

RIGOROUS IMPACT EVALUATION IN GERMAN DEVELOPMENT COOPERATION

Executive summary

Experimental and quasi-experimental impact evaluations can make an important contribution to more effective development cooperation. In Germany, however, such rigorous impact evaluations (RIE) have so far tended to be conducted infrequently and unsystematically. Moreover, there is no explicit strategy for using the findings they deliver. At the same time growing trends towards the use of such evaluations and related analyses are evident. We see six strategic tasks for systematically using the potential offered by rigorous impact evaluation in German development cooperation: (1) Policy-makers should anchor rigorous impact evaluation and impact-oriented accompanying research across the German development cooperation system. (2) Corresponding financial resources should be made available. (3) The needed capacities and incentives should be created within development cooperation organisations. (4) Existing evidence should be aggregated. (5) The research community should also be involved in implementation. (6) Finally, capacities for conducting rigorous impact evaluations self-reliantly should also be developed in partner countries.

Impact evaluation in German development cooperation

The demand for sound impact analyses of development interventions is increasing. This is partly because topics related to development policy have gained greater significance within society. It is also due to the fact that in times of migration, violent conflicts and climate change, questions concerning demonstrable results of humanitarian aid and development cooperation have become even more important.

In this context the evaluation of development cooperation serves both to provide accountability to parliament and society, and to support learning processes among both policy-makers and practitioners in German development cooperation. Evaluations thus contribute to the evidence-based improvement of strategies, programmes and projects – and in doing so to the legitimacy of development policy.

Such evidence-based policy-making in development cooperation places high methodological demands on evaluation and other forms of applied research. Analyses of portfolios, structures and processes have therefore become more sophisticated. This also applies to impact evaluation at the level of individual projects and programmes, where rigorous (i.e. experimental and quasi-experimental) methods have become more important. The importance of rigorous impact evaluation in poverty alleviation was recently demonstrated by the award of the Nobel Prize in Economics in 2019 to Esther Duflo, Abhijit Banerjee and Michael Kremer, who made a key contribution towards establishing these methods in development research.

Rigorous impact evaluation aims to causally attribute possible changes that occur within the target group (e.g. individuals, households, schools, economic actors) to a development intervention. The core of the method involves making a comparison between an intervention group that has received a certain intervention, and a control or reference group that is as similar to it as possible. This enables the evaluator to draw empirically well-founded conclusions concerning how the target group of a development intervention would have developed, both with and without implementation of the intervention (see Box 1).

Box 1:**In brief: Experimental and quasi-experimental methods**

Similar to experimental methods in medicine, health economics or educational sciences, field experiments in development are based on randomised controlled trials (RCTs) using randomly assigned intervention and control groups. A development intervention is conducted on a group of actors randomly selected from the population (the intervention group). The group not included in the random selection then serves as the control group. This procedure ensures that the two groups can be expected to be the same in terms of both their observable and non-observable characteristics, so that the difference in an outcome variable between the two groups after implementation of the intervention represents an undistorted estimate of the net effect. Quasi-experimental methods can be used with similar degrees of reliability by applying statistical methods to form comparison groups and systematically eliminating unwanted variables.

fundamental or sector-specific limitations on the applicability of such methods, insurmountable ethical problems or approaches that lack sufficient theoretical foundations have proved either unfounded, solvable or only relevant to restricted areas. Rigorous impact evaluations can be used in different sectors on a theoretically and ethically sound basis.

Having said that, rigorous impact evaluations are not suitable for all impact-related questions, nor would they be the preferred method for testing any kind of intervention. Good applied research is characterised by the fact that the method is derived from a practically relevant question or concern, not vice versa. Furthermore, useful evaluations that deliver robust findings usually comprise a combination of methods. This often means that quantitative experimental methods are combined with qualitative methods. Rigorous impact evaluations are particularly expedient when questions concerning the impact of an intervention are the focus of interest, and when the tested development intervention is designed to reach a large number of beneficiaries at the same point in time.

How widespread are rigorous impact evaluations?

In the social sciences these methods are now part of the standard repertoire of causality-based field research. They have also become established as an important component of the evaluation portfolio both at the multilateral level and among some bilateral donor organisations. RIEs are conducted both to provide accountability and as a tool for impact-oriented accompanying research to obtain timely information for improving interventions. It is also possible to test several interventions against each other – and in combination – in multi-arm impact evaluations. Rigorous evaluations are also being used increasingly in several emerging economies such as India, Mexico and Brazil to study the impact of state-run social programmes.

There are now also methods which organise and aggregate the findings from different rigorous impact evaluations that look at similar questions (see Box 2). Overall, reservations concerning

In German development organisations, however, the proportion of the entire portfolio subjected to rigorous impact evaluation is relatively low. So far the KfW Development Bank has gained some experience with rigorous impact evaluations, the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) has applied (quasi-)experimental methods in approximately 40 interventions, and some non-governmental organisations have conducted a few rigorous impact evaluations.

Moreover, to date there has been no systematic use of the findings from rigorous impact evaluations in line with an overarching strategy for learning or accountability. In fact, in many cases individual actors drove the implementation of rigorous impact evaluations in order to generate insights that enabled them to improve particular projects they were working in. Accordingly, in most cases the insights from rigorous impact evaluations remain confined to the project and organisation in question. Finally, so far there are no incentive systems to systematically encourage the application of rigorous methods and use of the resulting findings.

Box 2: The four phases of rigorous evidence in development cooperation

Internationally, the spread of rigorous evidence in development cooperation can be broken down into four phases:

- 1) Since the 1990s, several donors have been introducing results-based management schemes, where outcome-oriented target indicators are formulated, monitored and used as steering devices.
- 2) Since the 2000s, more and more individual rigorous impact evaluations have been conducted. These are designed to reliably identify outcomes and impact of development cooperation interventions.
- 3) To enable cross-project learning, since the 2010s more and more systematic reviews have been conducted to aggregate the findings from individual rigorous impact evaluations. This aggregation increases the validity of findings. Furthermore, the existing evidence in specific thematic areas is being captured in so-called evidence gap maps.
- 4) In an incipient fourth phase the utility of existing findings is being improved. Evidence portals are being used to make existing knowledge available to the global community in a simple and expedient way.

See White 2019.

Strategic tasks for the future

At Germany's Federal Ministry for Economic Cooperation for Economic Cooperation and Development (BMZ), KfW and GIZ, as well as in some non-governmental organisations, trends towards a change in the way rigorous impact evaluation is used are currently evident. This is manifested for instance in the systematic analysis of existing lessons learned with these methods, and an increased interest in rigorous impact evaluation, systematic reviews and evidence gap maps in selected thematic areas (e.g. conflict prevention and transitional development assistance; training and employment).

To systematically and fully capture the potential of rigorous impact evaluation for learning processes in development policy-making and practical implementation, we propose six strategic tasks for the future.

- Firstly, policy-makers should anchor the concept and the usefulness of rigorous impact evaluation and accompanying impact-oriented research at the strategic level of German development cooperation. The forthcoming evaluation policy for German development cooperation should mark a first important step in this direction.
- Secondly, more financial resources will be required for the conduct and use of rigorous impact evaluations in the official and non-governmental development organisations, for

Diagram 1: Tasks for harnessing the potential of RIE



Source: authors' own graphic.

instance – but not exclusively – within the framework of a dedicated funding programme.

- Thirdly, it will be necessary to create the necessary technical and human capacities as well as incentive schemes within German development organisations and in partner countries to enable the implementation of rigorous impact evaluations, and ensure that existing evidence is used.
- Fourthly, when rigorous impact evaluations and accompanying research are conducted, closer cooperation should be sought with national and international universities and research institutions – including those in partner countries. Important tasks for evaluators and researchers are to combine different methodological approaches according to the research purpose and to align the latter with the need of practitioners.
- Fifthly, the lessons learned to date should be systematically analysed on a cross-organisational basis. This will include storing the findings from rigorous impact evaluations in a repository, and investing more in corresponding syntheses.
- Sixthly, when conducting rigorous impact evaluations and impact-oriented accompanying research it will also be important to support the capacity development of actors and

institutions in the partner countries. Ultimately, those actors should be enabled to analyse and evaluate their own programmes and/or those of donors self-reliantly.

Above and beyond the aforementioned basic tasks, a BMZ-funded research project at DEval will elaborate specific recommendations for more systematic and appropriate implementation and use of rigorous impact evaluation in German development cooperation. It will also foster networking and offer advice to German development cooperation organisations.

Box 3: RIE at DEval

At DEval, quasi-experimental methods are becoming an increasingly important component of the methodological portfolio. They are not applied in isolation. They are used in selected theory-based evaluations, where they form one methodological element that is combined with other approaches. Evaluation syntheses and evidence gap maps are also being used more often as part of the institute's portfolio. Moreover, DEval sees itself as a platform where development cooperation organisations, universities and research institutions can share lessons learned with rigorous impact evaluation, both from a practical and from a research/evaluative perspective.

References

Funk, E., L. Groß, J. Leininger and A. v. Schiller (2019), 'Erkenntnisse aus der wirkungsorientierten Begleitforschung: Potential und Grenzen der rigorosen Wirkungsanalyse von Governance-Programmen', Discussion Paper 13/2019, Deutsches Institut für Entwicklungspolitik, Bonn.

White, H. (2019), 'The twenty-first century experimenting society: the four waves of the evidence revolution', Palgrave Communications 5, no. 47, available online at <https://www.nature.com/articles/s41599-019-0253-6>, accessed on 30 September 2019.



Dr Martin Bruder
Head of Department



Prof Dr Jörg Faust
Director of DEval



Dr Marion Krämer
Team Leader

The German Institute for Development Evaluation (DEval) is mandated by the German Federal Ministry for Economic Cooperation and Development (BMZ) to independently analyse and assess German development interventions. Evaluation reports contribute to the transparency of development results and provide policy-makers with evidence and lessons learned, based on which they can shape and improve their development policies.