Digital Development Dialogue (3D) Seminar Blog

"How and Why to Involve Policy Makers More Structurally in Impact Evaluations?"

For the implementation of policy interventions, both institutional partners and evaluators can benefit from setting clear expectations about the outcomes of interest, receiving feedback on the research design, and updating the beliefs of the key stakeholders in terms of the reach and effect of the programs after the assessment. **David McKenzie**, Lead Economist in the Development Research Group at the World Bank, and **Gabriela Macoveiu**, director for Communication, Innovation and External Cooperation at the Regional Development Agency (RDA), discussed these issues and shared their experiences from real world applications. This event was co-organized with the World Bank Competitiveness Policy Evaluation Lab (ComPEL). **Alexandra Avdeenko**, Evaluation Specialist at the World Bank, moderated the seminar.

The initial presentation introduced the Bayesian Impact Evaluation with Informative Priors developed by McKenzie, Iacovone and Meager, as well as its application to a Colombian export improvement program. According to the authors, the objective of the tool is to identify how policy makers set their goals for the programs and how this can be systematically implemented in impact evaluations. Besides the well-known challenges in terms of financial and time constraints and of small sample size of randomized controlled trials (RCTs), communicating the results from the evaluation and their relevance for decision making to the key stakeholders pose an additional obstacle to the research team. McKenzie illustrated these issues through several scenarios, for example, finding positive but not statistically significant results, getting sizable results for secondary outcomes, or finding effects in the opposite direction than the expected. How can the goals and beliefs from both parts be reconciled? McKenzie proposes the following four steps:

1. Agree on what does success look like for the project partners and researchers: what is the main outcome, how to measure it, and in what timeframe (the more specific, the better)?
2. Elicit prices from policy makers, researchers and treatment units about the program and their expectations towards it. The data should be collected from as many relevant stakeholders as possible to average out "the noise" and to find out what the program is really meant to achieve to be continued or extended.
3. Conduct the experiment and follow up on the sample and outcomes of interest in the year after the intervention.
4. Compare with baseline results, including to the elicited expectations (i.e., the "priors"), and estimate the posterior probabilities to get a sense of how much policy makers updated their beliefs about the effectiveness of the program, given the results of the intervention.

Digital Development Dialogue (3D) Seminar Blog

"How and Why to Involve Policy Makers More Structurally in Impact Evaluations?"

These steps are exemplified with an application for small and medium enterprises (SMEs) in Colombia's exporting sector, where there is a desire to diversify and extend the export base by improving the management and technical capacities. The ex-ante calculations suggested that it would probably be possible to get results for the business practices and effects on binary outcomes like whether firms export, but that the statistical power might be too low to detect effects on other outcomes such as exports' value.

To address this issue, the researchers worked with focus groups to elicit their expectations regarding the program and calculate all possible combinations to obtain a distribution of priors. In a second step, policy makers allocated how confident they were that the impact was going to lie on a certain range within this distribution, and an updated (posterior) distribution was obtained, which could imply even updating expectations fully.

The main benefits of the methodology are that it allows to conduct an "expectations vs. reality" comparison. Even if priors fit well the actual data, the approach allows to narrow down the confidence intervals to get more precise estimates. This way, it can be used to make decisions regarding, for instance, how much the average treatment effect (ATE) needs to increase for the program to be continued.

The presentation concluded by reaffirming the usefulness of the tool, particularly in settings where the sample size is small and it is costly to increase it or to repeat the experiment. Further, adopting the tool is promising when there is treatment heterogeneity, and when it is necessary to ensure that the project measured the key outcomes for which expectations might need to be updated after the intervention.

The second input was provided by Gabriela Macoveiu, who discussed how useful these ideas are for her current work, and how she and her agency usually form their prior beliefs. She shared her experience with a project that is still under development in North-East Armenia that consists of an RCT for the digital transformation of small SMEs. Digital transformation is a priority in a large-scale investment program of the European Bank for Reconstruction and Development (EBRD) for 2021-2027, aiming to develop a more digitalized region.

This intervention was included in the program because there was some evidence showing a very limited performance in terms of digitalization – Armenia is in the last position in the European Index for Digital, Economic and Social Development – although the investments are above the regional average. There are

Digital Development Dialogue (3D) Seminar Blog

"How and Why to Involve Policy Makers More Structurally in Impact Evaluations?"

also challenges in the sense that companies are not using digital technologies and therefore have a limited capacity to operate online, often losing business opportunities.

The vision for 2030 is that about 75% of SMEs use cloud computing with artificial intelligence systems and achieve a basic level of digital maturity. To allocate funding, there were two calls for proposals, with the first one offering EUR 25.50mn for companies to invest in equipment and services to develop websites, digital promotion, or ecommerce platforms.

The indicators to measure these outcomes in a study among 400 SMEs include how many companies received the grant, or how much they are expected to invest. The results are attached to a standard set of indicators to quantify the progress from the initial moment of the project until the end: in the short run, how many new digital technologies were implemented by the companies, and in the long term, whether those companies were able to develop their business with those investments. Additionally, the evaluation relies on two critical assumptions: that SMEs show an actual interest in the project, and that the current offer of information and communications technology services is large enough.

Lastly, Macoveiu points out that expectations are set based on statistical data evolution, official reports, and data collected through focus groups. At the same time, she stresses that even though the demand from the authority is to answer what and how the program was implemented, it is very important, as McKenzie mentioned, to take these types of evaluations to look in the broader context of the data evolution. This would help to detect what kind of indicators are sensitive for measuring the results and what are the indicators that one should consider tracking indirect channels. Another challenge is to explain the budget constraints of public funded projects to the research team for them to fit the RCT, so that the authorities embrace the research methodology.

During the **Q&A segment**, the audience focused on getting deeper insights regarding the tool and the experience of the research team during their interactions with the project managers. McKenzie started by pointing out that specifying the setting from the beginning does not mean that we cannot learn in the process and start seeing some outcomes that we were not expecting, and that even though the researchers can also look at those variables, it is imperative not to miss the initial goal. Regarding the estimation of the distributions, the averaging process captures the uncertainty at the institutional level also in decision making teams, so that collecting these individual perspectives (optimistic vs. pessimistic stakeholders) allows to estimate more realistic distributions.

Digital Development Dialogue (3D) Seminar Blog

"How and Why to Involve Policy Makers More Structurally in Impact Evaluations?"

Regarding whom to target to get the priors, he shared that getting people who know the program well, as well as decision makers, are key for obtaining relevant point estimates. For example, in Colombia it was a data driven decision-making process with high support from the implementing partner (the ministry).

Concerning heterogeneous treatment effects, he mentioned that one of the challenges in Colombia was to make sure that everyone was talking about the same treatment parameter (ATE or ITT), and that for heterogenous treatment effects, you also need to specify the measures of the heterogeneity and how much of it is expected, for instance, by gender.

Macoveiu was asked whether she would tend to express desires or actual expectations if she was approached with such methods. She shared that people inside the organization have not only the elicitations but also the data, so that they are looking at the sources, in order to formulate their expectations (for instance, the funding). Moreover, she expressed that it is not hard to find the right people, but instead to choose the kind of indicator that is sensitive enough to quantify the elicited expectations.

She also asked McKenzie whether he considers that this tool is demanded by policy makers or mainly by the academia. He expressed that hopefully it has benefits for both sides and that it is a good approach to make sure that both parts are on the same page from the beginning in terms of what matters, what they think the program is going to do, and how to measure it. But that even though this kind of discussion can help guide the research design and implementation, its applicability is dependent on the context and requirements.

The discussion also turned back to the Bayesian evaluation tool and additional insights regarding the plausibility of beliefs' updating. From the Colombian study, they learned that policy makers tend to be more optimistic about their programs than researchers, but that sometimes, implementing partners and funders are more realistic. Moreover, after seeing the results, there was less optimism among stakeholders, and they were more careful before taking further steps.

To conclude, the presentations and discussion suggest that this tool has the potential for being implemented in other contexts, mainly due to its innovativeness in methodological aspects but also because it addresses one of the main demands from policy makers: hand-in-hand work between academia and practitioners. Moreover, providing a range for updating the expectations beyond contrasting them

against the initial prospects is also promising for increasing the monitoring and accountability of the programs based on objective margins. This is particularly important for decisions concerning budget allocation or scalability of an intervention.

Lastly, it is also worth to analyze whether this tool could also be used to assess the feasibility of rolling out an intervention in the first place, in case elicited prices at the initial stage are too high or stakeholders seem over-pessimistic about the outcomes. A replication study or re-assessment of a project with this new toolkit might be a promising avenue to analyze how it compares to traditional methodologies and whether it can make a difference for policy making.

Author: Lucero Carballo Madrigal

Research Assistant at the Chair of Development Economics (Prof. Fuchs) and Master Student in Development Economics at the University of Göttingen (lucero.carballomadrigal@uni-goettingen.de)