Corporate Unit Evaluation Cross-Section Analysis

KNOWING WHAT WORKS

Rigorous Impact Evaluation

Cross-section analysis – Main Report

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Impressum

As a federally owned enterprise, GIZ supports the German Government in achieving its objectives in the field of international cooperation for sustainable development.

The Corporate Unit Evaluation of GIZ reports directly to the Management Board. She is separate from and independent of GIZ´s operational business. This organisation strengthens the independence. The Corporate Unit Evaluation is mandated to generate evidence-based results and recommendations for decision-making, to provide a credible proof of effectiveness and to increase transparency on the results.

The evaluation was conducted, and the evaluation report prepared by internal evaluators from the Corporate Unit Evaluation.

Evaluators and authors of the report:
Michael Florian, GIZ Evaluation Unit, Senior Expert Corporate Strategic Evaluations
Franziska Krisch, GIZ Evaluation Unit, Senior Expert Corporate Strategic Evaluations
Tatjana Till, GIZ Evaluation Unit, Expert Corporate Strategic Evaluations
Sophie Hermanns, GIZ Evaluation Unit, Junior Expert Project Evaluations

Contributors:
Wolfgang Stuppert, Camino gGmbH
Carroll Patterson, SoCha llc

Acknowledgements:
Gerald Leppert, DEval
Martin Noltze, DEval

Conception, Coordination und Management:
Franziska Krisch, Michael Florian, Tatjana Till,
GIZ Evaluation Unit, Corporate Strategic Evaluations

Responsible:
Albert Engel, GIZ, Director of Coporate Unit, Evaluation Unit

Published by:
Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

Friedrich-Ebert-Allee 36 + 40
53113 Bonn, Deutschland
T +49 228 4460-0
F +49 228 4460 - 1766

E evaluierung@giz.de
I www.giz.de/evaluierung
www.youtube.com/user/GIZonlineTV
www.facebook.com/gizprofile
https://twitter.com/giz_gmbh

Design/Layout etc.:
DITHO Design GmbH, Köln

Printing and Distribution:
GIZ, Bonn

Printed on 100 % recycled paper, certified to FSC standards.

Bonn 2019

This document can be downloaded as a PDF file from the GIZ website at www.giz.de/evaluierung. For printed copies, please contact evaluierung@giz.de

Content

List of tables

List of figures

List of abbreviations

3ie	International Initiative for Impact Evaluation
AusAID	Australian Agency for International Development
BMZ	Federal Ministry for Economic Cooperation and Development
CPE	Central Project Evaluation
DeGEval	(German) Evaluation Society
DEval	Deutsche Evaluierungsinstitut der Entwicklungszusammenarbeit
DFAT	Australian Department of Foreign Affairs and Trade
DFID	Department for International Development
DIME	Development Impact Evaluation Initiative
FGD	Focus Group Discussion
GDI	German Development Institute
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit
GloBe	Sector and Global Programmes Department
KfW	Kreditanstalt für Wiederaufbau
KII	Key Informant Interview
QCA	Qualitative Comparative Analysis
OECD-DAC	Organisation for Economic Co-operation and Development - Development Assistance Committee
RBM	Results-based Monitoring
RCT	Randomised Controlled Trial
RD	Results Data
RIE	Rigorous Impact Evaluation
ToC	Theory of Change
USAID	United States Agency for International Development

1 Executive Summary

Rigorous Impact Evaluation (RIE) in this review refers to a class of evaluations that provide clear answers to questions of whether and to what extent an intervention caused a change in specific context (net-effect). In the international development context, the strategic use of RIE has significantly increased over the past decade.

Review object

RIE relies upon (quasi-)experimental designs to attribute impacts to various interventions. This attribution focus allows RIE to substantially supplement GIZ´s two globally formalised instruments/ processes for results measurement: i) Results-based Monitoring (RBM) and ii) Central Project Evaluations (CPE).

Review objectives

The strategic purpose of this review is (1) to determine the corporate value of RIE for GIZ and (2) to assess the enabling environment needed to achieve this value. The review centrally examines the suitability of RIE to support the key functions of evaluations.

Hypothesis: RIE contributes to strengthening the evidence base for operational learning, conceptual learning and accountability.

In addition to the central hypothesis, the review investigates the potential of RIE to add corporate value by increasing the efficiency and effectiveness of CPE, mobilising resources and strengthening partnerships.

Methodological approach

A multiple case study design was applied following an organisational case mapping. The following staged process was applied:

- **Stage 1:** 39 RIE cases were identified in a case mapping exercise.
- **Stage 2:** For 22 RIE cases, qualitative key informant interviews (KIIs) among key internal stakeholders, as well as with bilateral and multilateral peer organisations, were used to capture the utility and experience of RIE.
- **Stage 3:** For 19 RIE cases, Qualitative Comparative Analysis (QCA) was performed to assess under which conditions RIE facilitates operational learning, conceptual learning and accountability.

Findings

Use of RIE at GIZ

At least 39 RIEs were conducted between 2014 and 2018 at GIZ. The RIEs unfolded across different regions and sectors. The organisational embeddedness of RIE can be divided into three types of structures: (1) decentralised (2) network and (3) hierarchical.

Most users at GIZ viewed RIE as a value-adding and cost-effective measure. However, project managers commonly expressed their need for a specialised, global support structure at GIZ to facilitate the implementation of RIE that is tailored to the intended use.

Operational learning

Ten out of 19 RIE cases successfully supported decision making at the project level based on evaluation findings or process learning (operational learning). An early integration of RIE into the project design and monitoring system was pivotal to achieve operational learning. If RIE was not properly integrated into the management framework, central support was required to facilitate a purpose-sensitive design aligned with short-term orientation to facilitate operational learning. Even when results were not available in a timely manner, RIE still supported decision making based on process learning

Conceptual Learning

13 out of 19 RIE provided a basis for further exchange of RIE findings across different countries (conceptual learning). The demand for rigorous data by commissioning parties and cooperation partners

of GIZ was an important driver for conceptual learning – such as central support. Central support facilitated systematic and strategic learning at the organisational level and was crucial for holistic learning approach without introducing bias towards only significantly positive or highly replicable interventions.

Accountability

Five out of 19 RIEs significantly improved the overall quality of available evidence on the project as perceived by the commissioning party (accountability). The achievement of increased accountability depended on positive results, demand for data, adequate integration into the monitoring system and results measurement at the project's outcome or impact levels.

Improving evaluability

Initial findings indicated benefits for evaluability aspects through complementing monitoring data and strengthening evaluative capacity among project staff and partners. Though more data is needed, RIE seems to be limited in fostering evaluability for CPE.

Mobilising partners and resources

RIE led to resource mobilisation as well as to strengthened cooperation with project partners in several cases. RIE was jointly implemented by the GIZ project and political partners in seven cases. In some of these cases, co-leading RIE contributed to a more frequent exchange between political partners and the GIZ project team, as well as in a wider range of consultation services for the partners. Even when an RIE was not co-led, in one case it strengthened cooperation with political partners as RIE results were used for policy recommendations in which political partners scaled up the project's interventions to a national level. Moreover, RIE led in at least three cases to a mobilisation of further resources.

Conclusion

RIE offers significant corporate value for GIZ that has yet to be realised to its full extent. Opportunities for adding corporate value through RIE systemically exist on the demand and supply side as well as within the enabling environment.

RIE require purpose-sensitive facilitation. RIE can serve conceptual learning, operational learning and accountability. However, these purposes require different sets of conditions for their achievement. This is strikingly displayed by the different paths or sets of conditions QCA identified for each purpose. Hence, purpose-sensitive facilitation that tailors RIE design as well as implementation strategy and monitors context conditions is critical for achieving intended use.

Global support structures are critical for effectively facilitating the corporate value of RIE. This is particularly evident for conceptual learning. A global support and coordination structure is pivotal to plan RIE strategically and discuss findings systematically across contexts to address evidence gaps in line with the corporate strategic needs and interest.

QCA results indicate that central support is less critical for operational learning and accountability. However, users are generally not aware of the varying conditions that apply to different RIE uses. Irrespective of the intended RIE use, central support can substantially add value by helping users clearly identify their intended RIE use from the start and by advising on the relevant conditions needed to achieve this intended use.

RIE is best promoted as an integrated service delivered to the project partners, as opposed to simply conducting RIE to assess partners' effects. Whilst adding considerable value that increases ownership and sustainability of results, RIE can equally mitigate risks, particularly, if the RIE indicates a negative intervention effect.

RIE is a vehicle for strategic positioning of GIZ. Though demand is not yet systematic, a trend of increasing demand and efforts to facilitate RIE is evident on multiple levels. Moreover, while the demand among commissioning parties may not always be existent at the on-set of an RIE, the investigated cases have proven that this can very much change once data is available.

2 Introduction

Rigorous Impact Evaluation (RIE) in this review refers to a class[1] of evaluations that provide clear answers to questions of whether and to what extent an intervention caused a change in specific context (net-effect). RIE relies upon (quasi-)experimental designs to attribute impacts to various interventions. The strategic purpose of this review is (1) to determine the corporate value of RIE for GIZ and (2) to assess the enabling environment needed to achieve this value**.**

Publications of RIE in the context of international development cooperation have increased steadily and significantly over the last decade (see figure 1). This is displayed in an online review across a wide range of relevant sources which Cameron et al. (2015) conducted in 2013. Meanwhile, bilateral and multilateral development agencies have expanded their capacities for (quasi-)experimental impact evaluations. Multiple global initiatives have been launched to foster the use of (quasi-experimental) impact evaluations, such as the World Bank´s Development Impact Evaluation Initiative (DIME) or the International Initiative for Impact Evaluation (3ie).

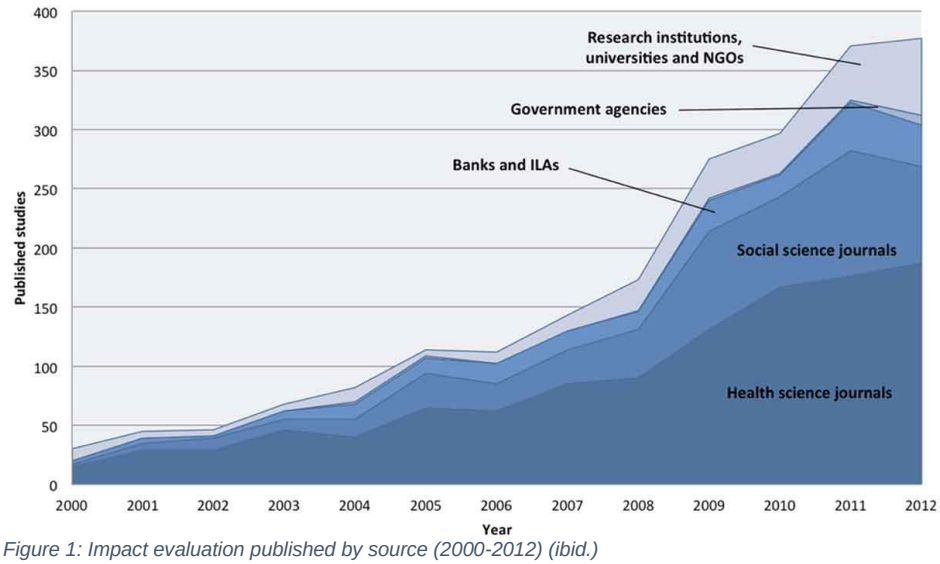

Figure 1: Impact evaluation published by source (2000-2012) (ibid.)

This global trend sets certain standards for German development cooperation. At GIZ, the value of RIE for results measurement has been an object of on-going discussion (Krapp & Dinges, 2012). Indeed, the interest in and the need for using counterfactual experimental approaches has recently increased. For example, GIZ´s (2018b) recently updated evaluation policy stipulates a specific role for (quasi-) experimental impact evaluation (IE). RIE is useful to provide clear answers to questions of whether and to what extent an intervention caused a change in the specific context (net-effect). Still, RIE is likely to be limited to specific circumstances. Occasionally, GIZ commissions such experimental methods when it wishes to examine the impact of innovative interventions or to support large-scale projects that are particularly relevant from a political point of view (GIZ, 2018b). GIZ fosters the use of mixed method approaches to increase the validity of the results and to facilitate more robust learning at the project and conceptual levels (external validity).

[1] Hence, the term rigorous in this review does not imply any normative statement of general superiority over other type of evaluations.

Initial dialog with multiple peer development organisations revealed that, while the use of (quasi-)experimental designs for impact evaluations differs substantially among agencies, there is a general trend towards a more strategic use of RIE. Whereas some organisations have dedicated programmes to facilitate the use of RIE; others make very limited use. All organisations with currently limited use of RIE have launched or are planning to launch initiatives for fostering RIE use. Meanwhile, organisations that had made more extensive use of RIE in the past – such as DFID or USAID – are currently scaling back its use. To this end, it appears that the methodological discourse around (quasi-)experimental and other evaluation approaches has become less ideological, and is instead increasingly driven by a general acknowledgement for the value of different methodological approaches.

In line with this general trend, this review takes on an institutional perspective to investigate how RIE can be strategically embedded in the organisational context of GIZ to add corporate value.

3 Review object and objectives

Review object

RIEs are impact evaluations that apply (quasi-)experimental designs for determining the net effect of an intervention. Definitions of impact evaluation vary significantly among international development agencies (see annex 8.2 for examples). In line with GIZ (2018a: 16), this review understands impact evaluations as the "[e]xamination of the degree to which development measures have brought about changes in an initial situation by analysing causal links […] in terms of the results logic (e.g. of a project)." Various evaluation designs and methods are deemed suitable for impact evaluations. To this end, RIE defines a subset of impact evaluations that apply (quasi-)experimental designs.[2] The use of a counterfactual, (quasi-)experimental design allows one to precisely determine the extent – quantified as net-effect – to which a specific intervention has caused a result.[3] This net-effect is *solely* attributed to the intervention. The understanding of impact evaluation stipulated in this review makes no distinction as to which type of effects *impact* evaluation refers; in contrast to the OECD-DAC impact criterion.[4]

RIE in relation to the formal GIZ results evidence functions

RIE can substantially supplement GIZ´s two globally formalised instruments/ processes for results measurement: i) Results-based Monitoring (RBM) and ii) Central Project Evaluations (CPE). RIE is distinct from these instruments mainly through its specific measurement approach (attribution). Through the focus on attribution, RIE differs from RBM's focus on correlations, i.e. trends as a description of the factual. RIE can support RBM through causal analysis, baseline data collection and the quality of indicators, but RIE does not deliver infor-

[2] This commonly includes randomised controlled trials, difference-in-difference estimates, propensity score matching, regression discontinuity and instrumental variables (White & Raitzer 2017).

[3] GIZ understands results as the intended or unintended, positive or negative changes in a situation or behaviour as the direct or indirect consequence of an intervention. Results include impacts, outcomes and outputs. The OECD-DAC and BMZ also define results as the overallterm for impacts, outcomes and outputs.

[4] Hearn & Buffardi (2016) provide a useful overview of the different definitions and uses of the term impact. Also see annex 8.2 for relevant definitions.

mation as frequent and comprehensive as required for regular reporting. While not limited, CPE most commonly applies theory-based approaches for (causal) contribution. Contribution also implies causality to the extent that an intervention has helped to cause an effect (contribution claim) in terms of the consistency of a causal relationship with the intervention and/ or of the elimination of alternative explanations (Rogers, 2014).

While RIE can support CPE through precise attribution and evidence related to specific interventions, it does not provide a comprehensive assessment of all OCED-DAC criteria across all project interventions as required by federal guidelines. Moreover, as measurement mostly focusses on specific interventions and often takes place during project implementation, RIE is more limited in capturing the full impact of a project. While the measurement of results may be complemented and thereby strengthened by embedding RIE into the project's implementation process, the boundaries between the different forms of results measurement at GIZ remain fluid as RIEs constitute an integral part of the monitoring and evaluation framework of the project.

Review objectives

The strategic purpose of this review is (1) to determine the corporate value of RIE for GIZ and (2) to assess the enabling environment in terms of an organisational support structure for achieving this value. The corporate board, the evaluation unit and the sectoral department are the primary users of this strategic review. Potential secondary users are all operational units in GIZ and key external evaluation partners, such as the Federal Ministry for Economic Cooperation and Development (BMZ) and local partners.[5] For achieving its purpose, this review pursues the following specific objectives:

- Map the existing use of RIE within GIZ
- Capture the utility of and experiences with RIE within GIZ
- Map the interest in RIE among key internal stakeholders

The use of evaluation for conceptual learning and accountability leads to critical value creation processes for GIZ. Against this background, the review investigates the value of RIE by contributing to:

- ***Operational learning****,* which supports on-going project management for results-oriented implementation and adaptation of activities during the current project cycle.
- ***Conceptual learning***, which supports project-overarching promotion and adaption of intervention designs.
- ***Accountability*** to commissioners by providing evidence for strengthened communication of development results.

Hypothesis: RIE strengthens the evidence base for operational learning, conceptual learning and accountability.

Based on its specific measurement approach and time of measurement, RIE supports the key functions of evaluations by providing complementary evidence that:

- Minimises potential bias from confounding factors for the quantification of the (net-)effect attributable to an intervention, and
- Maximises relevance and timely available data for use.

In addition to the central hypothesis stated above, it is assumed that RIE can add corporate value by increasing the efficiency and effectiveness of CPE through its complementarity. Moreover, it is assumed that RIE can sup-

[5] However, for effectively reaching this secondary audience further tailored communication outputs will be needed.

port mobilising resources and strengthening partnerships as it becomes more widely used across all development agencies and their partners. These value aspects were also investigated but not to the same extent as the central hypothesis.

Value aspect 1: RIE strengthens the efficiency and effectiveness of CPE.

Based on its prospective evaluation design, RIE is integrated into the project's monitoring framework and may support the key functions of CPE through strengthening these key evaluability aspects (Peersman, 2015):

- Adequacy of intervention designs and the quality of RBM is supported, particularly, by the early engagement of the evaluator(s) and the emphasis on process use of evaluation;
- Availability of timely and valid data is supported, as discussed for the central hypothesis, and
- Conduciveness of institutional context is supported, particularly, by the close engagement of the evaluator(s) with evaluation stakeholders.

Value aspect 2: Through meeting an increasing demand among key development partners, RIE creates increased opportunities for resource mobilisation and partnerships.

The relevance of RIE to GIZ´s core commissioning parties and partners is investigated in terms of serving public and political accountability, and of ensuring comparability in the global discourse. As such, the potential future demand for RIE will be assessed. A key for this hypothesis is the ability of RIE to support the communication of development results. This externally focused function indirectly supports the enabling environment for evaluation by strengthening legitimacy for resource mobilisation and partnerships. In addition to RBM and CPE, "Results Data "(RD) is another formalised *communication* instrument at GIZ (2018e). RD do not measure impacts, but data (monitored or plausibly estimated) aggregated to report on results of GIZ's work in the various sectors on global and regional levels to the general public. RIE can support this process by allowing for more precision.

4 Methodological approach

The scope of this strategic review includes formative and summative elements. Within its primary formative scope, the review aims at identifying pathways to a high corporate value of RIE, which in turn will facilitate the implementation of RIE. For this purpose, Qualitative Comparative Analysis (QCA) was conducted on the most relevant outcomes: operational learning, conceptual learning, and accountability. Within its summative scope, the review captured existing corporate experiences with, and assesses current demand for, RIE in order to substantiate the central hypothesis. Due to the only recently established reform of CPE and the limited extent to which RIE and CPE have been conducted for the same interventions, information relating to value aspect 1 is less readily available.

To examine the central hypothesis and RIE's corporate value, a multiple case study design was applied following an organisational field mapping and an in-depth analysis of RIE. The period under review was limited to RIEs conducted between 2014 and 2018 at GIZ to maximise access to relevant stakeholders. The former served to identify relevant stakeholders within GIZ (internal field) and beyond (external field), and their interest

in and capacity for RIE. On this basis, this review identified trends for RIE practice and relationships in the organisational fields.

The data was collected and analysed in the following stages and steps:

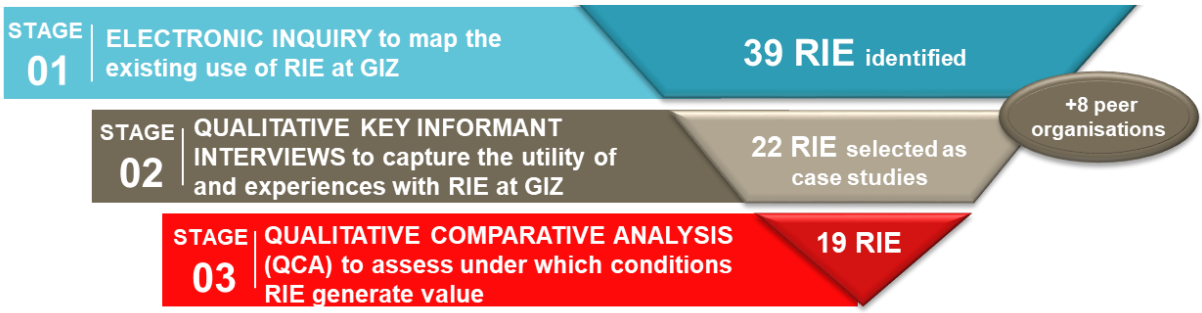

Figure 2: Methodological approach

Stage 1

- ***Electronic inquiry*** sent to regional evaluation focal points for obtaining basic information on conducted RIEs and related documents from organisational units within GIZ
- ***Desk review*** of a) publicly available documents from key development agencies to determine their position towards and experiences with RIE and b) GIZ documents obtained through electronic inquiry to verify the applicability of each case for the review and in-depth case study. A standardised quality checklist was applied for this purpose.

Stage 2

- ***Qualitative key informant interviews (KIIs)*** for a) 22 case studies of RIEs at GIZ and b) for mapping interest in and capacity for RIE among key internal and external stakeholders, as well as eight bilateral and multilateral peer organisations. The interviews at the organisational level included GIZ counterparts, the sectoral department, sector and global programmes and the evaluation unit. At project level, interviews included GIZ project managers and evaluation implementers. Each RIE, or each country in which an RIE was implemented, represents a case. Based on the assumption that global support structures play a critical role for RIE use, the 22 cases were purposively selected to represent three types of internal structures.
    - Hierarchical: Global programmes follow a hierarchical structure by implementing comparable interventions across different contexts (country packages). A purposeful sampling of ten (10) RIEs (country packages) as part of three global programmes was conducted to the point of theoretical saturation.
    - Network: Regional or sectoral networks exchange lessons learned on different challenges in the area of service delivery. Four bilateral projects conducting RIE as part of a global or regional initiative were considered. One RIE was excluded as the project manager was not available for an interview.
    - Decentralised: Eight projects conducting an RIE on the initiative of project managers were considered.

  In comparison to the total 39 RIE cases identified, 15 RIE cases of global programmes were excluded due to theoretical saturation. In addition, one RIE case was excluded for decentralised structures and one for network structures because project managers were not available for interviews.

Interviews with representatives from bilateral and multilateral peer organisations also aimed to initiate dialogue around the topic. The selection of organisations occurred less systematically and was often driven by the interests of organisational counterparts. Hence, findings are understood to provide a more general sense for the discussion around RIE only.

- ***Coding of interviews*** was conducted using the software MAXQDA.[6]

Stage 3

- ***Qualitative Comparative Analysis (QCA)*** was conducted to assess under which conditions RIE is used for operational learning, conceptual learning and accountability (central hypothesis). The methodology was chosen for its ability to provide answers to the main research questions of this Strategic Review and to generate robust findings, given the limits of 19 selected RIE cases (see annex 8.3). In its ability to identify different combinations of conditions that lead to the same type of RIE use, QCA identifies parallel cause-to-effect chains that provide an exhaustive explanation of the outcome for all cases. Thus, QCA is well equipped to deal with causal complexity and to provide evidence for context-sensitive recommendations. Finally, it increases the transparency and retrace-ability of qualitative research

The empirical basis for the present QCA is formed of 19 case studies consisting of assessments for each condition and outcome under study. In order to evaluate the 19 RIEs with respect to each of the case characteristics, the review team analysed documents and a minimum of two interviews per case to allow for triangulation within each case. Three of the 22 cases investigated in stage 2 had to be excluded for QCA due to availability of relevant additional interview partners. The detailed definitions of each case characteristic of the RIEs and of each score are documented in annex 8.5.

THE EXPLANATORY MODELS

By using qualitative comparative analyses (QCA), circumstances under which RIE leads to each respective outcome were identified. Separate QCAs were conducted for each of the three outcomes, as it is assumed that not only different combinations of the same factors may enable a specific type of use, but also that for each evaluation use, a different set of factors may be causally relevant. The conditions for an outcome may include structural factors, such as the setting in which an RIE takes place, or proximate factors, such as aspects of RIE design and implementation (Schneider/Wagemann 2006).

For each of the analyses, an explanatory model was developed. The model contains conditions that are assumed to be the most relevant ones to explain how RIE contributed to operational learning, conceptual learning or accountability. The number of conditions was limited to five for each analysis in order to exclude the possibility that findings resulting from the QCA are a product of chance. The five conditions for each outcome were selected by the review team from a pool of case characteristics that were assessed for each RIE. The selection process took place during a workshop in November 2018. A re-evaluation of the selected conditions took place at a second workshop in February 2019.

The detailed definitions of the selected conditions can be found in annex 8.5. Moreover, the percentage and the absolute number of cases in which the condition was present can be found in the graphical representation of each explanatory model.

For more information about the application of QCA in this review, please refer to annex 8.4.

[6] The coding scheme derives from the main evaluation questions and emergent conditions for the attainment of pre-defined outcomes based on the QCA.

5 Findings

Almost all users at GIZ see RIE as a value adding measure to promote operational learning, promote conceptual learning and strengthen accountability. The support by a specialised, central unit at GIZ is pivotal to facilitate an RIE that is tailored to the intended use.

5.1 Use of RIE at GIZ

The initial case mapping indicated that (quasi-) experimental impact evaluations are presently commissioned and steered by projects or staked out initiatives. Consequently, knowledge management was often restricted to the project, sectoral or regional levels, which limits organisation-wide access to findings and knowledge exchange regarding methodologies used. This also reflects the fact that multiple cases were only identified at later stages of the review.

Between 2014 and 2018, 39 RIEs were applied across all regions and various sectors (see figure 3), such as agriculture (14), rural development (11), governance (4) and employment (4). Most of the identified RIEs were implemented in the region of Africa (29), followed by Asia (7), Europe (2) and Latin America (1). As further cases were identified over the course of the review, it is likely even further relevant cases exist. Moreover, Leppert et al. (2018) presents an externally managed RIE of GIZ interventions,[7]

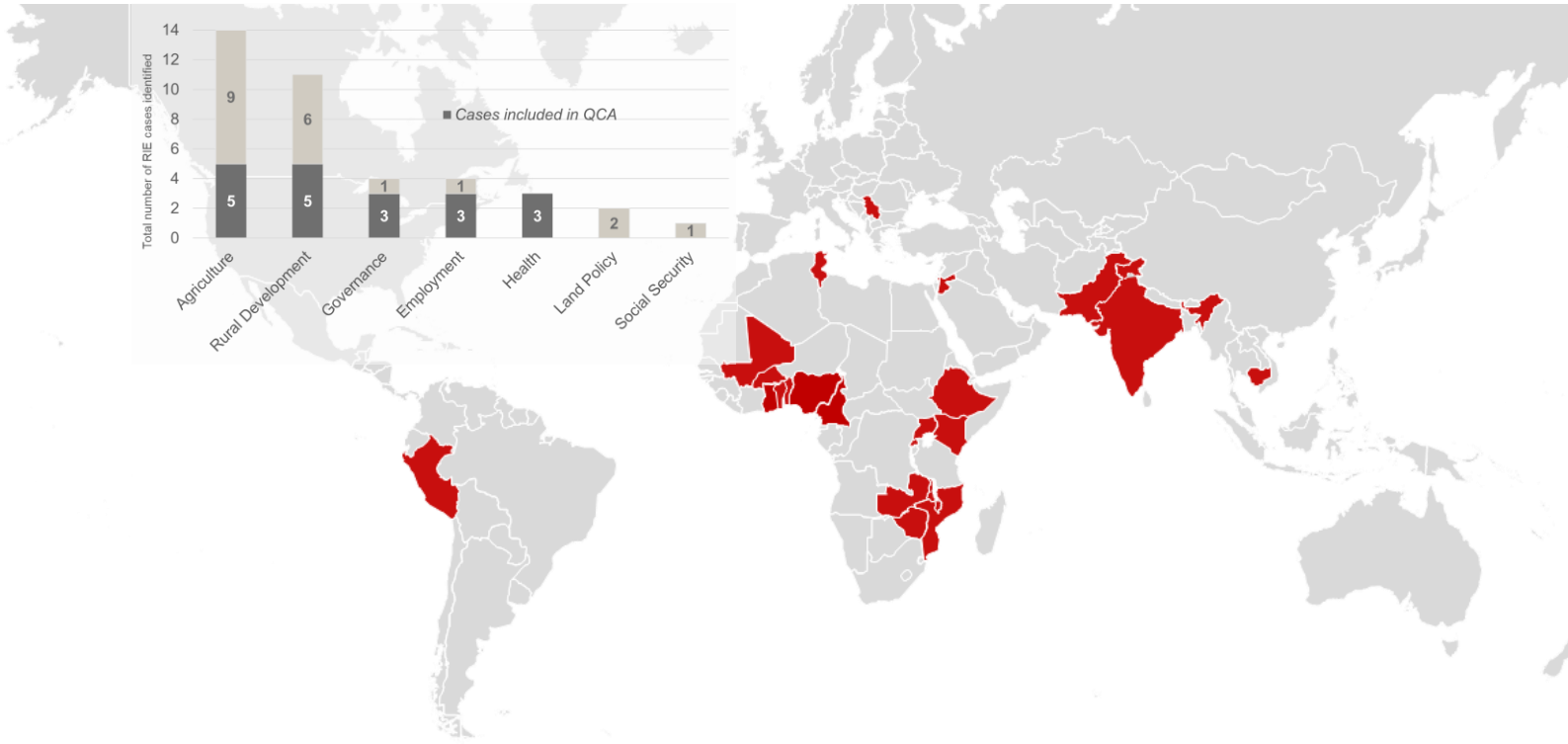

Figure 3: Distribution of 39 RIE cases between 2014 and 2018 identified in case mapping.

[7] The RIE assesses a technical approach within an environmental and rural development programme in the Philippines managed by GIZ. This case was excluded from further analysis in this review, as the RIE was neither initiated nor managed by GIZ.

Results of the case mapping suggest that RIE is usually conducted at the following organisation level: (1) hierarchical structure, (2) network structure and (3) decentralised structure. In approximately half of the cases, RIEs were conducted in bilateral projects (decentralised structure). In those cases, RIE implementation was mostly driven by the interest of local project stakeholders rather than by institutional demand. In the sectors of employment and governance, however, RIEs were initiated through regional or sectoral networks. The 'Wirkungsinitiative in Governancevorhaben' (results initiative for governance interventions) was initiated by the sectoral department and the Africa department in 2016 in order to explore opportunities to apply RIE more systematically in governance interventions (Funk et al., 2018). Furthermore, the German Federal Ministry for Economic Cooperation and Development (BMZ) commissioned a study in 2014 to develop evaluation designs that are capable to measure net employment impacts (Kluve/Stöterau, 2014). To gain experience with RIE and to generate knowledge about the effectiveness of labour market interventions, the sector project 'Employment Promotion in Development Cooperation' initiated several pilot evaluations of representative approaches across different GIZ focal countries in 2016. Finally, two global programmes prospectively embedded RIE into their intervention design to evaluate the impact of an overarching approach across different countries and contexts (hierarchical structure). As part of a special initiative, these programmes were characterised by high political relevance.

Independently of the internal structures, project managers and monitoring and evaluation (M&E) experts commonly expressed a need for a specialised, central support structure. Functions that support structures need to include: i) consultation on the potential use of RIE, ii) evaluability and feasibility assessment, iii) commissioning the RIE, i.e. development of the terms of reference (ToR) and selection of RIE implementer, iv) methodological support, v) networking for peer-to-peer learning and vi) dissemination of findings.

5.2 Supporting operational learning, conceptual learning and accountability

HYPOTHESIS

RIE strengthens the evidence base for decision-making, conceptual learning, and accountability.

In the 19 analysed cases, RIEs were used for operational learning (in 53% of the cases), conceptual learning (68%) and strengthening accountability (31%). A deep integration of RIE into the project's monitoring framework and the provision of RIE results in a timely manner form a combination of factors leading to the use of RIE. The demand for rigorous data emerges as an important driver for conceptual learning. Nonetheless, the results indicate that support by a specialised, central unit at GIZ is pivotal to facilitate systematic and strategic learning from RIE. Furthermore, RIE strengthens accountability if the demand for more rigorous data by commissioning parties of GIZ is met by positive results on the project's outcome or impact level whilst being deeply integrated into the project's monitoring framework.

Operational learning

RIE was used for operational learning in ten out of 19 cases. Decisions by project management regarding the implementation of activities during the current project phase were guided either by the results of the RIE (four out of the ten cases) or by learnings that occurred during the implementation process of the RIE (six cases). Interview partners emphasised that taking part in the RIE process improved their understanding of the project context and helped to refine the results logic of the project. Moreover, in most of the cases, the baseline data of the RIE already provided useful information to project management to effectively adjust the intervention design prior to its implementation. In five out of nine cases in which the RIE did not lead to management decisions, it was reported that the RIE still contributed to a better understanding of the project's results logic.

THE EXPLANATORY MODEL FOR OPERATIONAL LEARNING

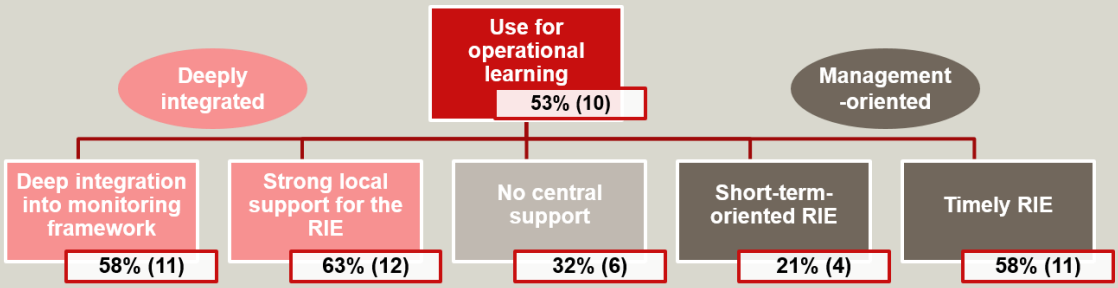

Figure 4: Explanatory model for operational learning use.
Note: Percentages of all cases (N=19) and absolute numbers (provided in brackets) indicate the presence of the outcome and conditions.

Five conditions are assumed to influence the use of RIE for operational learning: deep integration into the project's monitoring framework, strong local support, short-term orientation, timeliness and central support. *Deep integration into the project's monitoring framework* presupposes that the implementation of RIE starts at the same time as the implementation of the project/component (investigated by the RIE) or at the same time the monitoring system is established, whichever starts later. Moreover, it is only assumed to be present if RIE complements monitoring data, i.e. RIE or RBM do not separately produce data redundancy. *Strong local support* describes a high capacity to conduct the RIE. A high project capacity can be guaranteed either by evaluators who were continuously engaged or by the project staff itself, i.e. that the project had sufficient workforce, methodological knowledge and strong leadership to enable the project to implement RIE. Both conditions – a deep integration into the project's monitoring framework and strong local support – are aspects of deeper integration of the RIE into the project implementation, which is hypothesized to ensure that collected data is useful to the project. The conditions *short-term orientation* and *timeliness* are aspects of an RIE's management orientation. An RIE is assumed to be management oriented if the project results investigated by the RIE are below the level of a project objective, and if the RIE provided results at least half a year before the end of project implementation. A stronger focus on short-term results is expected to foster the use of RIE for operational learning as it allows one to assign RIE results to specific project activities. *Central support* refers to support that the project received in commissioning an RIE from a central unit at GIZ, such as a sectoral department or a coordination unit of a global programme. The high number of projects in which this condition is present can be explained by the number of country packages as part of a global programme and projects that are part of a sector initiative among the cases considered.

None of the cases applied “nimble” RIE designs that are specifically tailored towards supporting operational learning. For promoting operational learning, nimble RIE designs particularly differ from traditional RIE designs in three aspects as displayed in the table 1 below.

	Nimble RIE	Traditional RIE
Intervention	Usually a low-cost modification to an existing product	A modification to an existing product or an entirely new product
Data source	Largely administrative (internal data-bases)	Administrative & survey
Outcomes	Limited to first-order outcomes	Higher order outcomes

Table 1: RIE designs.
Adopted from: Dibner-Dunlap & Rathore 2017

The results of the QCA show that **in almost all cases in which RIEs were used for operational learning, a deep integration of RIE into the monitoring framework of a project was given** (see figure 5 below). Hence, an early integration of RIE into the project design or the monitoring system seems essential to ensure operational learning use.

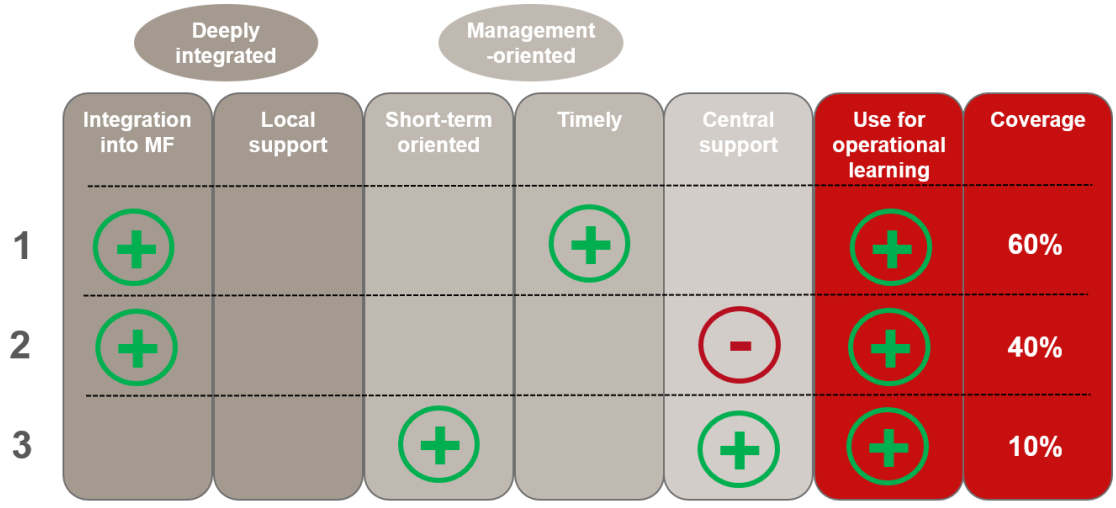

Figure 5: Paths to operational learning use.

In 60% of all cases in which an RIE was used for operational learning, it was deeply integrated into the project’s monitoring framework and provided results at least six months before the end of the current project phase (path 1). Here, the usefulness of the results was ensured, while the opportunities to use RIE results for decision-making were maximised. If integration into the monitoring framework is not present, central support must facilitate RIE to focus on short-term results in order to produce evidence that is useful to project management and to influence decision-making on project level (path 3).

The empirically most important path requires RIE results to be available in a timely manner (path 1). The second and third path, in contrast, include cases in which the RIE did not provide results before the end of the project phase. The reasons for not being able to provide RIE results in a timely manner were manifold. In some cases, the contracting of evaluators or ethical clearance processes led to delays of RIE results. In others, increased security measures were cited. Moreover, RIEs were sometimes simply not planned to provide results at least six months before the end of the project; in particular in cases with a regular project phase of up to

three years. If timeliness are absent, the decision-making process can only profit from insights that were gained during the implementation of the RIE. Even in such a situation, it is still sufficient when an RIE is well-integrated into a project's monitoring framework to influence local management decisions – but only if central support is absent (path 2). An explanation for this could be that if central support is absent, then the RIE's design and implementation is most effectively tailored towards the interests of local project stakeholders. However, for those cases in which evaluation results were not provided in a timely manner and in which central support was provided, the RIE must concentrate its data collection efforts on short-term results to produce evidence that is useful to project management (path 3).

While QCA served to identify combinations of conditions which lead to the use of RIE for operational learning, **two strategies to further increase the usefulness of RIE for operational learning were identified in the qualitative analysis: Testing of implementation alternatives and a successive implementation of project activities.** In some cases, implementation alternatives of an activity (treatment arms) have been tested against each other, and both have been tested against a control/comparison group (counterfactual). Testing alternative intervention approaches has enabled the project teams to choose the most effective implementation alternative in the current project phase and to base the design of follow-up projects on these findings. Another approach to make the RIE more management-friendly was to implement interventions as well as the RIE successively. This procedure allowed the project team to learn from the RIE findings of the first round of implementation and to inform the implementation of the second round.

HOW TO READ A PATH DIAGRAM

Paths represent parallel ways in which RIE have achieved a high corporate value by having operational learning, conceptual learning or increasing accountability. The path diagram provides all paths that lead to the outcome under study. Each path is numbered and put in a separate row. The two columns on the right side of the diagram (with a grey background) inform regarding the outcome under study and the percentage of all RIE reaching the outcome on the respective path. As some paths are not mutually exclusive, an individual RIE may reach the outcome on several of these paths. The plus symbol in the first of the two grey columns indicates that the outcome was achieved. The columns with coloured background in the middle of the diagram represent the conditions that were employed in the analysis. A plus in the column of a condition means that the respective condition must be present to achieve the outcome on the respective path. A minus means that the path leads to the outcome only when the condition is not present. If in one of the rows there is neither a plus nor a minus for a given condition it means that this condition is irrelevant for the respective path to the outcome. For example, the first row in Figure 5 indicates that 60% of the RIEs that contributed to operational learning achieved this outcome because they were deeply integrated in the monitoring framework and provided results at least six months before the end of a project phase.

Conceptual Learning

Conceptual learning is approximated here as a cross-country exchange on the findings of the evaluation that has taken place within GIZ. **Such a cross-country exchange has taken place for 13 out of the 19 cases**. Conceptual learning refers to learning by GIZ and typically includes the fact that learnings from one project are applied in another project. The presentation and communication of RIE findings within GIZ and across countries is a critical aspect of conceptual learning. The type of exchange that took place differed across cases. While country packages of global programmes shared their results in working groups of the global programme, bilateral projects exchanged information on RIE findings via presentations within GIZ headquarters or within their respective sector network. Notably, in hardly any of the investigated cases were RIEs strategically planned to address institutional evidence gaps through cross-country comparison.

THE EXPLANATORY MODEL FOR CONCEPTUAL LEARNING

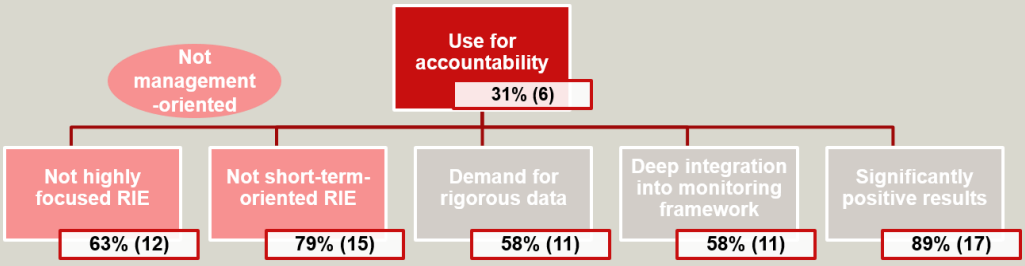

Figure 6: Explanatory model for conceptual learning.
Note: Percentages of all cases (N=19) and absolute numbers (provided in brackets) indicate the presence of the outcome and conditions.

Central support, demand for rigorous data, the replicability of the intervention, significantly positive results, and the absence of a short-term orientation were assessed to lead to the use of RIE for conceptual learning. While *central support* ensures that a corporate unit or department is directly involved in the implementation of an RIE, and could act as knowledge brokers inside GIZ, a *demand for rigorous data* by commissioning parties or partners increases pressure to learn from the RIE. Furthermore, GIZ itself may be more inclined to learn from the findings of an RIE if the evaluated intervention is *replicable*. In light of the fidelity discussion: replicable refers here to extremely standardised interventions such as a safe delivery check list for nurses. If the evaluated intervention is replicable the perceived utility of the RIE's findings may be higher, as there are opportunities to directly apply the newly gained knowledge whenever the intervention is reconsidered. To arrive at a rating of replicability two aspects of an intervention were assessed. First, the number of countries was determined in which the intervention was implemented in a comparable manner. Second, the standardisation of the intervention was assessed. The degree of standardisation was rated rather high if activities were standardised not only regarding their type (e.g. workshop or financial support) but also regarding their specific format or content (e.g. type of workshop or mode of delivery of financial support). Also, for a rating of "rather highly standardised" these activities had to be delivered identically to all beneficiaries that were supposed to receive the respective activities. An example of a replicable intervention is a health checklist with a fixed content that is provided to all beneficiaries. Further, GIZ may not only be more inclined to learn from the findings of an RIE if the intervention is replicable, but also if the RIE's findings are significantly positive and provide information on outcomes and impacts. If RIE findings show a positive effect of project activities on project results project managers are expected to act as knowledge brokers and disseminate information on the successes of their project inside GIZ. Finally, RIE *results on outcome and impact*

level are assumed to be of more interest for cross-country learning than results on outputs. The condition that is met most frequently is *significantly positive results*. It must be mentioned that RIE results were mixed in several cases. If the results were mainly positive, significantly positive results were assessed to be present.

The demand by GIZ's commissioning parties and cooperation partners for rigorous or additional data that allows for plausible contribution emerged as an important driver for conceptual learning. Its presence is required by the two empirically most significant paths (1 and 2, see figure 6). In 61% of the cases in which an RIE has contributed to conceptual learning, the involvement of a specialised, central unit in the implementation of the RIE ensured that newly gained knowledge was disseminated within GIZ, provided that the demand for rigorous data on project results is met by the RIE through the provision of information on outcomes and impacts (path 1). Irrespective of an existing central support, significantly positive effects can facilitate conceptual learning in cases where commissioning parties or partner of GIZ demand rigorous data on project results (path 2). As mentioned earlier, it is assumed that in these cases, the project management acts as a knowledge broker. Path 3 is the only path leading to a dissemination of RIE results across countries within GIZ though central support is not provided, and an external request for rigorous data is not necessarily required. This path, however, *only* leads to conceptual learning if the RIE evaluates a highly standardised, multi-country intervention and if the effects are at least measured on the outcome level. In these cases, the implementation of the same standardised intervention across different countries (beyond GIZ) sparked interest among project managers to further engage in an exchange on the methodological approach to measure the (net-) effect attributable to the intervention or to share evidence for future project designs.

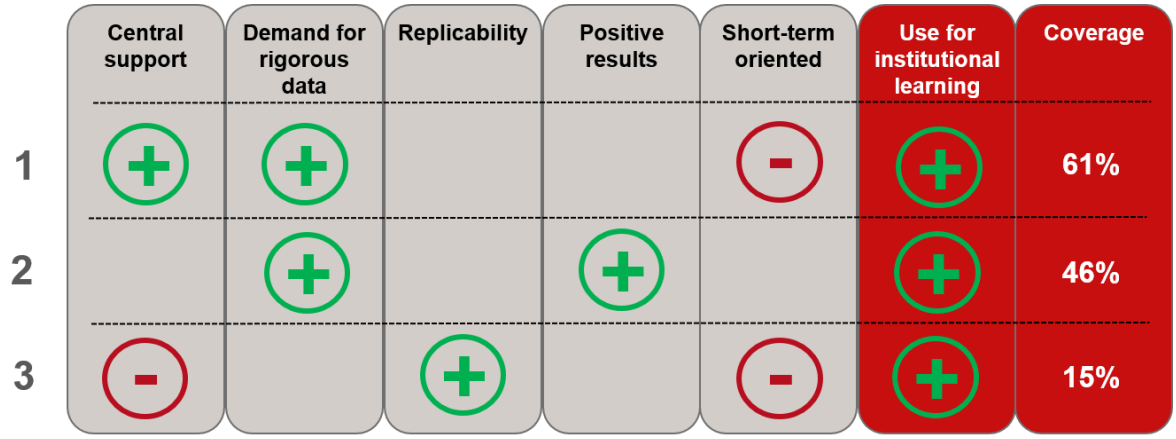

Figure 7: Paths to institutional learning

While demand for rigorous data emerges as an important condition, **the results indicate that support by a specialised, central unit at GIZ facilitates systematic learning from RIE**. This condition is part of the empirically most important path (path 1). If there is no central support, GIZ limits its opportunity to learn from RIE to cases where interventions are successful (path 2) or highly standardised (path 3). While the former path cannot be planned, the latter path might be only available for a limited number of projects which use highly standardised interventions. Roughly half of the interventions that were assessed for the study were considered highly replicable. RIE cases with highly replicable interventions generally contributed to conceptual learning. The qualitative analysis identified that the involvement of a central support unit, when absent, could have facilitated the RIE beforehand; particularly, in terms of knowledge management, methodological advice on design and implementation, feasibility and timeliness of the RIE.

In addition to disseminating findings within GIZ, RIE findings were also shared outside of GIZ. This might also lead to conceptual learning by GIZ as project teams can access published RIE results. **The evaluators published scientific papers or presented their results at a scientific conference in 47% of all cases** with potentially more publications or conference participations coming. Moreover, findings were shared with international (donor) organisations. Finally, **while conceptual learning is operationalised here as an exchange with regard to RIE findings, also an exchange regarding the methodology seems to have taken place within GIZ** and thus, a methodological capacity building. Similar to the exchange about findings, an exchange about methods took place in form of jour fixes of global programmes or within sector networks. Moreover, in one case, the design of an RIE explicitly built on a previous RIE that was conducted in another country for the same intervention.

Accountability

RIE increased accountability in 5 (26%) out of the 19 cases, i.e. project managers used the results for reporting purposes and commissioners affirmed that the evidence produced by the evaluation significantly improved the overall quality of the available evidence concerning the project. RIE supported the communication of development results to the commissioning parties of GIZ by providing more rigorous data on project results and by improving the transparency and reliability of the reported outputs, outcomes or impacts. Additionally, an evaluation conducted by an independent evaluator is seen as beneficial to the annual reporting to commissioning parties as it increases credibility. Moreover, RIE enabled one project team to argue that the project interventions have a positive effect on project outcomes even though regular monitoring indicated a negative trend regarding the project outcomes for the group of beneficiaries. In that case, the trend for the group of beneficiaries was less negative than the trend which was observed for the control groups. Generating knowledge on project interventions' efficacy by means of RIE is also part of some projects' service delivery as agreed upon with commissioning parties. In these cases, the main purpose of RIE was not only to increase accountability on project interventions but also formed part of the assistance service to partner institutions as well as generated new knowledge on behalf of the commissioning party.

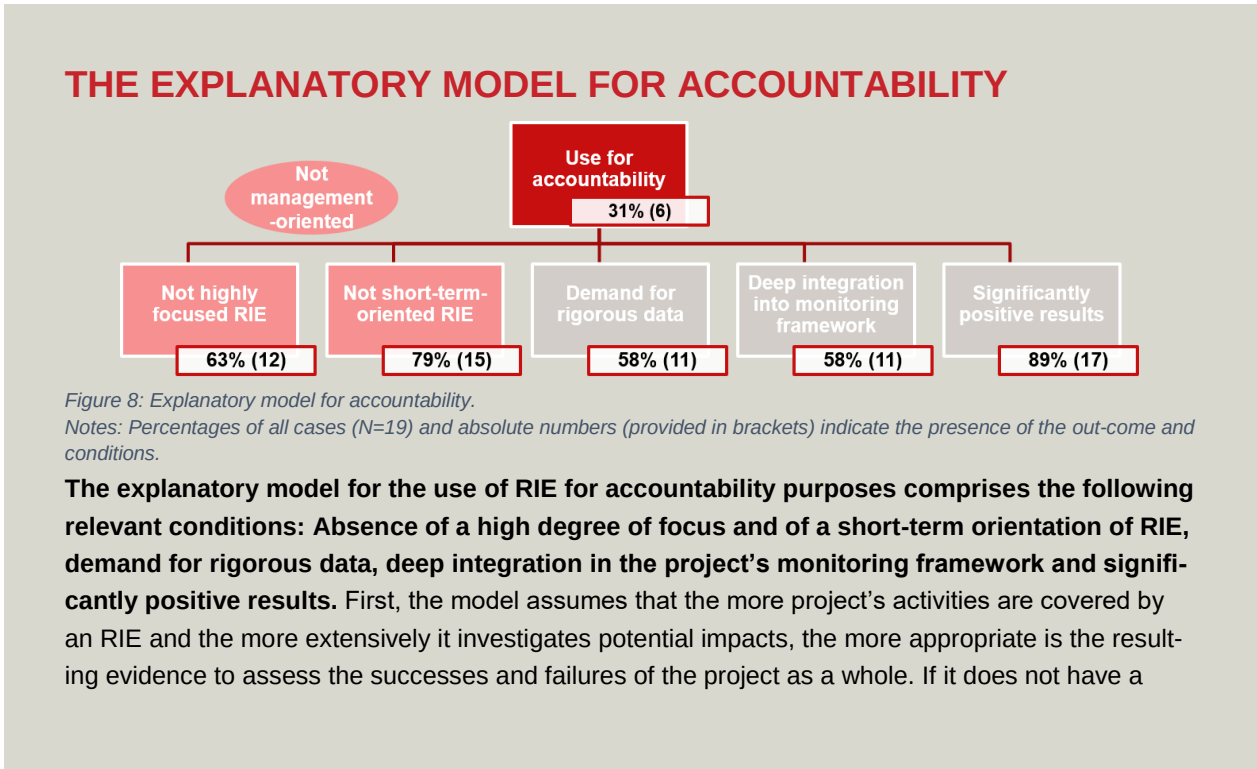

Figure 8: Explanatory model for accountability.
Notes: Percentages of all cases (N=19) and absolute numbers (provided in brackets) indicate the presence of the out-come and conditions.

The explanatory model for the use of RIE for accountability purposes comprises the following relevant conditions: Absence of a high degree of focus and of a short-term orientation of RIE, demand for rigorous data, deep integration in the project's monitoring framework and significantly positive results. First, the model assumes that the more project's activities are covered by an RIE and the more extensively it investigates potential impacts, the more appropriate is the resulting evidence to assess the successes and failures of the project as a whole. If it does not have a

high degree of focus and is not short-term oriented, RIEs produce evidence that covers outcomes or impacts of a large part of the project's activities. This makes the evidence suitable for reporting on the project's achievements and thus, for increasing accountability to GIZ commissioning parties. Second, a demand for rigorous data and significantly positive results are both expected to increase the motivation to use available evidence. While a request for rigorous data is expected to increase the pressure to report the results of an RIE, project management is presumed to use the results of the RIE more willingly if they are positive and significant. Third, a deep integration of RIE into the project's monitoring framework ensures its complementarity to monitoring data and hence, increases its usefulness for reporting on project implementation and results.

Figure 9 shows that all RIEs that supported the communication of development results to commissioning parties of GIZ reached this outcome on the same path. In these cases, **commissioning parties of GIZ had requested more rigorous data on the project, RIE were deeply integrated into the monitoring framework and it provided significantly positive results at the outcome or impact level of the project.** While this implies that negative results did not support accountability, it should be noted that, apart from two evaluations, all RIEs investigated delivered at least mainly significantly positive results. Furthermore, while in some cases, the commissioning parties demanded more rigorous data beforehand, the qualitative analysis shows that in other cases RIEs were able to trigger their interest later on in the implementation process of RIE, which also contributed to an increased accountability.

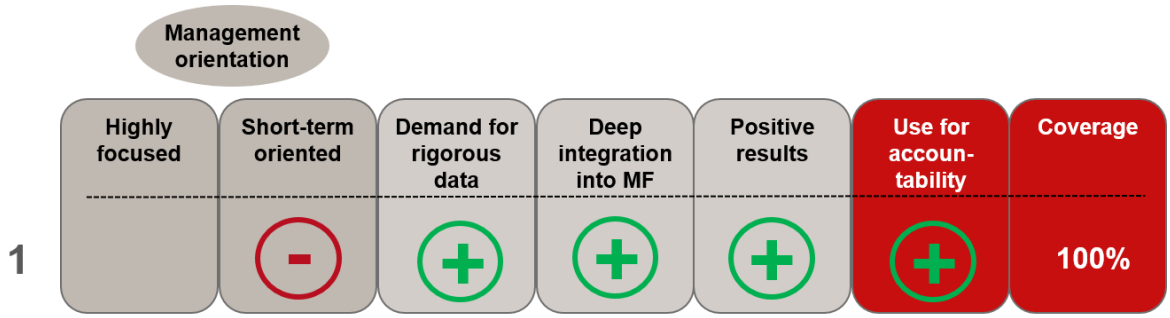

Figure 9: Paths to accountability

5.3 Improving evaluability

VALUE ASPECT 1

RIE strengthens the efficiency and effectiveness of CPE.

Based on its prospective evaluation design, RIE is embedded into the project's implementation process and hereby assumed to support the key functions of CPE, through strengthening key evaluability aspects (see chapter 3, value aspect 1). **However, up to now, no CPE and only one decentral project evaluation has been conducted in a project in which an RIE was also conducted**. In the case of the decentral project evaluation, RIE results were available only after the end of the project and after the decentral project evaluation was conducted. Therefore, the potential of RIE to strengthening the efficiency and effectiveness of CPE has yet to be determined. As there are CPE planned for two of the projects in which an RIE was conducted, first results regarding the second hypothesis are expected to be available soon.

The evaluability aspects that an RIE is assumed to strengthen such as data availability and quality are also critically served by planning and results-based monitoring. Therefore, the complementarity of the RIE to the monitoring system was assessed. **The complementarity of RIE and monitoring systems seem to differ widely among the investigated cases.** In approximately half of the cases subject to the in-depth analysis, RIEs effectively complemented the monitoring system. In these cases, RIEs particularly supported the ascertainment of verifiable baseline data and contributed to an improved monitoring system in terms of indicator quality as external consultation was provided. In other cases, RIEs generated datasets useful for project indicators and allowed for comparisons with the dataset generated by the monitoring system. The monitoring data can hence be checked and verified. Moreover, as the monitoring data and the results produced by an RIE do not always match, discussions about the monitoring system were facilitated. However, in some cases, RIEs produced data that did not serve the reporting of project indicators and was largely independent of the monitoring system. Regardless of whether an RIE reported on project indicators or not, interview partners emphasised that RIE allowed them to analyse spillover effects to control groups which could not be analysed with monitoring data. Additionally, the implementation of an RIE increased the understanding of a project staff for evaluations. However, it was also criticised that while RIEs improve data availability and quality, they also increase the amount of time and effort that must be spent on the data collection considerably and according to some interview partners excessively.

It is important to note that RIEs in many cases only examined specific subsets of a project intervention but did not evaluate the whole project in its entirety. Depending on the project type,[8] QCA results related to the factor "short-term orientation" and "focus of evaluation" also displayed that RIE is limited in its ability to effectively assess outcome or impact level measures for the overall project activities – at least in the case of bilateral programmes. However, while limited, RIE findings may still support CPE by delivering rigorous evidence for individual causal links.

In summary, more evidence for value proposition 1 is needed. However, initial findings indicate that RIE is limited in fostering evaluability for CPE. RIE is best viewed as complementary evaluation instruments within GIZ´s evaluation system that can create significant corporate value in its own right, rather than mediated by supporting other instruments.

[8] bilateral programme or global programme

5.4 Mobilising partners and resources

VALUE ASPECT 2

Through meeting an increasing demand among key development partners, RIE creates increased opportunities for resource mobilisation and partnerships.

RIE has led to resource mobilisation as well as to strengthened cooperation with project partners in several cases (see chapter 3, value aspect 2)**.**

RIEs were co-led by the GIZ project and political partners in 7 cases. Co-led means that either GIZ and its partners jointly decided about the design, implementation and use of RIE, or that partners of GIZ had sole responsibility for parts of the implementation of the RIE. RIEs were never co-led on projects that were part of a global programme. This finding supports the hypothesis that a consequence of the presence of central support could be that RIE is less tailored to the local needs of the project. **Co-leading RIEs resulted in some cases in a more frequent exchange between political partners and the GIZ project team as well as in a wider range of consultation services for the partners.** In one case, two representatives of political partner institutions as well as the GIZ project manager and the evaluator formed an RIE management team that held monthly meetings. The political partners in this case were also the only ones among all cases who explicitly requested the RIE and more specifically a randomised control trial. The RIE not only led to a frequent exchange between partners and the GIZ project team but was also part of service delivery to the partner. Moreover, co-leading RIEs might result in widening consultation services for project partners. Consultation on how to plan data collection was provided to partners by GIZ in yet another case which led to the initiation of a capacity building training undertaken by the partners themselves. Finally, while co-leading RIEs might strengthen the cooperation with partners, which is valuable in itself, it was mentioned that **co-leading RIEs may mitigate risks, especially in cases involving negative results.** Partners were assumed to be more willing to work on adjusting the project in case of negative results when they are deeply involved in the evaluation process. A strong ownership, again, promotes sustainability, which can be evidenced by the political intent to further adapt the tested intervention beyond GIZ support in some cases or by a nationwide take-up of the intervention after it proved to be effective in another project.

Even when RIEs are not co-led, it was reported in some interviews that RIE can strengthen the cooperation with political partners as the results can be used for policy recommendations. The results of an RIE were used for policy recommendations in one case in which providing consultation services to political partners is part of the project's service delivery as agreed upon with commissioning parties. Moreover, due to RIE findings, political partners scaled up the project's interventions to a national level in another case. Partners might be more confident in project activities based on RIE findings. Nonetheless, while some previous examples show that partnerships can still be strengthened even when RIEs are not co-led, **RIEs seemed to not influence partnerships in several of the cases analysed.** RIE findings were frequently presented together with findings concluded from monitoring data, but no further interaction occurred explicitly because of the RIE findings. Moreover, in some cases, the data produced by the RIE did not match with national data collected by the political partner, which stressed the partnerships of projects with political partners.

RIE findings have been presented and discussed at conferences of international development partners in two cases. These RIEs contributed to the improvement of internationally used indicators or checklists. Thus,

RIE might not only strengthen partnerships with the project partner but also with other development partners. Moreover, **in at least three cases, RIEs contributed to a further mobilisation of project resources provided by international, multilateral or national development partners.**

5.5 Cost effectiveness

RIE often requires considerable resources (time, effort and money). Against this background, it is important to compare an RIE's value against the resources invested. A follow-up questionnaire, which was filled by 15 out of 19 cases (79% response rate), indicated that **all project managers except one (93%) considered the (prospective) cost-effectiveness of the RIE to be rather high, thus assessing the RIE worth the cost and effort to implement it.** By contrast, 74% of respondents to the qualitative interviews of the 19 analysed cases indicated that RIEs were believed to be cost effective. This discrepancy can be explained by a combination two factors. On the one hand, the qualitative interviews were likely to have generated a better understanding for the value of RIE. Hence, when subsequently answering the questionnaire users were able to factor into their assessment additional value that their RIE had generated. On the other hand, the judgement by users is likely mainly driven by the 'absolute' value generated than the investigators' 'relative' judgement against the potential value. While RIE generated substantial value for its users, the review also displayed additional potentials.

The reasons to consider RIE highly cost-effective are manifold. The provision of specific external funding rendered the costs more likely such that RIE was evaluated as cost-effective: While **projects that financed RIE with their own resources reported to have spent at least 200.000€** on it, several projects had access to external funding, for instance by a sector programme or regional initiative. However, **not all RIEs with external funding were evaluated as cost-efficient by project managers.** The same holds for outcomes. While all three outcomes – the use of RIE for operational learning, conceptual learning and accountability – as well as strengthened partnerships were mentioned as reasons to rate the cost effectiveness of RIE as high, there are also cases in which the outcome was achieved yet the cost-efficiency rate was low. Therefore, there seems to be no single recipe to achieve high cost-effectiveness, as assessed by project managers.

However, this might also be due to the subjective nature of the variable itself and because of the different conditions under which RIE is conducted. **RIE is pre-dominantly conducted in a highly decentralised manner.** The project manager or an M&E specialist initiated RIEs in almost all cases as they were convinced these RIEs are useful for specific purposes. **For some of those project managers, RIEs provided the basis of an efficiency analysis**. They recognised RIE as the only way to attribute effects to project interventions and thereby to assess the value of an intervention compared to its costs. **In other cases, generating knowledge is part of a project's service delivery as agreed upon with the commissioning parties** (see chapter 5.2). Therefore, the cost-effectiveness of RIE is not only assessed by comparing the usefulness of the process and the findings to their costs but also by assessing whether RIE serves the goal of knowledge generation. Finally, **in those cases in which an RIE was designed to assess the impact of project pilots, the RIE cannot be separated from the project itself** such that the cost-effectiveness could not be assessed separately for the project and the RIE.

6 Conclusion

RIE offer significant corporate value for GIZ that has yet to be realised to its full extent. Nearly all investigated users at GIZ view RIE as a cost-effective measure. Moreover, the evidence clearly indicates that further substantial corporate value for GIZ has yet to be seized from RIE. On the one hand, additional value can be generated through better exploiting the already existing evidence-base established through RIE. On the other hand, a more strategic implementation of RIE in the future can substantially increase the utility of RIE for generating evidence to continuously increase effectivity of GIZ interventions.

The key conclusions of this review display that opportunities for adding corporate value through RIE systemically exist at the demand and supply side as well as within the enabling environment, as displayed by figure 10. Opportunities can be seized i) on the demand side by utilising RIE as a vehicle for strategic positioning with partners and ii) on the supply side, as a more purpose-sensitive use of RIE can substantially increase its utility. Systematically facilitating iii) the delivery of RIE as a service and iv) global support structures within GIZ can create a strengthened enabling environment for seizing demand and supply side opportunities.

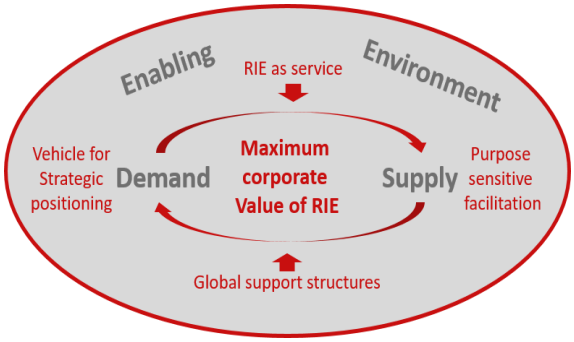

Figure 10: Maximum corporate value of RIE

RIE require purpose-sensitive facilitation. RIE can serve conceptual learning, operational learning and accountability. However, these purposes require different sets of conditions for their achievement. This is strikingly displayed by the different paths or sets of conditions which the QCA has identified for each purpose. Hence, purpose-sensitive facilitation that tailors RIE design as well as implementation strategy and monitors context conditions is critical for achieving its intended use. It is important to note that this, in turn, implies that not all purposes can be equally achieved, at least not to the same extent. For example, if a RIE measures the project's outcome or impact level, it may well serve conceptual learning and accountability, but is less likely to generate results that can be used for operational learning during the current project phase. So far, the use of nimble RIE designs, particularly for operational learning, is completely unexploited. None of the RIE cases identified at GIZ for the purpose of this review applied a nimble RIE design. Meanwhile, most cases did specify operational learning as an intended use. Among the cases that successfully used RIE for operational learning, a very limited number did so through using the RIE findings. On the one hand, this shows that process use can be an effective driver for operational learning. On the other hand, it also confirms that an increased RIE utility for operational learning can be gained through comprehensive facilitation of process use and use of findings.

Furthermore, in many cases, RIE is not strategically planned to address evidence gaps and to generate findings that allow for synthesis across multiple context. Cross-contextual synthesis, however, substantially increases the value of RIE, as it can help to reduce critical limitations to external validity and generalisation, respectively, inherent to RIE. Fostering cross-contextual synthesis particularly aids conceptual learning and accountability. However, the review findings suggest that the potential for synthesis across multiple contexts was only successfully exploited by one global programme. Whilst existing sectoral or regional initiatives bared the potential to generate comparable knowledge across different countries and projects, exploring the methodological approach appeared to be the key incentive to initiate RIE within the given sectors.

Global support structures are critical for effectively facilitating the corporate value of RIE. This is particularly evident for conceptual learning, which can only be achieved without central support in cases of highly replicable interventions or if results display the effectiveness of the intervention. However, systematic conceptual learning must not be limited to such cases, as this is likely to introduce bias. Effective conceptual learning depends on identifying all interventions of merit – not just already standardised, implemented and documented ones – for their corporate wide promotion. Meanwhile, excluding evidence on interventions for them not being successful also substantially diminishes conceptual learning. However, so far, RIEs are used almost exclusively in isolation. Even in case of similar RIEs that bare the potential for further exchange, results are not systematically discussed or analysed for cross-contextual learnings. Hence, already existing potential is not fully exploited. Moreover, a global support and coordination structure is pivotal to strategically plan RIE to address evidence gaps[9] in line with the corporate strategic needs and interest.

QCA results indicate that central support is less critical for operational learning and accountability. In case of operational learning central support can even hinder the use of RIE. This can be explained through the fact that central support units are deliberately promoting global interest that may hinder the primary project focus on the function of operational learning. However, users were generally not aware of the varying conditions that apply to different RIE uses. Independent of the intended RIE use, central support can substantially add value by helping users to clearly identify the intended RIE use from the start, and by advising on the relevant conditions for achieving it. Subsequently, a further engagement of the central support unit may depend on the identified, intended use.

RIE is best promoted as an integrated service delivery. Rather than conducting RIE exclusively for assessing the effects of services to partners, RIEs are seen more useful if they are provided as a service delivery to the project partners and are firmly integrated into the project design as agreed upon with commissioning parties and partners. Whilst bearing a considerable value to increase ownership and sustainability of results, it can equally mitigate risks, particularly, if the RIE indicates a negative intervention effect.

RIE present a vehicle for strategic positioning of GIZ. While demand is not yet systematic, a trend of increasing demand is evident on multiple levels. Among GIZ's commissioning parties demand for RIE remains mainly driven by individuals within the agencies. However, efforts to more systematically facilitate RIE are underway. For example, DEval has initiated a research project to support more adequate and systematic use of RIE within German development cooperation as far as both implementation of RIE and the utilisation of results (e.g., systematic reviews) are concerned. Some of the divisions at the German Federal Ministry for Economic Cooperation and Development (BMZ), such as the division for transitional development assistance, increasingly foster the systematic use (quasi-) experimental evaluation designs in German development cooperation. While demand among commissioning parties may not always be existent from the on-set of an RIE, the investigated cases have proven that this can very much change once data is available. Finally, the cases of RIE identified for this review illustrate that there is also an internal demand. Internationally this trend is particularly displayed by organisations that have so far made little use of RIE. Investments in RIE among these organisations primarily focus on the systemic use of RIE to complement the existing evaluation system. Hence, rather than increasing the scale of such evaluations, a highly strategic use of RIE is the main aim.

[9] Recent evidence gap maps have been conducted in the area of climate change adaptation (DEval/Green Climate Fund) and building peaceful societies (3ie/DEval/GIZ/BMZ). In the latter case, a systematic review is being commissioned by GIZ to consolidate results of the mapping exercise.

7 References

Befani, B. (2016): Pathways to Change: Evaluating Development Interventions with Qualitative Comparative Analysis (QCA). Expertgruppen för biståndsanalys (EBA). Rapport 05.

BMZ (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung) (2011): Micro-Methods in Evaluating Governance Interventions. BMZ Evaluation Division: Evaluation Working Papers. https://www.bmz.de/en/zentrales_downloadarchiv/erfolg/BMZ_WP_Micro.pdf [Accessed 01.04.2019].

Cameron D. B., Mishra A. & Brown A. N. (2015): The growth of impact evaluation for international development: how much have we learned?, Journal of Development Effectiveness 8 (1), S. 1-21. DOI: 10.1080/19439342.2015.1034156.

Funk E., Groß L., Leininger J. & von Schiller A. (2018): Lessons learnt from impact-oriented accompanying research – Potentials and limitations to rigorously assessing the impact of governance programmes. Bonn: German Development Institute [Accessed 01.08.2019].

GIZ (2018a): GIZ´s Evaluation System. General Description. Eschborn: Deutsche Gesellschaft für internationale Zusammenarbeit (GIZ) GmbH. https://www.giz.de/en/downloads/GIZ_EVAL_EN_general%20description.pdf [Accessed 01.04.2019].

GIZ (2018b): GIZ's evaluation policy. Principles, guidelines and requirements of our evaluation practice. Eschborn: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. https://www.giz.de/en/downloads/GIZ_EVAL_EN_evaluation%20policy.pdf [Accessed 01.04.2019].

GIZ (2018c): GIZ's evaluation system. Central project evaluations for BMZ business. https://www.giz.de/en/downloads/GIZ_EVAL_EN_ZPE_BMZ%20business.pdf [English version accessed 01.04.2019].

GIZ (2018d): GIZ's evaluation system. Theory of change for GIZ's evaluations. Eschborn: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. https://www.giz.de/en/downloads/GIZ_EVAL_EN_Theory%20of%20Change.pdf [Accessed 01.04.2019].

GIZ (2018e): GIZ Results Data. https://www.giz.de/en/aboutgiz/63962.html [Accessed 01.04.2019].

Grofman, B. & Schneider, C. Q. (2009): An Introduction to Crisp Set QCA, with a Comparison to Binary Logistic Regression. Political Research Quarterly 62 (4), S. 662-672.

Hearn, S. and Buffardi, A.L. (2016): 'What is impact?'. A Methods Lab publication. London: Overseas Development Institute.

Kluve, J. & Stöterau, J. (2014): A Systematic Framework for Measuring Aggregate Employment Impacts of Development Cooperation. Bonn: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

Krapp, S. & Dinges, S. (2012): Measuring Impacts Using Experimental Designs – and Their Applications at GIZ. Eschborn: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.

Legewie, N. (2013): An Introduction to Applied Data Analysis with Qualitative Comparative Analysis (QCA). FQS 14(3), Art. 15. http://www.qualitative-research.net/index.php/fqs/article/view/1961 [Accessed 01.04.2019].

Leppert, G., L. Hohfeld, M. Lech and T. Wencker (2018): Impact, Diffusion and Scaling-Up of a Comprehensive Land-Use Planning Approach in the Philippines. From Development Cooperation to National Policies, German Institute for Development Evaluation (DEval), Bonn.

Marshall, P. (2018): Evidence-Oriented Approaches in Development Cooperation. Experiences, Potential and Key Issues. DIE Discussion paper 08/2018. https://www.die-gdi.de/uploads/media/DP_8.2018.pdf [Accessed 01.04.2019].

Marx, A. & Dusa, A. (2011): Crisp-Set Qualitative Comparative Analysis (csQCA), Contra-dictions and Consistency Benchmarks for Model Specification. Methodological Innovations 2, S. 103-148.

Peersman G., Guijt I., & Pasanen, T. (2015): Evaluability Assessment for Impact Evaluation. A Methods Lab publication. London: Overseas Development Institute. https://www.odi.org/publications/9442-evaluability-assessment-impact-evaluation-guidance-checklists-and-decision-support [Accessed 01.04.2019].

Ragin, C. C. (2008): Redesigning Social Inquiry. Fuzzy Sets and Beyond. Chicago and London: University of Chicago Press.

Rogers P. (2014): Overview: Strategies for Causal Attribution. Methodological Briefs: Impact Evaluation 6. UNICEF Office of Research, Florence. https://www.unicef-irc.org/KM/IE/impact_6.php [Accessed 01.04.2019].

Schneider, C. Q. & Wagemann, C. (2006): Reducing complexity in Qualitative Comparative Analysis (QCA): Remote and proximate factors and the consolidation of democracy. European Journal of Political Research 45, S. 751-786.

Schneider, C. Q. & Wagemann, C. (2012): Set-Theoretic Methods for the Social Sciences. A Guide to Qualitative Comparative Analysis. Cambridge: Cambridge University Press.

Schwellnus, G. (2013): Eliminating the Influence of Irrelevant Cases on the Consistency and Coverage of Necessary and Sufficient Conditions in Fuzzy-Set QCA. Paper prepared for the 7th ECPR General Conference, Bordeaux, 4-7 September 2013. https://ecpr.eu/Filestore/PaperPro-posal/99d7fbea-b604-4251-847c-f8e33c82cb41.pdf [Accessed 01.04.2019].

White, H. & Sabarwal, S. (2014): Quasi-experimental design and methods. Methodological Briefs: Impact Evaluation, 8, S.1-16.

White, H. & Raitzer D. A. (2017): Impact evaluation of development interventions. A practical guide. Manila: Asian Development Bank.

8 Annexes

8.1 Glossary

Expression	Definition
Difference in Differences	Difference-in-differences (DID) is a quasi-experimental research design "also known as the 'double difference' method, compares the changes in outcome *over time* between treatment and comparison groups to estimate impact. DID gives a stronger impact estimate than single difference, which only compares the difference in outcomes between treatment and comparison groups following the intervention (at t+1). Applying the DID method removes the difference in the outcome between treatment and comparison groups at the baseline. […] The major limitation of the DID method is that it is based on the assumption that the indicators of interest follow the same trajectory over time in treatment and comparison groups. This assumption is known as the 'parallel trends assumption'. Where this assumption is correct, a programme impact estimate made using this method would be unbiased. If there are differences between the groups that change over time, however, then this method will not help to eliminate these differences." (White & Sabarwal 2014: 9, 10).
Formative Impact Evaluations	A Formative Impact Evaluation is understood in this review as an impact evaluation that focusses on causality using attribution approaches and is conducted at discrete point(s) of measurement during the implementation of an intervention. Hereby impact evaluations refer to the "[e]xamination of the degree to which development measures have brought about changes in an initial situation by analysing causal links and assumptions related to causal links and assessing causality in terms of the results logic (e.g. of a project)" GIZ (2018a: 16).
Instrumental Variable	The instrumental variables approach is a quasi-experimental method. "In simple terms, this approach involves identifying a special variable that affects outcome and intervention, but without the two having any causality. IV approach is needed if the causal relationship between intervention and outcome run the opposite direction. Such a scenario occurs when (i) the intervention has been deliberately targeted or (ii) participants have joined the programme for specific reasons. This implies that potential unobserved characteristics or omitted variables have not been taken into account." (BMZ 2011 :17)
(Quasi-)experimental methods	"(Quasi-)experimental research designs […] test causal hypotheses. In both experimental and quasi-experimental designs, the programme or policy is viewed as an 'intervention' in which a treatment – comprising the elements of the programme/policy being evaluated – is tested for how well it achieves its objectives, as measured by a pre-specified set of indicators. (Quasi-)experimental designs identify a comparison group that is as similar as possible to the treatment group in terms of baseline (pre-intervention) characteristics. The comparison group captures what would have been the outcomes if the programme/policy had not been implemented (i.e., the counterfactual). Hence, the programme or policy can be said to have caused any difference in outcomes between the

	treatment and comparison groups." (White & Sabarwal 2014: 1). For experimental methods see randomised control trials. "A quasi-experimental design by definition lacks random assignment, however. Assignment to conditions (treatment versus no treatment or comparison) is by means of self-selection (by which participants choose treatment for themselves) or administrator selection (e.g., by officials, teachers, policymakers and so on) or both of these routes." (White & Sabarwal 2014: 1)
Randomised Control Trials	An experimental "research or evaluation design with two or more randomly selected groups (an experimental group and control group) in which the researcher controls or introduces an intervention (such as a new programme or policy) and measures its impact on the dependent variable at least two times (pre- and post-test measurements). In particular RCTs – which originated in clinical settings and are known as the 'gold standard' of medical and health research – are often used for addressing evaluative research questions, which seek to assess the effectiveness of programmatic and policy interventions in developmental settings." (White & Sabarwal 2014: i).
Regression Discontinuity Design	Regression discontinuity design is a quasi-experimental design that "[…] can be used when there is some kind of criterion that must be met before people can participate in the intervention being evaluated. This is known as a threshold. A threshold rule determines eligibility for participation in the programme/policy and is usually based on a continuous variable assessed for all potentially eligible individuals. For example, students below a certain test score are enrolled in a remedial programme […] Those just either side of the threshold are not very different, however. If the threshold for being enrolled in a remedial study programme is a test score of 60, students enrolled in the programme who get a score of 58 to 59.9 are not very different from those who get a score of 60 to 60.9 and are not enrolled. Regression discontinuity is based on a comparison of the difference in average outcomes for these two groups." (White & Sabarwal 2014: 7)
Rigorous methods	GIZ believes that a rigorous approach includes not just experimental evaluation designs but also any methodological approach that systematically deals with the attribution of results to measures. These include experimental, quasi-experimental, statistical, theory-based and participatory approaches (GIZ 2018).
Output	Outputs describe the changes and conditions achieved by the project's activities, such as products, goods, services, acquired technical skills and knowledge, institutions and regulations.
Outcome	Outcomes are the probable or actually achieved short-term and medium-term changes that occur as a consequence of the use of an intervention's outputs. They correspond more or less to the achievement of the programme/project objective and include both the use of the outputs delivered by the project by the intermediaries and target groups, and the direct benefit.
Impact	Impacts are the long-term, overarching development results (is in line with OECD-DAC). They are usually located at the level of the DC programme objective.

8.2 Differences in the understanding of impact evaluations

Agency	Definition
GIZ	Examination of the degree to which development measures have brought about changes in an initial situation by analysing causal links and assumptions related to causal links and assessing causality in terms of the results logic (e.g. of a project). It offers a range of benefits, including the analysis of assumed causal mechanisms and the provision of information on the modification of results models and hypotheses. (2018: 16)
DFID	Impact evaluation approaches focus on establishing cause and effect. (…) Impact evaluation can be conducted for pilot initiatives (…). DFID-funded impact evaluations use an appropriate design which is able to measure impact credibly given the type and context of the intervention. This must include a counterfactual or similar analysis to address attribution and establish causality. It should be noted that where a randomised trial or use of a control group is not appropriate or feasible, alternative approaches can and should be used. (DFID, 2013: 13)
AusAID/ DFAT	[I]mpact evaluation links cause and effect: it assesses the direct and indirect causal contribution of the intervention to change in people's lives. (…) AusAID recognises that a range of methodological approaches can be used for impact evaluation. (2012: 3)
World Bank	IE [impact evaluation] assesses the causal effects (impacts) attributable to specific interventions, where the outcomes of interest are compared with a counterfactual situation. (…)IEs are structured to respond to one question in particular: the causal effect of a program on short-, medium-, or long-term outcomes. (IEG, 2012: 3)
USAID	Impact evaluations measure the change in a development outcome that is attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. (2016: 3)
3ie	A study of the attribution of changes in the outcome to the intervention. Impact evaluations have either an experimental or quasi-experimental design. (2012)
Better Evaluation	An impact evaluation provides information about the impacts produced by an intervention - positive and negative, intended and unintended, direct and indirect. This means that an impact evaluation must establish what has been the cause of observed changes (in this case 'impacts') referred to as causal attribution (also referred to as causal inference). (2018)

8.3 Methodological remarks to QCA use

Qualitative comparative analysis (QCA) is based on the assumption that different configurations of conditions may lead to the same outcome, e.g., that depending on the setting, different aspects of the design and the implementation of RIE may combine to produce a specific type of evaluation use. These configurations are called "paths to the outcome" (Schneider & Wagemann 2012, 330). The assumption that several paths may lead to the same outcome is called "equifinality" (Schneider & Wagemann 2012, 78). In its ability to identify several paths to the same outcome, QCA's approach to causal analysis is in stark contrast to that of statistical analyses (Grofman & Schneider 2009). While statistical procedures such as linear regression analyses reduce causal complexity to a single model that provides a partial explanation of an outcome across cases, QCA identifies parallel cause-to-effect chains that provide an exhaustive explanation of the outcome for all cases. Thus, statistical analyses derive net-effects of individual variables regardless of the context in which they operate, while QCA determines which causal conditions are decisive in which contexts (Ragin 2008, 182). The results of a configurational analysis are therefore particularly suitable for context-sensitive recommendations.

QCA's ability to identify sets of paths to an outcome rests on an asymmetrical understanding of causation. Instead of positing that a factor correlates with a dependent variable, QCA hypothesises that a factor is either a necessary or a sufficient condition for an outcome (or part of a configuration of conditions that is necessary or sufficient). If it is posited that a factor correlates positively with a dependent variable, it is assumed that the higher its value is, the higher is the value of the dependent variable, and the lower its value is, the lower is the value of the dependent variable. If the hypothesised correlation is negative, the inverse is expected, but in both cases, the posited relationship is the same for high and low values on both sides of the equation. Such a hypothesis is one of causal symmetry. QCA assumes a different kind of causation. Necessary and sufficient conditions are both forms of asymmetric causality. If, for example, it is hypothesised that the support of a central, specialised unit at GIZ in implementing an RIE is a sufficient condition for conceptual learning, this hypothesis does not imply that conceptual learning does not occur when the RIE was not centrally supported. Likewise, if it was posited that central support is a necessary condition for conceptual learning, this does not mean that conceptual learning could not be absent when central support is present.

A logical minimisation process identifies the configurations of conditions that consistently lead to the outcome among the set of conditions that feature in the explanatory model. In the present study, the resulting paths of each of the three QCA represent the circumstances under which RIE were used for operational learning, conceptual learning, or accountability. Together, these paths also represent an answer to the question why some RIE have provided cooperate value in the forms of evaluation use, while others have not. The robustness of this answer depends on the number of conditions that are included in the model, and the number of cases analysed (Marx & Dusa 2011). In order to exclude the possibility that the resulting paths of our QCA are a product of chance, the number of conditions was limited to five for each analysis. If five conditions are included in the configurational analysis of 19 randomly generated cases, the analysis leads in 93% of the cases to contradictory results (Befani 2016, 154). This means that if the three QCA of evaluation use in our study lead to logically coherent sets of sufficient configurations, these results are almost certainly not a product of chance (Marx & Dusa 2011).

A QCA thus represents a systematic and transparent cross-case analysis. The empirical basis for the QCA are detailed case studies. For the present study, each RIE, or each country in which an RIE was implemented, represents a case. These case studies consist of assessments for each condition and each outcome under study. Rules must be defined that allow researchers to perform assessments of each condition for each case according to the same criteria. Thus, each choice that a researcher makes in characterising a case is made explicit, and the analysis is retraceable for the reader (Legewie 2013).

The assessment of conditions differs for crisp-set QCA (csQCA) and fuzzy-set QCA (fsQCA), which are the two most common methods of formalised set-theoretic analysis (Schneider & Wagemann 2012, 13). Crisp-set QCA operates on sets with membership scores of either 0 (full non-membership) or 1 (full membership). Fuzzy-set QCA allows cases to have gradations of their set membership between 0 and 1. For example, in fsQCA, a case could receive one of four scores for its membership in a given set. 0.00 would express that the case is not a member, 0.33 that it is rather not a member, 0.66 that it is rather a member, and 1.00 that it is a full member of the set. Thus, a condition could be said to be fully absent, rather absent, rather present, or fully present in a given case.

In this review conditions have first been assessed on a four score scala, which was reduced to a two score scala for the analysis. That is, “fully absent” and “rather absent” were recoded as “absent,” while “rather present” and “fully present” were recoded as “present.” The case characteristics “Significantly positive results” and “Conceptual learning” are exceptions. Here, only “fully present” was recoded as “present”, while all other ratings were recoded as “absent”. Assessing conditions on a four score scala allows to reflect on which differences between the projects are regarded as the most important ones. However, csQCA is then used for the analysis as the additional value of fsQCA for explaining the different outcomes does not match the additional difficulties in analysis and interpretation. First, the added precision of the measurements does typically not lead to the identification of different paths to the outcome (Schwellnus 2013). Second, consistency and coverage of necessary conditions and sufficient paths have no easy interpretation in fsQCA. Whereas in csQCA a coverage value of a sufficient configuration of conditions can be translated directly into the number of cases that reach the outcome through that path, in fsQCA, cases have partial memberships in a configuration, and the coverage value represents the sum of these partial memberships. Given that familiarity with QCA is often low among the intended audiences of applied research, this represents an additional complication for the practical interpretation of results. Third, the formulae to calculate consistency and coverage in fsQCA include cases that are irrelevant for the hypothesised set relationship (Schwellnus 2013). Thus, consistency and coverage values may indicate that a condition is sufficient and that it covers a substantial part of the outcome, while there is no typical case confirming the assumption of sufficiency. Such erroneous evaluations of consistency and coverage are impossible with csQCA.

8.4 Short definitions of conditions

Condition	Definition	Expected sign		
		Operational learning	Conceptual learning	Accountability
Deep integration in monitoring framework	The implementation of the RIE started at the same time as the implementation of the intervention (or the component the RIE evaluated, respectively) or at the same time the monitoring system started operating, whichever came later.	+		+
Strong local support for RIE	Evaluators were continuously engaged with the project, or the project itself had a high capacity to support the RIE.	+		
Short-term oriented RIE	The project results the RIE investigated where below the level of a project objective.	+	-	-
Timely RIE	The RIE provided results at least half a year before the end of project implementation.	+		
Central support	The project received support from a specialised, central unit at GIZ in at least two of the following tasks: (1) Joint reflection of the suitability of RIE for the intervention and the intended purpose, (2) support in developing the ToR for the RIE, (3) selection of evaluators, and (4) briefing of evaluators.	+	+	
Demand for rigorous data	Commissioners of the GIZ or cooperation partners requested additional data on the project's contributions to results or on the attribution of results to the project at the beginning of the evaluation.		+	+
Replicability of intervention	The intervention was implemented in more than two countries, activities were specified regarding their specific format and content, and all beneficiaries in a country received identical activities.		+	
Significantly positive results	The results of the RIE indicated that the evaluated project or activities were overall successfully implemented and produced the intended results.		+	+

8.5 Definitions of conditions and thresholds

High project volume	
Project volume refers to the budget available for the project in the current intervention cycle. The more budget the project has available the higher the project volume.	
Fully absent	Budget is less than three million Euros.
Rather absent	Budget is between 3 to 6 million Euros.
Rather present	Budget is between 6 to 10 million Euros.
Fully present	Budget is higher than 10 million Euros.

High political relevance	
Highly politically relevant are projects that are part of the special initiatives “One World, No Hunger”; “Stability and development in the MENA region” and “Tackling the root causes of displacement, reintegrating refugees” as well as those within the title “Transitional development assistance”	
Fully absent	The project is not part of a special initiative/transitional development assistance
Fully present	The project is part of a special initiative/transitional development assistance

High number of project countries	
The investigated activity/activities are implemented in several countries in a comparable manner.	
Fully absent	The investigated activity/activities are implemented in one country.
Rather absent	The investigated activity/activities are implemented in two countries.
Rather present	The investigated activity/activities are implemented in three to five countries.
Fully present	The investigated activity/activities are implemented in more than five countries.

High degree of intervention replicability	
The degree of intervention replicability refers to the ability to reproduce activities in a comparable manner. This ability depends on the documentation of technical specificities of activities and is fostered when activities have been delivered identically to all beneficiaries.	
Fully absent	Activities are not specific **AND** not delivered identically to all beneficiaries that are supposed to receive a given type of activity.
Rather absent	Activities are specific at the level of the mechanism (e.g. type of training or type of agricultural innovation), but not delivered identically to all beneficiaries that are supposed to receive the respective activities.
Rather present	Activities are specific at the level of the mechanism (e.g. type of training or type of agricultural innovation) **AND** delivered identically to all beneficiaries that are supposed to receive the respective activities.
Fully present	Activities are fully specified through a handbook, technological blueprint or standard product **AND** delivered identically to all beneficiaries that are supposed to receive the respective activities.

Deep integration in monitoring framework	
RIE is an integral part of the monitoring and evaluation framework of the project. Synergies with results-based monitoring are sought and/or achieved. Note: In the case of conflicting views, information given by the evaluator is subsidiary to the perception of the project manager.	
Fully absent	RIE starts after the first year of project/component implementation **OR** the RIE and the monitoring systems are almost complete substitutes.
Rather absent	RIE starts after the beginning of project monitoring, but within the first year of project/component implementation.
Rather present	RIE starts at the same time as project/component implementation (and potentially after the monitoring system has been set up) **OR** RIE starts at the same time as the monitoring system is being set up (and potentially after project/component implementation has started).
Fully present	RIE is already mentioned in the impact matrix as a means to collect data on indicators **AND** starts at the same time as the project/component.

Continuous engagement of external evaluators	
To assess the continuity of the engagement of external evaluators in the evaluation, we distinguish three phases of the evaluation: (1) Design, (2) baseline data collection, (3) data analysis and reporting. Engagement can mean physical presence at the implementation site, personal exchanges with project stakeholders via telephone or video chat, or written consultation via, e.g., e-mail. The evaluator may vary with each phase. Note: In the case of conflicting views, information given by the evaluator is subsidiary to the perception of the project manager.	
Fully absent	External evaluators were engaged in the evaluation in only one phase of its implementation.
Rather absent	External evaluators are engaged in the evaluation in two of the three phases of its implementation, namely the design and the data analysis and reporting phase.
Rather present	External evaluators are engaged in the evaluation in all three phases.
Fully present	At least one external evaluator is embedded in the project team **OR** at least one evaluator is consulting with the project team between phases as well.

High capacity of project to support RIE	
A high capacity of the project to support the RIE means that there is adequate knowledge, work force, and leadership to support its implementation. It is implied that knowledge, work force, and leadership are not only available, but also used in support of the RIE. The project is considered to have “adequate knowledge” when project staff responsible for the evaluation have previous experience in commissioning and managing or in implementing an evaluation with a QED. The project is considered to have “adequate work force” if the GIZ staffer responsible for the project was not described as being strapped for time and resources. If work force is not explicitly mentioned by the interview partner when s/he is asked about challenges during the RIE implementation in general, an adequate work force is assumed. “Leadership” is assessed as adequate if it is not explicitly emphasised as challenging. Note: In the case of conflicting views, information given by the evaluator is subsidiary to the perception of the project manager.	
Fully absent	None of the three aspects (adequate knowledge, work force, and leadership) are present.
Rather absent	One aspect is present.
Rather present	Two of the three aspects are present.
Fully present	All three aspects (adequate knowledge, work force, and leadership) are present.

High degree of focus of the evaluation	
The objectives of an evaluation may cover an entire project/country package or focus on parts thereof. If an evaluation's objectives and questions concentrate on one specific project activity **OR** an area of project activities within one specific project component, a high degree of focus is present. Note: In the case of conflicting views, information given by the evaluator is subsidiary to the perception of the project manager.	
Fully absent	The project or country package has several components in which several activities are implemented **AND** the objectives and research questions of the RIE cover several activities in more than one of these components.
Rather absent	**EITHER:** The project or country package has several components in which several activities are implemented **AND** the objectives and research questions of the evaluation focus on one entire component. **OR**: The project or country package consists of several activities **AND** the research questions of the evaluation focus on all of these activities.
Rather present	**EITHER:** The project or country package has several components in which several activities are implemented **AND** the objectives and research questions of the evaluation focus on several activities of one component. **OR**: The project or country package consists of several activities **AND** the research questions of the evaluation focus on several of these activities.
Fully present	**EITHER:** The project or country package has several components in which several activities are implemented **AND** the objectives and research questions of the evaluation focus on one activity of one component. **OR (1)**: The project or country package consists of several activities **AND** the research questions of the evaluation focus on one of these activities. **OR (2):** The project or country package consists of one activity **AND** the research questions of the evaluation focus on this activity.

Short-term oriented evaluation	
The project results investigated by the evaluation can be located at different levels along the results chain. This also presents a time hierarchy. Direct results of activity are the most immediate. Changes on subsequent levels require additional time. Hence, the lower the level on which the result is located along the project´s results chain, the more short-term orientated the evaluation. Note: In the case of conflicting views, information given by the evaluator is subsidiary to the perception of the project manager.	
Fully absent	The results investigated are beyond the project objective (Modulziel).
Rather absent	The results investigated are a project objective (Modulziel).
Rather present	The results investigated are below the project objective (Modulziel) and beyond specific outputs, i.e. outcome at the project component level.
Fully present	The results investigated are immediate effects of activities, i.e. outputs.

Timely evaluation	
The timeliness of an evaluation is measured relative to the duration of the project. The more time in the intervention cycle remains after evaluation results become available, the timelier the evaluation is. The evaluation results do not necessarily have to be available in form of a final report, but can also be available in form of presentations, memos and the like. Note: In the case of conflicting views, information given by the evaluator is subsidiary to the perception of the project manager.	
Fully absent	The evaluation is an end-of-project or ex-post evaluation.
Rather absent	There is less than half a year between the time at which results of the evaluation become available and the end of project implementation.

Rather present	There is more than half a year between the time at which results of the evaluation become available and the end of project implementation.
Fully present	There is more than a year between the time at which results of the evaluation become available and the end of project implementation.

Comprehensive partner involvement in evaluation	
Comprehensive partner involvement in the evaluation is achieved when political partners have decision-making power with regard to the design, implementation, and use of the evaluation or if they are involved in the implementation of the RIE. It implies that political partners are in a positon to veto changes to the evaluation. Note: In the case of conflicting views, information given by the evaluator is subsidiary to the perception of the project manager.	
Fully absent	Partners have not been informed in the design phase of the RIE **OR** are not regularly briefed.
Rather absent	Partners were informed in the design phase of the RIE **AND** are regularly briefed about the progress of its implementation, as well as the further use of its results.
Rather present	The RIE is co-led by GIZ and partners. GIZ and partners jointly decide about design, implementation and the processing of results **OR** partners are responsible for parts of the implementation.
Fully present	The RIE is partner-led. Responsibility for design, implementation, and processing of results lies with partners.

Central support	
Central support refers to support received by the project and given by a specialised, central unit at GIZ during the commissioning phase of the evaluation. It comes mainly in the form of advice and quality assurance. Four tasks are differentiated: (1) Reflection of suitability of RIE for the intended purpose and intervention at stake; (2) development of Terms of Reference, (3) selection of evaluator(s) (including dissemination of ToR); and (4) briefing of evaluator(s) after selection.	
Fully absent	Project received no central support.
Rather absent	Project received central support in one of the tasks.
Rather present	Project received central support in at least two tasks.
Fully present	Project received central support in all four tasks.

Demand for rigorous data	
Demand for rigorous data refers to requests for additional or more rigorous data on project outputs, outcomes or impacts put forward by commissioners or cooperation partners of GIZ.	
Fully absent	There was no demand for more rigorous or additional data.
Rather absent	There was a demand for plausible additional data on project results, but no specific demand for more rigorous data.
Rather present	There was a demand for more rigorous or additional data on project results. This demand was predominantly for additional data on the project's contributions to results.
Fully present	There was a demand for more rigorous or additional data on project results. This demand was specifically for additional data on the attribution of results to the project.

Significantly positive results
The RIE produced positive **AND** statistically significant results regarding the project's outputs, outcomes or

impacts. Note: In the case of conflicting views, information given by the evaluator is subsidiary to the perception of the project manager.	
Fully absent	The RIE produced predominantly negative **AND** statistically significant results.
Rather absent	The RIE produced predominantly negative results that are not statistically significant.
Rather present	The RIE produced predominantly positive results that are not statistically significant.
Fully present	The RIE produced predominantly positive **AND** statistically significant results.

Use of RIE for operational learning	
The evaluation supported decision-making on the project level during the current intervention cycle of the project. Specifically, a use for operational learning is present when a decision by project management with regard to the implementation of the project has been influenced by insights generated in the process of the evaluation (process learning) or derived from the evaluation's findings and conclusions.	
Fully absent	Project implementation did not change during the current intervention cycle **OR** changes in project implementation did not take place in reaction to the evaluation process or its findings and conclusions.
Rather absent	Project implementation did not change during the current intervention cycle **OR** changes in project implementation did not take place in reaction to the evaluation process or its findings and conclusions. However, the evaluation has reportedly contributed to a changed understanding of the intervention and/or its impact.
Rather present	Project implementation has been affected by a management decision. In consequence, implementation of project activities changed. The decision was taken in reaction to the evaluation process.
Fully present	Project implementation has been affected by a management decision. In consequence, project implementation or outputs changed. The decision was taken in reaction to the findings and conclusions of the RIE.

Use of RIE for conceptual learning	
Conceptual learning refers to learning by GIZ. This is understood to have taken place when an exchange about the findings of the RIE has taken place within GIZ.	
Fully absent	The results of the RIE have not been disseminated.
Rather absent	An exchange has taken place, but it was exclusively about the methodological aspects of the RIE.
Rather present	An exchange about the findings of the RIE has taken place within GIZ, but not across countries.
Fully present	A cross-country exchange about the findings of the RIE has taken place within GIZ.

Increased accountability	
The evaluation supported the communication of development results to the commissioners of GIZ. Because the evaluation provided more rigorous data on project implementation or impacts, the perception of the project regarding its transparency and the quality of its reporting improved.	
Fully absent	The evaluation did not have a positive effect on the perceived transparency or quality of reporting of the project from the perspective of the commissioners of GIZ.
Rather absent	The evidence produced by the evaluation increased the ease of reporting for the project. This is indicated by the project's use of the evaluation's findings and conclusions in reports.

Rather present	Commissioners perceive that the data produced by the evaluation on the project's implementation or impacts significantly increased the amount of available evidence on the project.
Fully present	Commissioners perceive that the data produced by the evaluation on the project's implementation or impacts significantly improves the overall quality of the available evidence on the project.

8.6 Data tables

Table 2: Significance and replicability of intervention

ID	High project volume	High political relevance of project	Highly replicable intervention	High no. of implementing countries	High degree of standardization
1	1	0	0,33	0	1
2	1	0	0	0	0,33
3	1	0	0,33	0	0,66
4	1	0	1	1	1
5	1	1	0	0	0
6	0,33	0	0	0	0,33
7	0,66	0	0,33	0	0,66
8	1	1	1	1	1
9	0,66	1	1	1	1
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	0,33	0,66	0,33
13	1	1	0,33	1	0,33
14	1	1	0,33	1	0,33
15	1	1	1	1	1
16	1	1	0,33	1	0,33
17	0,33	0	1	1	1
18	0,33	0	1	1	1
19	0	0	1	1	1

Table 3: Deep integration of RIE in project

ID	Deep integration in monitoring framework	Strong local support for RIE	Continuous engagement of evaluators	High capacity of project to support RIE
1	0	0,33	0,66	0,33
2	0,33	1	1	0,66
3	1	0,33	0,33	0,66
4	0,66	0,66	0,33	1
5	0,66	0,33	0,33	0,66
6	1	0,66	0,33	1
7	0	1	1	0,66
8	1	0,33	0,66	0,33
9	1	0,33	0,66	0,33
10	1	1	0,66	1
11	1	1	0,66	1
12	0,66	1	0,66	1
13	0	1	0,66	1
14	0	1	0,66	1
15	0	0,33	0,66	0,66
16	0	1	0,66	1
17	1	1	1	0,66
18	0	0,33	0,66	0,33
19	1	1	1	0,66

Table 4: Management orientation of RIE

ID	High degree of focus of RIE	Short-term oriented RIE	Timely RIE
1	1	0,66	0,33
2	0,66	0,66	1
3	0	0	0,33
4	1	0,33	0,33
5	0	0,33	1
6	0,33	0,33	0,33
7	1	0,66	0
8	0	0,33	1
9	0,33	0,33	1
10	0	0,33	1
11	0	0,33	1
12	0,33	0,66	1
13	0,33	0,33	1
14	0,33	0,33	1
15	0,33	0,33	1
16	0,33	0,33	1
17	1	0,33	0,33
18	0,66	0,33	0,33
19	1	0,33	0

Table 5: Other aspects of RIE implementation

ID	Comprehensive partner involvement	Central support	Demand for rigorous data	Significantly positive results
1	0,66	1	0,33	1
2	0,66	0	1	1
3	0,66	0	1	1
4	0,66	1	1	0,66
5	0	1	1	0,66
6	0,66	1	1	0,66
7	0	0	0	1
8	0	1	1	1
9	0	1	1	1
10	0,33	1	1	1
11	0	1	1	0,66
12	0,33	0,33	0,33	1
13	0	0,66	0,33	0,66
14	0	1	0,33	0,66
15	0	1	0,33	0,33
16	0	1	0,33	0,66
17	0,66	0,33	1	1
18	0,33	1	1	0,66
19	0,66	0	0	0,33

Table 6: Types of evaluation use

ID	Operational learning	Conceptual learning	Accountability
1	0,66	0,66	0,33

2	0,33	1	0,33
3	0,66	1	1
4	0	1	0,33
5	0,66	1	0,33
6	0,33	1	0
7	0	0	0,33
8	1	1	1
9	1	1	1
10	1	1	1
11	1	1	0,66
12	0,66	0,66	0,33
13	0	0,33	0,33
14	0,33	1	0,33
15	0	0,33	0
16	0,33	0,33	0,33
17	0,66	1	1
18	0,33	1	0,33
19	0,66	1	0,33

Photo credits and sources

Photo credits/sources:
© GIZ / Ranak Martin, Carlos Alba, Dirk Ostermeier, Ala Kheir

URL-Verweise:
This publication contains links to external websites. Responsibility for the content not contain any URL links. The listed external sites always lies with their respective publishers. When the links to these sites were first posted, GIZ checked the third-party content to establish whether it could give rise to civil or criminal liability. However, the constant review of the links to external sites cannot reasonably be expected without concrete indication of a violation of rights. If GIZ itself becomes aware or is notified by a third party that an external site it has provided a link to gives rise to civil or criminal liability, it will remove the link to this site immediately. GIZ expressly dissociates itself from such content.

Maps:
The maps printed here are intended only for information purposes and in no way constitute recognition under international law of boundaries and territories. GIZ accepts no responsibility for these maps being entirely up to date, correct or complete. All liability for any damage, direct or indirect, resulting from their use is excluded.

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

Friedrich-Ebert-Allee 36 + 40
53113 Bonn, Deutschland
T +49 228 44 60-0
F +49 228 44 60-17 66

Dag-Hammarskjöld-Weg 1-5
65760 Eschborn, Deutschland
T +49 61 96 79-0
F +49 61 96 79-11 15

E info@giz.de
I www.giz.de