

Another example is J-PAL MENA's work with the Micro Small and Medium Development Agency (MSMEDA). In 2015, the World Bank and MSMEDA reached out to J-PAL affiliated

professors to start three impact evaluations on improving the effectiveness of job fairs and training programs for both wage employment and self-employment opportunities. Based on this, J-PAL affiliated professors along with Dr. Mona Said from the American University in Cairo, and Dr. Ahmed el Sayed from the Institute of Labor Economics in Germany agreed to develop three RCTs. The job fairs RCT focused on tackling the information barrier, while the two other RCTs that are employment-focused compared training only to training, coaching and counseling for both self and wage employment opportunities. This experience helped MSMEDA realize the importance of participating in impact evaluations. It also helped them see the precise impact of these interventions on different outcomes of their interest such as income, interest to migrate, and of course employment.

It is important to consider that the philanthropic sector has two key features that should be put to the right use. The first is agility and flexibility, which, unlike the public sector, gives it the space to adopt a learning mentality and evidence-based culture. The second is a much closer connection to people on the ground, which gives it the advantage, and the responsibility, of constantly striving to improve the quality of their programs and interventions, help them better understand the workings of their projects and avoid the unintended consequences of well-intentioned programs.

Reference List

- Atkin, D., Khandelwal, A., Osman, A., & Varma, P. (2019). Egyptian rug firms weave their way into foreign markets. Retrieved from <https://www.povertyactionlab.org/blog/9-9-19/egyptian-rug-firms-weave-their-way-foreign-markets>
- Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., & Sulaiman, M. (2017). Labor Markets and Poverty in Village Economies. *The Quarterly Journal of Economics*, 132(2), 811–870. doi: 10.1093/qje/qjx003
- Banerjee, A., & Duflo, E. (2008). *The Experimental Approach to Development Economics*. doi: 10.3386/w14467
- Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., ... Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236), 1260799–1260799. doi: 10.1126/science.1260799
- Banerjee, A., Duflo, E., Hornbeck, R., (2018). How Much Do Existing Borrowers Value Microfinance? Evidence from an Experiment on Bundling Microcredit and Insurance. *Economica*. 85. 10.1111/ecca.12271.
- Banerjee, A., Duflo, E., & Hornbeck, R. (2018). How Much do Existing Borrowers Value Microfinance? Evidence from an Experiment on Bundling Microcredit and Insurance. *Economica*, 85(340), 671–700. doi: 10.1111/ecca.12271
- Bourguignon François, Pleskovic, B., & Duflo, E. (2004). Scaling up and Evaluation. In *Accelerating development: Annual World Bank Conference on Development Economics, 2004*. Washington, DC: World Bank.
- Fahey, A. (2015, September). Retrieved from <https://www.povertyactionlab.org/sites/default/files/publications/TUP.pdf>
- Gertler, P., Wilde, S. M., Premand, P., Rawlings, L., & Vermeersch, C. (2011). *Impact evaluation in practice*. Washington, DC: World Bank Group.
- Glennerster, R., & Takavarasha, K. (2013). *Running Randomized Evaluations: a practical guide*. Princeton, NJ: Princeton University Press.
- Hartnell, C. (2018). *Philanthropy in the Arab Region. Introduction to Evaluations*. Retrieved January 2020, from <https://www.povertyactionlab.org/research-resources/introduction-evaluations>.
- Karlan, D., & Appel, J. (2012). *More than good intentions: improving the ways the worlds poor borrow, save, farm, learn, and stay healthy*. New York: Plume.
- Mckenzie, D. (2017). Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition. *American Economic Review*. doi.org/10.1257/aer.20151404
- Osman, Adam, et al. 2020, *Is It Who You Are or What You Get? Comparing the Impacts of Loans and Grants for Microenterprise Development*, www.adam-osman.com/wp-content/uploads/2020/05/Loans-vs-Grants.pdf.
- Stewart, R. J., & Van, R. C., & Dickson, K., Majoro, M., & De Wet, T. (2010). *What is the Impact of Microfinance on Poor People? A Systematic Review of Evidence from Sub-Saharan Africa (Protocol)*. Teaching at the Right Level to improve learning. (n.d.). Retrieved from <https://www.povertyactionlab.org/case-study/teaching-right-level-improve-learning>
- World Bank. (n.d.). *Graduating the Poor: Cgap completion report*, from <http://documents.worldbank.org/curated/en/988811549967967067/pdf/Graduating-the-Poor.pdf>
World Microfinance Barometer 2019 report, from http://www.convergences.org/wp-content/uploads/2019/09/Microfinance-Barometer-2019_web-1.pdf

