IMPACT ASSESSMENT IN COMPLEX EVALUATIONS

Photo: Cornelis Gollhardt/laif

With the aspect of impact gaining importance in development co-operation, there is a growing demand for evaluations. This article's authors maintain that a more systematic integration of evaluation methods is needed to reflect the complexity of the real world and to make results more meaningful.

By Martin Noltze, Gerald Leppert & Sven Harten

In times of scarce resources and mounting public interest in questions around global development, there is a growing demand for impact evaluations as a means of measuring whether public resources are spent effectively and efficiently. Policy-makers, development partners and implementing agencies want and need to show that they make decisions based on evidence and that they learn from what works and what does not. Stakeholders, including funders, beneficiaries and the general public, increasingly ask for information whether spending was meaningful and effective. Besides this issue of aid effectiveness, accountability and transparency are central to development co-operation. Comprehensive global development agendas, such as the 2030 Agenda for Sustainable Development, emphasise the role of impact evaluations in assessing the achievement of highly aggregated development targets. The past decade has also seen the advent of new actors in international co-operation, such as philanthropic organisations, private sector companies or new forms of social investment funds. What these actors have in common is a firm belief in measuring success (i.e. return on investment) through quantifiable indica-

tors. These developments have multiplied the demand for rigorous impact evaluation, meta-analysis and evidence mapping.

A CALL FOR A COMPREHENSIVE APPROACH

Yet, despite a long history of interdisciplinary work and many interesting recent developments, the evaluation profession has not sufficiently lived up to the challenge of presenting a comprehensive approach. We therefore call for a more systematic integration of methods with a view to bringing together quantitative measures of results achieved and a thorough understanding of the underlying causal mechanisms. In other words, evaluations must get out of the trenches where either answering the question of "how much" was achieved or "why" and "how" is the predominant focus. Indeed, by building on encouraging theoretical developments and making better use of new types of data, it will be possible to answer both questions in a rigorous fashion. In order to make this argument, we will first recapitulate the main contemporary challenges of

evaluating development co-operation and the response of the profession to them. Next we will point briefly to the opportunities presented by "big data" and then make the case for an integrated, comprehensive approach.

With impact evaluations moving more into the spotlight of development co-operation, a range of new challenges are emerging. Firstly, evaluations can no longer hide in a niche of either measuring impacts of individual projects very rigorously or assessing broad programme implementation at a higher level. Indeed, evaluations are expected to go both deep and wide. Secondly, development programmes are becoming more and more complex. Typical interventions include a variety of instruments to reach multi-dimensional development targets implemented by multiple stakeholders. On the one hand, global agendas such as the Paris Agreement and the 2030 Agenda for Sustainable Development defined a large number of detailed impact indicators at a highly aggregated level (see also the article on pages 26–28). On the other hand, both agendas also raise the demand for disaggregated impact statements since the

Testing of a programmed questionnaire for a quantitative household survey in the Philippines.
Photo: Malte Lech/DEval

Validation of information from satellite

"leave-no-one-behind" principle advocates measuring effects at an individual or household level. Thirdly, impact evaluations are supposed to deliver results in a timely manner and thus enhance policy relevance. Whilst all impact evaluations aim to (quantitatively) answer the question "to what extent" results were achieved, the focus has broadened over the past few years to also include questions on "how", "why" and "under which circumstances" an intervention caused an effect. Evidence-based policy making requires both, knowing the impact and understanding the underlying causal mechanisms.

REFLECTING THE COMPLEXITY OF THE REAL WORLD

Hence evaluations are facing the challenge of higher expectations regarding the number and types of questions that have to be answered while also backing the answers with quantifiable evidence. In response to the difficulty to meet all these requirements, the evaluation profession deepened the trenches between ostensibly opposing methodological (and epistemological) camps. Part of the profession concentrated on a narrow set of well measurable (i.e. "to what extent") questions. Others chose a wider set of (i.e. "why"/"how") evaluation questions in an attempt to better reflect the complexity of the real world. In this view, interventions are part of a causal package and only work in combination with other factors such as the cultural background of the ben-

eficiaries, stakeholder behaviour, institutional settings, environmental context, etc. Building on this, over the last decade, evaluators have constantly been working on broadening the range of impact evaluation methods. Besides randomised controlled trials (RCTs) – the former "gold standard" of the development evaluation community (see also the article on pages 9–11) – other econometric, theory-based, case-based and participatory approaches have gained ground and experienced enormous improvements in the field of systematic testing procedures, a prerequisite for rigorous causal inference.

Beyond the increased variety of rigorous methods, the field of impact evaluation benefits from new forms of data collection and analysis which emerged in the digital revolution era. Both monitoring and evaluation build increasingly on information gathered by mobile technologies, social media and remote-sensing data (see also the article on pages 24–25). On the side of data analysis, so called "machine learning" is especially innovative. Setting out from computer algorithms, machine learning predicts trends based upon the processing of large data sets. On their own, new and larger data sets are not a panacea. Often, they only reflect major trends and probabilities without useful contributions to the questions of attribution and causality. However, great potential lies in the integration of big data and machine learning in complex evaluations and in the triangulation of such data sets with case studies and cross-case analysis.

MORE THAN COMBINING EXISTING METHODS

In light of the broad range of impact evaluation methods, their strengths and weaknesses, as well as new forms of data, there is a huge untapped potential of integrating multiple methods in complex evaluations. The basic idea of mixed- or multi-method approaches is to overcome weaknesses of one method with the strength of another. However, thus far most of the literature has concentrated on how to combine (or "mix") quantitative and qualitative methods. In our view, this is too narrow and has resulted in RCTs being complemented by a few focus group discussions or the implementation of a survey as part of a qualitative case study. While this is certainly meaningful for the individual studies, the overarching aim is to achieve more comprehensive evaluation designs that improve the measurement of causal inference and answer complex and multi-dimensional evaluation questions. In this sense, the authors regard comprehensive multi-method approaches as going beyond combining qualitative and quantitative data. Rather, they combine theorising about how activities lead to outcome and impacts (as in theories of change) with different types of causal inference.

For instance, RCTs or quasi-experiments rely on the counterfactual logic, comparing the outcome of one or more treatment groups to the outcome of controls, in other words comparing the beneficiaries of an intervention with

data by aerial photos from a drone in the Philippines.
Photo: Courtesy of Julian Barth/Kigali-Films

Focus group discussions as part of a realist mixed-method design in Ghana.
Photo: Marcus Kaplan/DEval

those not having received the intervention. Statistical models such as longitudinal studies or most econometric techniques draw causal inference from the statistical relationship between cause and effect or between variables. Theory-based approaches include causal process designs that build on the identification of causal processes or causal pathways (e.g. process tracing or contribution analysis) and causal mechanism designs that consider supporting factors and causal mechanisms (such as the realist evaluation paradigm or congruence analysis). Case-based approaches include grounded theory or ethnographic approaches and rely mainly on within-case analysis. Cross-case analysis of several case studies can be managed by configurational approaches such as qualitative comparative analysis (QCA) with the analytic generalisation based on theory. Thereby, integrated approaches are able to combine exploratory and explanatory approaches in sequence or in parallel, nested or balanced with different conceptual frameworks to causal inference. Thanks to their integrative character mixed- or multi-method approaches are also open for new data types and analytical approaches such as geographic data, big data or text mining.

This article argues that no methodological approach is best or even sufficient on its own. Complex development challenges require complex interventions and consequently more complex evaluation designs. Multi-dimensional questions and the need to not only measure the impacts, but also to understand the under-

lying causal mechanisms, require an extension of the toolbox of researchers and evaluators. Facing these challenges for impact evaluation in the field of development co-operation today, only the systematic integration of different forms of causal inference can sufficiently address this demand. Certainly, (quasi-) experimental designs that allow for high levels of rigor and attribution are one important piece in the evaluators' toolbox in complex impact evaluations. However, they are best understood as one of the elements in a complex evaluation design.

The design of advanced mixed-method approaches explicitly follows the particular epistemological interest of the evaluation questions. Through the smart and systematic integration of methods, they are superior to single-method approaches, as they can better address the complexity of interventions, making any discussion on "gold standards" obsolete.

Thus, future development of impact evaluation designs should focus on enhancing the capacities of multi/mixed-method approaches beyond simply sequential or triangulation strategies. At the same time, evaluators should not hesitate to improve systematic testing procedures of single methods to improve the robustness of the overall design. New forms of data collection and analysis raise the bar for possible mixed-method approaches and thus significantly contribute to the further development of future impact evaluation designs.

Martin Noltze is a senior evaluator at the Competence Centre for Evaluation Methodology of the German Institute for Development Evaluation (DEval). He holds a PhD in agricultural economics from Georg August University of Goettingen, Germany. Together with Gerald Leppert and Sven Harten, he is co-ordinating DEval's methodology project on Integrating Multiple Methods in Complex Evaluations.
Contact: martin.noltze@deval.org
Gerald Leppert is a senior evaluator at the Competence Centre for Evaluation Methodology of DEval, holding a PhD in economics and social sciences from the University of Cologne, Germany.
Contact: gerald.leppert@deval.org
Sven Harten heads the Competence Centre for Evaluation Methodology and is Deputy Director of DEval. He holds a PhD in political science from the London School of Economics.

From 2018–2021, the German Institute for Development Evaluation (DEval) is implementing a four-year research programme on the integration of multiple methods in complex evaluations. The methodology project will be accompanied by empirical testing of new forms of method integration in several DEval evaluations.