Rigorous Impact Evaluation: Evidence generation and take-up in German Development Cooperation

Krämer, Marion; Jechel, Lea; Kretschmer, Timo; Schneider, Elisabeth

Veröffentlichungsversion / Published Version
Forschungsbericht / research report

RIGOROUS IMPACT EVALUATION: EVIDENCE GENERATION AND TAKE-UP IN GERMAN DEVELOPMENT COOPERATION

Research Report

2021

DEval
GERMAN INSTITUTE FOR DEVELOPMENT EVALUATION

This report presents the findings of a BMZ-funded research project on rigorous impact evaluation (RIE) in German development cooperation (GDC). The research project examined the status quo of RIE, relevant barriers and potential measures to increase both the initiation of RIEs and the take-up of (rigorous) evidence.

GDC actors have implemented some RIEs and have used rigorous evidence in some cases. Yet, the approach to the implementation of RIEs and the take-up of rigorous evidence is not systematic so far. The report summarises a number of barriers hindering RIE initiation and take-up of (rigorous) evidence in GDC along a theory of change (ToC). The report also identifies a number of potential measures to address the barriers.

Our analysis revealed that barriers and potential measures are interdependent and often hierarchical. Therefore, we propose an integrated systemic approach to fostering the initiation of RIE and the take-up of rigorous evidence.

RIGOROUS IMPACT EVALUATION: EVIDENCE GENERATION AND TAKE-UP IN GERMAN DEVELOPMENT COOPERATION

Research Report

2021

IMPRESSUM

Authors

Dr Marion Krämer
Lea Jechel
Timo Kretschmer
Elisabeth Schneider

Responsible

Dr Martin Bruder

Design

MedienMélange:Kommunikation!, Hamburg
www.medienmelange.de

Editing

Jannet King

Photo credits

Cover: Rad Radu, Shutterstock

Bibliographical reference

Krämer et al. (2021), *Rigorous Impact Evaluation: Evidence generation and take-up in German development cooperation*, German Institute for Development Evaluation (DEval), Bonn.

ISBN 978-3-96126- 156-7 (PDF)

Published by

German Institute for Development Evaluation (DEval)
Fritz-Schäffer-Straße 26
53113 Bonn, Germany

Phone: +49 (0)228 33 69 07-0
E-Mail: *info@DEval.org*
www.DEval.org

The German Institute for Development Evaluation (DEval) is mandated by the German Federal Ministry for Economic Cooperation and Development (BMZ) to independently analyze and assess German development interventions.

This report can be downloaded as a PDF-file from the DEval website:
https://www.deval.org/en/publications

ACKNOWLEDGEMENTS

In its work on this report, the DEval team was supported by many individuals and organisations. We would like to express our cordial thanks to all of them.

We would like to thank the many employees and experts in Germany and its partner countries whom we interviewed and who completed our online surveys. Their knowledge and experiences were indispensable in this research. The consultancy firms Scio and Sattva provided an insightful analysis of international experiences regarding the institutionalisation of rigorous impact evaluation. Birte Snilstveit (3ie), Howard White (Campbell Collaboration), Laurenz Langer (Africa Centre for Evidence), Tatjana Till and Andrew Kaiser-Tedesco (both GIZ) and Jochen Kluve, Jörg Langbein and Luciane Lenz (all KfW) provided external peer reviews that greatly helped to improve the report. The same is true for our DEval colleagues and internal peer reviewers Jörg Faust, Miriam Amine and Sven Harten.

We would like to thank members of the thematic team on RIE for their feedback on initial results of our research project.

Finally, we thank our colleagues Christian Richter, Anna Grimminger and Constantin Grywatz for their support throughout the research process.

EXECUTIVE SUMMARY

Introduction

This report presents the findings of a research project on rigorous impact evaluation (RIE) in German development cooperation (GDC), funded by the Federal Ministry for Economic Cooperation and Development (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, BMZ). RIEs have the potential to contribute to a more effective and efficient development cooperation by answering questions about which interventions work – and to what extent – and which do not. The research project examined the status quo of RIE, and relevant barriers and potential measures to increase both the initiation of RIEs and the take-up of (rigorous) evidence.

Method

The study employed a mixed-methods research design combining quantitative and qualitative methods of data collection and analysis. It drew on six different sources of data: interviews, a stocktaking survey of existing RIEs in GDC, an evidence survey among all development professionals in GDC, a review of international RIE experiences, a portfolio and document analysis, and a literature search. We triangulated our findings across the different data sources, methods and researchers within the team.

Findings

Research question 1: What is the status quo of (a) the initiation of RIE and (b) the take-up of RIE evidence in GDC?

GDC actors have implemented some RIEs and have used rigorous evidence in some cases. However, there is clear potential for more RIEs, more systematic implementation of RIEs, and a greater take-up of evidence resulting from such evaluations. There have been at least 97 RIEs in GDC since 2014. Although this is more than anticipated, the overall incidence of RIEs in the German Official Development Assistance (ODA) portfolio is low. The approach to implementing RIEs is not yet systematic. It is not embedded in an overarching learning strategy and does for instance not align with ODA flows or other possible indicators of relevance (Section 3.1). Roughly the same is true for evidence take-up. Whereas our data indicate that organisations often use evidence from their own RIEs at the project level, these RIEs do not seem to be used for strategic decision-making. This is despite GDC employees' positive view on evidence-informed decision-making. Project documents outline causal links between the intervention and its intended impacts or outcomes, yet these are rarely backed up by rigorous evidence. In addition, the global RIE evidence base is not systematically used or consulted when designing projects (Section 4.1).

Research question 2: What are existing barriers to (a) the initiation of RIE and (b) the take-up of RIE evidence in GDC?

The report summarises a number of barriers hindering RIE initiation (Section 3.2) and take-up of rigorous evidence in GDC (Sections 4.2.1 and 4.2.2) along a theory of change (ToC). These barriers are of varied relevance in hindering RIE initiation and evidence take-up. The highly relevant barriers for initiating RIE are, amongst others, unclear benefits of RIE and uncoordinated timing between RIEs and project cycles. The highly relevant barriers for evidence take-up are, amongst others, low priority of evidence for decision-making and scarce and unsystematic dissemination of findings. Further, barriers often interact with each other.

Research question 3: What are potential measures for overcoming barriers for (a) the initiation of RIE and (b) the take-up of RIE evidence in GDC?

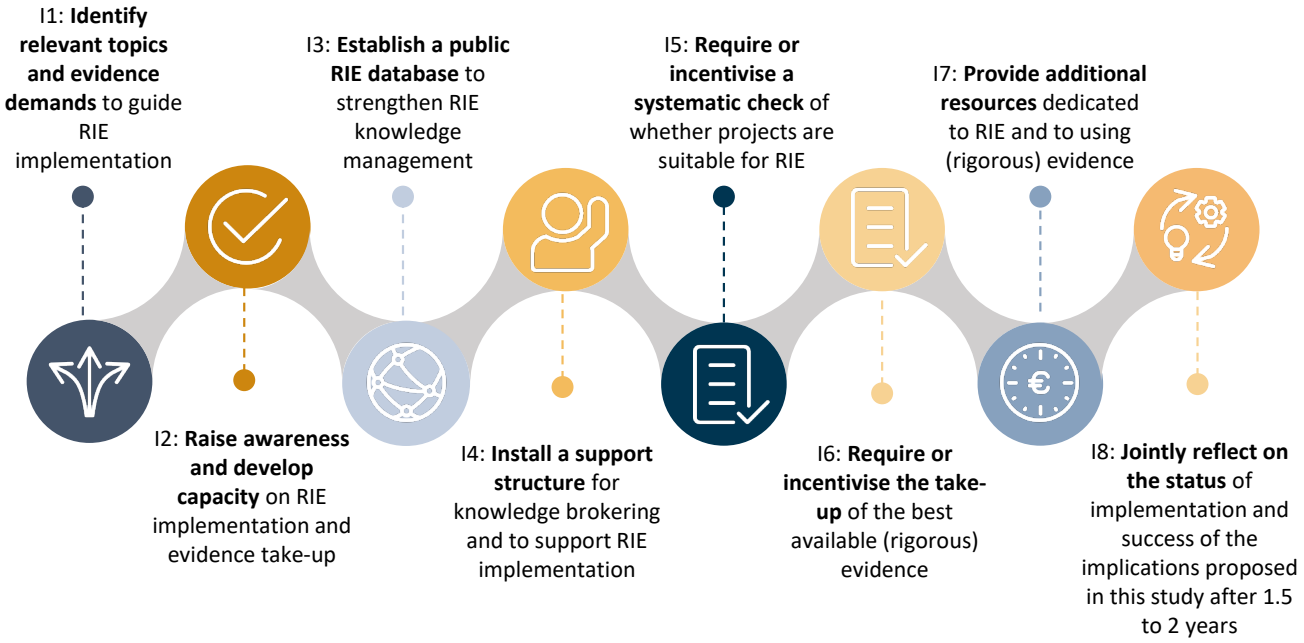
The report identifies a number of potential measures that may help to address the barriers. Some of these have already been tested in practice, either within GDC or by other development partners (Sections 3.3 and 4.3).

When **linking highly relevant barriers and potential measures**, we observe that for most of the barriers several potential measures exist (Sections 3.4 and 4.4). For instance, the low priority given to RIE evidence can be addressed not only through capacity development but through formal requirements or the definition of (more) relevant evaluation questions. Conversely, one potential measure can address a number of different barriers. Yet, we did not identify any "silver bullet" that alone will suffice in addressing all different barriers to RIE initiation or take-up, nor the most relevant barriers.

Implications

Our analysis revealed that barriers and potential measures are interdependent and often hierarchical. Therefore, we propose an integrated systemic approach to fostering the initiation of RIE and the take-up of rigorous evidence. We have identified a set of eight **implications** (see Figure 1), illustrated below, that are intended to trigger a systemic effect.

Figure 1 Implications



Source: own illustration

ZUSAMMENFASSUNG

Einleitung

Dieser Bericht präsentiert die Ergebnisse eines Forschungsprojektes zu rigoroser Wirkungsevaluierung (rigorous impact evaluation, RIE) in der deutschen Entwicklungszusammenarbeit (EZ), das durch das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) finanziell gefördert wurde. Da RIE Auskunft darüber geben, welche Interventionen wirken – und in welchem Umfang – und welche nicht, können sie zu einer effektiveren und effizienteren Entwicklungszusammenarbeit beitragen. Das Forschungsprojekt untersuchte den Status quo von RIE, relevante Hürden sowie potenzielle Maßnahmen, um sowohl die Initiierung von RIE als auch die Nutzung von (rigoroser) Evidenz zu erhöhen.

Methoden

Für die Studie wurde ein Mixed-Methods-Forschungsdesign angewandt, das quantitative und qualitative Methoden der Datenerhebung und -analyse kombiniert. Hierfür wurden sechs Datenquellen herangezogen: Interviews mit nationalen und internationalen EZ-Fachkräften, eine Bestandsaufnahme durchgeführter RIE in der deutschen EZ, eine Online-Befragung unter allen Fachkräften der deutschen EZ, eine Analyse internationaler RIE-Erfahrungen, eine Portfolio- und Dokumentenanalyse sowie Literaturrecherchen. Die Erkenntnisse wurden über die verschiedenen Datenquellen, Methoden und Mitglieder des Forschungsteams trianguliert.

Ergebnisse

Forschungsfrage 1: Wie stellt sich der Status quo (a) der Initiierung von RIE und (b) der Nutzung von RIE-Evidenzen in der deutschen EZ dar?

Akteure in der deutschen EZ haben einige RIE durchgeführt und rigorose Evidenzen wurden vereinzelt verwendet. Es besteht jedoch ein deutliches Potenzial für mehr RIE-Durchführungen, eine systematischere Initiierung von RIE und eine stärkere Nutzung der aus RIE gewonnenen Ergebnisse. Seit 2014 wurden mindestens 97 RIE in der deutschen EZ durchgeführt. Wenngleich dies mehr ist als erwartet, kann diese Anzahl von RIE gemessen an den deutschen sogenannten ODA-Mittel (official development assistance, ODA) als gering bewertet werden. Es gibt keine systematische Herangehensweise für die Initiierung und Durchführung von RIE. Die RIE-Initiierung folgt derzeit keiner übergeordneten Lernstrategie und orientiert sich beispielsweise nicht an ODA-Strömen oder anderen Relevanzindikatoren (Abschnitt 3.1). Ein ähnliches Bild zeigt sich hinsichtlich der Nutzung von Evidenzen. Wenngleich unsere Analysen darauf hindeuten, dass Organisationen die Erkenntnisse aus ihren eigenen RIE auf Projektebene oftmals nutzen, scheinen diese RIE nicht für die strategische Entscheidungsfindung verwendet zu werden. Dies steht entgegen der positiven Einstellung der deutschen EZ-Fachkräfte zu evidenzinformierter Entscheidungsfindung. In Projektdokumenten werden kausale Zusammenhänge zwischen der Intervention und den angestrebten Wirkungen auf Impact- oder Outcome-Ebene dargelegt, die jedoch in der Regel nicht durch rigorose Evidenzen untermauert werden. Darüber hinaus werden internationale RIE-Evidenzen in Projektdokumenten nicht systematisch konsultiert und beachtet (Abschnitt 4.1).

Forschungsfrage 2: Welche Hürden behindern (a) die Initiierung von RIE und (b) die Nutzung von RIE-Evidenzen in der deutschen EZ?

Die Studie identifiziert eine Reihe von Hürden, die der Initiierung von RIE (Abschnitt 3.2) und der Nutzung rigoroser Evidenz in der deutschen EZ (Abschnitte 4.2.1 und 4.2.2) entgegenstehen. Die Hürden werden anhand einer so genannten Theorie des Wandels (ToC) dargestellt. Die Analyse zeigt, dass Hürden die Initiierung von RIE und die Nutzung von Evidenzen in unterschiedlichem Ausmaß behindern. Die relevantesten Hürden für die Initiierung von RIE sind unter anderem der unklare Nutzen von RIE und eine unkoordinierte zeitliche Abstimmung zwischen RIE und Projektzyklen. Die relevantesten Hürden für die Nutzung von Evidenzen sind unter anderem die geringe Priorität von Evidenzen bei Entscheidungsprozessen und die geringe und unsystematische Disseminierung von Evidenzen. Zudem stehen Hürden häufig in Wechselwirkung zueinander.

Forschungsfrage 3: Welches sind potenzielle Maßnahmen, um die Hürden für (a) die Initiierung von RIE und (b) die Nutzung von RIE-Evidenzen in der deutschen EZ zu überwinden?

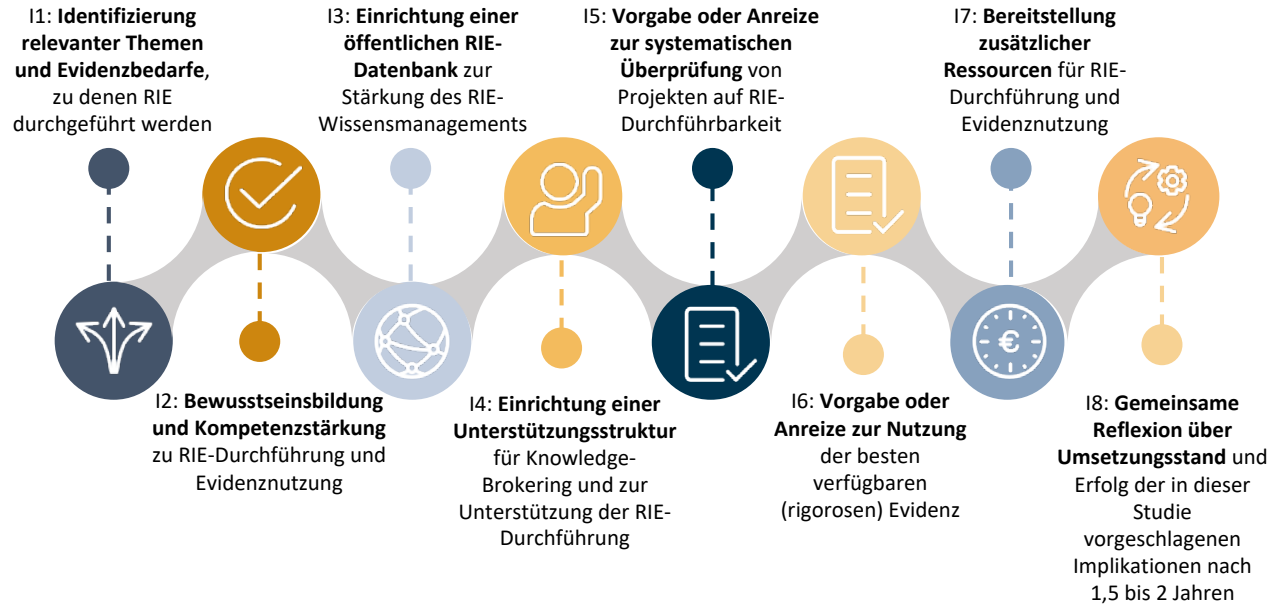
Die Studie identifiziert eine Reihe potenzieller Maßnahmen, die zur Überwindung der Hürden beitragen können. Einige davon wurden bereits in der Praxis erprobt, entweder innerhalb der deutschen EZ oder von anderen internationalen Entwicklungspartnern (Abschnitt 3.3 und 4.3).

In der **Gegenüberstellung der relevantesten Hürden mit den potenziellen Maßnahmen** zeigt sich, dass die meisten Hürden durch mehrere potenzielle Maßnahmen adressiert werden können (Abschnitte 3.4 und 4.4). So kann beispielsweise die geringe Priorität, die RIE-Evidenzen eingeräumt wird, nicht nur durch Kapazitätsentwicklung, sondern auch durch formale Anforderungen oder die Definition (relevanterer) Evaluierungsfragen adressiert werden. Umgekehrt kann eine potenzielle Maßnahme eine Reihe verschiedener Hürden adressieren. Gleichwohl haben wir kein „Allheilmittel" identifiziert, das allein alle verschiedenen oder gar nur die relevantesten Hürden für die Initiierung von RIE und die Nutzung von (rigorosen) Evidenzen überwinden kann.

Implikationen

Unsere Analyse zeigt, dass Hürden und potenzielle Maßnahmen voneinander abhängig und oft hierarchisch miteinander verknüpft sind. Daher regen wir einen integrierten systemischen Ansatz an, um die Initiierung von RIE und die Nutzung von rigorosen Evidenzen zu fördern. Die folgende Kombination von acht Implikationen (siehe Abbildung 1) zielt darauf ab, diesen systemischen Wandel anzustoßen.

Abbildung 1 Implikationen



Quelle: Eigene Darstellung

CONTENTS

Figures

Boxes

ABBREVIATIONS AND ACRONYMS

3ie	International Initiative for Impact Evaluation
ADB	Asian Development Bank
AFD	*Agence française de développement*
AV	Project manager (*Auftragsverantwortliche*r*)
BHO	Federal Budget Regulations (*Bundeshaushaltsordnung*)
BMZ	Federal Ministry for Economic Cooperation and Development (*Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung*)
C4ED	Center for Evaluation and Development (*Zentrum für Evaluation und Entwicklungsforschung*)
CGD	Centre for Global Development
CRS	Creditor Reporting System
CSO	civil society organisation
DEP	(3ie's) Development Evidence Portal
DEval	German Institute for Development Evaluation (*Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit*)
DFID	Department for International Development
DiD	difference-in-differences
DIE	German Development Institute (*Deutsches Institut für Entwicklungspolitik*)
DIME	Development Impact Evaluation (World Bank Group)
EAP	East Asia and Pacific
ECD	evaluation capacity development
EGM	evidence gap map
FCDO	Foreign, Commonwealth & Development Office (formerly DFID)
GDC	German development cooperation
GIZ	*Deutsche Gesellschaft für internationale Zusammenarbeit*
GVR	Joint procedural reform (*Gemeinsame Verfahrensreform*)
IO	implementing organisation
IDB	Inter-American Development Bank
IPA	Innovations for Poverty Action
J-PAL	The Abdul Latif Jameel Poverty Action Lab
KfW	*Kreditanstalt für Wiederaufbau*
LAC	Latin America and the Caribbean
M&E	monitoring and evaluation
MCC	Millennium Challenge Corporation
MENA	Middle Eastern and North African
NONIE	Network of Networks for Impact Evaluation
NORAD	Norwegian Agency for Development Cooperation
ODA	Official Development Assistance
OECD	Organisation for Economic Co-operation and Development

PEGNet	Poverty Reduction, Equity and Growth Network
PSM	propensity score matching
RIE	rigorous impact evaluation
RCT	randomised controlled trial
SFF	Studies and Experts Fund (*Studien- und Fachkräftefonds*)
SR	systematic review
SSA	sub-Saharan Africa
ToC	theory of change
ToR	terms of reference
UN	United Nations
USAID	United States Agency for International Development
WACIE	West Africa Capacity-building and Impact Evaluation
WWN	What Works Network

GLOSSARY

Term	Definition
Culture of learning from failure	There is no universal definition of a culture of learning from failure (sometimes shortened to "learning culture"). However, there are several concepts that most definitions share: within a culture of learning from failure, (1) learning is integrated into the systems, values and resources of an organisation and allowed to shape strategy and process, (2) learning is promoted and rewarded at an individual, team and organisational level so that individuals feel empowered to fail and learn from mistakes and (3) leaders embrace open dialogue and reflection. Generally, a culture of learning from failure applies to the individual, team, organisational and inter-organisational level (CPID, 2020).
Evidence	By evidence, we mainly refer to studies, evidence syntheses, such as systematic review (SR) and evidence gap map (EGM), and evaluation reports, that generate evidence using scientific standards and methods.
Evidence culture	A culture of evidence means that evidence is valued, understood and taken-up within the system or organisation. This implies that evidence plays an important role in policy decisions. A culture transcends political ideology and different leaderships (Carter et al., 2018; Cochrane, 2021).
Evidence gap map (EGM)	Evidence gap maps (EGMs) provide a visual representation of scientific studies, often SRs and RIEs, in a sector or thematic area. They are structured around a framework of interventions and outcomes. By mapping evidence in this framework, evidence gaps, in which few or no studies exist, become visible. Likewise, areas with a higher concentration of evidence are identified (Snilstveit et al., 2017).
Evidence-informed decision-making (EIDM)	Evidence-informed decision-making (EIDM) is an operating principle for public officials to use the best available scientific evidence when designing policies and programmes or defining institutional strategies. We make a distinction between evidence-*informed* decision-making from evidence-*based* decision-making. Whereas in EIDM, evidence is used to inform decision-making alongside other sources of information, evidence-based decision-making describes an approach in which findings are transferred one-to-one to projects, programmes or strategies (Head, 2016; Stewart et al., 2018).
Evidence take-up	We define evidence take-up as the use of evidence when making decisions at project and organisational level for both operational and strategic decision-making.
Experimental design	In an experimental design, often called randomised controlled trial (RCT), units of observation from a population of interest (e.g. households) are randomly assigned to two groups: (1) the intervention group, which experiences a development intervention, (2) the control group, which does not experience the intervention. Random assignment of a sufficiently large number of observational units (e.g. households) ensures that the two groups

Term	Definition
	can be expected to be identical in terms of both their observable and non-observable characteristics prior to the intervention. The difference in the outcome of interest (e.g. household income) between the two groups after the intervention thereby represents an undistorted estimate of the true effect of the intervention (e.g. the microloan) (Duflo et al., 2007; Gertler et al., 2016). Together with quasi-experimental designs, experimental designs form the group of rigorous impact evaluation designs.
GIZ global project	A global project is an instrument of bilateral German development cooperation commissioned by BMZ and implemented by GIZ. Global projects implement (often similar) interventions in several partner countries and regions, together with on-site local partners. Also called “global programme”.
Global RIE evidence base	We understand the “global RIE evidence base” to comprise RIEs that have been conducted by other projects within one’s own organisation or by other organisations globally, including development partners or other donor countries. This includes peer-reviewed and rigorous grey literature.
GVR (= BMZ’s joint procedural reform)	The planning and implementation of bilateral projects follows mandatory procedural rules which are documented in internal guidance documents. A comprehensive reform of procedures (Gemeinsame Verfahrensreform, GVR) entered into force in 2017. The procedural rules defined during this revision process specify which steps must be taken by BMZ and the respective implementing organisation throughout the project cycle (Syspons, 2018).
Quasi-experimental design	An RIE design used to measure the impact of an intervention when assignment to the treatment group cannot be randomised. Quasi-experimental designs identify or statistically construct a comparison group whose baseline characteristics are as similar to the characteristics of the treatment group as possible. The comparison group captures the counterfactual outcomes, i.e. the outcomes that would have occurred in the treatment group if the intervention had not been implemented. Thereby, quasi-experimental designs allow the determination of the causal impact of an intervention with high confidence (Gertler et al., 2016; White and Sabarwal, 2014). Together with experimental design, quasi-experimental designs form the group of rigorous impact evaluation designs.
Randomised controlled trial (RCT)	See “Experimental design”.
Rigorous impact evaluation (RIE)	RIE is an evaluation approach comprising impact evaluation designs that allow the causal attribution of a mean change in an outcome of interest (e.g. household income) to a specific intervention (e.g. a microloan). To do so, it is necessary to compare what actually happened with the so-called “counterfactual situation”. This involves, for example, comparing incomes of households that received microloans with what the incomes of the very same households would have been if they had not received the microloan.

Term	Definition
	Because such an observation is logically impossible, the counterfactual condition is approximated using experimental and quasi-experimental study designs. Experimental designs include different variations of randomised controlled trials (RCTs); quasi-experimental designs include regression discontinuity designs, different matching techniques, difference-in-differences estimation, interrupted time series, instrumental variable approaches and fixed effects models (Bruder et al., 2019).
Systematic review (SR)	Systematic reviews (SR) synthesise the best available (rigorous) evidence on a specific research question. They have clear criteria for inclusion and exclusion of primary studies, an explicit and transparent search strategy, systematic procedures for data extraction, and offer a critical appraisal and analysis of included studies. By statistically integrating the results of a number of quantitative studies, a meta-analysis identifying average effects of interventions across large numbers of studies can be carried out and included in SRs. In cases where the evidence base is large enough, SRs give the best possible, generalisable statements about what is known about interventions. SRs can therefore reliably inform decision-makers which interventions work and why (Waddington et al., 2012).

1. INTRODUCTION

1.1 Background and motivation

> Each year billions of dollars are spent on thousands of programs to improve health, education and other social sector outcomes in the developing world. But very few programs benefit from studies that could determine whether or not they actually made a difference. This absence of evidence is an urgent problem: it not only wastes money but denies poor people crucial support to improve their lives. (Savedoff et al., 2006)

This quote comes from the pioneering report *When Will We Ever Learn?*, published by the Centre for Global Development (CGD) in 2006. The authors make a strong case for the need for rigorous impact evaluations (RIEs) to be conducted more frequently in international development. RIEs can answer questions about which interventions work – and to what extent – and which do not, enabling limited resources to be invested effectively and efficiently. RIE is an evaluation approach that aims to establish causality between an intervention and an outcome of interest. To achieve this, the change in an outcome of interest in a treatment group is compared to the change of the same outcome in a control or comparison group. The control or comparison group is constructed in such a way that it can be expected to be as similar as possible to the intervention group. Box 1 describes how experimental and quasi-experimental designs can be used to construct control or comparison groups.

Box 1 What is rigorous impact evaluation (RIE)?

RIE is an approach comprising impact evaluation designs that allow the causal attribution of a mean change in an outcome of interest (e.g. household income) to a specific intervention (e.g. a microloan). To do so, it is necessary to compare what actually happened with the so-called "counterfactual situation". This involves, for example, comparing incomes of households that received microloans with what the incomes of the very same households would have been if they had *not* received the microloan.

Because such an observation is logically impossible, the counterfactual condition is approximated using experimental and quasi-experimental study designs. Experimental designs include different variations of randomised controlled trials (RCTs); quasi-experimental designs include regression discontinuity designs, different matching techniques, difference-in-differences estimation, interrupted time series, instrumental variable approaches and fixed effects models.

In an RCT, units of observation from a population of interest (e.g. households) are randomly assigned to two groups: (1) the intervention group, which experiences a development intervention, (2) the control group, which does not experience the intervention. Random assignment of a sufficiently large number of observational units (e.g. households) ensures that the two groups can be expected to be identical in terms of both their observable and non-observable characteristics prior to the intervention. Thereby the difference in the outcome of interest (e.g. household income) between the two groups after the intervention represents an undistorted estimate of the true effect of the intervention (e.g. the microloan). Quasi-experimental methods do not use randomisation but instead apply other specific study designs or statistical methods to estimate the counterfactual situation. For a quick introduction to RIE also see this video: *https://youtu.be/2iVqBhooeA8.*

There are two discussions in the evaluation literature, in which we would like to situate RIE. First, there sometimes is confusion about whether RIE is an evaluation approach, an evaluation design, or an evaluation method (amongst other terms). Following Lemire, Peck and Porowski (2020), we consider RIE to be an evaluation *approach* that is mainly concerned with general questions of valid causal attribution. An experimental or quasi-experimental evaluation *design* specifies concrete strategies of how to provide answers to evaluative questions in the particular context of an evaluation. The evaluation design employs one or more evaluation *methods* (e.g. surveys) to collect empirical information.

Second, there is increasing emphasis on understanding not only whether an intervention has an effect, but also determine the way in which it produces the effect and under what circumstances it does so. When answering such questions, RIE are often most insightful if they are theory-based and are combined with qualitative components (White, 2009).

Due to the causal attribution of outcomes of interest to interventions, RIEs provide a unique opportunity to learn what works. When systematically taking up results of RIEs, decision makers can learn whether interventions actually caused the intended outcomes and impacts and therefore contributed to a more effective and efficient development cooperation. Accordingly, we define evidence take-up as the use of (rigorous) evidence when making operational or strategic decisions at a project or organisational level. Hence, all decision makers in government, implementing organisations (IOs), and civil society organisations (CSOs) can be engaged in evidence take-up. There is an urgent need for the take-up of results of RIEs, given that it is estimated that 80% of development programmes do not work (White, 2019). Without the use of rigorous evidence, funds are likely to be spent less effectively and efficiently than they could be.

Fortunately, decision makers do not have to rely solely on generating rigorous evidence within their own organisation – they can also learn from the findings that others have observed for similar interventions in similar contexts. The findings of a growing number of RIEs are collected in evidence databases and can often be accessed and used. Systematic evidence take-up can draw on this global public good, organised, for example, in 3ie's Development Evidence Portal (DEP). Evidence take-up is facilitated by synthesis products such as systematic reviews (SRs) and evidence gap maps (EGMs). Box 2 provides further details of these products.

Box 2 In short: SRs and EGMs

What is an SR?

Systematic reviews (SR) synthesise the best available (rigorous) evidence on a specific research question. They have clear criteria for inclusion and exclusion of primary studies, an explicit and transparent search strategy, systematic procedures for data extraction, and offer a critical appraisal and analysis of included studies. By statistically integrating the results of a number of quantitative studies, a meta-analysis identifying average effects of interventions across large numbers of studies can be carried out.

What is an EGM?

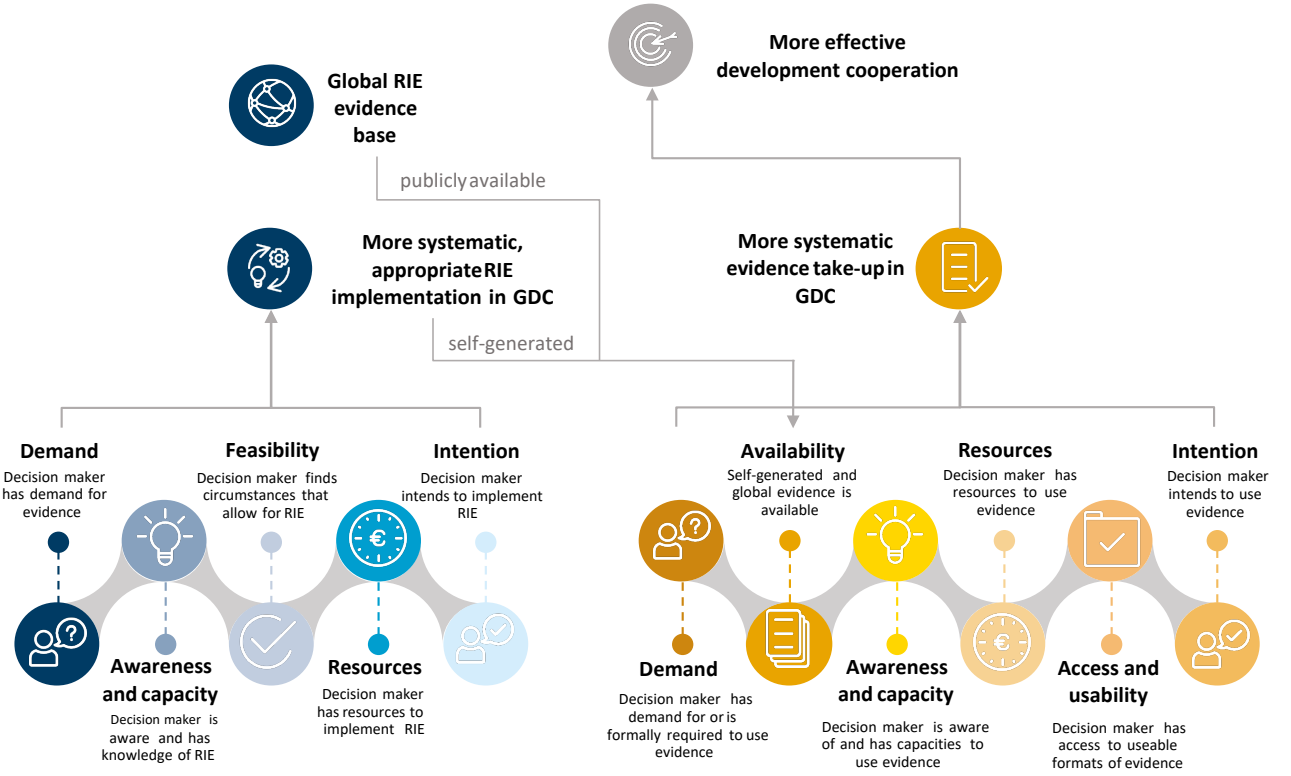
Evidence gap maps (EGMs) are compilations of (rigorous) evidence on the impact of policies and programmes in a specific sector or thematic area. EGMs provide a visual representation of SRs and primary studies (usually RIEs) available in a certain sector or field, structured around a framework of interventions and outcomes.

Source: 3ie (2017) and 3ie (2021)

Figure 2 depicts a simplified theory of change (ToC) that illustrates a potential causal chain from the generation of rigorous evidence through the take-up of evidence to a more effective development cooperation centred on the political or project decision maker.

The ToC has two main components. On the left-hand side, the ToC outlines important steps on the way to initiating an RIE. Decision makers need to be receptive to evidence, they need to be aware and have sufficient knowledge of RIE, they need to establish the feasibility of conducting an RIE, including having the resources to do so, and they need to put their intention into practice. This self-generated evidence, as well as publicly available evidence of RIEs from other organisations, feeds into the right-hand side of the ToC, which outlines important steps towards a systematic evidence take-up. Here, relevant evidence needs to be available and decision makers require awareness, capacities and resources to use it. The formats in which evidence is available need to be easily accessible and usable and, again, decision makers need to intend to use the available evidence. At the impact level, the systematic evidence take-up should lead to a more effective development cooperation in the long-term.

Figure 2 Theory of change



Source: own illustration

A better use of evidence has been called for in previous reports, such as in *When Will We Ever Learn?*, quoted above. But haven't we learned anything yet? Today – 15 years after the publication of that CGD report – the international evidence landscape has changed significantly. The number of RIEs has increased continuously. 3ie's DEP has identified fewer than 100 impact evaluations published before 2000, but counts over 9 000 published between 2000 and 2021 (3ie, 2021). This has been supported by newly established dedicated organisations such as the International Initiative for Impact Evaluation (3ie), the Abdul Latif Jameel Poverty Action Lab (J-PAL), Innovations for Poverty Action (IPA) and the World Bank's Development Impact Evaluation (DIME) unit. The increasing importance of RIEs is also reflected in the award of the Nobel Prize in Economic Science to Abhijit Banerjee, Esther Duflo and Michel Kremer in 2019 "for their experimental approach to alleviating global poverty", specifically their application of randomised controlled trials (RCTs) and to Joshua D. Angrist and Guido W. Imbens in 2021 for showing "what conclusions about cause and effect can be drawn from natural experiments".

As White (2019) points out, after the expansion of RIEs it became difficult to stay on top of the increasing body of evidence. Attention therefore turned to ways to synthesise existing evidence. 3ie and the Campbell Collaboration – among others – promoted SRs and developed EGMs in international development. In recent years, the international development community also acknowledged more and more that solely increasing the number of RIEs, SRs and EGMs does not necessarily result in increasing evidence take-up. The new generation of RIE focuses on their integration into a larger evidence and data ecosystem to "enhance their usefulness, responsiveness, and relevance for public policy decision making" (CGD, 2021). Hence, effective dissemination and communication of rigorous evidence and other forms of "knowledge brokering" (White, 2019) are receiving additional attention.

At the same time, there are still many evidence gaps, and the extent to which (rigorous) evidence is produced and used for policy making and development programming varies widely across countries and organisations. Whereas a number of bilateral development partners and international organisations, such as, for example, the Foreign, Commonwealth and Development Office (FCDO, formerly Department for International Development (DFID)), the United States Agency for International Development (USAID), the Millennium

Challenge Corporation (MCC), the World Bank and the Inter-American Development Bank (IDB), are known for their efforts to generate evidence and promote its use, the situation in Germany is less evident.

In the last two decades, a more impact-oriented approach to the design of development cooperation has also become a central issue within German development cooperation (GDC) (for more information see Faust, 2020). In line with the so-called Aid Effectiveness Agenda and further international agreements, such as the 2030 Agenda (United Nations General Assembly, 2015), the Paris Agenda (OECD, 2005) and the Accra Agenda for Action (OECD, 2008), stakeholders within GDC took action. The recent reorganisation "BMZ 2030" (BMZ, 2020) focuses on a stronger orientation towards the measurement of impact in the design of development cooperation and the allocation of funds. Furthermore, the newly published BMZ evaluation guidelines (2021), which describe the objectives, principles, standards and procedures in the GDC evaluation system, identify RIE as an approach to impact measurement (BMZ, 2021). In fact, BMZ's research and evaluation department, along with interested individuals working in GDC, have been fostering the implementation of RIE and take-up of rigorous evidence since the early 2000s, for instance in national workshops or through the Network of Networks for Impact Evaluation (NONIE). Yet, up to this point, the total number of RIEs, and their sectoral or geographical distribution within GDC, remain unclear. Further data on evidence take-up – both from self-generated and from publicly available RIEs – has been lacking (see Florian et al., 2019, for an initial effort to shed light on the situation within the *Deutsche Gesellschaft für Internationale Zusammenarbeit* (GIZ)).

1.2 Aim of the report

Recognising this lack of information and the potential for RIE to increase the effectiveness of GDC, the BMZ's research and evaluation unit started funding a DEval research project on RIE within GDC at the end of 2018. The project investigated the status quo, barriers to both the initiation of RIEs and the take-up of (rigorous) evidence, and potential measures to increase the use of RIE and the evidence it produces.

DEval is well-placed to carry out this project given (a) its experience in conducting RIE (in particular quasi-experimental designs implemented as a component of theory-based evaluations),[1] (b) its role in (co-)conducting three conferences that put the topic on the agenda of GDC,[2] (c) its international ties to organisations like 3ie, the Campbell Collaboration, DIME, IPA or J-PAL, and (d) its efforts to outline central tasks and challenges related to more systematic evidence generation and use within GDC (Bruder et al. 2019; Faust, 2020).

Jointly with BMZ, DEval developed three central research questions, which reflect the aspects of RIE initiation and the take-up of rigorous evidence:

1. What is the **status quo** of (a) the initiation of RIE and (b) the take-up of RIE evidence within GDC?
2. What are existing **barriers** to (a) the initiation of an RIE and (b) the take-up of RIE evidence within GDC?
3. What are potential **measures** to overcome barriers for (a) the initiation of RIE and (b) the take-up of RIE evidence within GDC?

Linking those questions to the ToC in Figure 2, question 2 examines barriers preventing the causal chain of outputs from operating smoothly and thereby hindering the systematic implementation of RIEs and take-up of rigorous evidence. Question 3 identifies inputs and activities necessary to reducing barriers and thereby achieving the outcomes and impact outlined in the ToC.

The intended audience for the results of the project are all those working in GDC – particularly decision makers within BMZ tasked with drafting overarching sector or country strategies and decision makers taking

[1] Polak, J.T., K. Guffler and L. Scheinert (2017), "Weltwärts volunteers and their civic engagement in Germany"; Roxin, H.; A. Kocks, R. Wedel, N. Herforth and T. Wencker (2021) "Effectiveness of German development cooperation in dealing with conflict driven migration crisis"; 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".

[2] "Impact evaluation in development cooperation in the tension between theory and practice", Mannheim, 30 November 2016; "(Quasi-)experimental impact evaluation in German development cooperation: Usefulness, stakeholder management and cooperation", Bonn, 30 November 2017; "Using evidence for a more effective development policy and practice: The role of evidence synthesis", Bonn, 9/10 May 2019.

operational decisions about project design and implementation (mainly in IOs and CSOs, but also within BMZ). Both groups are supposed to take evidence-informed decisions. The findings aim to foster such evidence-informed policy making and programming. Many implications focus on the systemic level and are therefore mostly addressed at BMZ and by senior management of IOs and CSOs. A smaller number of implications could also be implemented at the individual or project level. The report might also provide interesting insights for other bilateral funders and German government departments interested in fostering evidence-informed policy making, and for the scientific community and evaluators interested in conducting RIEs within GDC.

1.3 Scope of the report

This report focuses on RIEs that are in line with our definition (see Box 1). Our research considers German organisations operating in the context of development cooperation, such as IOs, CSOs and the scientific community. Yet, we also relied on international experiences (for more details see Chapter 2).

Our focus is on the initiation of RIE and the take-up of (rigorous) evidence. We therefore do not focus on the practical field implementation of RIEs and the associated barriers and possible measures. For details and best practices on RIE implementation, please refer to the large body of applied literature, including practical tool kits and training programmes (e.g. Duflo et al., 2007; Glennerster and Takavarasha, 2013; Leeuw and Vaessen, 2009; Peersman et al., 2015).

RIEs based on experimental methods are a contested territory, raising criticism about the increasing importance of rigorous evaluation approaches. There is an ongoing debate about methodological drawbacks, ethical challenges of RCTs and the global RCT “industry”. Critics further question the claim that RCTs should be considered as the “gold standard” of causal attribution. It is beyond the scope of this report to present these discussions in detail. This has been done by multiple authors elsewhere (see for example Murunga and Ogachi, 2020; Faust, 2020; Deaton, 2020, 2018; Muller, 2015; Bédécarrats et al., 2019; Kaplan et al., 2021). In this report, we do not assume rigorous evaluation approaches as being superior to other evaluation approaches per se. The choice of the evaluation approach and design should be determined by the evaluation question(s) at hand. However, we do consider RIEs to be the most reliable approach for answering questions of causal impact when a number of other conditions are fulfilled (see separate *DEval publication (2021)* for the feasibility and usefulness of RIEs). Hence, in the scope of this report we focus on RIE with experimental or quasi-experimental designs, while acknowledging the need for a diverse toolbox of methods and for method integration.

1.4 Structure of the report

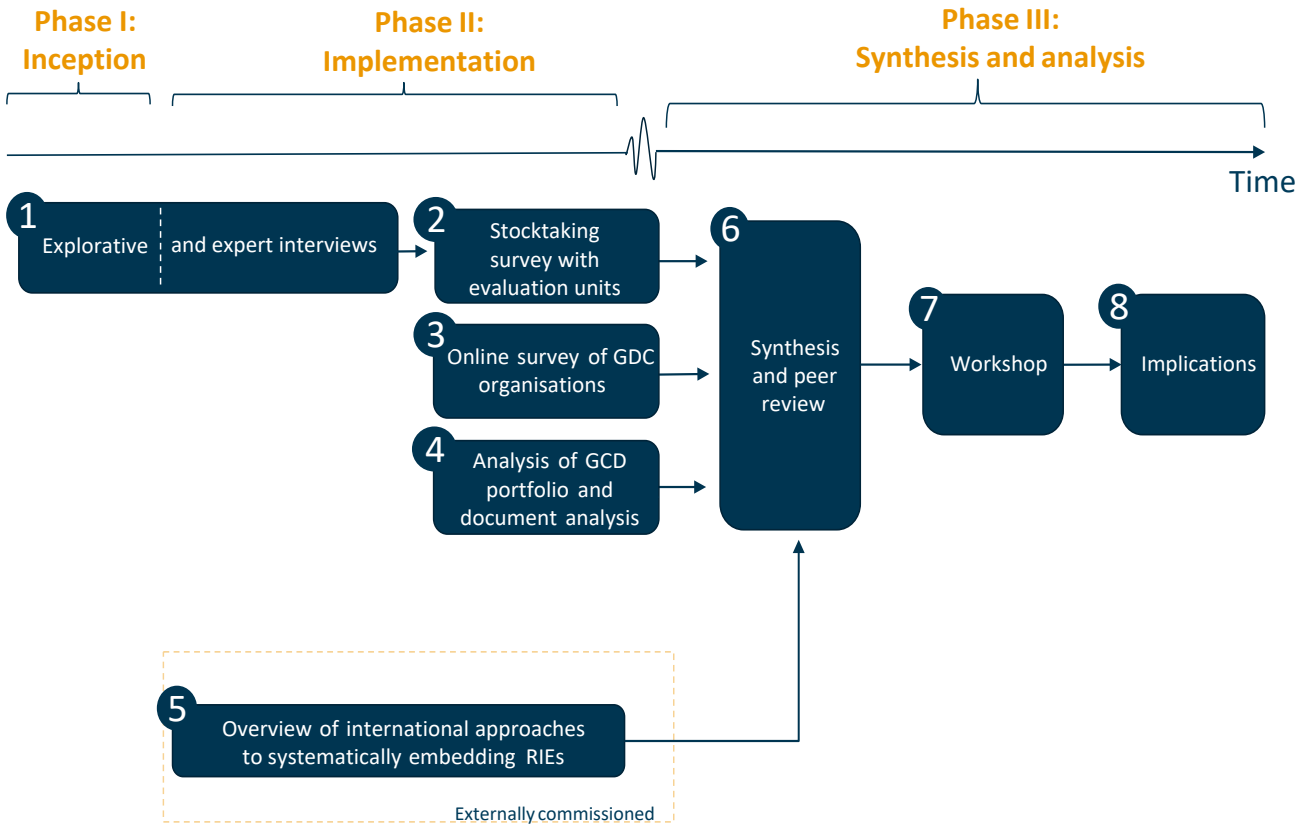
The report is structured according to our research questions. There are separate sections for the initiation of RIE and the take-up of rigorous evidence in order to allow the reader to focus on the specific challenges of each of these aspects. In this Chapter (Chapter 1), we present the motivation and background of the study and define the aim and scope of the report. In Chapter 2, we describe our methodological approach. Chapter 3 presents findings on the initiation of RIE, including a description of the status quo, barriers and potential measures to overcome these barriers. In Chapter 4, we turn to the take-up of rigorous evidence, while differentiating between self-generated RIE results and results from the global RIE evidence base. In Chapter 5, we offer implications for GDC concerning both the initiation of RIEs and the take-up of rigorous evidence.

2. METHODOLOGY

2.1 Overview of research process and design

The research process consisted of **three phases**: (I) Inception, (II) Implementation, (III) Synthesis and analysis (see Figure 3). An initial screening of the German RIE landscape (including explorative interviews) and the resulting definition of the research design characterised the inception phase. During the implementation phase, we carried out or commissioned different interlinked data collections. Finally, in the analysis phase, we analysed data and synthesised findings across all data sources. A **peer review** process with regular reflection sessions enhanced the quality of the research process.

Figure 3 Research process



Source: own illustration

The study followed a **mixed-methods** research design, comprising multiple primary and secondary data sources and combining quantitative and qualitative methods of data collection and analysis.[3] We drew on six sources of data (see Figure 4) and triangulated our findings across the different data sources, methods and researchers within the team. Throughout this report, we indicate data sources using the following icons:

[3] Please refer to the *online appendix* for a detailed analysis grid, which displays the research questions and corresponding data sources and analysis methods.

Figure 4 Data sources

Source: own illustration

The following section details the sampling, data collection and data analysis process, organised by **data sources**. The final section of this chapter addresses **limitations** of this research effort.

2.2 Data Sources

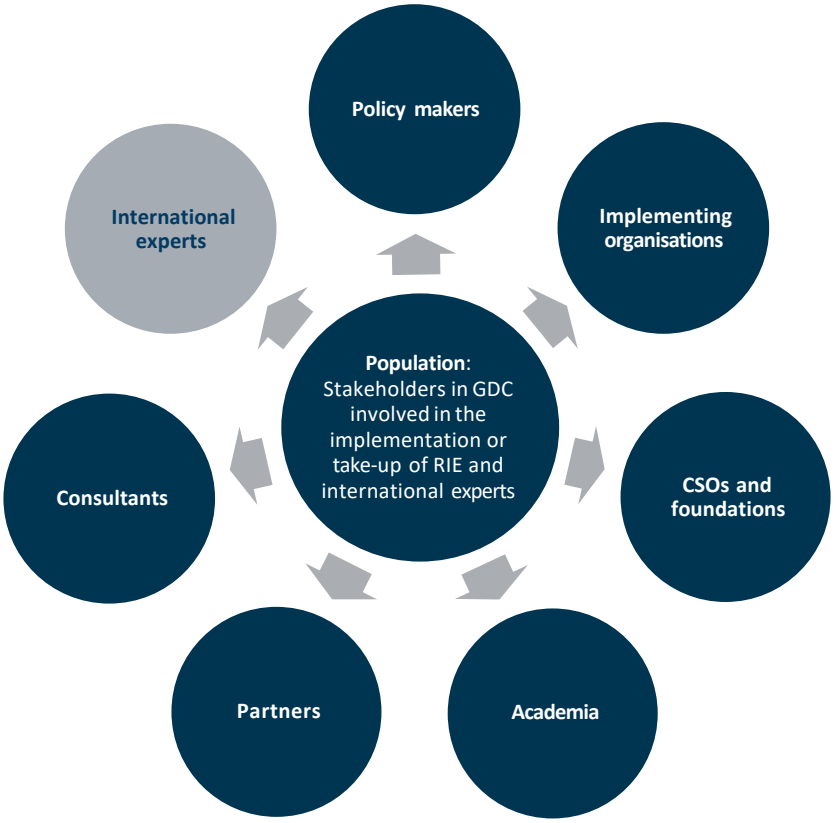
Explorative and semi-structured interviews (international RIE experts and people working in GDC)

We conducted interviews among (a) international RIE experts and (b) a comprehensive set of stakeholders with RIE knowledge within GDC.

For a comprehensive picture of RIE, **sampling** of interview partners followed a three-step approach. First, we conducted two exploratory interviews (not directed by a structured interview guideline) to better define the scope of the research field. On this basis, we mapped all relevant stakeholder groups and corresponding organisations within GDC and complemented this with the stakeholder group of international RIE experts (see Figure 5). Second, we made a purposeful selection of organisations within each stakeholder group. The minimum requirement was for each stakeholder group to be represented by at least one organisation, while also considering the organisations' size and relevance to GDC. Third, we selected interviewees from these organisations who were able to cover as many different perspectives and thematic areas as possible with regard to the use of RIEs in GDC. The semi-structured interviews comprised a largely identical set of open-ended questions regarding our central research topics but also left room for enquiries and spontaneous deliberations by the interview partners. We thus followed a diverse-case method within different strata as the sampling strategy (Seawright and Gerring, 2008). In total, we conducted 50 interviews with stakeholders from seven sub-groups between July 2019 and February 2021.[4]

[4] Please refer to the *online appendix* for an anonymised overview of interviewees.

Figure 5 Interview sampling



Source: own illustration

Interviews lasted one hour on average. They were usually conducted by two researchers and subsequently transcribed. For **data analysis**, we first developed a basic coding scheme. Second, we tested and refined the coding scheme and defined rules for inductive coding of the transcripts (e.g. for identifying barriers). We cross-checked all codings in a tandem setup (independent double coding for the first five transcripts, the following transcripts in pairs). During our analysis, we further clustered codes, studied frequencies and carved out patterns, for example by comparing sub-groups. Again, we did so in a tandem setup to increase reliability.

Stocktaking exercise for RIE in GDC

To take stock of RIEs conducted in GDC, we collected data of RIEs that had been conducted with the involvement of GDC since 2014. We defined involvement of GDC as follows: either the evaluation itself or the evaluated intervention received funding from BMZ. We also included currently ongoing or approved RIEs.

Box 3 Registration of RIE

The stocktaking survey page is still open and RIEs can be registered on an ongoing basis via *https://www.deval.org/en/methods-standards/rigorous-impact-evaluation.*

Data collection was carried out via an online survey. We sent out the survey to the evaluation departments of relevant organisations, such as BMZ, GIZ, KfW Development Bank, various development CSOs, foundations and to several research institutes and private consultancies. Evaluation departments either completed the survey autonomously after consulting with their colleagues within the organisations or directly forwarded it to the relevant teams. In addition to these organisational contacts, we conducted an online desk search for RIEs of relevant organisations. We directly contacted individuals related to RIEs that we knew about but that

had not yet been registered in our survey. In most cases, project managers of the implementing organisation filled in the survey. Data collection ran from October 2020 to April 2021. The survey included questions about the RIE itself (e.g. its methodology, sample size and use of results), and questions about the evaluated development project.

In total, 138 RIEs were registered. During **data analysis** we excluded twelve RIEs for which the endline data collection took place before 2014, and 26 RIEs for which we could not identify financial involvement of BMZ, as well as three other RIEs for further reasons such as insufficient information. In total, 97 RIEs fulfilled the inclusion criteria and were analysed using descriptive statistics. Open responses were coded and clustered using content analysis.

Even though this was the most comprehensive search for RIEs in GDC to date, it is very likely that the 97 RIEs identified by our survey do not constitute the full number of RIEs in GDC between 2014 and 2020. First, our desk research identified 18 potentially relevant RIEs for which no information was reported via our survey and for which we were unable to acquire further information. Second, the substantial effort necessary to detect 97 RIEs suggests that there may still be RIEs we do not know about.

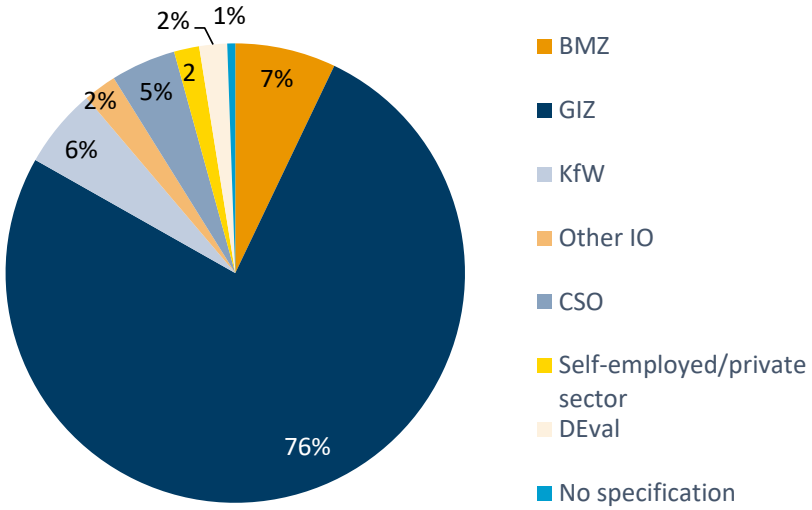
Evidence survey

The evidence survey targeted all employees of GDC organisations receiving BMZ funds, except administrative staff. It was designed to provide a picture of (a) the degree to which those working in GDC regard (rigorous) evidence as important, (b) what kind of experience they have with implementing RIE and using (rigorous) evidence, (c) which measures they envisage to better exploit the potential of RIE. We aimed at a full sample of this population.

Data collection was conducted online as a 30-minutes survey available in German and English. The survey was pre-tested by DEval experts as well as selected employees of GDC organisations. It was distributed to the evaluation departments of 58 GDC organisations, requesting them to forward the survey link to all colleagues (except for administrative staff). Data were collected between March 2021 and April 2021. In sum, 1 086 respondents from at least 23 different organisations participated in the evidence survey and provided information. Dropout rate during survey completion was 39.9%, with 653 participants reaching the final page of the survey. Figure 6 summarises the institutional affiliation of respondents. Given the snowball procedure, we do not know exactly how many people were reached by the invitation to participate in our evidence survey. Therefore, it is not possible to calculate a response rate. The following numbers may help to put the number of employees per organisation and the reach of our evidence survey into perspective: BMZ has approximately 1 230 employees, GIZ has approximately 23 600 employees and KfW Development Bank has approximately 1 000 employees.

Data analysis involved descriptive statistics as well as qualitative content analysis of open response fields. For more detailed information on our sample, see the *online appendix*.

Figure 6 Institutional affiliation of respondents to the evidence survey



Source: evidence survey; question: "Where are you currently working primarily?" N=1,071. The number of respondents differs from the total sample as 15 respondents dropped out before answering this question.

Review on international experiences with RIE

We commissioned a study on the institutionalisation of RIE implementation and evidence take-up among international peers in development cooperation. A consortium of the consulting firms Scio and Sattva carried out the study between April 2020 and December 2020.

Based on the criteria of (a) similarity to GDC organisations, (b) reputation regarding RIE implementation or evidence take-up, and (c) learning potential for GDC, we selected the following institutions for in-depth analysis: Asian Development Bank (ADB), Agence française de développement (AFD), FCDO, J-PAL, MCC, Oxfam, USAID and the What Works Network (WWN).

Data collection was based on key informant interviews and accompanied by a literature search and analysis. With the goal of capturing a spectrum of relevant perspectives from each of the selected organisations, we identified between three and five interviewees per organisation who were either concerned with implementing RIEs or using (rigorous) evidence. In total, 28 semi-structured interviews with individuals and one group interview with three persons were conducted. Interviews lasted one hour on average, were conducted by at least two researchers and transcribed for subsequent data analysis. Members of the DEval RIE Team attended most of the interviews.

A framework for institutionalisation guided the semi-structured interviews and subsequent data analysis. The framework covered different factors that potentially affected the systematic integration of the RIE approach into an organisation (e.g. strategic alignment and level of buy-in, funding and resources etc.). The report identifies organisational “archetypes”, collects best practices and common barriers and offers general recommendations for institutionalising RIE implementation and evidence take-up.

Analysis of GDC portfolio and document analysis

We analysed the GDC portfolio and official project documents to learn whether the global RIE evidence base is relevant to the GDC portfolio and to what extent the global RIE evidence base has already been utilised in GDC. Our analysis comprised four steps:

1. **Comparing German ODA flows with evidence from the 3ie development evidence portal (DEP):** We mapped a five-year average (2014–18) share of bilateral BMZ disbursements drawn from the OECD

Creditor Reporting System (CRS) database[5] on the available rigorous evidence from the 3ie's DEP – the most comprehensive repository of rigorous evidence in international development. This comparison reveals the sectoral overlap between globally available rigorous evidence and German ODA.

2. **Screening 124 project documents of randomly sampled projects from GIZ and KfW Development Bank with regard to evidence take-up:** We first implemented a typical-case method as sampling strategy (Seawright and Gerring, 2008) to select a sector for our analysis of GIZ and KfW Development Bank documents. "Government & Civil Society" was identified as a typical sector because it receives a large share of German ODA flows, the amount of rigorous evidence available in this sector is of an average level, and it is of high political relevance – as demonstrated by its role in BMZ's core topic areas (*Kernthemen*).
   Within this sector, we randomly drew five projects each from GIZ and KfW Development Bank for the period 2014–18. This resulted in a database of 124 documents, including, for example, project proposals, project reports and project evaluation reports. We screened all documents to analyse whether project documents made use of scientific evidence and what this evidence was used for. Furthermore, we expanded our analysis beyond scientific evidence for two of the ten projects to identify which non-scientific sources were used in addition or instead of scientific evidence in project documents.
3. **Checking availability of relevant RIE evidence for sampled projects:** Using the advanced search option of 3ie's DEP, we conducted a non-exhaustive keyword search to find thematically suitable RIEs, SRs and EGMs published after 2014 for each of the 10 projects. This was to check whether the global RIE evidence base was, in fact, relevant to these specific projects. In addition, we screened abstracts of those studies that appeared to be of potentially high relevance to the sampled projects.
4. **Screening of project documents for two additional projects from KfW Development Bank and GIZ where RIEs had been conducted:** We chose two additional projects in which RIEs had been conducted. As described in point 3, we screened all 67 project documents to see whether the RIEs conducted in the respective projects were mentioned. We also screened the documents for the use of other types of scientific evidence.

Literature search and further documents

The research process was accompanied by a review of the scientific literature as well as other secondary sources related to our research questions. The most relevant reports and clusters of documents were:

- Impact evaluation guidance documents: We identified and screened more than a dozen official guidance documents that set out criteria for conducting RIE. We first identified a set of international development organisations engaged in RIE. Second, we searched for available guidance documents on RIE and, third, we summarised criteria for RIE implementation that were proposed in these documents.
- GIZ RIE report: Florian et al. (2019) published an internal review on RIE in 2019. This report examined RIEs conducted within GIZ between 2014 and 2018 and outlined conditions under which RIE have proven particularly useful for GIZ.
- Research papers and flagship reports: We included insights from research papers and flagship reports on topics such as identifying relevant evidence, evidence take-up, experiences with RIE within GDC, research ethics and behavioural science.
- Government documents: To analyse the contracting procedures between IOs and BMZ (see Box 4), we screened official BMZ guidance documents. The analysis aimed at examining formal requirements and options regarding RIE initiation and evidence take-up when planning and implementing bilateral projects.

[5] For more information on OECD DAC data and sector classification, see http://www.oecd.org/dac/stats/purposecodessectorclassification.htm.

Box 4 In short: Procedures of GDC

The planning and implementation of bilateral projects follows mandatory procedural rules, which were revised in 2017 through a joint procedural reform (*Gemeinsame Verfahrensreform,* GVR) and are documented in internal guidance documents. These procedural rules specify which steps must be taken by BMZ and the respective IO when implementing a project.

2.3 Limitations

Despite the comprehensive research design and triangulation procedures, we cannot rule out some potential shortcomings:

- We ensured data triangulation for all our main research questions. However, there may be more marginal questions for which data was only available from one data source. We have made this transparent where it occurs and it should be taken into account when interpreting our findings.
- Our sampling of interview partners is susceptible to a positive bias towards RIE, as we interviewed people with a basic understanding of the concept of RIE. Yet, to fully understand and illuminate factors influencing RIE initiation and use, it was indispensable to interview individuals with previous RIE knowledge or experience. To help achieve a sufficiently balanced picture, we addressed all GDC practitioners in our online survey.
- Even though we analysed the basic contracting procedures between BMZ and IOs, we have not carried out an in-depth process analysis, for example of the commissioning of a new module. To formulate specific recommendations on how to better align these standard processes with RIE implementation and use, a more comprehensive process analysis would be an advisable subsequent step.
- This research focuses primarily on views from those working in GDC. This internal view is supplemented by international RIE experiences and selected partner views. However, strengthening the partner country perspective could be a useful step in future research on the topic, particularly in a potentially increasingly country-led evaluation culture.
- The majority of our data is based on self-report. Self-report may be subject to bias or lack of knowledge, in particular when it comes to identifying which designs are (quasi-)experimental in nature. Given that an in-depth check of all responses given in our stocktaking exercise can only be conducted when adding the respective study to the emerging DEval RIE repository, there is some scope for misidentification of some of the designs.
- We deliberately included different types of scientific evidence in the evidence survey, because the boundaries between rigorous and non-rigorous evidence are by definition fuzzy. At the core, we are also concerned with whether employees working within GDC use evidence in general, including rigorous evidence. In the interviews, we were able to ensure that our general findings also apply to RIE.

3. FINDINGS: INITIATION OF RIES

This chapter is structured in four sub-sections. Section 3.1 illustrates the status quo regarding RIE implementation within GDC. Section 3.2 presents barriers that intervene at the five steps, leading to a more systematic and appropriate RIE implementation defined in our ToC (Chapter 1). Barriers are clustered around these five steps and their relevance is assessed and rated. Section 3.3 provides a comprehensive list of potential measures to foster RIE initiation. This list is presented in simple alphabetical order to illustrate the entire range of potential measures. Section 3.4 brings together the most relevant barriers and potential measures addressing them.

Together with the analysis of the status quo, barriers and measures for evidence take-up (Chapter 4), this analysis will inform our selection and formulation of implications in Chapter 5.

3.1 Status quo in GDC

In this chapter, we present the current state of initiating and conducting RIE within GDC. It builds on findings from our stocktaking exercise and to a lesser extent on the explorative interviews and literature review. Because RIEs are mostly conducted decentrally and do not get registered systematically, the number and sectoral or geographical distribution of RIE within GDC was largely unknown until very recently. This also holds for many international development agencies. “Information on the number of [R]IEs and related products is often poorly documented at bilateral agency level […] because the commissioning of evaluations, including IEs, is largely decentralised“ (Manning et al., 2020).

A recent report by GIZ (Florian et al., 2019) shed some light on the number of RIEs within GIZ. It reported 39 RIEs between 2014 and 2018. Our stocktaking exercise builds on this effort but goes beyond it in that it both integrates all other GDC organisations and adds a substantial number of additional GIZ RIEs for the same time period.[6]

In total, we identified 97 RIEs fulfilling our criteria,[7] of which respondents reported 47 RIEs as completed and 50 RIEs as still being in progress or having been approved. Figure 7 shows the development of the number of RIEs in GDC over time. Years refer to (expected) dates of endline data collections. The high number of RIEs in 2023 may indicate a possible slight upward trend but with the cautionary note that it is unclear whether all planned RIEs will indeed be completed.

Thirteen RIEs were conducted as part of a GIZ global project (*Globalvorhaben*) in the field of rural development and 10 RIEs within another rural development global project. Within these global projects, the same or similar interventions were rigorously evaluated across different countries. Furthermore, an entrepreneurial training programme partially financed by a German CSO was rigorously evaluated in different contexts and at multiple points in time. This programme accounts for 13 RIEs.

Our stocktaking survey asked respondents to upload publicly available documents for respective RIEs. In total, scientific publications were only uploaded for 14 of the RIEs.

Since our interview partners frequently described GDC as lagging behind internationally in RIE, the total number of RIEs in GDC somewhat exceeded our prior expectations.[8] Taking into account the increasing recognition of RIE among GDC stakeholders (see for example Florian et al. (2019) and KfW Development Bank (2021)), we also expect the number of RIEs to increase in the future. However, the total number has to be interpreted with caution, given the fact that substantial numbers of these RIE were clustered within larger

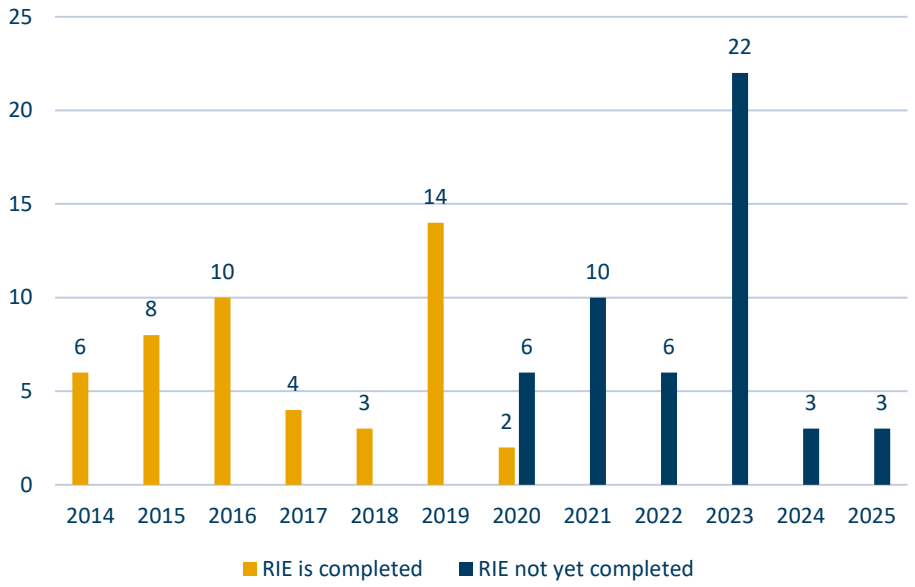
[6] Of the 39 RIEs of the GIZ (Florian et al., 2019) report, 34 were registered in our stocktaking survey (32 included in final sample), while five RIEs were not registered. In addition, our final sample includes 26 RIEs that had not been part of the GIZ (Florian et al., 2019) report.

[7] The specific criteria for including an RIE in our survey were that 1) the evaluation or the evaluated intervention receive or received any funding from BMZ and 2) that it had been conducted since 2014 or that it is currently being conducted or has already been approved.

[8] We do not give a direct comparison with other international donors, because data on the number of RIEs is often not publicly available or because comparability is not given (for example in the case of USAID).

projects and given that we relied on self-assessments of methodological rigour. Against the backdrop of the large volume of German ODA there is still substantial potential for increasing the effectiveness of GDC by conducting RIEs more systematically and learning from their results (also see von Schiller, 2021).

Figure 7 Number of RIEs over time

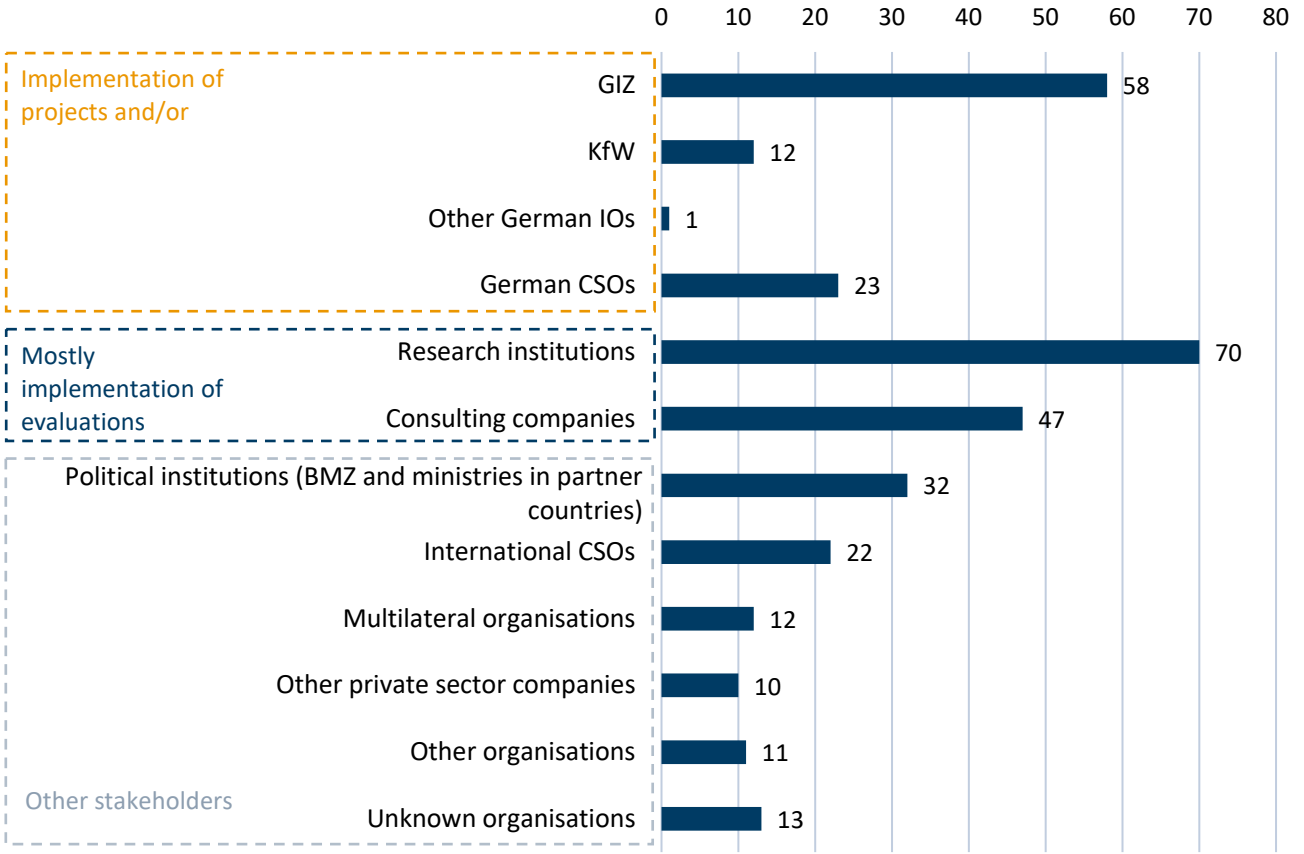


Source: stocktaking survey; questions: "(Expected) start of endline data collection (own data collection or secondary data)" and "Status of RIE"; N=97.

Figure 8 shows which types of organisations were involved in how many GDC RIEs. Different stakeholders often have different roles within an RIE, such as implementing the project, or implementing the evaluation, which includes data collection and data analysis. Our interviewees stated that the typical stakeholder setup of an RIE comprises an IO or CSO that usually implements the project, together with their local implementing partners and a research institution or consulting firm that conducts or supports the impact evaluation (source: interviews). Involvement of local stakeholders usually takes one of three forms: 1) government departments in partner countries involved as coordinator or facilitator, 2) local CSOs or public institutions involved in the implementation of projects, and 3) local private companies carrying out data collection. Other organisations in RIE implementation include international CSOs and multilateral organisations, which implement the project, as well as the World Bank, whose Development Impact Evaluation unit, DIME, conducted several of the RIEs.

GIZ, KfW Development Bank and German CSOs have all been involved in RIEs (for example as project implementer or by financing the RIE). With 58 RIEs, GIZ reported the highest number within this group, followed by CSOs with 23 RIEs, and KfW Development Bank, which was involved in 12 RIEs. In the relevant period, DEval has conducted four RIEs (DEval is listed included in "research institutions" in Figure 8). For about 70% of RIEs at least one organisation from the partner country was reported to be involved in the implementation of the development project or the RIE evaluating it, leaving a share of approximately 30% of RIEs with no reported partner involvement.

Figure 8 RIE involvement by organisation



Source: stocktaking survey; question: "Please list all organisations involved in the RIE and/or the project under evaluation. Please specify (a) the name, (b) the type (e.g. implementing organisation, CSO, multilateral organisation, university), and (c) the role of each organisation (e.g. "responsible for data collection")." N=97; multiple answers possible.

Figure 9 shows the number of RIEs by evaluation design. Experimental designs (i.e. RCTs) characterised 45% of RIEs, according to those reporting the RIE. RIEs using quasi-experimental designs make up 46% of the reported RIEs. For 8% of RIEs the design was unknown or not reported.

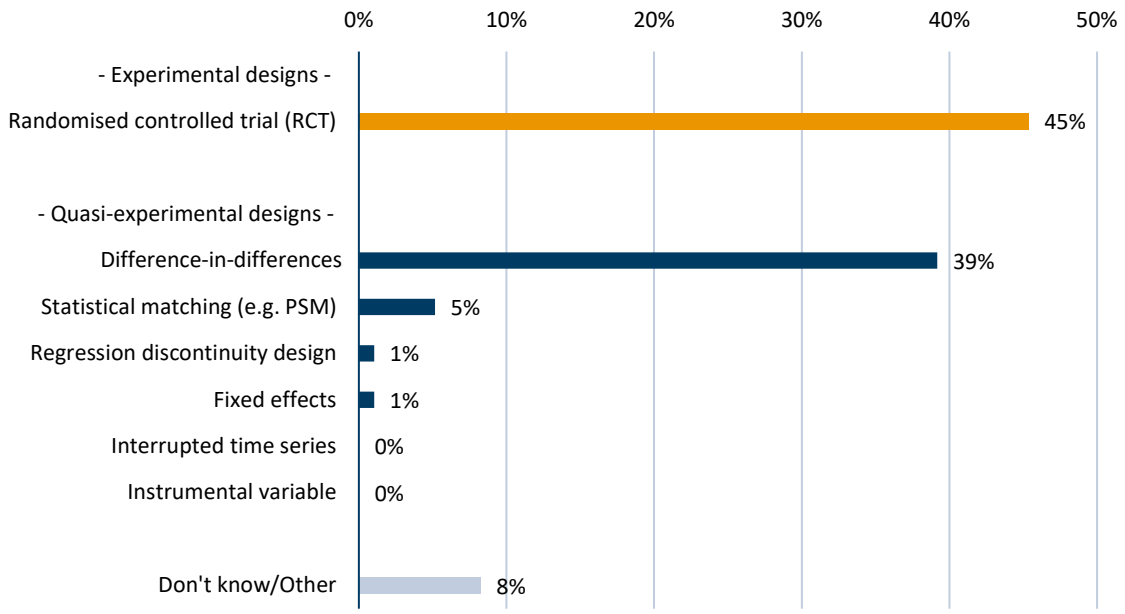
The most frequently applied quasi-experimental design was difference-in-differences (DiD) with 39% of all reported RIEs.[9] GDC uses other quasi-experimental designs rarely or not at all, even though these designs often allow for RIE implementation at a lower cost than RCTs (for example when existing data can be used) (Lech et al., 2018).[10]

The frequent use of DiD by those working in GDC is notable, compared to the global RIE evidence base, which we approximated using the 3ie DEP. RIEs in 3ie's DEP use fewer DiD (17%) and more fixed effects (16%) and other quasi-experimental designs. The share of RCTs in 3ie's DEP is 54% (3ie, 2021).

Of the RIEs in our stocktaking survey, 61% report that their RIE was accompanied by qualitative research.

[9] For a short summary of quasi-experimental designs see, for example, White and Sabarwal (2014).
[10] One example of such an approach is geo-spatial impact evaluations (see Lech et al., 2018).

Figure 9 RIEs by design



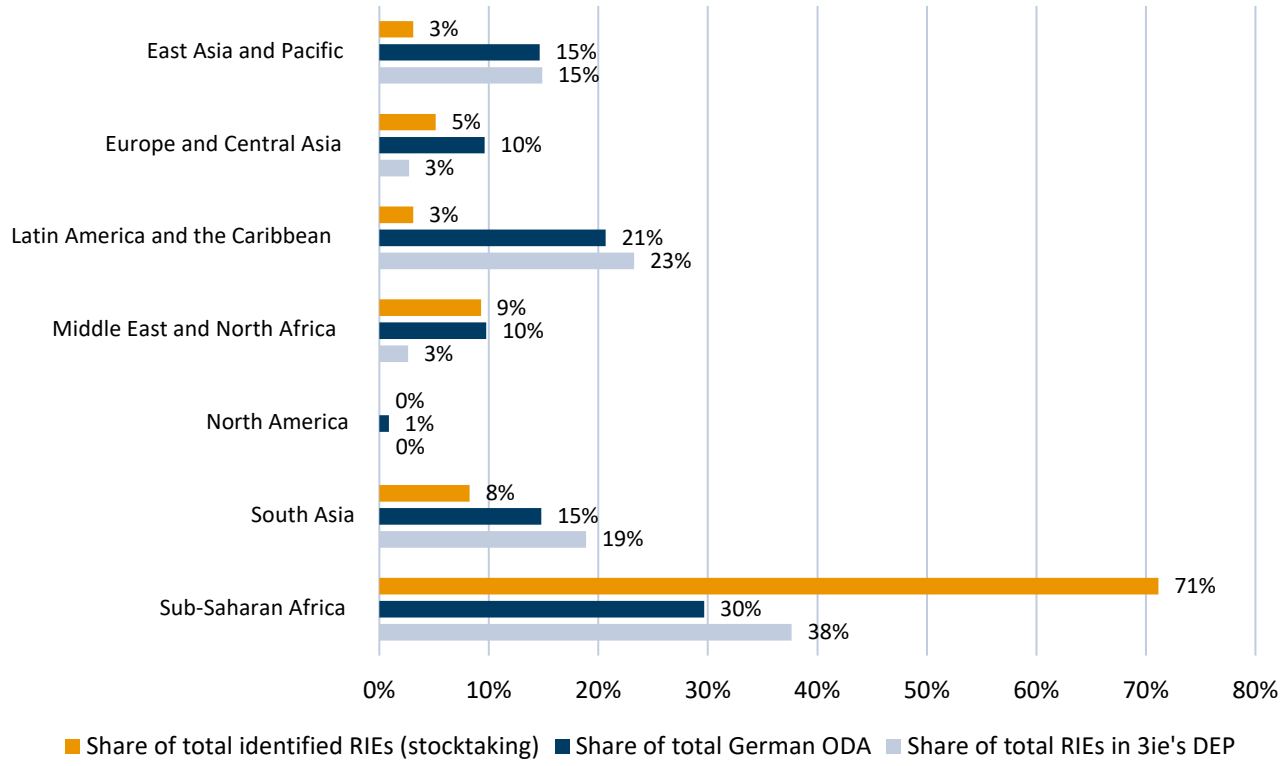
Source: stocktaking survey; questions: "Is assignment to the treatment and control group done through randomisation (i.e. is the RIE a randomised controlled trial; RCT)?" and "Which quasi-experimental method is (has been) used?"; N=97.

Figure 10 compares the regional distribution of RIEs that we identified in the stocktaking exercise to the regional distribution of German bilateral ODA flows and RIEs in the 3ie's DEP. Sub-Saharan Africa (SSA) is the region in which RIEs of our stocktaking were by far the most frequently conducted (71%), followed by the Middle Eastern and North African region (MENA, 9%) and South Asia (8%). Europe and Central Asia, Latin America and the Caribbean (LAC) and East Asia and Pacific (EAP) jointly account for only 11% of RIEs. The countries in which RIEs of our stocktaking were conducted most frequently are Uganda (13), Ethiopia (6), Malawi (5) and Zambia (5).

The RIE shares shown in Figure 10 are not aligned with the shares of German ODA flows. For example, although SSA has a share of 71% of RIEs within GDC, it only receives 30% of German ODA. The stocktaking survey does not allow for conclusions as to what drives the discrepancy between RIE shares and German ODA shares. One potential explanation is that RIEs are often initiated by motivated individuals and are therefore not embedded in an overarching ODA-oriented strategy. Another potential explanation is that RIE implementation is influenced by the stakeholders in partner countries and their different degrees of affinity with RIE. For example, various countries in SSA, such as Uganda, have strong local evidence ecosystems. Aligning RIEs regionally or sectorally with ODA flows is one possible strategy for identifying when and where to implement RIEs. Other examples might be conducting RIEs in pilot projects or systematically identifying and targeting evidence gaps. However, we did not find any particular strategy for RIE implementation within GDC during our research.

Even though SSA also accounts for the largest share of RIEs in 3ie's DEP (38%), its share is much lower than in our stocktaking sample. The share of LAC and South Asia is significantly higher in 3ie's DEP than in GDC. For the MENA region on the other hand, the share is much higher in GDC (9%) than in 3ie's DEP, where evidence on the MENA region is very scarce (3%).

Figure 10 Regional comparison between stocktaking RIEs, German ODA flows and 3ie impact evaluations



Source 1: stocktaking survey; question: "In which country/countries is (has) the RIE been implemented?"; Ex-post categorisation using World Bank regions; N=97; Source 2: OECD (2021), German bilateral ODA disbursements with BMZ as extending agency, 2014–18 average, own calculations; Source 3: 3ie (2021)

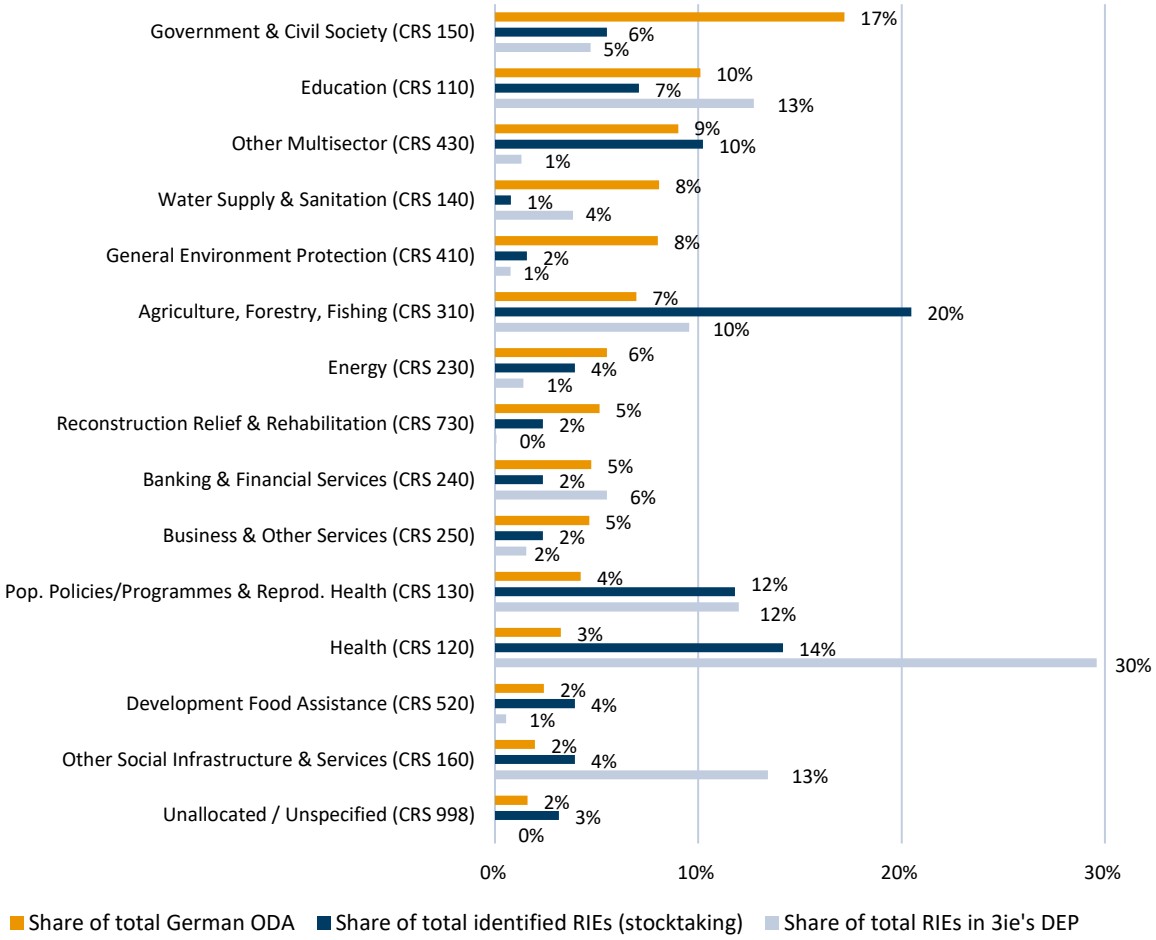
Figure 11 compares the sectoral distribution of German bilateral ODA flows and RIEs in 3ie's DEP to the sector of the project in which RIEs are or were conducted. "Agriculture, Forestry, Fishing" (20%), "Health" (14%) and "Population Polices/Programmes & Reproductive Health" (12%) were reported most frequently in our stocktaking survey.

Much as with the regional distribution of RIEs, the sectoral distribution of GDC RIEs does not closely reflect the distribution of German ODA flows. For example, the sector "Government & Civil Society" accounted for 17% of BMZ-related, bilateral ODA disbursements between 2014 and 2018 – a much larger share than the 6% share of stocktaking RIEs. The same holds true for sectors such as "Water Supply & Sanitation" (8% of German ODA and 1% of RIEs) and "General Environmental Protection" (8% of German ODA and 2% of RIEs).

For some sectors, such as "Education", the distribution of GDC RIEs is more aligned with the international RIE distribution than for other sectors, such as "Agriculture" or "Other Social Infrastructure and Services".

There may be some differences between sectors with regard to how well RIEs can be implemented on average. However, the RIE experts in our interviews agreed that it is not a question of sector whether an RIE can be implemented or not (source: interviews). Instead, this can only be analysed and decided individually on a project-basis. RIEs can be implemented in all sectors – for example Funk et al. (2018) documented RIEs in GIZ and KfW Development Bank government projects, a sector that is sometimes considered as difficult for RIE implementation.

Figure 11 Comparison of DAC sector distribution between stocktaking RIEs, ODA flows and 3ie impact evaluations (largest 15 sectors by ODA volume)



Source 1: stocktaking survey; questions: "Which CRS codes (OECD sector-codes) can be assigned to the overall project?"; N=80; multiple answers possible; Source 2: OECD (2021), German bilateral ODA disbursements with BMZ as extending agency, 2014-18 average, own calculations; Source 3: 3ie (2021)

Projects within GDC often comprise multiple sub-components. Several interviewees stated that RIEs normally evaluate the impact of a relatively small part of a project or component (source: interviews). Our stocktaking survey, however, shows that RIEs can also cover a larger share of the project. On average, RIEs in the stocktaking dataset evaluated about 60% of the respective project. More than 20% of all RIEs evaluated 90% of the overall project or more.

Most often project managers of IOs or CSOs decentrally initiated RIEs, alone (35 RIEs or 36% of the 97 RIEs) or together with the scientific community (4%). Central units of the IO or CSO also initiated a large share of RIEs – either alone (15%) or together with partners from the scientific community (14%). In some cases, central units and project managers jointly initiated RIEs (5%). The scientific community was responsible for 7% of RIE initiations. Other types of initiations accounted for 17% of RIEs and for 13% of RIEs no information on initiation was reported. In contrast, our interviewees reported that RIEs are more often initiated in decentral structures of the IOs (e.g. country offices) than in central structures (e.g. evaluation units or central sectoral units) and only rarely requested and commissioned by the donor (especially in cases of international co-financing). Motivated individuals usually play a role in initiating an RIE. However, our data on initiation should be interpreted with caution, as responses might be biased towards respondents' institutions.

In conclusion, various organisations within GDC have started to gather experiences with conducting RIEs, reflecting a growing relevance of RIE for GDC. However, these efforts do not seem to be embedded in an overarching strategy or a systematic approach but are rather driven by motivated individuals. Furthermore, compared to the overall size of GDC, the total number of RIEs is low and it is unlikely that their potential for

increasing GDC effectiveness has already been fully realised. This leads to the question, what main obstacles hinder GDC organisations from implementing RIEs more frequently. We will address this question in the next section.

3.2 Which barriers hinder the initiation of RIEs in GDC?

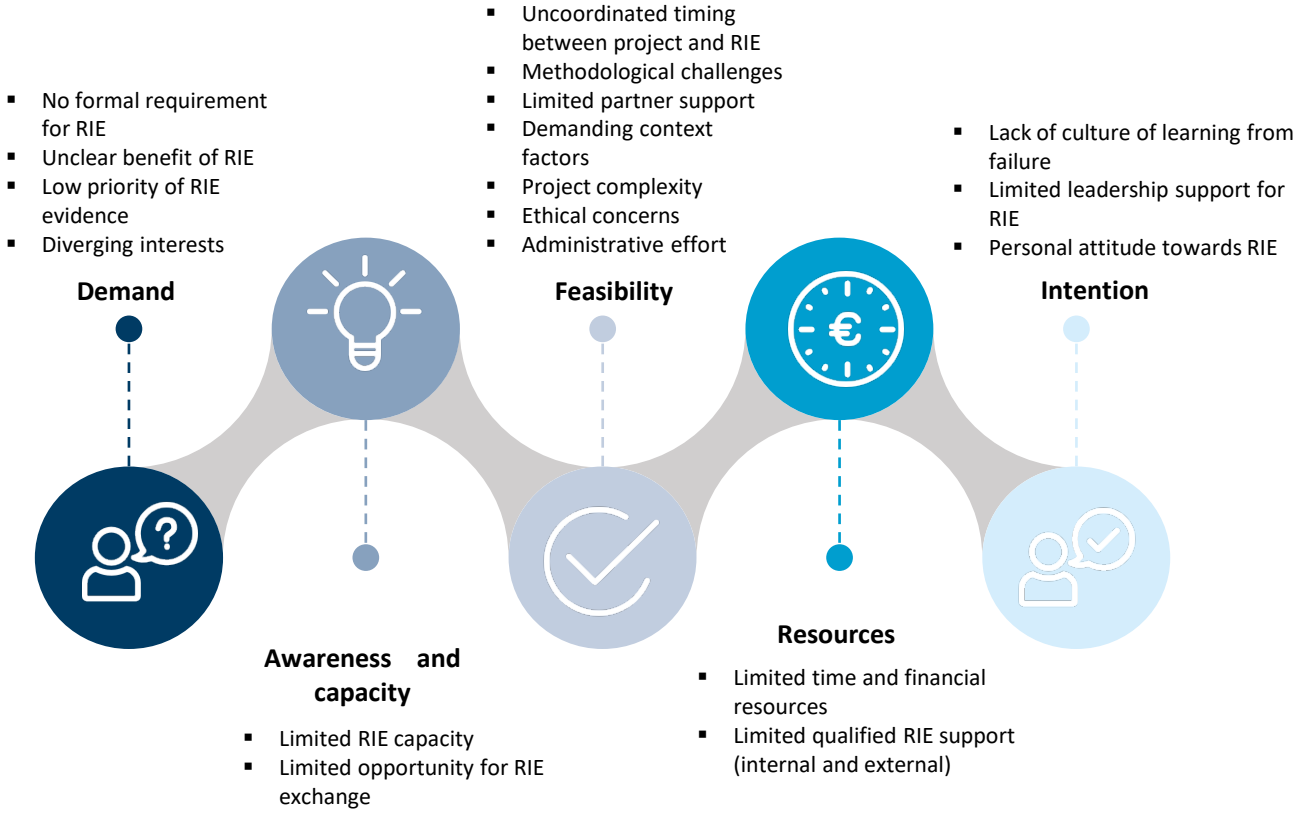
We now present barriers to RIE initiation within GDC. The analysis is based on data from the evidence survey, our interviews and the RIE stocktaking exercise.

In total, we identified 18 different barriers. The identified barriers reflect experiences and perceptions of interviewees and survey respondents. Our analysis shows that different barriers have varying relevance in affecting the initiation of RIEs. We express the relevance in the following way: Three stars (***) indicates a barrier of high relevance, two stars (***) one of medium relevance and one star (***) one of marginal relevance. The classification is based on:

- the triangulation of information on the barriers from different data sources
- our evaluative assessment of the reliability of interviewees' statements.

Based on the simplified ToC presented in Chapter 1, we connect the barriers we identified to the five steps leading to more systematic and appropriate RIE implementation (see Figure 12). Participants in our evidence survey also rated these barriers (see Figure 13).

Figure 12 Overview of barriers for initiating RIE, mapped to the five steps leading to more systematic and appropriate RIE implementation

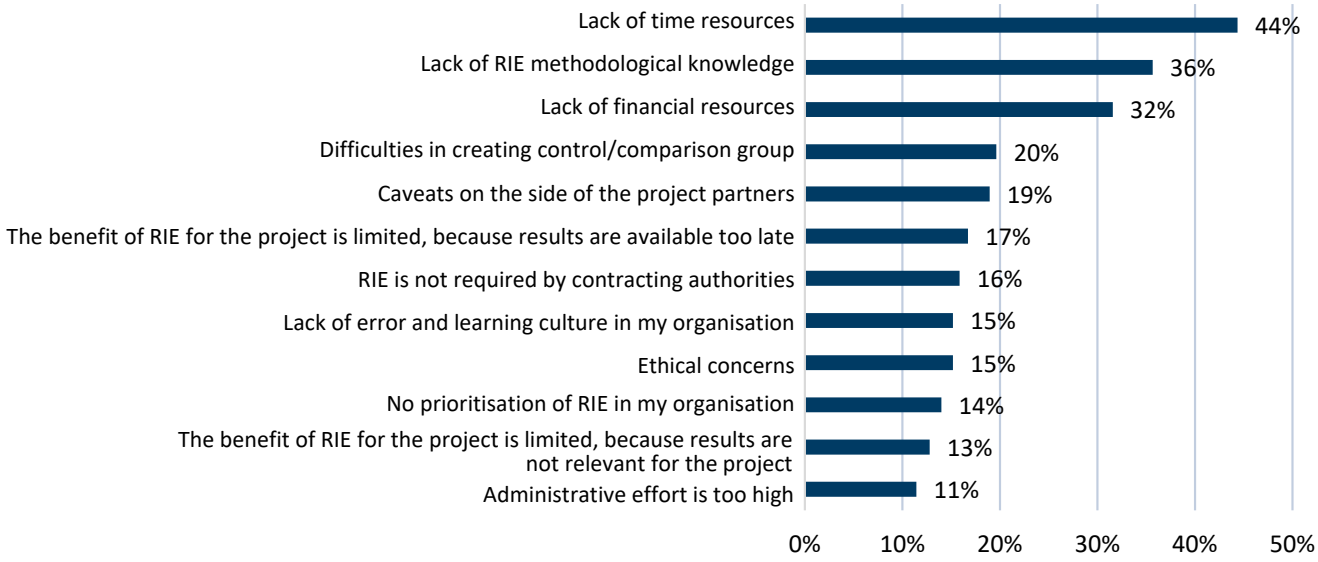


Source: own illustration

Demand

***** No formal requirement for RIE:** A majority of our interviewees reported a lack of formal requirement or procedural anchoring (including respective guidelines) for initiating RIE within GDC. The analysis of practical guidance documents also confirms this and the evidence survey identifies "RIE is not required" as being a barrier of moderate relevance (see Figure 13). Actually, the Federal Budget Regulations (*Bundeshaushaltsordnung*, BHO) formulates high requirements for performance assessments (*Wirksamkeitskontrolle*). In particular, effects should be causally linked to interventions (Administrative Regulation to the BHO § 7). Yet, the regulations do not require a specific evaluation approach. Hence, this regulation is not currently interpreted as necessitating RIEs. In fact, our data suggest that reporting requirements are perceived as being focused on achieving predefined project goals and respective indicators without much investment in testing the causal relationships between project activities and goal achievement.

Figure 13 Barriers to RIE initiation



Source: evidence survey; question: "What are considerations against the implementation of an RIE?"; N=586; % of respondents; multiple answers possible.

***** Unclear benefit of RIE:**

- Respondents reported a lack of incentives and (career) benefits for individuals for conducting RIEs. Implementing RIEs is reportedly generally not rewarded and comes hand in hand with additional personal effort. Indeed, interviewees even reported fear of negative consequences for their own career when conducting an RIE.
- Respondents in all data sources stated that the lack of clarity about the benefits of RIEs for the project, for example for operational project steering, could hinder the initiation of an RIE (see also Figure 13). This may be because findings are available too late for relevant project decisions, researchers do not address operational questions as a standard, or the value added compared to "regular" project evaluations was unclear.
- A lack of clarity and awareness about how to transfer or pool single RIE findings and use RIE for policy making, steering and strategy definition can also hinder RIE initiation. This includes difficulties in judging whether and how to use single RIE findings, doubts about the generalisability and transferability of findings of individual RIEs, and questions about the usefulness of RIE findings for policy advice and political communication (see Figure 13). As a consequence, the results of RIE are not fully exploited and RIEs are not encouraged (see also sections 4.2.1 and 4.2.2).

*** **Low priority given to RIE evidence:** Interviews further suggest that there is little (though it is increasing) demand for (RIE) evidence from different stakeholders, most importantly from donors (i.e. BMZ) and partners. This is accompanied by the observation of interviewees that other types of evidence (e.g. from process evaluations or expert opinions) are often ascribed equivalent or higher utility and quality. In the evidence survey, however, low priority of RIE is rarely seen as a relevant barrier to RIE initiation (see Figure 13).

*** **Diverging interests:** Project implementers, policy makers, researchers and project partners have different interests and expectations, resulting in diverging motivations for implementing an RIE. Researchers, for instance, pursue scientific publications, whereas project implementers may be more interested in practical questions of project steering. This tension remains mostly unsolved within GDC to date. This barrier was primarily identified by respondents with RIE experience.

Awareness and capacity

*** **Limited RIE capacity:** Limited RIE capacity is a barrier for RIE initiation among both project implementers and policy makers. A lack of RIE capacity makes it difficult for RIE commissioners (both at project and policy level) to anticipate what kind of insights and benefits RIEs can generate, how to commission them and how to conduct quality assurance. It also hampers RIE initiation in very practical terms, for example, when faced with questions concerning evaluation methods. Due to limited RIE capacity, RIEs are often not considered at the beginning of a project. However, this phase is critical to the implementation of a sound evaluation design within an RIE approach. All stakeholder groups mention this barrier as one of the most prominent ones hindering RIE initiation. Lack of RIE knowledge is among the three most frequently selected barriers of the evidence survey (see Figure 13).

*** **Limited opportunity for RIE exchange:** Limited opportunities for inter- and intra-organisational peer exchange as well as lack of (knowledge about) positive RIE examples can hinder RIE initiation. Without such insights it is more challenging for project implementers and policy makers to decide on where and how to apply RIE.

Feasibility

*** **Uncoordinated timing between project and RIE:** According to a significant number of interviewees and respondents to the evidence survey, lack of harmonisation in the timing of the project and the RIE was a critical barrier to RIE implementation. Interviewees emphasised that RIE and projects usually have different time horizons (e.g. RIEs can only be contracted after the project has been formally approved and intervention design has already started, and completing them often takes more time than the standard project cycle). This substantially reduces incentives to conduct RIEs, because RIE cannot be financed from project funds beyond the project duration. In addition, it may be unclear whether RIE results can still be utilised for the current project cycle. Our stocktaking exercise identified this barrier as the most challenging one.

*** **Methodological challenges:** Methodological challenges when setting up an RIE were identified in all data sources and by all interviewed stakeholder groups as one of the most important factors hindering RIE initiation. Challenges regarding methodological feasibility include:

- "contamination" of the control group (meaning that a fraction of the control group also receives the treatment even though it should not) and spill-over effects (meaning that positive effects of the treatment among the treatment groups have positive knock-on effects on the control group)
- availability of, access to and quality of data
- lack of statistical power due to few available units of observation ("small or medium N") and
- difficulties in constructing treatment and control groups.

The evidence survey identifies one of those methodological challenges (i.e. difficulties creating treatment and control groups) as one of the five most frequently selected barriers (see Figure 13).

**** **Limited partner support:** Limited support for RIE by political and implementing partners in partner countries can hinder RIE initiation. For example, partner governments may not approve data collection efforts in certain districts due to domestic political reasons. Given that development projects within GDC are implemented jointly with, or even solely through, partners (especially in financial cooperation) this may be a prohibitive barrier according to all data sources (see Figure 13).

**** **Demanding context factors:** All data sources identified demanding context factors as a barrier to initiating RIE, as these factors endangered the stable delivery of interventions over a period of time, thereby making impact assessment of a specific form of intervention challenging or impossible. Relevant context factors are:

- fragile implementing contexts (e.g. Yemen or Afghanistan)
- general political circumstances in the partner country but also the donor country (e.g. changes in government and resulting changes/discontinuation in policy priorities)
- missing or pending approval of local authorities for data collection
- the sudden closure of local implementation sites (e.g. hospitals).

**** **Project complexity:** Project complexity can be a barrier to conducting an RIE, as project implementers found it difficult or impossible to isolate single interventions from a complex set of (multi-level) project activities to be evaluated with an RIE (source: interviews, evidence survey).

**** **Ethical concerns:** All data sources mention ethical concerns as a barrier to RIE initiation[11] (see Figure 13). The main challenge is how to treat the comparison group in an ethically responsible way (e.g. how to deal with the fact that one group receives treatment and the other – at least initially – does not). Also, there is a question as to whether the additional workload for partners and target groups caused by an RIE is ethically justified. Also, distress for researchers when working in low- and middle-income countries has recently been added to the field of ethical concerns (Steinert et al., 2021).

**** **Administrative effort:** The additional administrative effort related to RIE (e.g. regarding procurement) can pose a barrier to RIE initiation. The evidence survey identifies this barrier as being of comparatively low relevance (see Figure 13). Further, the prospect of managing a complex stakeholder coordination and communication process (on top of the regular stakeholder management for project implementation) can hinder RIE initiation at project level.

Resources

*** **Limited time and financial resources:** A lack of resources – be it money or time – is one of the primary barriers to RIE initiation.

- Lack of **time** was mentioned as the most relevant factor hindering RIE initiation among respondents to the evidence survey, even more so than financial resources (see Figure 13). The stocktaking exercise also mirrors this priority of time resources over financial resources. In concrete terms, interviewees mentioned issues such as high workloads or lack of dedicated RIE staff or time allotments.
- The evidence survey identified **financial** resources as the third most frequently indicated barrier to RIE initiation. RIEs are relatively expensive compared to other types of evaluations. Therefore, they often cannot be financed out of the general monitoring and evaluation (M&E) budget but require additional funding. This might also mean that money would need to be diverted from project activities.

[11] The most frequently raised ethical concern that the control group is denied the intervention has been debated in the literature for many years and has been largely refuted by the fact there is never anything taken away from any group. Rather, it takes advantage of the fact that, with limited resources, not everyone in need can receive an intervention (see also *https://www.povertyactionlab.org/sites/default/files/research-resources/2017.04.14-Real-World-Challenges-to-Randomization-and-Their-Solutions.pdf*).

**** **Limited qualified RIE support (internal and external):** The implementation of RIE requires specific skills (e.g. pertaining to research design and statistical data analysis), which are not widely available among internal and external evaluators within GDC. Those skills are present at both international and national universities and research institutes. National universities and research institutes, however, have so far mostly worked with international and in particular multilateral donors. The increasing capacities and skills of German universities and research institutes to conduct RIE are therefore not yet brought to bear on relevant research questions of GDC organisations.

Intention

**** **Lack of culture of learning from failure:** On a systems and organisational level, a lack of an actively practised and vivid culture of learning from failure (see Box 5) hinders the initiation of RIE. The interviews revealed the perception of many that GDC evaluation in general – and RIE in particular – is primarily considered a means for realising a simple understanding of accountability rather than an opportunity to learn and improve (although accountability – understood more adequately – actually includes learning and improvement). Interviewees even reported fear of negative consequences for the project or organisation if they had to report negative results from an RIE. However, both the stocktaking exercise and the evidence survey did not identify the lack of a culture of learning from failure among the highly relevant barriers (see Figure 13).

Box 5 In short: Culture of learning from failure

There is no universal definition of a culture of learning from failure (sometimes shortened to "learning culture"). However, there are several concepts that most definitions share: within a culture of learning from failure, (1) learning is integrated into the systems, values and resources of an organisation and allowed to shape strategy and process, (2) learning is promoted and rewarded at an individual, team and organisational level so that individuals feel empowered to fail and learn from mistakes and (3) leaders embrace open dialogue and reflection. Generally, a culture of learning from failure applies to the individual, team, organisational and inter-organisational level (CPID, 2020).

**** **Limited leadership support for RIE:** A lack of leadership support can hinder RIE initiation. A number of interviewees and respondents to the stocktaking exercise with experience in RIE implementation mentioned that they had to overcome concerns in the hierarchy within their organisation and felt they had to take a certain risk when initiating an RIE.

**** **Personal attitude towards RIE:** A general, often nonspecific, personal disapproval of RIE as an Anglo-American trend, and the perception of RIE as a technocratic and purely quantifying venture that does not focus on the qualitative impacts for beneficiaries, can be a general barrier to initiating RIE. Few interviewees mentioned this barrier.

3.3 What are potential measures to foster the initiation of RIEs?

In this chapter we present a comprehensive list of potential measures to foster RIE initiation:

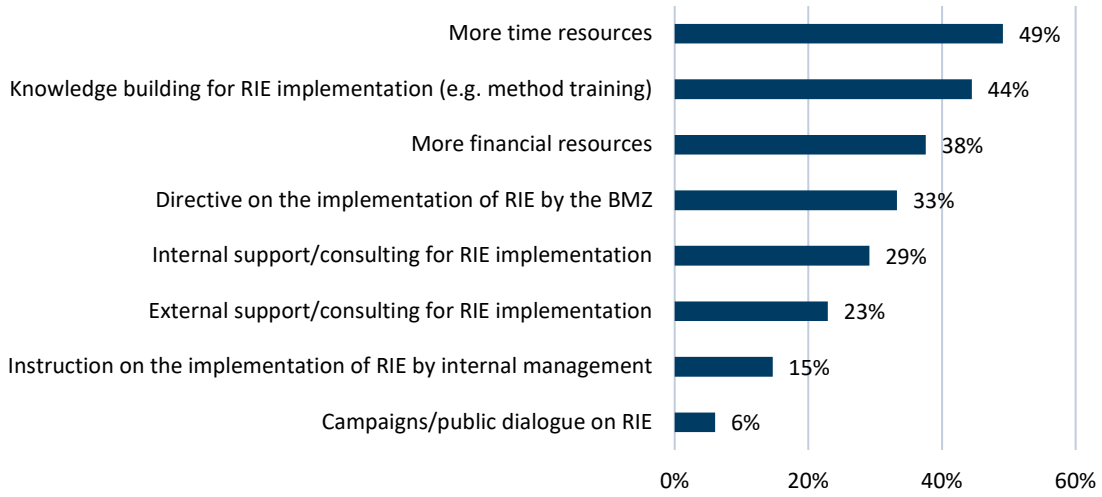
- novel measures that have not yet been trialled in GDC or by other development partners.
- measures that have turned out to be actual facilitating factors for the initiation of RIE in concrete past instances within GDC.
- measures that have been implemented by other development partners and therefore plausibly have the expected effect.

In total, we identified 17 potential measures (presented below in alphabetical order). Measures can be explicit (e.g. issuing a directive) or implicit (e.g. changing culture), push (e.g. issuing legislation) or pull (e.g. creating RIE awareness), demand-side oriented (e.g. building awareness) or supply-side oriented (e.g. establishing an RIE support structure).

Allocate financial resources and time to RIE:

- Allocating **more time** and sufficient personnel to RIE is one of the primary potential measures to foster RIE initiation within GDC. With 49% this is the most frequently mentioned measure in the evidence survey (see Figure 14). Interview partners stated that additional personnel or larger shares of the available working time dedicated to M&E could facilitate RIE initiation.
- Providing **additional financial resources** is another primary measure to facilitate RIE initiation. All data sources name it among the most relevant potential measures. In the evidence survey, 38% of respondents to this question perceive additional financial resources as particularly helpful (see Figure 14). Financial resources can be provided by topping up project budgets (without taking away resources from planned activities) or by establishing a central fund or support programme for RIE that guides financial resources into strategically relevant topics. Partnering with other donors ("co-funding") or research institutions can also be an internationally tested option (e.g. at USAID, FCDO, AFD, IDB or Oxfam GB) either on a project basis or by establishing a central fund.

Figure 14 Rating of potential measures to increase RIE initiation



Source: evidence survey; question: "What measures do you consider to be particularly helpful to conduct more RIE in German development cooperation?"; N=580; % of respondents; multiple answers possible.

Allow for less costly and more pragmatic RIE solutions: The willingness of project teams, administrative staff and researchers to find pragmatic solutions to challenges during RIE implementation is a facilitating factor for initiating RIE. Finding pragmatic RIE solutions can mean realising an RIE at short notice and generating findings more quickly (e.g. through real-time data analysis or rapid evaluations) or deviating from the experimental "gold standard", while pursuing high-quality statistical analyses (e.g. using quasi-experimental designs). It can also mean collecting context-relevant data remotely or exploiting synergies regarding data. For example, data collection needs for RIE can be integrated with existing project or client demands (e.g. project monitoring, market research). It may also be possible to make use of existing administrative data (e.g. from management information systems of partner financial institutions, insurers, etc.) or other secondary data (e.g. geodata, mobile phone data). These measures can reduce financial and time resources required for conducting an RIE.

Build RIE capacities: All data sources and all stakeholder groups name building RIE capacities as a high-priority measure. About 45% of respondents in the evidence survey perceive knowledge building for RIE as particularly helpful (see Figure 14). Capacity development measures should not only target decision makers and employees within GDC, but also include evaluation capacity development (ECD) in partner countries. At the same time, GDC can learn from partner countries like Mexico, Costa Rica and South Africa that have a long history of, and strong expertise in, evidence-based policy making. Building RIE capacities should follow the systemic approach, targeting individuals, networks, institutions and the enabling environment. Topics can cover the following aspects: fundamental conceptual knowledge ("What is impact evaluation and what is it good for?"), methodological knowledge, RIE procurement and management, practical RIE implementation knowledge and knowledge on evaluation management and dissemination.

> But of course you also have to meet the people in such an exercise, so that they know why this is actually important, and also give them a little help. How does it work? I think many people are also overwhelmed by the debate and don't feel comfortable. So the simple toolbox for everyone is still missing. How do I actually do that in concrete terms? I think one knows that when people are empowered to participate, they change. (Int_35)

In recent years a number of ECD initiatives have included RIE in their programmes, including the new Global Evaluation Initiative by the World Bank and UNDP with funding from BMZ, or DEval's FOCELAC project. GDC can take advantage of those activities. Having (or hiring) staff with previous RIE knowledge on the project team can – according to our interview partners – contribute to building RIE capacities and thereby be a supporting factor for the initiation of RIE.

Create awareness of RIE: To foster the initiation of RIEs within GDC, it is important to create awareness of RIE among GDC stakeholders.

> When you're not aware, you've got a problem…. Promoting the use of evidence by decision-makers requires behaviour change, so we can use a behaviour change framework. At the start decision-makers are not aware there's any problem. "We don't need evidence…. We know what we're doing. We've always done it this way. It works. So why should we change?" So the first step is to create awareness. So decision-makers think, oh, actually, maybe there's a point in using evidence. (Int_Exp_3)

This can mean pointing out specific uses of RIE, particularly the causal attribution of impacts to interventions. It can also mean pointing to the potential of a solid impact communication to the public or to the fact that the costs of projects with low actual impacts are higher in the long run than costs associated with RIE. Creating and publicising positive RIE examples, emphasising the international relevance of RIE in the sectoral dialogue, and engaging in open and honest dialogue about ethical concerns can also contribute to awareness building. Awareness building among employees in leadership positions may be particularly beneficial as it signals support and creates demand (source: report on international experiences). Interviewees encouraged this measure more than respondents to the evidence survey did – only 6% of respondents considered campaigns and public dialogue as a helpful measure (see Figure 14).

Create opportunities for RIE exchange: Creating opportunities for networking and exchange of RIE experiences within but also beyond their own organisation is another potential measure to foster RIE initiation. This can take the form of peer exchange, research–practice exchange or policy–practice exchange, and includes reflections on practical implementation questions as well as existing evidence gaps.

Define roles and responsibilities: One potential measure to foster RIE initiation can be to clarify which actors are involved in the process of initiating and setting up an RIE. This means defining roles and the division of responsibilities, both between organisations within GDC (on a systems level) and within organisations (e.g. between central evaluation units and thematic units).

Demonstrate leadership commitment to RIE: Leadership commitment is a strong facilitator for RIE initiation within GDC and internationally (source: report on international experiences). Specifically, such leadership commitment can be expressed by giving priority to evidence generation and take-up, by functioning as role models and explicitly using RIE findings for decision making or by rewarding evidence generation and take-

up. Interviewees from different stakeholder groups pointed to the central role of evidence advocates in leadership positions.

> And when people realise that their evidence, which they have generated with a great deal of effort and time investment, is not being heard and that the political interests of the donor are always in the foreground, then they will very quickly come to the end of the line. Yes? So there must also be a willingness on the part of the donor to support evidence-based decisions. And this tension between political decisions and evidence-based decisions must be balanced. (Int_9)

Develop guidance tools for RIE initiation: Having clear guidelines on when and how to initiate an RIE (i.e. an RIE evaluability assessment) can support the initiation of RIE. Such guidance tools should cover aspects of practical and methodological feasibility as well as the usefulness of RIE in a given project and context. This measure is particularly useful to foster a more systematic initiation of RIE, as international experiences, for example at USAID, Oxfam GB, FCDO and IDB suggest. The DEval RIE project has developed such a practical guidance tool (see *DEval, 2021*).

Embed RIE in a strategy for evidence generation and take-up: Since take-up of results of an RIE is generally a core motivation for initiating an RIE, guaranteeing evidence take-up from the onset of an RIE is a relevant measure for its initiation. Evidence take-up in turn is facilitated when (rigorous) evidence is generated on operationally, strategically or politically relevant questions. In consequence, this means that such relevant questions must be defined prior to or at the beginning of an RIE, and that they should be supported by an explicit strategy for evidence take-up by both donors and partners. Such a strategy for evidence take-up may be limited to the project level, yet should ideally be embedded in an organisational or systems-wide evidence system. To implement this measure, a set of central questions with importance to GDC should be defined (e.g. within the scope of the core topic strategies or initiative topics) such that they can be systematically addressed with RIE and other evaluation approaches. The subject of measures for evidence take-up is discussed in more detail in Section 4.3.

Ensure close collaboration between researchers and other stakeholders: To foster RIE initiation, it is helpful if researchers engage in close collaboration with all other stakeholders at the beginning of an RIE and, in particular, with project partners in partner countries. In this process, evaluators and researchers should present themselves as learning partners and supporters rather than auditors and "judges". This approach can help emphasise the learning purpose of the RIE over that of accountability. Understanding stakeholder demands, and making use of and involving local capacities, can be concrete steps to generate relevant results, enhance support and build ownership for the RIE (source: report on international experiences).

Ensure early coordination of RIE and project cycles: To facilitate RIE initiation, it is important to systematically check at early stages in project cycles for the possibility of conducting an RIE (e.g. using an RIE evaluability assessment tool (*DEval, 2021*)). This can be done by integrating RIE reflection points in existing process flows and in guidelines about how to create module proposals or brief statements. Defining such entry points for RIE in existing process flows can foster a more systematic initiation of RIE. This also allows early involvement of RIE experts, which is a helpful measure to initiate methodologically sound RIE (e.g. by ensuring that project targeting of beneficiaries and requirements for an RIE research design are compatible and allow for the creation of valid comparison groups). In practice, early coordination between the project cycle and a possible RIE was achieved by either involving RIE experts already in the project proposal or design phases (after commissioning), or by the RIE being proposed as a component of the commissioned project.

Generate additional knowledge beyond the RIE: Interviewees highlighted the generation of additional knowledge beyond the results from the RIE as an effective measure to foster RIE initiation. This might be descriptive information, needs assessments, cost calculations and process evaluations. First, answering these additional questions has an added value in itself. Second, the additional knowledge can be useful when interpreting the results of an RIE.

> … if you do an impact evaluation… just having your identification strategy, but you have not informed it also by understanding how implementation is working, how the process is working,… what are the costs, so… all these other things around it, then you're… not in a position really to come with good recommendations. You're not fully informed if you don't also have those other parts. And I think a lot of

> the first-generation impact evaluations were not well informed by context, by implementation, by process. And so I think they lost a little bit of credibility.… We're no longer there. So there's the package of things one can do. (Int_Exp_4)

The additional knowledge was either generated by additional data (often qualitative data within a mixed-methods design) or by extensive analysis of baseline and follow-up data from the RIE. As exemplified by the report on international experiences with RIE, this approach is increasingly followed by USAID and J-PAL as well as IPA and DIME.

Include RIE in the organisation's lead strategy: To encourage RIE initiation it would be helpful for GDC organisations to commit to evidence-informed policy making in their strategies (source: report on international experiences). Those documents provide orientation and help anchor RIEs during changes in leadership. BMZ's recent strategy document for evaluation in GDC is a good example of such an effort in that it explicitly encourages impact orientation and the role of accompanying research such as RIEs within GDC (BMZ, 2021).

Pass formal legislation or requirement: Other potential measures to increase RIE initiation are formal legislation or requirements at the national level. The United States passed such legislation in December 2018. The US Evidence Act, amongst other things, requires a number of governmental agencies to generate their own, or draw on existing, evidence to answer relevant policy questions. The Biden administration reiterated its commitment to evidence-based decision making in a recent memorandum mandating federal agencies to "consider, supplement, and support their plans for forming evidence-based policies", which also includes evidence-building (Biden, 2021). Mexico's 2004 Social Development Law requires external evaluation of all government-funded social programmes (White, 2019), and South Africa has had a mandatory system of evidence assessment since 2015. Instead of national legislation, RIE initiation could also be formalised by requirements passed by the respective government department or by binding organisational instructions. As pointed out in Section 3.2on barriers, the Federal Budget Regulations have not yet been understood as requiring more RIE within GDC. In order to encourage the initiation of RIE, formal requirements could necessitate project implementers to conduct:

- RIEs in projects above a certain budget (threshold)
- a number of RIEs per year in each sector (quota)
- RIEs within (annually) changing thematic windows
- RIEs in all pilot projects or in any project that is intended to be scaled-up
- a standardised check for RIE feasibility and usefulness in existing process flows and respective internal guidance documents.

An RIE-related directive by BMZ was perceived as a helpful measure by 33% of respondents to the evidence survey, whereas internal instructions were perceived as less relevant and considered helpful by 15% of respondents (see Figure 14). Interviewees' opinions on the value of formal legislation and requirements varied. Whereas quota regulations or thresholds for RIE were largely deemed unhelpful, top-down requirements or clear donor signals were generally judged quite effective in establishing a more systematic use of RIE. "So, quite honestly, the first responsibility lies with the management of the BMZ… so if the management gives a clear direction that you want to become better there and want to expand there, then you trigger that down to the units anyway." (Int_25)

Set incentives for RIE: Next to formal regulations, setting incentives can enhance the initiation of RIEs, for example by including the initiation of RIEs or the take-up of evidence in annual staff targets, only scaling up interventions that have been rigorously evaluated, or giving visibility to evidence champions (source: report on international experiences). Evidence champions are influential members of staff from GDC institutions or well-respected international organisations who push for RIE and evidence-informed decision making. They function as role models and create interest in evidence generation and take-up. Giving visibility to evidence champions could be done by organising events in which best practices are presented, or by setting up project awards that honour evidence-generation through RIE. Interviewees also proposed anchoring RIEs in management board goals and including RIE as one criterion for assessing brief proposals when commissioning

new modules. To incentivise researchers to conduct policy-relevant RIEs, the academic reputation system would need to change such that policy-relevant research receives more appreciation. This is unlikely to happen in the short-term, but the UK research councils' "Impact Agenda" has shown that insisting on tracking the policy influence of research as an important criterion for judging research quality can promote a cultural change over time (Smith et al., 2020).

Strengthen a culture of learning from failure: Strengthening a culture of learning from failure can support RIE initiation. In the interviews, the international experts stressed this aspect in particular: "So the other big thing I would say what your agency needs to do is they need to accept failure and learn from failure. So a nice aphorism is you should penalize the failure to learn from failure, not failure itself, you know…." (Int_Exp_3)

This means, for instance, putting more emphasis on honest and transparent communication about errors – particularly among leaders who function as role models – and a change of mindset towards seizing the opportunity to learn from mistakes.

> Basically, you need to establish a mode of communication, or rather a relationship between donors and partner organisations, in which it is clear that recipients of funds can make errors, are allowed to detect weaknesses, and then jointly work on finding solutions to those weaknesses. Obviously, this will produce different results in comparison to when they [recipients of funds] feel that they have to keep the pretence and hope that nobody looks closely. (Int_19)

Explicitly allowing for phases of experimentation in which different interventions or implementation approaches are tested and refined using evidence from one or multiple RIEs could be an effective measure to foster RIE initiation. Although this will not be something easily applied to all GDC projects, it could be particularly useful for innovative and pilot projects.

Strengthen (existing) or build new support structures for RIE: Strengthening support structures for RIE, which can either provide direct support or link those looking for support with external structures (e.g. researchers or consultancies), is a highly relevant measure to facilitate RIE initiation (see Figure 14). Being able to turn to a support structure was a particularly relevant facilitating factor from the perspective of project teams who wanted to conduct an RIE. This can mean establishing new support units or strengthening existing units. Such support structures can either be internal or external and can offer services such as project-based consulting, design clinics or matchmaking workshops. GIZ's evaluation unit established a support structure for RIE in 2020 (Bräuer et al., 2020: 80). Similarly, KfW Development Bank's evaluation unit recently promoted the possibility of receiving support for conducting an RIE (KfW, 2021).

3.4 Linking highly relevant barriers and potential measures

This section links highly relevant barriers (Section 3.2) with corresponding potential measures (Section 3.3). We assessed the following barriers as **highly relevant (***)** to the hindrance of RIE implementation in GDC:

- diverging interests
- limited financial resources and time
- limited RIE capacity
- low priority given to RIE evidence
- methodological challenges
- no formal requirement for RIE
- unclear benefit of RIE
- uncoordinated timing between project and RIE.

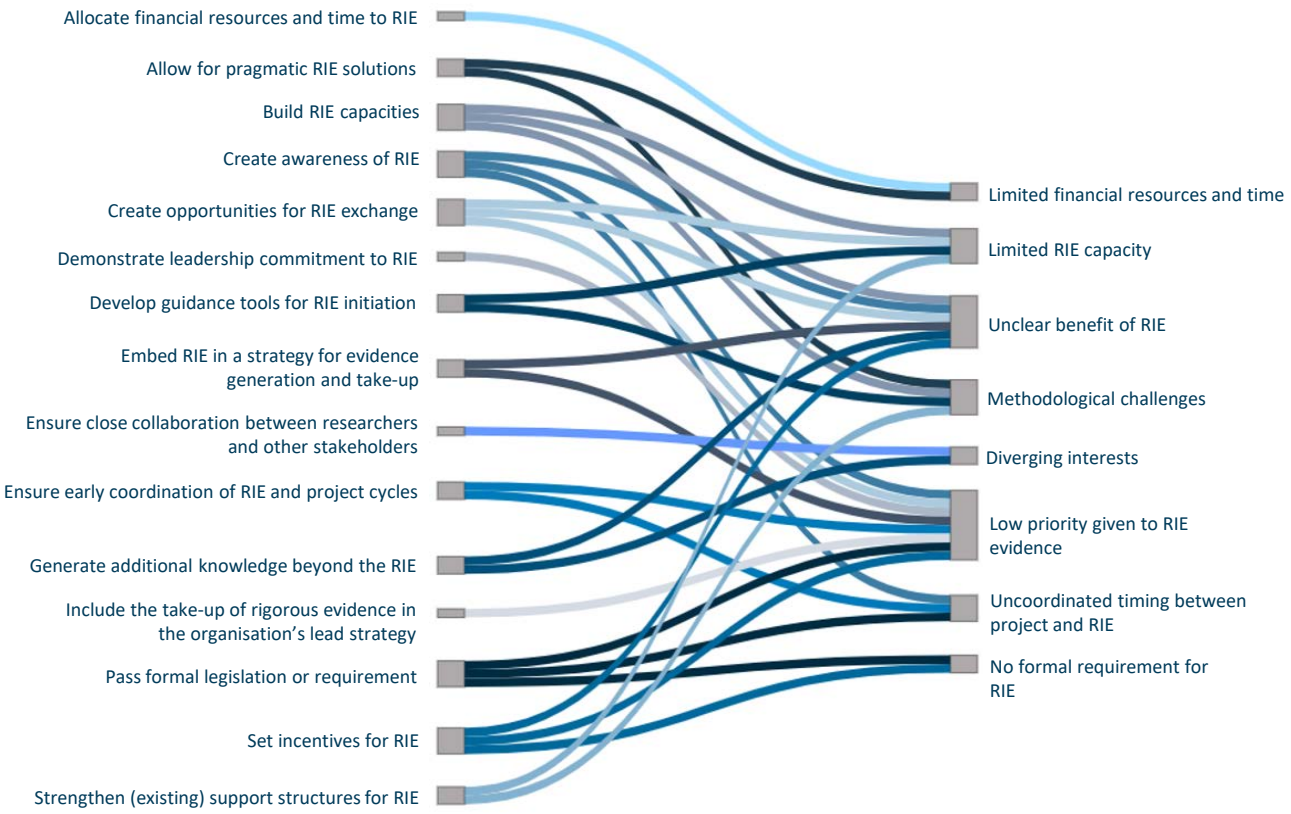
We also analysed whether the relevance of barriers varies across **stakeholder groups** (BMZ, GIZ, KfW Development Bank, other IOs, CSOs, private sector, DEval). We found only slight variations across stakeholder groups when comparing the ranking of barriers in the evidence survey across groups. Also, when disaggregating responses according to **RIE experience**, the barriers identified as most relevant are identical,

irrespective of the amount of experience. Hence, we present a unified matching of highly relevant barriers and potential measures.

We have presented barriers and potential measures in a clearly delineated way. Yet, during our analysis it has become clear that barriers and potential measures are in fact often **interlinked or even hierarchical.** For example, there is a close link between the barriers "low priority given to RIE evidence" and "limited financial resources and time" or a hierarchical relationship between the measures "build RIE capacities" and "generate additional knowledge beyond the RIE".

Finally, when linking highly relevant barriers with the potential measures of Section 3.3, it becomes apparent that one potential measure can often **address more than one barrier**. Figure 15 illustrates these links between potential measures and barriers. Here, we only present measures that can obviously and directly be linked to the barriers. Some potential measures might also indirectly impact relevant barriers. In sum, this analysis suggests that a set of interrelated and complementary measures is needed to strengthen RIE implementation within GDC.

Figure 15 Linking potential measures and highly relevant barriers for RIE initiation



Source: own illustration

4. FINDINGS: TAKE-UP OF RIGOROUS EVIDENCE

This chapter is divided into five sub-sections: Section 4.1 illustrates the current status quo regarding evidence take-up in GDC. Sections 4.2.1 and 4.2.2 present and assess barriers to the six steps to a more systematic take-up of self-generated and global evidence (see ToC in Chapter 1). Section 4.3 provides a comprehensive list of potential measures to foster evidence take-up. Section 4.4 brings together the most relevant barriers and potential measures to address these barriers.

Together with the analysis of the status quo, barriers and measures for RIE initiation (Chapter 3), this analysis will inform our definition of implications in Chapter 5.

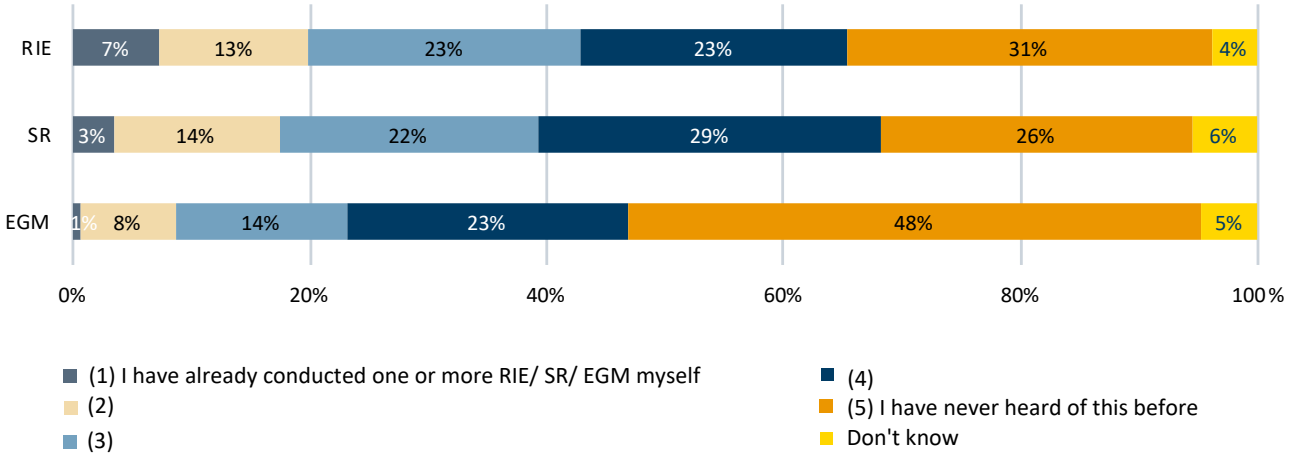
4.1 Status quo in GDC

In this section, we describe what people know about rigorous evidence, which types of evidence are currently used in GDC, to what extent they are used and for which purposes. We structure this sub-section according to the following three groups of evidence: **self-generated rigorous evidence, the global RIE evidence base,** and **any other type of evidence**. We define self-generated rigorous evidence as findings from RIEs that have evaluated a specific project within GDC and which are used in that context. We understand the global RIE evidence base to comprise RIEs that have been conducted by other projects within one's own organisation or by other organisations globally, including development partners or other donor countries. By any other type of evidence, we refer, for instance, to process evaluations, policy briefs or briefings from colleagues.

Knowledge about rigorous evidence

Since knowledge about RIE is a precondition for the take-up of rigorous evidence, respondents in our evidence survey indicated whether they were familiar with RIEs, SRs and EGMs (see Figure 16). Thirty-one percent of respondents had never heard of RIEs before and 23% had a vague idea about RIE. Almost 50% of respondents had never heard of EGMs before and 24% only had a vague idea of EGMs. Respondents were more familiar with SRs: 26% had never heard of SRs before – still, only 17% reported knowing SRs well. In sum, knowledge about these formats of rigorous evidence is rather limited among GDC employees (source: evidence survey).

Figure 16 Knowledge about RIE, SR and EGM



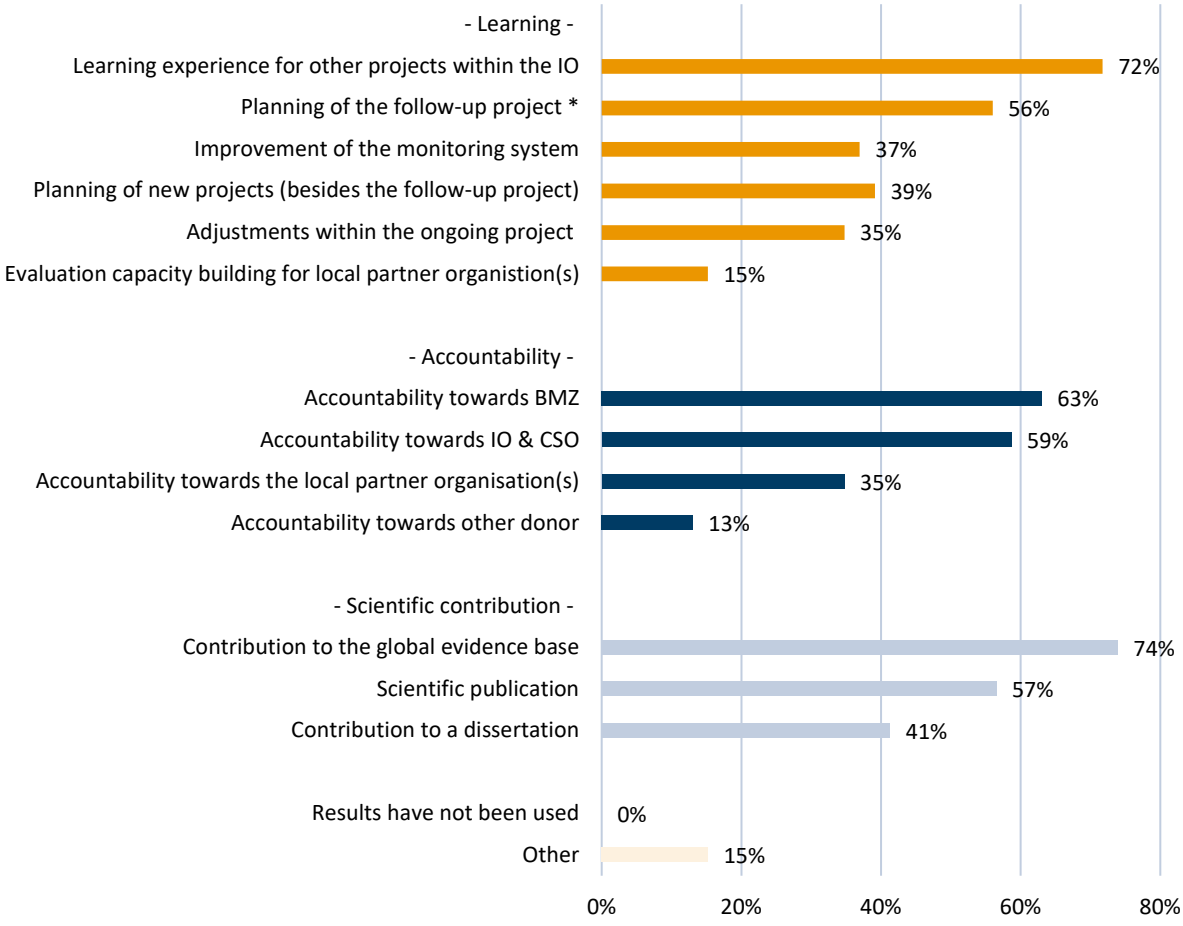
Source: evidence survey; Question: "How familiar are you with RIE, SR, EGM?"; Scale 1–5; N=839.

Take-up of self-generated rigorous evidence

We analysed the reasons for take-up of self-generated rigorous evidence along the classical functions of evaluation: accountability and learning (Højlund, 2015; van der Meer Edelenbos, 2006; Stockmann 2004), which both build on knowledge generation.

All respondents in our RIE stocktaking exercise reported that they used the results of their respective RIEs for at least one or more of learning, accountability and scientific contribution (see Figure 17).

Figure 17 Purposes of take-up of self-generated RIEs



*Source: stocktaking survey; question: "What have the results of the RIE been used for?"; N=51; multiple answers possible, only completed RIEs included. *For this response option we only included completed RIEs with existing or planned follow-up projects (N=9).*

Learning can take place at different levels. We examined a) operational learning in the project in which the RIE has been conducted, and b) learning at the organisational level, either for operational learning for other projects or for strategic decisions.

In the stocktaking exercise, only 35% of respondents indicated that findings of RIE were used for operational learning within the project in which the RIE had been conducted ("adjustment within the ongoing project"). Accordingly, our interviewees indicated that even though implementers aimed at using their RIE findings for project decisions, they could only realise this in a minority of all cases.

> Actually, almost all projects that decided to conduct an accompanying impact evaluation wanted to use the results for project steering... but when it comes to the results of this accompanying impact evaluation, hardly any project was able to implement these within a standard period of three years, because the entire implementation [of an RIE] took so long. (Int_4)

Our analysis of a small sample of project documents further indicates that even in projects that conducted an RIE, this RIE was not presented prominently in the project documents (e.g. within the yearly project reporting). Hence, findings of RIEs are not used very much at the project level.

In contrast, we see a higher percentage of take-up of findings from RIEs for follow-up projects. Respondents reported that in five out of the nine cases for which follow-up projects were implemented the RIE was used

in the design of the follow-up project (see Figure 17). Still, in four cases the RIE results did not inform the follow-up project. This is striking, as the use of previous evaluation findings for the planning of follow-up projects is a formal reporting requirement from BMZ. One possible explanation for this could be that the specific intervention or project activity was not continued in the follow-up project.

The most common purpose for using rigorous evidence (mentioned by 72% of respondents) is operational learning for other projects within the organisation. However, we do not know if the learning experiences lead to adjustments or discontinuations in other ongoing or planned projects or not. From our interviews we gather that learning may mainly involve the sharing of findings, for example in sectoral dialogue formats. Thirty-nine percent of respondents indicated that RIE findings were used for planning new projects (source: stocktaking exercise). Although this suggests some take-up of results of RIEs for other projects within the organisation, our interviews indicate that rigorous evidence is hardly used for strategic decision-making at the organisational level.

> Look at the strategies; they are full of normative theses. But they don't somehow say that this has proven itself. ... There is no empirical foundation for such things. I hardly know of any strategy that says: We've had great experiences, they're robust, and that's why we're going to continue. (Int_35)

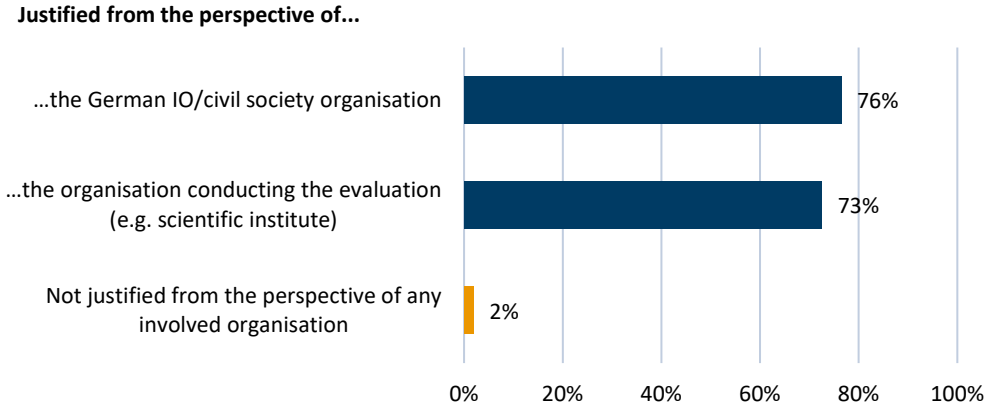
Sixty-three percent of respondents reported having used their RIEs for **accountability** towards BMZ and 59% for accountability towards their own organisation. Respondents less frequently mentioned accountability vis-à-vis local partners (35%) or other donors (13%) (source: stocktaking exercise). In some cases, interviewees used an RIE to establish cooperation with external stakeholders:

> And as I said in all these discussions, it was always the project manager and the head of cooperation [i.e. focal point in the embassy] who went from development partner to development partner and from donor coordination meetings to donor coordination meetings with the study in hand and the beautiful graphics waving and saying, here, the method works. (Int_20)

The majority of RIEs conducted in GDC made a **scientific contribution**. Seventy-four percent of respondents stated that their RIE contributed to the global RIE evidence base in a broader way, 57% of the RIEs were used for scientific publications and 41% of the respondents reported that the RIE was part of a dissertation (source: stocktaking exercise).

In our stocktaking survey we also asked whether the benefits of the RIE had justified its costs and effort (see Figure 18). Seventy-six percent of respondents indicated that the benefits had outweighed the costs for project implementers and 73% reported that it was worthwhile for the organisations conducting the evaluation. Only one respondent (2%) reported that the costs and benefits were not justified for any of the stakeholders involved. In sum, the majority of those with RIE experience in their own organisation held that the benefits of an RIE exceeded its costs (source: stocktaking exercise).

Figure 18 Benefits from RIEs vs. costs and efforts



Source: stocktaking survey; question: "Has the overall effort/cost of the RIE been justified by the generated benefit of the RIE?"; N=51; multiple answers possible, only completed RIEs included.

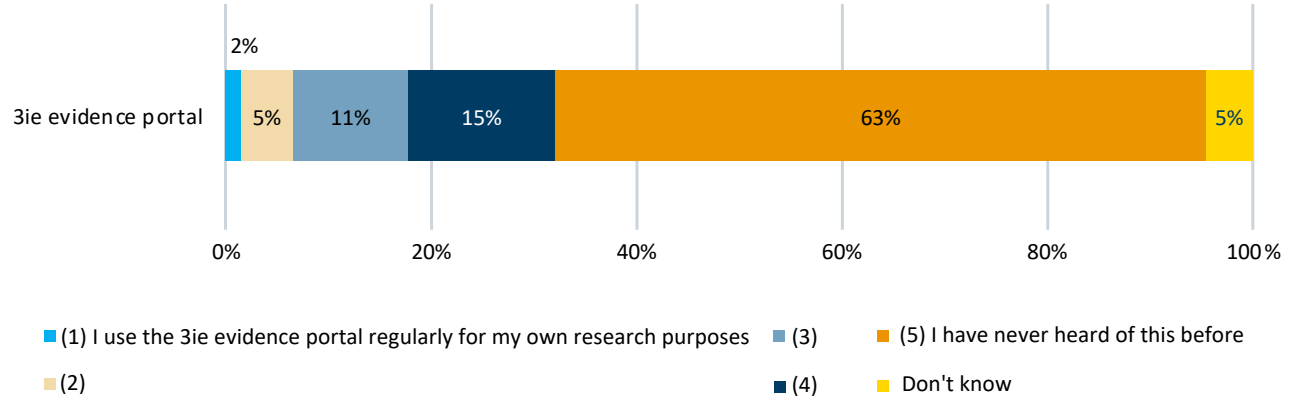
Overall, our different data sources show that RIEs were used partially for learning and accountability and that they contributed to knowledge generation. Evidence take-up for ongoing as well as for strategic decision-making in particular was rather low, despite the fact that implementers place strong emphasis on the benefits of their RIE.

Take-up of global RIE evidence

In this section, we examine the extent to which the global RIE evidence base is used in GDC. We approximate the global RIE evidence base with 3ie's DEP – the largest global database for high-quality RIE in international development. In September 2021, it comprised 6 588 RIEs, 732 SRs and 28 EGMs from diverse sectors. Whereas it is not completely identical with our definition of the global RIE evidence base (e.g. due to time lags in the inclusion of studies, the non-identification of relevant RIEs, and the high methodological requirements for including studies), there is a large overlap of the high-quality global RIE evidence base and studies included in 3ie's DEP.

A necessary precondition for using the DEP is knowledge of its existence. In our evidence survey, 63% of respondents reported that they had never heard of 3ie's DEP before and only 2% reported using it regularly (see Figure 19). The fact that 3ie's DEP is largely unknown in GDC is supported by our interviews. Accordingly, many interviewees shared the impression that the global evidence base is used very sparingly in policy-making and programming in GDC.

Figure 19 Knowledge about 3ie's DEP



Source: evidence survey; question: "How familiar are you with 3ie DEP?"; N=839; scale 1-5.

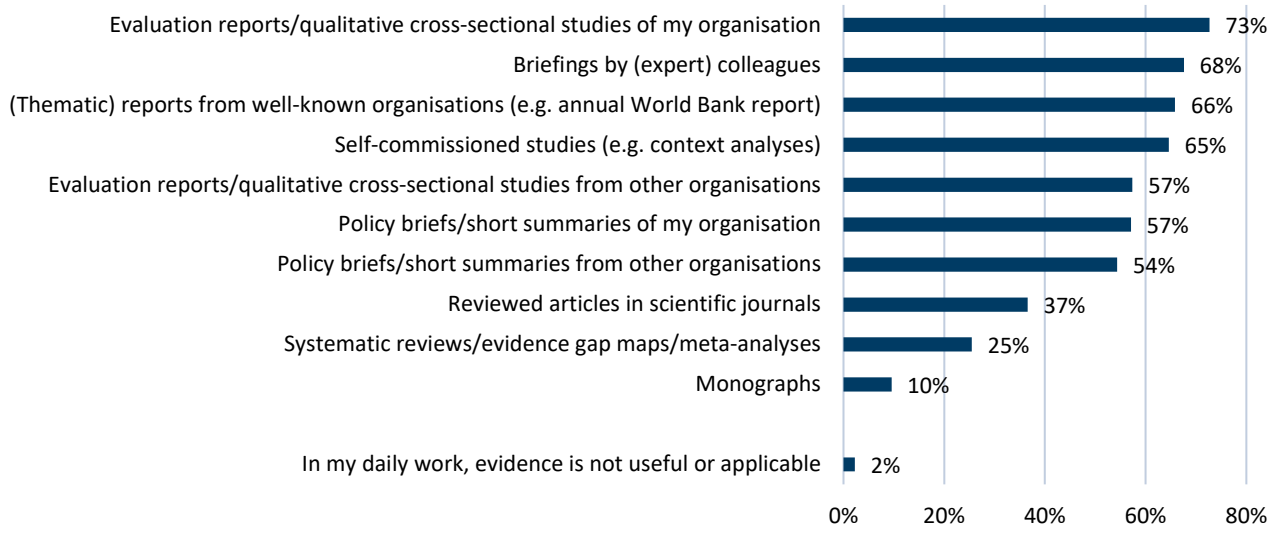
Our analysis of project documents revealed that scientific sources in general, and sources describing rigorous evidence in particular, are rarely referenced in project documents. In 124 documents, there were 14 explicit references to scientific publications in four different projects. These included five RIEs. Three of these RIEs were also registered in our stocktaking survey, and findings were published as scientific papers in cooperation with a policy-oriented research organisation. Information on the other two RIEs indicates that these studies demonstrate only limited rigour, for example due to very small sample sizes (one study had only 100 observations). RIEs are neither highlighted in the project documents nor named as such in most cases. Instead, we found descriptions such as "social experiment" or "control villages". Our different data sources strongly indicate that using external scientific evidence and explicitly highlighting those sources (e.g. providing a list of references) is not common practice, since it is not required in the formal reporting processes between BMZ and IOs or CSOs. At the same time, employees from BMZ mentioned that they expect the state of the art of the scientific literature to be (implicitly) represented in project proposals and other formal reporting documents.

In sum, 3ie's DEP is largely unknown to employees in GDC. Against this background, it is not surprising that our data indicates that results of external RIEs are rarely used and that rigorous evidence is hardly cited in formal project documents.

Take-up of any other type of evidence

Employees working in GDC regularly make use of non-rigorous sources of information. Respondents to our evidence survey most often reported using evaluation reports (73%), briefings by colleagues (68%) and reports from well-known organisations (66%; see Figure 20). Forty-four percent of respondents who reported using own evaluation reports, use this evidence format at least once a month. As for briefings by colleagues, 40% of respondents who use this format use them at least once a week. Two percent of respondents stated that evidence is not useful or applicable to their daily work (source: evidence survey).

Figure 20 Take-up of any other type of evidence



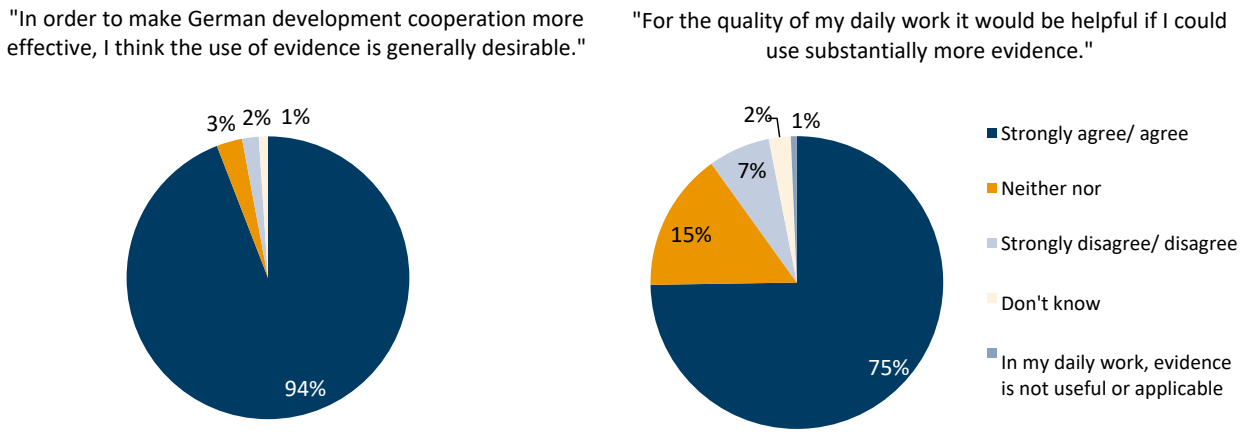
Source: evidence survey; question: "What specific types of evidence do you use for your current work?"; N=1,025; multiple answers possible.

On average, we found references to about five non-rigorous sources of information per document in our analysis of project documents. Most frequently, these sources were local government documents, followed by official (national) databases and other reports such as background studies conducted by consultants. Even though none of these sources uses rigorous evaluation designs allowing reliable causal inference, they were still used to make statements on causal links between the project and its intended outcomes or impacts in more than half of the cases. Likewise, formal reporting documents, such as progress reports, frequently made claims about the impact of projects but did not support these claims with credible evidence.

Despite this low priority of (rigorous) evidence in the project documents, we see a very positive attitude of employees in GDC towards using evidence. A clear majority of respondents (strongly) agreed that evidence take-up is key to making GDC more effective (94%). Moreover, about 75% of GDC employees stated that more capacity to use evidence would substantially improve the quality of their work (see Figure 21). This high demand for evidence differs neither across organisations nor across different functions within an organisation. Respondents agreed that more evidence should be used during the planning and design stages of projects in particular.

The findings on take-up of any other type of evidence speak for a strong demand by GDC professionals for an evidence-informed approach to development cooperation. However, actual use and visibility of evidence in current formal procedures are low.

Figure 21 Attitude towards evidence take-up



Source: evidence survey; question: "To what extent do you agree with the following statement? In order to make GDC more effective, I think the use of evidence is generally desirable"; N=1054; scale 1-5. Evidence survey; question: "To what extent do you agree with the following statement? For the quality of my daily work it would be helpful if I could use substantially more evidence"; N=907; scale 1-5.

Overall, the potential of evidence from **self-generated RIEs** is not fully exploited in GDC. Lessons from RIEs are shared with other projects within GDC organisations. In some cases, they are also used for designing follow-up projects and for adjustments within the current project. Self-generated evidence from RIEs is not used for strategic decision-making. **The global RIE evidence base** is rarely used and very rarely explicitly referenced. In contrast, other **non-rigorous types of evidence** play a more important role in GDC employees' daily work. Claims about impact are typically made without referencing evidence at all or using non-rigorous evidence with limited validity for making strong causal claims. Despite the potential availability of rigorous evidence that is relevant to GDC (see Box 7, p. 49) and GDC employees' positive view on evidence-informed decision-making, they do not exploit the potential for the take-up of rigorous evidence. In the following sections, we analyse reasons why the take-up of rigorous evidence in GDC is hindered.

4.2 Which barriers hinder the take-up of rigorous evidence in GDC?

In the following section, we present barriers to the take-up of rigorous evidence in GDC. The analysis is divided into two sub-sections: one on barriers to take-up of **self-generated evidence** (4.2.1) and one on barriers to take-up of insights from the **global RIE evidence base** (4.2.2).

For self-generated RIE evidence, the analysis is mostly based on our interviews and the document analysis. For the global RIE evidence base, the analysis is based on interviews, the evidence survey and the document analysis.

Again, we express the relevance in the following way: Three stars (***) indicates a barrier of high relevance, two stars (**) one of medium relevance, and one star (*) one of marginal relevance. The classification is based on:

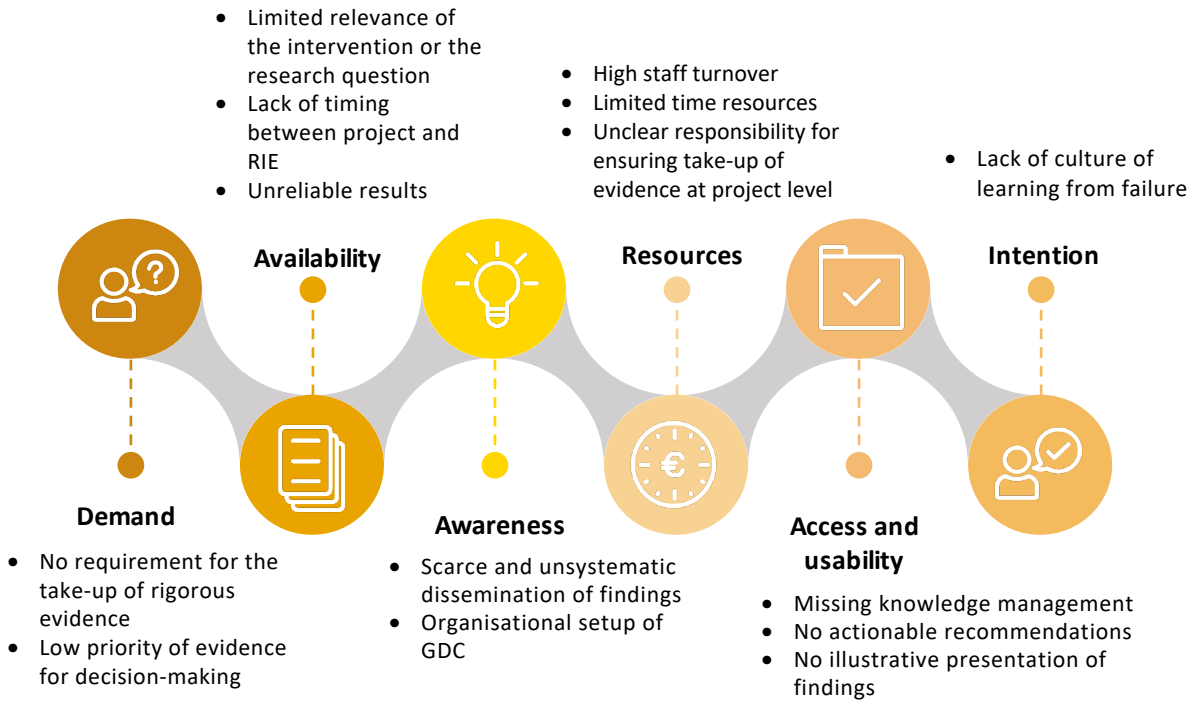
- the triangulation of information on the barriers from different data sources, and
- our evaluative assessment with respect to the reliability of interviewees' statements.

4.2.1 Barriers for the take-up of self-generated evidence

In our interviews we identified 35 code segments that mentioned a barrier to take-up of results of self-generated RIE evidence. Those code segments were distributed over 18 different interviews. All interviewed stakeholder groups, such as employees from BMZ, from IOs, CSOs, private sector and the scientific community mentioned barriers.

Based on the simplified ToC presented in our introduction (Chapter 1), we list barriers along the six steps leading to more systematic evidence take-up (see Figure 22):

Figure 22 Overview of barriers to take-up of self-generated evidence, mapped to the six steps leading to a more systematic evidence take-up



Source: own illustration

Demand

***** Low priority of evidence for decision-making:** Project and political decisions are based on a number of factors, such as (domestic) political priorities, priorities of partner countries or the need to spend funds quickly. Evidence is only one of those factors, which – according to our interviewees – is given low priority. In particular, a lack of support from the leadership level hinders the take-up of rigorous evidence.

> In my opinion, only a very, very small fraction of project decision is based on evidence. … At least I derive that from the fact that the evidence we provide through evaluations, of whatever kind, is already not used. So why would you use other evidence? Because projects and approaches are the result of a negotiation process between many stakeholders, and evidence, just as in politics, only accounts for a small part, if any. (Int_23)

In the perception of some interviewees from IOs, private sector and CSOs, there was a lack of interest in results of evaluations in BMZ. They reported the impression that evaluations are often purely administrative tasks or were used for formal accountability at best but not as a tool for learning. The lack of interest in

evidence is even stronger when it does not fit into the political or organisational agenda or the respondent's own beliefs.[12] In such cases, the evidence can be questioned, interpreted wrongly or ignored entirely: "Evidence is always appreciated if it fits into the political decision. If this is not the case, then it is just perceived as annoying" (Int_21).

***** No requirement for the take-up of rigorous evidence:** There is only a limited requirement for take-up of self-generated RIE evidence. For the yearly project reporting it is required to confirm or disprove the underlying project ToC by making use of lessons learnt and evaluations, if available. In our analysis of project documents, we observed that in the single case for which an RIE had been conducted, it was not reported very prominently (source: portfolio and document analysis). For module proposals of follow-up projects, guidance documents state that learning experiences from previous projects should be taken into account. Although it is not defined what the basis of the learning experiences should be, it would be appropriate to include evaluations, including RIEs, if they have been conducted. We did not find any requirement concerning the take-up of results from (rigorous) evaluations from other projects within the respective organisation nor concerning the use of the global (rigorous) evidence base. This is confirmed by our interviews and the open responses in the evidence survey. "Evidence [is] not a requirement of the client [BMZ] and is not rewarded [...]. There is no place for it in the format of progress reports and final reports." (open response from evidence survey). The use of evidence – if it happens at all – therefore depends on the preference of individual employees. "Well, there is no standard process that ensures that such evidence is used. It is very much dependent on the motivation of individuals" (Int_4) (see also Section 4.1).

Availability

***** Limited relevance of the intervention or the research question:** The take-up of RIE results was hindered when the intervention evaluated or the research questions asked were not sufficiently relevant. The perception of relevance of interventions or research questions might differ across stakeholders. For instance, at organisational and policy level, the take-up of evidence was hindered when the tested intervention was not of political or strategic relevance, when political priorities had changed over time, or when the intervention was not sufficiently innovative to create interest. At the same time project implementers judged questions on short-term project implementation that could be used for project steering as most relevant.

***** Lack of timing between project and RIE:** The lack of timing between project cycles and RIEs was a barrier to the take-up of RIE results. RIEs in general and RCTs in particular often take several years from their conceptualisation to the availability of final results. Project cycles that last about three years were often shorter than the period an RIE required. Results were often only available after projects had ended and could therefore not be used at the project level. Sometimes even follow-up projects were already planned and results could not be used in the design of this follow-up. In some cases, the RIE took longer than initially planned due to a multitude of practical obstacles (e.g. national holidays or delay in signing of contracts). The long period of time until results from RIEs were available was also a barrier at the organisational level because of the change in political priorities over time.

***** Unreliable results:** Few interviewees mentioned that sometimes methodological or practical implementation problems (e.g. invalid control group) emerged, which rendered the results of the RIE invalid and hence made it unusable. Also, sample sizes were often quite small and smaller effects could therefore not be detected. Methodological problems arose even though project implementers partnered with experienced researchers.

[12] This tendency to misinterpret, question or ignore evidence if it does not support one's own agenda or beliefs is related to a well-known cognitive bias: confirmation bias. A scientific publication by the authors of this report on confirmation bias and other cognitive tendencies in political decision-making is forthcoming.

Awareness and capacity

*** **Scarce and unsystematic dissemination of findings:** Interviewees reported that results of self-conducted RIEs were scarcely and unsystematically disseminated within the organisation and externally. For most of the RIEs no dissemination concept existed. Any dissemination activities of self-generated RIE findings were driven by the motivation of individual GDC employees or researchers.

*** **Organisational setup of GDC:** The organisational structure of GDC, in particular the division between political steering (BMZ) and practical implementation (IOs, CSOs etc.) and between financial and technical assistance, can operate as a barrier for RIE use. This division hinders self-generated RIE evidence to find its way from project level into strategic decision-making and from one organisation to another.

Resources

*** **High staff turnover:** High staff turnover has frequently been mentioned as a barrier to the take-up of self-generated evidence. Employees with intrinsic motivation for evidence-based programming, who had initiated an RIE, were sometimes not working in the respective project anymore when the RIE was completed. Hence, there was a lack of responsibility and ownership of take-up of RIE findings when they became available.

**** **Limited time resources:** Employees across all GDC organisations reported having very little time to engage with research findings. Whereas lack of time is a substantial barrier when engaging with the global evidence base (see Section 4.2.2), even employees that worked in the evaluated project and for which RIE results were potentially highly relevant, mentioned lack of time as a substantial barrier.

**** **Unclear responsibility for ensuring take-up of evidence at project level:** Interviewees reported that the core responsibility for the process to put findings from RIEs into actionable project adjustments – be it with the project implementers, the researchers or a third party – was often not clear.

Access and usability

*** **Missing knowledge management:** The take-up of self-generated RIE evidence at an organisational level was hindered because of the lack of an easily accessible knowledge management system for completed, ongoing and planned RIEs. RIEs were mostly initiated by decentral units (e.g. in partner countries) and central units rarely kept track of the initiated RIEs and their results. Hence, employees could not know about the existence of RIEs unless they heard about it by chance. Organisational knowledge management systems are particularly relevant for organisations with high staff turnover. The lack of such a system not only hinders the take-up of RIE evidence but also makes project implementers and policy makers ask the same impact evaluation questions over and over again. A missing knowledge management system also hinders sharing lessons learned in the initiation or implementation of RIEs.

> That's why the question for us is: Where is our institutional memory? ... Where is it actually stored in BMZ? It can't be that we keep commissioning the same thing and no one knows. ... People are hired and have to commission things in [country X]. How are they supposed to know whether it works or not, what the GIZ writes down? Absurd. … It's funny, isn't it? No one asks. Everyone thinks that whoever is hired here can do it. (Int_35)

**** **No actionable recommendations:** The take-up of self-generated RIE evidence was hindered when recommendations formulated by researchers were not directly actionable in the given context. Interviewees mentioned that a good knowledge of the implementing organisation and its processes as well as of the local context was needed to come up with actionable recommendations. At an organisational level, this aspect is also linked to limited transferability (external validity), since a different context can be the reason why recommendations are not directly actionable.

**** **No illustrative presentation of findings:** There is a perceived deficit in the illustrative presentation of research findings. Findings from RIEs are typically presented in scientific papers or evaluation reports. Those documents are rarely user-friendly, sometimes difficult to read due to scientific jargon and do not allow the

main take-away messages to be grasped quickly. Given the time restrictions that policy makers and programme implementers face, this hinders the take-up of evidence. “But maybe that also has something to do with the tiresome topic of how evaluation and study results are communicated. ... No one at any rate, except for a few people, read these boring research articles that are just needed for the publication list. Maybe 100 people on this planet.” (Int_30)

Intention

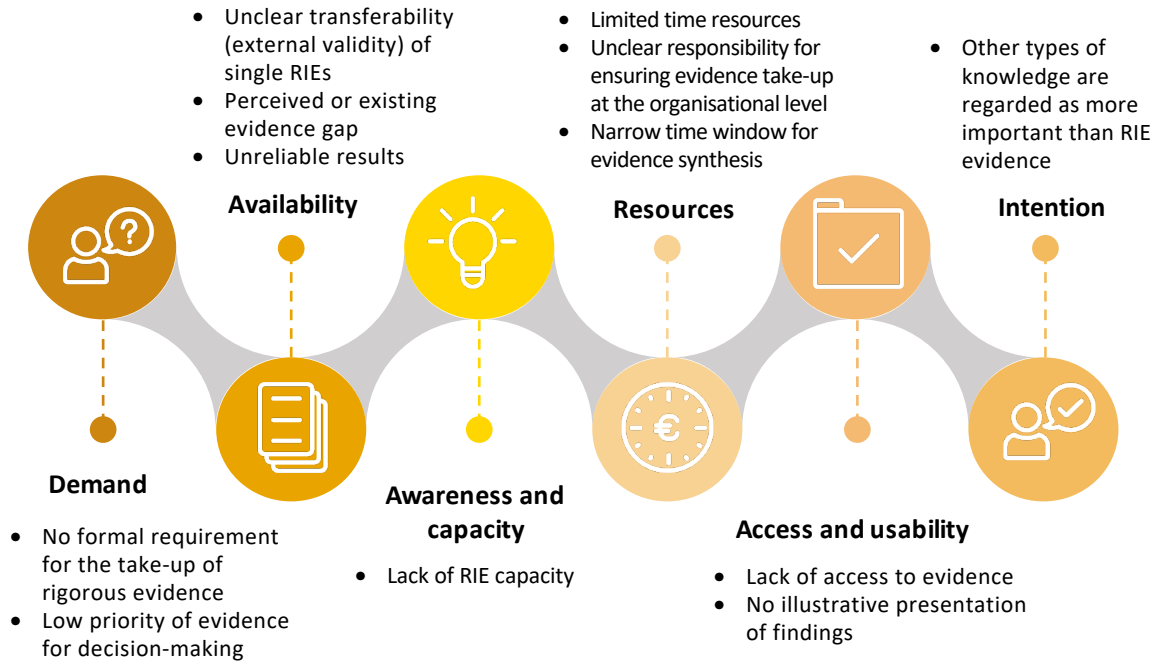
*** **Lack of culture of learning from failure:** The lack of a culture of learning from failure (see Box 45) was frequently mentioned as a barrier for evidence take-up. Respondents perceived pressure by BMZ and felt the need to report positive results of projects for domestic political communication. Head of projects are required to meet 10 to 15 indicators in a three years project cycle, which is often not realistic. “We have built an attitude to make the impossible possible in an absurdly short period of time” (source: open response field from evidence survey). In addition, interviewees mentioned the challenge of dealing with negative results from RIEs. At root there is the fear of not getting additional funding (in which practitioners have a personal interest) if indicators are not fulfilled or the evaluation shows negative results. Furthermore, interviewees highlighted a culture of reluctance to change and limited openness to experimentation and innovation as barriers to the use of evidence.

4.2.2 Barriers for the take-up of global RIE evidence

In 38 interviews, 175 code segments mentioned barriers. Compared to the number of code segments that identified barriers for self-generated RIE evidence this number is very high. All interviewed stakeholder groups mentioned barriers.

Based on the simplified ToC presented in our introduction (Chapter 1), we have allocated barriers to the six steps leading to more systematic evidence take-up (see Figure 23):

Figure 23 Overview of barriers to take-up of global RIE evidence, mapped to the six steps leading to a more systematic evidence take-up

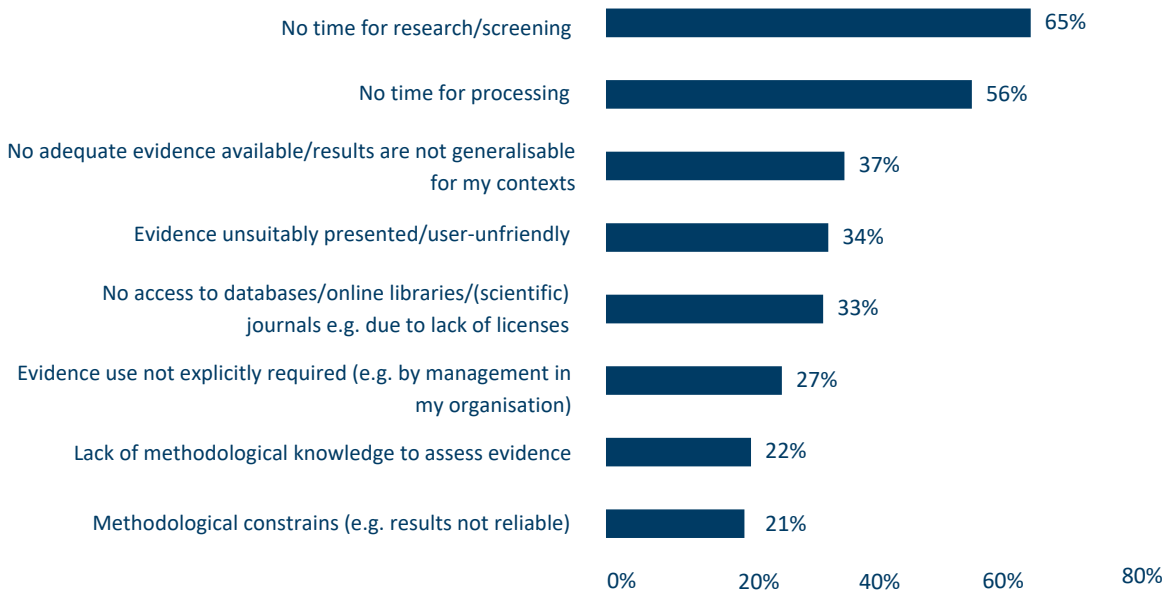


Source: own illustration

Demand

*** **No formal requirement for the take-up of rigorous evidence:** There is no explicit requirement or formally structured process to use the global RIE evidence base for policy making, programming and project implementation. Though the BMZ guidance documents implicitly require the take-up of self-generated evidence, which we understand to include RIE evidence, there is no requirement to make use of the global RIE evidence base (source: interviews; portfolio and document analysis). In our evidence survey 27% of respondents identified "No explicit requirement for evidence use" as a barrier (see Figure 24). This barrier overlaps with the barrier for self-generated evidence on "Limited formal requirement for the use of RIEs". Please refer to Section 4.2.1 for more details.

Figure 24 Barriers to the take-up of RIE



Source: evidence survey; question: "What makes it difficult for you to use evidence in your daily work?"; N=882; multiple answers possible.

*** **Low priority of evidence for decision-making:** This barrier has already been described in detail in Section 4.2.1.

Availability

A frequently mentioned barrier to the take-up of global RIE evidence at the project level, and partially for strategic decision-making, was the non-availability of relevant evidence. Reasons were as given below.

*** **Unclear transferability (external validity) of single RIEs:** In our interviews, the limited external validity of single RIEs was repeatedly mentioned as a barrier to the take-up of global RIE evidence. Interviewees assumed that results from RIEs conducted in one context and on one population were not transferable to other contexts or populations and therefore existing RIEs – even if the same intervention was looked at – were not relevant. This was a prominent barrier both at the project level and for strategic decision making. At the strategic level, in particular, general lessons are of interest rather than context-specific insights. Respondents to the evidence survey gave this barrier lower priority: 37% mentioned "No adequate evidence available" as a barrier to evidence take-up (see Figure 24).

*** **Perceived or existing evidence gap:** Evidence gaps were mentioned as a barrier to the use of global RIE evidence. Interviewees reported that they did not know about any rigorous studies that would be relevant to the project they worked on or that existing studies were outdated.

*** **Unreliable results:** Some interviewees mentioned that they did not trust the existing RIEs, either because they were aware of technical issues during implementation, doubted that the methodology was applied well or because they had reason to believe that findings were polished in a certain way (e.g. to please the donor). Twenty-one per cent of the respondents to the evidence survey marked the methodological constraints of the RIE as a central barrier to the take-up of rigorous evidence.

Awareness and capacity

**** **Lack of RIE capacity:** Interviewees reported a lack of methodological knowledge on RIE and on evidence synthesis tools such as SRs and EGMs as barriers to searching for and assessing global RIE evidence. In the evidence survey, 22% of respondents mentioned the lack of methodological knowledge as a barrier to assess evidence (see Figure 24). As reported in Section 4.1, across all our data sources it became evident that RIEs, EGMs and SRs, as well as RIE databases such as the 3ie's DEP, are largely unknown by GDC staff.

Resources

***** **Limited time resources:** Employees across all organisations reported being structurally overworked and having no time to search, read, assess, and process the existing evidence. "Yes, one of the main obstacles is certainly time. We work under extreme time pressure. And I think that many colleagues would like to use much more evidence and take much more time to read studies and so on when proposing projects. That's what everyone actually says." (Int_9). Even research-oriented employees, who have a good evidence-literacy and know where to find relevant scientific evidence, reported not having time to engage with the existing evidence base. This barrier was the most frequently mentioned barrier in the interviews. In the response categories of the evidence survey we differentiated between the searching/screening for and the processing of evidence. With 65% and 56% of the respondents agreeing, those two categories were the most frequently chosen categories in the evidence survey, confirming the findings from the interviews (see Figure 24).

***** **Unclear responsibility for ensuring evidence take-up at the organisational level:** Unclear responsibilities hinder the take-up of rigorous evidence at the organisational level. Interviewees from BMZ consider themselves to be generalists and argued that the IOs – particularly the thematic and methodological back offices – should be on top of the latest scientific evidence. In contrast, employees of IOs reported that they were no experts in research but in project implementation and that due to their high workload, keeping track of the latest and growing body of scientific evidence within broad and multiple topics would not be possible. International experts argue that it is neither the core competence of policy makers nor of project implementers to keep track of the scientific evidence but that this responsibility should lie with researchers.

> I find it's very unfair to project teams to expect them to be experts in research. Researchers are expert in research. That's the truth of it. Research evolves very quickly. There's so many studies at all times. You have to keep up with the literature. You have to read and be informed. And researchers are professionals of evidence. And so the idea that somebody who works in development should know all the evidence and be able to incorporate that evidence in their own project I think is unfair. (Int_Exp_1)

Though sporadic university cooperation exists, researchers are not an integral part of the GDC system, and explicit positions of "applied researchers", "chief economists", "chief scientific advisors", "scientific knowledge brokers" or research departments do not exist in GDC organisations.

*** **Narrow time window for evidence synthesis:** Employees of thematic and technical back offices of the IOs reported that time windows to produce evidence synthesis products are too narrow. Requests for quality assurance of new project proposals are often made at short notice such that it is impossible to systematically prepare the evidence for those requests. "Relevant evidence can only be collected and prepared if we would have longer lead times. ... We are not given this time. Overall, the duration of the projects is too short to be accompanied by relevant research." (Open response field in evidence survey).

Access and usability

**** **Lack of access to evidence:** Interviewees reported that the lack of easy and fast access to RIE evidence was a factor hindering its uptake. Evidence is scattered over numerous websites and literature databases. In addition, often studies are non-accessible behind paywalls. Some studies are not accessible at all because they have never been published. Language barriers were only mentioned rarely, and predominantly by francophone employees. In the evidence survey, the lack of "access to databases/online libraries/(scientific) journals e.g. due to lack of licences" was perceived as of medium relevance and marked as a barrier by 33% of the respondents (see Figure 24). According to the interviewees, the lack of access to RIE evidence not only hinders evidence-informed policy making but also means that policy makers and programme implementers ask the same research questions over and over again.

**** **No illustrative presentation of findings:** This barrier is equivalent to the barrier "No illustrative presentation of findings" for self-generated evidence. Please refer to Section 4.2.1 for more details. In the evidence survey, 34% of the respondents marked "non-illustrative presentation of findings" as a barrier (see Figure 24).

Intention

*** **Other types of knowledge are regarded as more important than RIE evidence:** Some interviewees reported that other types of knowledge, particularly expert knowledge, own experiences and experiences of colleagues (tacit knowledge) are valued more highly for policy making and programming than are results from scientific evidence, including RIE evidence (see also Yanguas, 2018). Others highlighted that the underlying method was not of relevance at all:

> Evidence is evidence for project implementation. So as soon as I enter into dialogue, it is a fact when a result is presented. How one arrives at the result is generally irrelevant to the person responsible for the project. So if I can state, yes, with 96.6% probability it is the case, or I simply say, no, in my opinion it is the case, it doesn't matter to the project manager for the time being. (Int_23)

4.3 What are potential measures to foster take-up of rigorous evidence?

This section presents a comprehensive list of potential measures to foster the take-up of rigorous evidence. Since there is some overlap between potential measures for self-generated RIE and global RIE evidence, we present potential measures for those two types of evidence jointly. We indicate in brackets whether it addresses self-generated RIEs or global RIE evidence or both. There is also some overlap to measures that foster RIE initiation, which have been presented in Section 3.3. To avoid repetition, we refer to Section 3.3 in those cases and only add specificities related to evidence take-up in this section.

Similar to Section 3.3, potential measures to foster take-up of rigorous evidence comprise:

- novel measures that have not yet been trialled in GDC or by other development partners.
- measures that have turned out to be actual facilitating factors for the take-up of evidence in concrete past instances within GDC.
- measures that have been implemented by other development partners and therefore plausibly have the expected effect.

In total, we identified 25 potential measures. As described in Section 3.3, measures for take-up of RIE evidence can be explicit (e.g. including the take-up of rigorous evidence in the organisation's lead strategy) or implicit (e.g. changing the evidence culture), push (e.g. issuing a formal legislation) or pull (e.g. building awareness) or demand-side (e.g. identifying interventions and evaluation questions of operational, strategic or political relevance) or supply-side oriented (e.g. establishing an RIE support structure). We have listed potential measures alphabetically below.

Allocate financial resources and time to evidence take-up *(global and self-generated RIE evidence):* It is evident that most of the presented measures to foster evidence take-up would need financial and time resources. Our interviewees identified this as an important measure. With 58% of respondents supporting "more time to search" in the global RIE evidence base, it was perceived as the most helpful measure in the evidence survey (together with the illustrative presentation of evidence) (see Figure 25). According to our interviewees, more time would also be beneficial for engaging in exchanges with researchers and experienced colleagues, for example by attending conferences and meetings in which scientific evidence is presented and which hence facilitate the access to evidence. Some of the measures presented in this section, such as evidence synthesis products or an improved knowledge management system, would also reduce the time needed to get on top of the current state of the evidence. This potential measure was also mentioned in Section 3.3.

Figure 25 Potential measures to foster evidence take-up

Source: evidence survey; question: "What measures do you consider to be particularly helpful to integrate more evidence into your work?"; N=863; multiple answers possible.

Build a network of knowledge brokers *(global RIE evidence):* Building a network of knowledge brokers (see Box 6) is by far the highest ranked potential measure for increasing take-up of global evidence. Knowledge brokers can take on many of the measures that foster take-up of global evidence mentioned in this section, including education and training, production or commissioning of evidence synthesis products or illustrative presentation of findings. "My sense is that researchers should do more to communicate their findings. But in practice good researchers are not always good communicators. That's when knowledge brokers can have a huge impact, putting together infographics, videos, columns, to communicate ideas more effectively than researchers are able to." (Int_28). There are already examples of existing structures of knowledge brokers. For example, FCDO have dedicated positions for knowledge brokering (source: report on international experiences). The Norwegian Agency for Development Cooperation (NORAD) has recently established a position of a "director for knowledge and evaluation" (Evans et al., 2021). Knowledge brokers should keep a close relationship to the potential end-users, thereby gaining a good understanding for whom different pieces of evidence might be relevant and what the most pressing evidence questions are. Knowledge brokers can be organised in a central unit or in a decentralised network.

Box 6 In short: Knowledge brokers

A knowledge broker is an individual, group or organisation that functions as intermediary between research producers and end-users. Knowledge brokers are familiar with the culture and language of both actors. He or she collaborates with the end-user to identify pressing evidence questions or problems and facilitates the access, screening, assessment, interpretation and translation of scientific research into actionable recommendations for the given context (Dobbins et al., 2009).

Build evidence take-up capacities *(global and self-generated RIE evidence):* Interviewees frequently mentioned capacity development as a potential measure to increase the take-up of both self-generated and global RIE evidence. In the evidence survey, "knowledge building" ranks fourth amongst potentially effective measures to increase evidence take-up (see Figure 25). This potential measure has been described in detail in Section 3.3 for RIE initiation. Capacity development for self-generated and global evidence should comprise:

- **Self-generated RIE evidence:** According to our interviewees, trainings for interpreting the results of RIEs and their integration into project planning and policy advice would be particularly helpful to facilitate their actual take-up.
- **Global RIE evidence base:** Capacity development measures for increasing the take-up of global RIE evidence should focus on how to find, assess, interpret and integrate relevant RIEs and evidence synthesis products into project planning and policy making. For researchers and evaluators, the focus should be on the adequate communication of findings. Our interviewees emphasised that for policy makers and programme implementers, capacity development measures should be very practical and tailored to needs and working realities (e.g. formats should be short). They also mention that capacity development would be most relevant for employees in the methodological and thematic back offices. GIZ has recently started to do RIE capacity work in the form of short online trainings and webinars, e.g. on 3ie's DEP. Some interview partners, particularly from IOs, consider such an "enabling" approach to be more effective than a "top-down" legislation or rule.

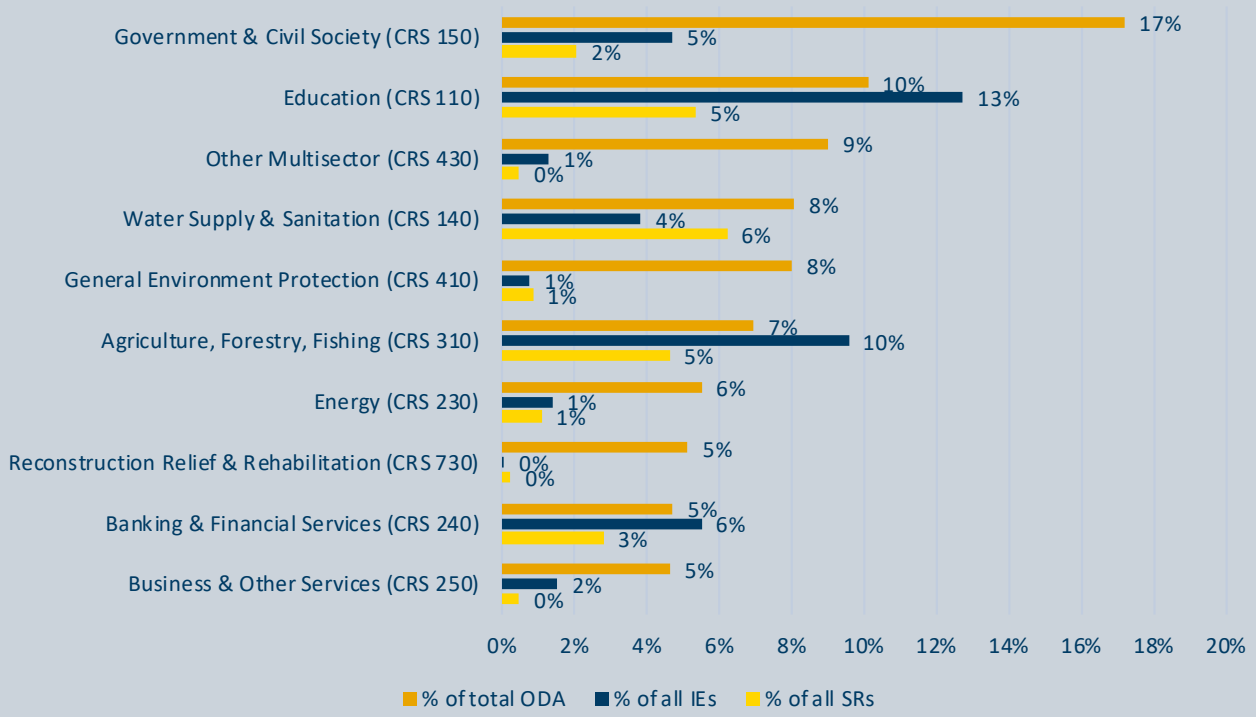
Conduct a number of similar RIEs in different contexts *(self-generated RIE evidence):* To foster evidence-take up at an organisational level, interviewees proposed conducting a number of similar RIEs in different contexts in order to synthesise their findings together with RIE findings from the global RIE evidence base. Synthesising RIEs allows general conclusions that apply to different contexts. This type of evidence is particularly useful at an organisational level, such as for strategic decisions, but also for informing other projects. WFP's thematic windows and EGAP's Meta Keta initiative are good examples of such an approach. Within GDC, this approach was reported as highly beneficial for GIZ global initiatives. As mentioned by some of our interviewees, a central **RIE fund** that finances and coordinates a sequence of similar RIEs in specific sectors could also enable evidence synthesis and hence evidence take-up at the organisational level. DIME, CEDIL and 3ie manage such central RIE funds on a global level.

Create awareness for evidence take-up *(global RIE evidence):* Creating awareness for RIE, SRs and EGMs is central to increasing take-up of rigorous evidence (sources: interviews; report on international experiences). Creating awareness is also a potential measure for RIE initiation (see Section 3.3 for more details). When creating awareness for the take-up of global evidence, it is particularly important to address the widespread belief that the global RIE evidence base does not lend itself to the GDC portfolio. Our analysis of 3ie's DEP found that it contains a large number of studies that are indeed potentially relevant to the GDC portfolio (see Box 7).

Box 7 Does the global RIE evidence base lend itself to the GDC portfolio?

To approach the question of whether the global RIE evidence base lends itself to the GDC portfolio, we applied a two-step approach. In a first step, we mapped the ten sectors with the highest German ODA shares against the number of IEs and SRs published on 3ie's DEP (see Figure 26 below). Sectors with the largest ODA shares are not necessarily the sectors with the largest share of evidence. Nevertheless, a substantial volume of IEs and SRs are available in most sectors (except Reconstruction Relief & Rehabilitation). Although this aggregate analysis cannot rule out the possibility that studies in the 3ie DEP are still different from projects implemented in GDC, this analysis indicates a substantial likelihood that relevant evidence is available within the global RIE evidence base.

Figure 26 German ODA vs IEs and SRs



Source: ODA dataset, German bilateral ODA disbursements with BMZ as extending agency, 2014–2018 average; top 10 sectors by percentage share. Data source 2: 3ie DEP, amount of IEs & SRs, absolute numbers, May 2020 German.

In the second step, we further investigated the existence of RIEs relevant to the GDC portfolio at the project level for one sector: "Government & Civil Society". To do so, we took a random sample of five GIZ and five KfW Development Bank projects from this sector and conducted a non-exhaustive keyword search for relevant studies in 3ie's DEP. We restricted the search to publications since 2014 and identified potentially relevant evidence for all projects that were part of our analysis. On average, the keyword search showed more than 200 IEs and about 20 SR hits per project (see *online appendix).* We then screened abstracts and used filters (e.g. by country, sector or fragility dimension) to find a maximum of five studies for which the abstracts indicated high relevance to the projects at hand. For eight out of the ten projects we were able to identify those five studies, for one project we found four and for another we found three studies.

Although it is certainly true that there are still many evidence gaps and that the global RIE evidence base does not contain relevant studies for each project within GDC, our document analysis suggests that a large amount of rigorous evidence is available within 3ie's DEP that is potentially relevant to the GDC portfolio. Even at a project level, we were able to find a number of potentially relevant studies.

Conversely, the fact that a substantial number of RIEs have already been conducted in sectors relevant to GDC indicates that **conducting** RIEs in projects of GDC might be possible more often than is currently the

case. The DIE initiative on impacts in governance already showed this in four concrete projects in the governance sector (Funk et al., 2018). This is particularly relevant since several interviewees described the sector "Government & Civil Society" as particularly unsuitable for conducting RIEs.

Create opportunities for RIE exchange *(global and self-generated RIE evidence):* Interviewees indicated that creating opportunities for RIE exchange could be a potential measure to increase evidence take-up. They emphasised that direct, personal interaction was often more effective in fostering evidence take-up than disseminating written documents. Exchange can happen as peer-exchange, research-practice-exchange or policy-practice-exchange. In concrete terms, this can imply the planning of conferences or events where research and practice meet to present and discuss new RIEs. The yearly BMZ-funded conference of the "Poverty reduction, equity and growth network (PEGNet)" is an example of such an event. This measure was also mentioned in Section 3.3.

Define roles and responsibilities *(self-generated RIE evidence):* Clearly defining responsibilities for take-up of self-generated evidence can increase evidence take-up. This can simply involve selecting a person or group of people in charge of knowledge translation and dissemination at project, organisational or global level.

Demonstrate leadership commitment to RIE *(global and self-generated RIE evidence):* Leadership commitment and giving priority to evidence-informed decision making is an important measure to foster evidence take-up (sources: interviews; report on international experiences). Please refer to Section 3.3. for more details.

Enable access to academic journals by buying subscriptions (*global RIE evidence*): One measure to increase the access to RIEs is buying subscriptions to academic journals or joining a network offering such subscriptions.

Ensure close cooperation between researchers and other stakeholders *(global and self-generated RIE evidence):* Close cooperation between researchers and other stakeholders is a central measure to enhance take-up of self-generated and global RIE evidence (sources: interviews; report on international experiences). This potential measure has been described already in Section 3.3 for RIE initiation.

- **Self-generated RIE evidence:** Whereas close cooperation is particularly relevant in the early phase of an RIE to enhance RIE initiation, it is important at all stages during project implementation to enhance evidence take-up. Continuous interaction enables the formulation of relevant and realistic research questions and facilitates researchers' real-time responses to needs of project implementers. In addition, it creates ownership and helps to formulate actionable recommendations. This method is followed by the World Bank DIME unit.
- **Global RIE evidence:** Close cooperation with local experts is indispensable for contextualising the existing rigorous evidence to development programmes in a new geography and thereby addressing the problem of a possible lack of external validity. Further close cooperation with stakeholders who have good organisational knowledge is important for formulating recommendations that fit the abilities and structures of the organisations involved.

Establish a comprehensive and active dissemination strategy *(global and self-generated RIE evidence):* The illustrative presentation of findings from RIEs (as well as SRs and EGMs) should be combined with a diverse dissemination strategy that responds to different needs and preferences of employees in GDC. A dissemination strategy can include the translation of products into different languages, including the language of the partner country. J-PAL and WWN, for instance, combine presentations at sector events and one-on-one evidence sharing with targeted emails, webinars, conferences or social media activities both for internal and external users.

Extend project cycles *(global and self-generated RIE evidence):* Extending project cycles can foster evidence take-up for global and self-generated RIE evidence. First, during longer project design phases, the global RIE evidence base could be consulted and synthesised more thoroughly. Second, since self-generated RIEs are sometimes only available after projects have already ended, extending project cycles could be a potentially

effective measures to increase the take-up of self-generated RIE evidence. Extending project cycles might, however, come at the expense of reducing the opportunity of project and portfolio steering from BMZ.

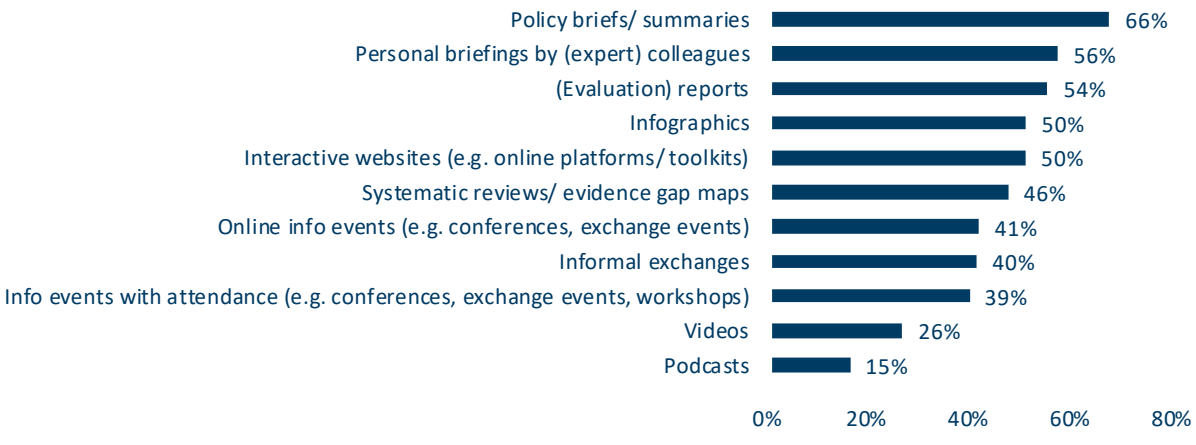
Generate additional knowledge beyond the RIE *(self-generated RIE evidence):* Interviewees highlighted the generation of additional knowledge beyond the results from the RIE as effective measure to increase evidence take-up. This potential measure has already been described in Section 3.3. International experts, in particular, highlighted this measure.

Give support and visibility to "evidence champions" *(global and self-generated RIE evidence):* Giving support and visibility to "evidence champions" can increase evidence take-up (source: report on international experiences). Please refer to Section 3.3 for more details.

Identify interventions and evaluation questions of operational, strategic or political relevance *(self-generated RIE evidence):* According to our interviewees – and quite intuitively – conducting RIEs on interventions and questions of relevance is a prerequisite for evidence take-up. Questions and interventions can be relevant from an operational, strategic or political perspective. Unfortunately, these three perspectives are not necessarily aligned with each other. Interventions or questions of operational relevance can be identified together: "… with the head of the projects. They will tell you what they do not know yet and in which field they need support… and [then] make a ranking according to relevance" (Int_9). Interviewees named, for instance, targeting or testing different implementation options against each other in a multi-armed RCT as operationally relevant. Questions of strategic or political relevance will instead be identified at the political level or at the thematic and methodological back offices. For instance, as part of the 2030 strategy, BMZ has decided to include a section in which central questions and the outlines of an evidence agenda can be formulated in its strategies for core topics and initiative topics. To what extent this possibility will be used, remains to be seen.

Improve (intra-)organisational knowledge management systems *(self-generated RIE evidence):* Interviewees highlighted a better knowledge management system as a potential measure to improve take-up of self-generated RIE evidence at the organisational level. They envisioned a platform or unit that collects and displays data, results and experiences from RIEs that have been generated in GDC in an easily accessible and user-friendly way. IOs strongly highlighted this measure. KfW Development Bank has recently launched an internal database for their general evaluations that gives quick access to KfW Development Bank's evaluation reports. GIZ is considering developing a similar tool. BMZ has anchored the development of an RIE database in its BMZ 2030 reform processes. DEval will launch an initial version of this database by the end of 2021.

Figure 27 Preferred formats for evidence presentation



Source: evidence survey; question: "Thinking ahead to the future: What formats for presenting evidence do you find particularly useful in your work?"; N=865; multiple answers possible.

Include the take-up of rigorous evidence in the organisation's lead strategy *(global and self-generated RIE evidence):* Organisations should commit to evidence-informed decision-making in their lead strategy to facilitate the take-up of self-generated and global RIE evidence (source: report international experiences). This measure is described in Section 3.3.

Monitor evidence take-up *(self-generated RIE evidence):* Regularly monitoring the implementation of recommendations from RIE is a potential measure for increasing actual evidence take-up.

Pass formal legislation or requirements *(global and self-generated RIE evidence):* Another potential measure to increase evidence take-up is formal legislation or requirements. This measure has been described in more detail in Section 3.3. In concrete terms, this could be done by explicitly including the obligation to draw on rigorous evidence – be it self-generated in one's own organisation or from the global RIE evidence base – in existing process flows and respective guidance documents. Currently, there are only vague hints in the guidance documents to draw on self-generated evidence. Referring to scientific studies in the formal reporting documents would allow stakeholders, including BMZ employees, to verify the lines of reasoning. Formal requirements could also regulate the dissemination of results. Donor requirements often foster innovative activities to increase evidence take-up in organisations such as Oxfam GB or J-PAL (source: report on international experiences). Formal legislation and requirements for evidence take-up are judged heterogeneously. Although many interviewees think an explicit requirement would be highly effective ("one does what the Ministry says" (Int_13)) others think that a requirement would not be practically feasible or not taken seriously and hence evidence reviews – if done at all – would be of bad quality: "… requirements do not necessarily lead to improvements. There is the risk that requirements are perceived as tick boxes only, meaning that an intern quickly summarises five studies, … which is still far from indeed using evidence for project planning" (Int_4). In the evidence survey, only 24% of respondents identified this potential measure as effective (see Figure 25). Potentially, formal legislations need to be combined with other measures such as resources and capacity development to be effective.

Present findings in an illustrative way *(global and self-generated RIE evidence):* Our data sources concur that presenting findings in a user-friendly, innovative format and in plain language can facilitate their take-up (see for example Figure 27). The presentation should catch the interest of the potential end-user and be fun to engage with. AFD, IDB, Oxfam GB, J-PAL, and the WWN, amongst others, all translate rigorous evidence into simpler formats (source: report on international experiences). Examples of such formats are WWN's "intervention toolkits"[13] or WWN's "evidence finder"[14]. Also, in GDC there are some good examples of illustrative presentation of findings. For instance, KfW Development Bank has recently redesigned the cover page of their evaluation reports. The cover page now lists three to four central lessons learnt from the evaluation. Similarly, GIZ produces short summaries and one-pagers, in addition to their central project evaluation reports. In the evidence survey, we asked for preferred formats for presenting evidence (see Figure 27). Policy briefs and summaries are ranked highest (66% of respondents). Personal briefings by colleagues (56%) and evaluation reports (54%) are also ranked high. Lesser relevance was given to podcasts and videos. The low ranking of videos is surprising since it contradicts findings from our interviews. One reason for the low ranking of videos might be that it is not commonly used yet and experiences are lacking. According to the open-response fields in the evidence survey, another potentially effective format can be newsletters clustered by sector, country or region. In the response fields of the online survey, respondents highlighted the high quality of the different formats as being extremely important to the facilitation of evidence take-up.

Provide evidence on a demand-driven basis *(self-generated RIE evidence):* Demand-driven provision of evidence is a potential measure to foster take-up of global RIE evidence. In the evidence survey the demand-driven provision of evidence by an internal body was ranked third highest, with 52% of respondents selecting this measure (see Figure 25). This demand-driven model is, for instance, implemented in Demark, Norway

[13] See for example https://whatworks.college.police.uk/toolkit/Pages/Toolkit.aspx

[14] See for example https://www.homelessnessimpact.org/evidence-finder

and Sweden. They have publicly funded research centres for education, health and social welfare, the regular job of which is compiling evidence for policy making and programming. Those units actively and regularly engage with policy makers to discuss findings and identify evidence needs (White, 2019). South Africa, Uganda and Zimbabwe also have rapid response services embedded within their governments. Another example is the West Africa Capacity-building and Impact Evaluation (WACIE) Helpdesk that provides rapid evidence synthesis to help policy makers in West Africa answer targeted policy questions. The timely provision of evidence is a precondition for the success of this measure, meaning that either rapid syntheses are required or that the demand is expressed early (see also measure "Synthesise and further condense the body of evidence").

Set incentives for evidence take-up *(global and self-generated RIE evidence):* Interviewees pointed to incentives as potential measures to foster evidence take-up. Incentives could be the inclusion of evidence take-up in annual staff targets, or giving visibility to "evidence champions" who push for evidence take-up. This could happen, for example, by granting evidence awards, such as the Africa Evidence Leadership Award, in which the Africa Evidence Network awards leaders who have significant "influence in advancing EIDM in their contexts, and their relative reach" (Africa Evidence Network, 2021). Section 3.3 describes this measure in more detail. To incentivise researchers to conduct evidence syntheses, it would be beneficial to recognise SRs as an adequate contribution to a PhD thesis. As an incentive on a national level, one international expert also raised the idea of making ODA accounting conditional on evidence-based proof of effectiveness – either by drawing on existing evidence or generating new evidence.

Strengthen a culture of learning from failure *(self-generated RIE evidence):* Strengthening a culture of learning from failure can support take-up of self-generated RIE evidence (see Section 3.3 for more details). This also includes the open and unbiased interpretation of evaluation results by researchers and evaluators. Respondents to the evidence survey asked for "independent and professional interpretations and recommendations that are unpolished and do not please the commissioner or any particular stakeholder group" (open response field).

Synthesise and further condense the body of evidence *(global RIE evidence):* According to our interviewees, synthesising the global evidence base is the most relevant measure to increase take-up of global evidence. White (2019) describes the steps from individual RIEs to highly condensed evidence products, such as guidelines or checklists. SRs, followed by EGMs, were mentioned as effective synthesis tools in our interviews. Individual interviewees pointed at the portals of the WWN, such as the teaching and learning toolkit from the Education Endowment Foundation, as highly useful tools.[15] Unfortunately, such high-quality synthesis tools, including SRs and EGMs, are largely unknown in GDC (see Section 4.1). In addition, the production of synthesis products needs time. Hence, either project design phases need to be significantly extended or – and maybe more realistically – evidence synthesis products need to be produced independently from project cycles. For the latter option, interviewees proposed a synthesis strategy of several years that gives priority to a certain sector each year (see also the measure "Identify interventions and questions with political priority"). Some interviewees even argued that evidence syntheses should be coordinated internationally in order to provide the entire development community with these products in a cost-effective manner.

Systematically consult the RIE evidence base when designing projects *(global and self-generated RIE evidence):* A systematic consultation of the global and self-generated RIE evidence base early on when designing projects is a potential measure to foster take-up of global RIE evidence. A systematic anchoring could happen by integrating reflection points in existing process flows and in guidelines about how to create module proposals or brief statements. Those reflection points should ensure that state-of-the-art RIE evidence supports the proposed project. AFD, IDB, Oxfam GB, WWN and J-PAL perform such reflective learning sessions (source: report on international experiences). Similarly, USAID, Oxfam GB and FCDO

[15] See here *https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/*

reported having long-term formalised partnerships with universities and think tanks to provide inputs for project design using global evidence.

4.4 Linking highly relevant barriers and potential measures

This section links highly relevant barriers (Sections 4.2.1 and 4.2.2) with corresponding potential measures (Section 4.3). Recall that we assessed the following barriers as **highly relevant (***)** for currently hindering evidence take-up in GDC:

For self-generated RIE evidence:

- high staff turnover
- lack of culture of learning from failure
- limited relevance of the intervention or the research question
- missing knowledge management
- scarce and unsystematic dissemination of findings.

For the global RIE evidence base:

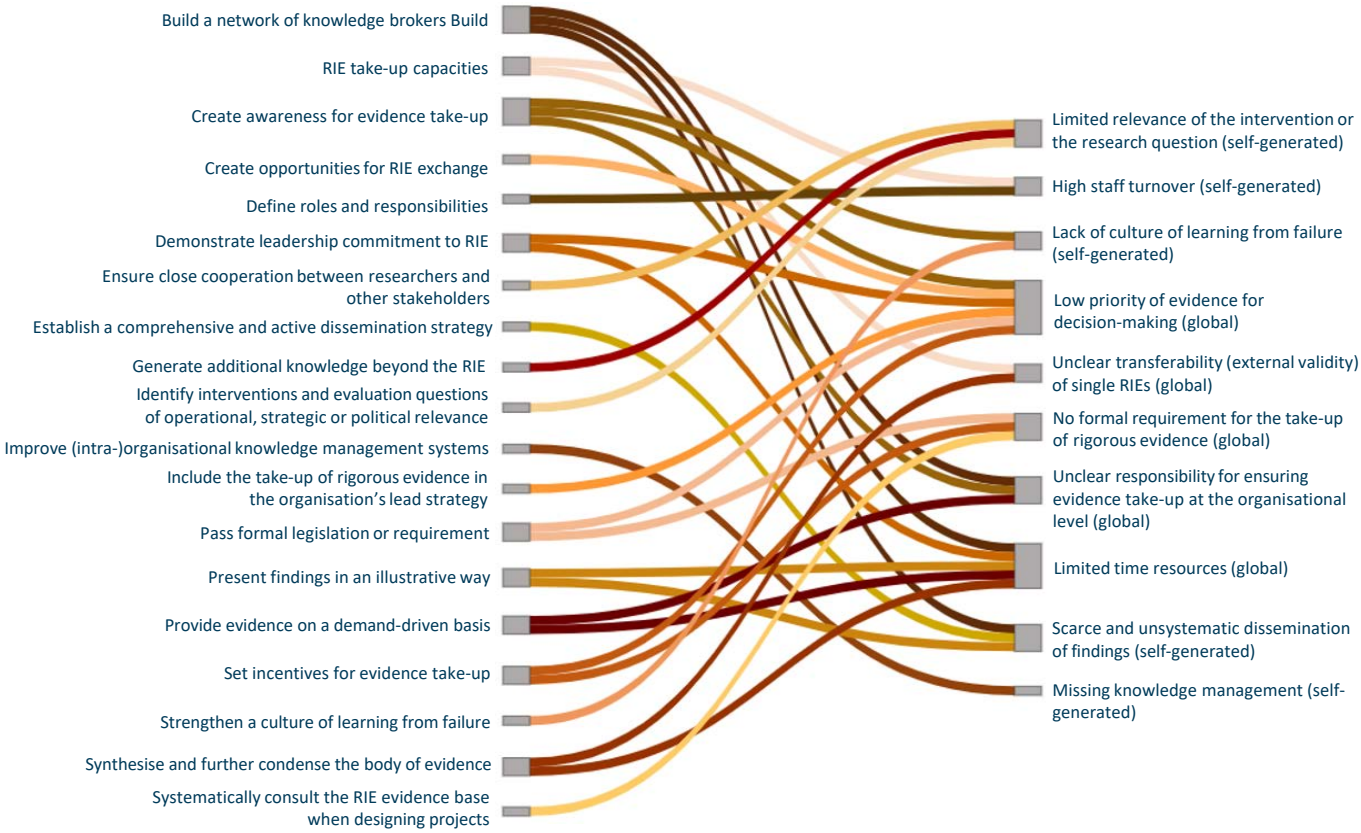
- limited time resources
- low priority of evidence for decision-making
- no formal requirement for the take-up of rigorous evidence
- unclear responsibility for ensuring evidence take-up at the organisational level
- unclear transferability (external validity) of single RIEs.

We, again, examined whether the relevance of the barriers varies across **stakeholder groups** (BMZ, GIZ, KfW Development Bank, other IOs, CSOs, private sector, DEval). According to the evidence survey, this ranking is very similar across all stakeholder groups of GDC. This picture is largely confirmed by the interviews.

For evidence take-up, we have again presented barriers and potential measures in a clearly delineated way. Yet it has become clear that barriers and potential measures are in fact often **interlinked or even hierarchical**. For example, we can see a close link between the measures “Build a network of knowledge brokers” and “Present findings in an illustrative way”, or a hierarchical relationship between the measures “Pass formal legislation or requirements” and “Systematically consult the evidence base when designing projects”.

Finally, when linking highly relevant barriers with potential measures of Section 4.3, we see that one potential measure can often **address more than one** barrier. For example, this is true for measures like capacity development and awareness creation. Figure 28 illustrates these links. Some potential measures might indirectly impact relevant barriers. In Figure 28 we only present measures that can obviously and directly be linked to the barriers. Allocating financial resources and time is a cross-cutting measure that is a precondition for most of the other measures. Therefore, we do not explicitly mention this measure in Figure 28. In sum, this analysis of links suggests that a set of interrelated and complementary measures is needed to strengthen evidence take-up in GDC.

Figure 28 Linking potential measures and highly relevant barriers for evidence take-up



Source: own illustration

5. IMPLICATIONS

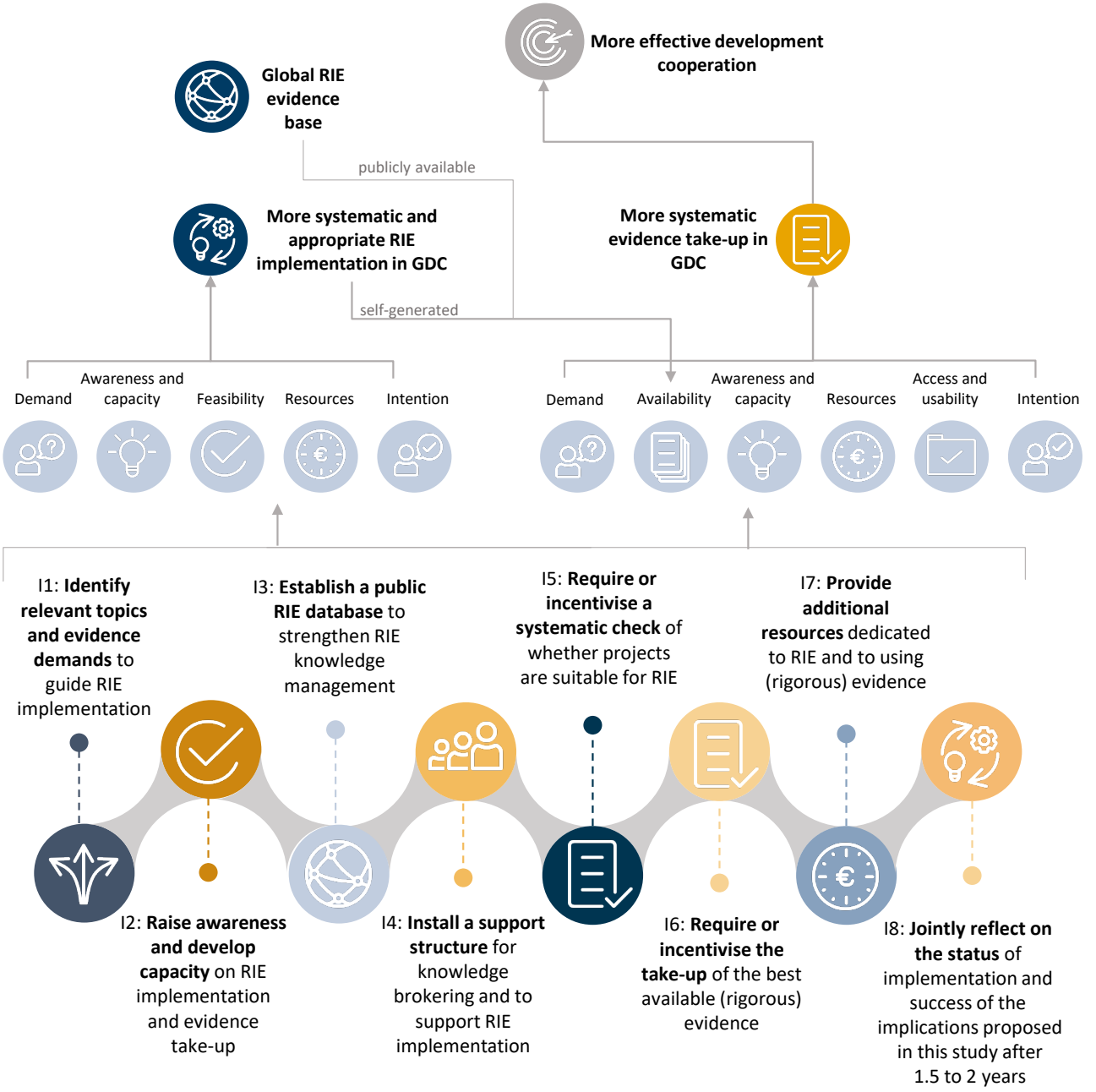
Due to the strong interdependence and hierarchy of potential measures (see sections 3.4 and 4.4), we propose an integrated approach to fostering the initiation of RIE and the take-up of rigorous evidence. Although specific barriers can be addressed by implementing isolated measures, such an approach will not suffice if the system of evidence generation and take-up is to be strengthened as a whole and the processes outlined in the ToC is to be improved. Acknowledging that there are no "silver bullets", we have identified a set of implications[16] based on our findings that is intended to have a systemic effect. Changing the quantity and quality of implementation of RIE and of take-up of rigorous evidence requires, we believe, changes at different points in the ToC in order, ultimately, to contribute to a more effective development cooperation. The implications are:

- empirically grounded in the findings presented in this report
- interdependent, such that implementing one implication also supports the implementation of others
- focused on addressing the most relevant barriers in parallel to facilitate systemic change.

These implications build on and further develop six strategic tasks that have been formulated in a policy brief by Bruder et al. (2019). Figure 29 illustrates the ToC, with a focus on our proposed set of implications.

[16] We use the term "implications" to distinguish them from official "recommendations" of DEval evaluations. For DEval evaluations, there is a formalised process for its conceptualisation and follow-up, including feedback from Reference Groups, a BMZ management response and an implementation monitoring after two years. Instead of this standardised process, the implications formulated here are open to be discussed and considered by BMZ and other actors.

Figure 29 Theory of change for more systematic RIE implementation and evidence take-up



Source: own illustration.

In the following, we present the proposed implications of our study in more detail.

Implication 1: Identify relevant topics and evidence demands to guide RIE implementation.

Currently, there is no systematic approach or guidance as to when to conduct an RIE in GDC. We therefore propose identifying topics and evidence needs that are strategically relevant to GDC such that a number of RIEs could be conducted in these relevant topics. Since RIE initiation and evidence take-up are very closely connected and mutually dependent, systematically conducting RIE in relevant areas will also make evidence take-up more likely (e.g. in the course of identifying evidence needs or while preparing an RIE).

The following options illustrate how this implication could be implemented:

- BMZ, together with other stakeholders, could define strategically relevant questions in a collaborative process. The recently defined strategies for core topics and initiative topics are viable starting points. These strategies include a section on impact measurement which could provide the basis for the

formulation of questions central to future impact evaluations and the outline of an evidence agenda. This section and the consultative process implemented in writing the strategy could be used to formulate two to three concrete evidence needs to be addressed over the course of the strategy lifetime.

- The IOs could define questions strategically and operationally relevant to their organisations in order to guide RIE initiation. If implementing an RIE within a specific project would be feasible and could contribute to answering these questions, the RIE could already be included in the project proposal.
- BMZ or the support structure(s) for RIE (see Implication 4) could routinely or selectively commission EGMs for those strategically relevant topics named in the sector strategies (core topics and initiative topics, see above).
- BMZ, DIE and DEval could strengthen existing opportunities for RIE exchange (such as the thematic team on RIE) or create additional exchange platforms for exchange between policy makers, development practitioners and researchers and evaluators to make sure that evidence creation is oriented towards actual demands.
- RIE experts (academia and consultancies) could actively develop and offer integrated research designs that address both the strategically relevant questions posed by BMZ or senior management in IOs and CSOs and project-specific questions that are of interest to project staff.

Box 8 GDC can learn from and build on these activities when it comes to defining relevant topics for RIE

- The GIZ programme on transitional development assistance in cooperation with 3ie has assembled and synthesised evidence on the topics of "Intergroup social cohesion" and "Gender empowerment in fragile contexts" in two SRs. Based on these findings, the programme is currently preparing an initiative to generate evidence in the thematic field of peace/social cohesion.
- In South Africa, EGMs are produced by government departments based on policy priority to understand and anticipate evidence needs and evidence gaps for policy development.
- FCDO's (former DFID's) central evaluation unit and the country office's leadership prioritise the commissioning of RIEs based on whether the RIE will address knowledge gaps.

Implication 2: Raise awareness and develop capacity on RIE implementation and evidence take-up.

We encourage raising awareness and developing capacity for RIE and evidence take-up. Our evidence shows that individual awareness and capacity are critical for the initiation of RIEs and take-up of (rigorous) evidence. Given that RIE is only one (important) element of the evaluation toolbox, we propose combining RIE capacity development with building "evidence literacy" in general. This starts with strengthening the ability to concisely formulate evidence needs. These activities would foster a culture of evidence and learning, and create demand for rigorous evidence where it can be helpful, without overextending the tool to areas where it is not the first choice.

The following options illustrate how this implication could be implemented:

- The research and evaluation departments of BMZ, IOs, CSOs, and DEval, could conduct or commission awareness-raising workshops for senior officials and other senior management on RIE and evidence-based policy making. These sessions could introduce the global evidence base and highlight the strengths and opportunities as well as the limitations and challenges of RIE and rigorous evidence. They could also communicate the risks of not conducting RIE and not using rigorous evidence (e.g. ignorance about the actual effectiveness of interventions and the consequential risk of channelling limited financial resources to ineffective development interventions).

- The research and evaluation departments of BMZ, IOs, CSOs and DEval could conduct or commission technical RIE and evidence take-up trainings. Capacity-building activities should include modules on method integration. Such programmes could be developed and conducted either for each organisation separately or jointly.
- GDC could, for instance through the BMZ-financed FOCELAC+ project, include specific evaluation capacity development (ECD) on RIE and rigorous evidence in partner countries. Conversely, countries from the global South, such as Mexico or South Africa, which have a long history of evidence-informed policy making, could function as important learning partners.

Box 9 GDC can learn from and build on these activities when it comes to capacity development for RIE

- GIZ conducted an internal RIE-training in cooperation with J-Pal.
- GIZ initiated several (virtual) events to present RIE results and to create peer exchange.
- DEval conducted an RIE-training for GDC staff in cooperation with C4ED.
- In South Africa, the national school of governance offered courses on evidence-informed decision making for government officials.

Implication 3: Establish a public RIE database to strengthen RIE knowledge management.

Based on our analysis, we propose establishing a publicly available RIE database to strengthen RIE knowledge management within GDC. Such a database should pool RIEs conducted within GDC but also link up to the global RIE evidence base (in particular the 3ie DEP). Currently, such a central database is prepared at DEval.

Box 10 GDC can learn from and build on these activities when it comes to public RIE databases

- The Social Systems Evidence (SSE) database collects syntheses of rigorous research evidence about programmes, services and products available in a broad range of government sectors and programme areas.

Implication 4: Install a support structure for knowledge brokering and to support RIE implementation.

A support structure should be set up to support both practitioners and decision-makers in GDC along the cycle of evidence generation and take-up. We propose to build up this support structure as a cooperative model. In such a model, existing decentralised structures in the IOs and CSOs would be further strengthened where they are necessary and economical. The advantage of a decentralised support structure is that it is positioned close to practitioners such that evidence needs can be identified more easily. At the same time, a central structure could (a) coordinate efforts, (b) promote inter-organisational learning, (c) provide advice to those organisations that do not have support units themselves, and (d) serve as a point of contact for BMZ. Such a structure could build on experiences at DIE and DEval.

Such a support structure could support the implementation of all implications. In particular, it could be responsible for the implications 4.1 to 4.3:

Implication 4.1: Synthesise existing global and self-generated evidence.

Our analysis has shown that time constraints among all actors and limited relevance of single RIEs for strategic decisions are major barriers to evidence take-up. Synthesis products, such as EGMs or SRs, can address these barriers by collating existing knowledge and making it more readily accessible. We therefore propose that BMZ and IOs, in coordination with the RIE support structure, should invest in the synthesis of existing global and self-generated evidence. Such products could also help to make GDC more visible in the international sectoral discourse. Such evidence synthesis could be guided by strategically relevant topics and

questions identified under Implication 1. Besides EGM and SRs, evidence synthesis could also include rapid reviews or policy briefs to guide decision-making in a timelier manner. In the first year of the COVID-19 pandemic, four policy briefs provided rapid synthesis of the RIE evidence on pressing issues during the global crisis for the GDC (Kuhnt and Schüttler, 2020; Freudenreich et al., 2020; Kluve et al., 2020; Avdeenko and Heesemann, 2020).

Implication 4.2: Convert evidence into simple and user-friendly formats.

Our analysis has further revealed that evidence needs to be transformed into innovative and user-friendly formats and presented in an illustrative way. Edited in this way, evidence can be used with little effort. Researchers and RIE support structures (Implication 4) should experiment with ways to make evidence more accessible and learn from international best practices in doing so. For example, evidence can be summarized in "intervention tools",[17] structured along a theory of change ("outcomes and evidence framework")[18] or regionally mapped in an "evidence finder".[19]

Implication 4.3: Provide practical support and tools for RIE initiation and implementation.

The support structure should provide direct technical and administrative support (e.g. individual project consulting), link those looking for support with external structures or provide material that simplifies the RIE implementation process for development practitioners. This can include checklists, guidelines, explanatory videos or templates for terms of references and other administrative documents.

Box 11 GDC can learn from and build on these activities when it comes to installing a support structure for RIE

- Both KfW and GIZ have taken initial steps to establish RIE support units at an organisational level (Bräuer et al., 2020 and KfW, 2021).
- The GIZ Sector Program on Transitional Development Assistance creates an evidence unit.
- The Evidence into Action team at DFID used to support evidence processing and take-up. Uganda has a rapid response unit to support evidence processing and take-up.
- The What Works Centres in the UK are good examples for knowledge brokering organisations that bundle evidence synthesis and processing functions in a centralized structure.
- The International Rescue Committee undertook a series of EGMs to provide content for the Outcomes to Evidence Framework (OEF), which is intended to enable IRC programme managers to make evidence-informed decisions.

Implication 5: Require or incentivise a systematic check of whether projects are suitable for RIE.

Currently, there is no defined point in time to check the possibility or necessity of conducting an RIE when projects are established or extended. We therefore propose defining a requirement for examining the possibility of initiating an RIE within the process of planning and implementing GDC projects. Alternatively or additionally, there should be strong positive incentives for conducting such a feasibility and utility check on RIE implementation.

[17] For an example see *https://www.homelessnessimpact.org/intervention-tool*.
[18] For an example see *https://www.rescue.org/resource/outcomes-and-evidence-framework*.
[19] For an example see *https://www.homelessnessimpact.org/evidence-finder.*

The following options illustrate how this implication could be implemented:

- BMZ could formulate a requirement to check the feasibility and utility of conducting an RIE. Such a requirement could be incorporated in existing process descriptions to ensure that it is followed by default and to avoid process misalignments between project and RIE cycles. The ideal entry point in existing process flows would be during the project design phase (e.g. during the creation of module proposals or brief statements). The reflection on conducting RIE should be facilitated by an RIE evaluability assessment guidance tool (see *DEval guidance tool*; DEval, 2021).
- BMZ could formulate a requirement to accompany pilot interventions, interventions that should be scaled up or that have a large budget yet are untested by RIEs.
- BMZ or IOs could additionally or alternatively incentivise RIE by opening (annually) changing thematic windows for RIE evidence generation (e.g. aligned with strategically relevant topics as defined in Implication 1), by tying the continuation of a project to an evaluation of its impact, or by showing that RIE implementation can bring career benefits by giving visibility to evidence champions.

Box 12 GDC can learn from and build on these activities when it comes to requirements for evidence take-up

- South Africa's national policy framework includes an official mandate that makes the consideration of evidence legally compulsory and provides a strong institutional structure to support evidence use.
- At DFID (now FCDO) it was required to cite (rigorous) evidence in new business cases, and where such evidence was not available to include its generation as part of the programme.

Implication 6: Require or incentivise the take-up of the best available (rigorous) evidence.

Currently, there is no clearly defined and systematic approach as to when, in the process of planning and implementing GDC projects, available evidence should be consulted and used. We therefore propose defining a requirement or rule for evidence review and take-up. The goal should be to solidly base GDC projects on the best available (rigorous) evidence. Alternatively or additionally, there should be strong positive incentives for taking up the best available (rigorous) evidence.

The following options illustrate how this implication could be implemented:

- BMZ could formulate a requirement to reflect on the availability of relevant (rigorous) evidence when designing new projects. Such a requirement could be incorporated in existing process descriptions to ensure that it is followed by default and to avoid process misalignments between project and RIE cycles. Module proposals or brief statements could be viable reflection points early in the module commissioning process, yet the ideal entry point could be examined by a detailed process analysis. Guidelines for the preparation of module proposals or brief statements could be expanded to include an evidence review.
- The IOs and CSOs could formulate or further specify internal requirements to consult the best available evidence (including rigorous evidence) when designing new projects.
- BMZ or IOs could additionally or alternatively incentivise take-up of (rigorous) evidence by tying a project approval to the presentation of available evidence or by showing that RIE implementation can bring career benefits by giving visibility to evidence champions.

Box 13 GDC can learn from and build on these activities when it comes to requirements or incentives for RIE initiation and implementation

- The US Evidence Act from 2018 requires a number of governmental agencies to generate their own, or draw on existing, evidence to answer relevant policy questions.
- In addition, USAID's Evaluation Policy, requires to conduct RIEs for all new, untested approaches that are anticipated to be scaled up (USAID, 2016).
- Mexico's 2004 Social Development Law requires external evaluation of all government-funded social programmes.
- South Africa has a mandatory system of evidence assessment since 2015.

Implication 7: Provide additional resources dedicated to RIE and to using (rigorous) evidence.

Limited resources are both a fundamental barrier to RIE initiation and to evidence take-up. We therefore propose allocating additional resources: (a) to implement the proposed set of measures (implications 1–8) and (b) to fund the implementation of RIEs. This will also signal leadership support.

The following options illustrate how the latter (b) could be implemented:

- BMZ could install a central fund dedicated to RIE implementation. Such a fund could be open to applications by projects and researchers, could link projects and RIE experts, and could strategically guide RIE generation in the direction of relevant topics and questions (see also Implication 1). The implementation of such a fund has been discussed for some time and a proposal for such a fund is currently being considered by BMZ.
- The IOs and CSOs could set aside and provide financial resources to fund RIEs, EGMs or SRs in a central evaluation budget. Some CSOs already have comparable models with central evaluation budgets that have been used to finance meta-studies, or even RIEs in some cases.
- IOs and CSOs could include the funding of RIEs and dissemination activities in project proposals, in accordance with BMZ funding and financing guidelines. This would necessitate a more detailed examination of these guidelines and possible adaptations than was possible in this research project.

Box 14 GDC can learn from and build on these activities when it comes to providing resources for RIE

- BMZ occasionally directly commissions RIE.
- GIZ funds RIE preparation via the Studies and Experts Fund (SFF) and sector projects.
- The French government is starting a Fund for Innovation in Development providing 15 million euros per year for a programme combining innovative approaches with rigorous evaluation.

Implication 8: Jointly reflect on the status of implementation and success of the implications proposed in this study after 1.5 to 2 years.

In line with a culture of evidence and learning, we propose to jointly reflect on the implementation status and success of implications 1 to 7 after 1.5 to 2 years. This is intended to allow for timely adjustments in BMZ's efforts towards more systematic and appropriate RIE implementation and evidence take-up.

6. REFERENCES

3ie (2021), "3ie Development Evidence Portal", 3ie Development Evidence Portal, *https://developmentevidence.3ieimpact.org/* (accessed 31.05.2021).

Africa Evidence Network (2021), "Africa Evidence Leadership Award", *https://www.africaevidencenetwork.org/en/learning-space/article/6/* (accessed 18 October 2021).

Avdeenko, A. and E. Heesemann (2020), "Evidence on Covid-19 pandemic control interventions and their impacts on health-related outcomes", *Policy Brief*, Center for Evaluation and Development (C4ED), Mannheim and Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

Bédécarrats, F., I. Guérin and F. Roubaud (2019), "All that glitters is not gold. The political economy of randomized evaluations in development", *Development and Change*, Vol. 50/3, pp. 735–762.

Biden, J. (2021), "Memorandum on restoring trust in government through scientific integrity and evidence-based policymaking", The White House, *https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/memorandum-on-restoring-trust-in-government-through-scientific-integrity-and-evidence-based-policymaking/* (accessed 16 August 2021).

BMZ (2020), *BMZ 2030 Reform Strategy: New Thinking – New Direction*, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ), Bonn/Berlin.

BMZ (2021), "Evaluierung der Entwicklungszusammenarbeit: Leitlinien des BMZ", *BMZ Papier No. 4*, BMZ Strategien, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ), Bonn/Berlin.

Bräuer, B. et al. (2020), "Evaluation Report 2020: Using knowledge", *Evaluation Report*, Deutsche Gesellschaft für internationale Zusammenarbeit (GIZ), Bonn/Eschborn.

Bruder, M., J. Faust and M. Krämer (2019), "Rigorous impact evaluation in German development cooperation", *Policy Brief*, No. 5/2019, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

Carter, S. et al. (2018), *Creating a Culture of Evidence Use – Lessons from J-PAL's Government Partnerships in Latin America*, J-PAL, Cambridge, USA.

Deaton A. and N. Cartwright, N. (2018), "Understanding and misunderstanding randomized controlled trials", *Social Science & Medicine*, Vol. 210, pp. 2–21.

CGD (2021), "Working group on new evidence tools for policy impact", Center for Global Development (CGD), *https://www.cgdev.org/working-group/working-group-new-evidence-tools-policy-impact* (accessed 27 September 2021).

Cochrane (2021), "Cochrane is supporting a culture of evidence", *https://www.cochrane.org/news/cochrane-supporting-culture-evidence* (accessed 18 October 2021).

CPID (2020), *Creating Learning Cultures: Assessing the Evidence*, Chartered Institute of Personnel and Development (CPID), London.

Deaton, A. (2020), "Randomization in the tropics revisited: A theme and eleven variations", *Working Paper*, No. 27600, National Bureau of Economic Research, doi:10.3386/w27600.

Dobbins, M. et al. (2009), "A description of a knowledge broker role implemented as part of a randomized controlled trial evaluating three knowledge translation strategies", *Implementation Science (IS)*, Vol. 4/1.

Duflo, E., R. Glennerster and M. Kremer (2007), "Using randomization in development economics research: A toolkit", in Schultz, T.P., J. Strauss and H.B. Chenery (eds.), *Handbook of Development Economics, Volume 4*, Elsevier, North-Holland, Amsterdam, pp. 3895–3962.

Evans, A. et al. (2021), "Evidence-informed development – Starting with the man in the mirror – Evidence dialogue", presented at the 3ie #EvidenceDialogues webinar, online.

Faust, J. (2020), "Rigorose Wirkungsevaluierung – Genese, Debatte und Nutzung in der Entwicklungszusammenarbeit", *Der Moderne Staat – Zeitschrift für Public Policy, Recht und Management*, Vol. 13/1–2020, pp. 61–80.

Florian, M. et al. (2019), "Rigorous Impact Evaluation: Cross-Section Analysis - Main Report", Knowing What Works, Deutsche Gesellschaft für internationale Zusammenarbeit (GIZ), Bonn/Eschborn.

Freudenreich, H. et al. (2020), "Effective interventions to increase food and nutrition security in response to Covid-19", *Policy Brief*, Thematic Team on "Rigorous Impact Evaluation", Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) and Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

Funk, E. et al. (2018), "Lessons learnt from impact-oriented accompanying research: Potentials and limitations to rigorously assessing the impact of governance programmes", *Discussion Paper*, No. 28, German Development Institute/Deutsches Institut für Entwicklungspolitik (DIE), Bonn.

Gertler, P. J. et al. (2016), *Impact Evaluation in Practice*, World Bank, Washington, D.C.

Glennerster, R. and K. Takavarasha (2013), *Running Randomized Evaluations: A Practical Guide*, Princeton University Press, Princeton, doi:10.2307/j.ctt4cgd52.

Head, B. W. (2016), "Toward More 'Evidence-Informed' Policy Making?", *Public Administration Review*, Vol. 76/3, pp. 472–484.

Højlund, S. (2015), "Evaluation in the European Commission: For accountability or learning?", *European Journal of Risk Regulation*, Vol. 6/1, Lexxion Verlagsgesellschaft mbH, pp. 35–46.

Kaplan, L., J. Kuhnt and J. Steinert (2019), "Ethics in Development Research: 'Doing No Harm' to research staff when conducting research in low and middle-income countries", German Development Institute/Deutsches Institut für Entwicklungspolitik (DIE), Bonn.

KfW (2021), *Evaluierungsbericht 2019–2020 Begleiten. Bewerten. Lernen.*, No. 16. *https://www.kfw-entwicklungsbank.de/Bilder/Evaluierungsbericht-2021/Startseite/KfW_Evaluierungsbericht_2019-2020.pdf*.

Kluve, J., J. Langbein and M. Weber (2020), "Protecting workers and firms in times of crisis: Key labour market policies for low- and middle-income countries", *Policy Brief*, Thematic Team on "Rigorous Impact Evaluation", Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ), Bonn, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn, Kreditanstalt für Wiederaufbau (KfW), Frankfurt and World Bank, Washington D.C.

Kuhnt, J. and K. Schüttler (2020), "Economic integration into host communities in times of crisis", *Policy Brief*, Thematic Team on "Rigorous Impact Evaluation", Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ), German Development Institute/Deutsches Institut für Entwicklungspolitik (DIE) and Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

Lech, M. et al. (2018), "Improving international development evaluation through geospatial data and analysis", *International Journal of Geospatial and Environmental Research*, Vol. 5/2.

Leeuw, F. and J. Vaessen (2009), *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*, NONIE Guidance on Impact Evaluation, World Bank, Washington, D.C.

Lemire, S., L. R. Peck and A. Porowski (2020), "The growth of the evaluation tree in the policy analysis forest: Recent developments in evaluation", *Policy Studies Journal*, Vol. 48/S1, pp. S47–S70.

Leppert, G. et al. (2018), *Impact, Diffusion and Scaling-Up of a Comprehensive Land-Use Planning Approach in the Philippines: From Development Cooperation to National Policies*, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

Manning, R., I. Goldman and G. Hernández Licona (2020), "The impact of impact evaluation", *WIDER Working Paper*, No. 2020/20, UNU-WIDER, Helsinki, Finland.

van der Meer, F.-B. and J. Edelenbos (2006), "Evaluation in multi-actor policy processes: Accountability, learning and co-operation", *Evaluation*, Vol. 12/2, SAGE Publications Ltd., pp. 201–218.

Muller, S. M. (2015), "Causal interaction and external validity: Obstacles to the policy relevance of randomized evaluations", *World Bank Economic Review*, World Bank Group, Vol. 29/1, pp. 217–225.

Murunga, G. R. and I. O. Ogachi (2020), "Randomised Control Trials and Development Research in Africa", Council for the Development of Social Science Research in Africa (CODESARIA), *Bulletin*, No. 1/2020., Special Issue.

OECD (2005), *The Paris Declaration on Aid Effectiveness*, Organisation for Economic Co-Operation and Development, Paris, France, *https://www.oecd.org/dac/effectiveness/34428351.pdf*.

OECD (2008), *The Accra Agenda for Action*, Organisation for Economic Co-Operation and Development, Paris, France, *https://www.oecd.org/dac/effectiveness/34428351.pdf*.

Peersman, G., I. Guijt and T. Pasanen (2015), *Evaluability Assessment for Impact Evaluation*, Method Lab Publication, ODI, London.

Polak, J. T., K. Guffler and L. Scheinert (2017), *weltwärts Volunteers and Their Civic Engagement in Germany*, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

Roxin, H. et al. (2021), "Effectiveness of German development cooperation in dealing with conflict driven migration crises: Executive summary", Deutsches Evaluierungsinstitut der Entwicklungs-zusammenarbeit (DEval), Bonn.

Savedoff, W. D., R. Levine and N. Birdsall (2006), *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, Report of the Evaluation Gap Working Group, Center for Global Development (CGD), Washington, D.C.

von Schiller, A. (2021), "Why GIZ Governance programmes should conduct impact assessments and how to get the most out of them", *How-to note*, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), Bonn.

Schmitt, J. (2020), "The causal mechanism claim in evaluation: Does the prophecy fulfill?", *New Directions for Evaluation,* Vol. 2020/167, *Special issue: Causal Mechanisms in Program Evaluation.*

Seawright, J. and J. Gerring (2008), "Case selection techniques in case study research: A menu of qualitative and quantitative options", *Political Research Quarterly*, Vol. 61/2, pp. 294–308.

Smith, K. E. et al. (2020), *The Impact Agenda – Controversies, Consequences and Challenges*, Policy Press.

Snilstveit, B. et al. (2017), "3ie evidence gap maps: A starting point for strategic evidence production and use", *3ie Working Paper Series*, No. 28, International Initiative for Impact Evaluation (3ie), New Delhi.

Steinert, J. I. et al. (2021), "A systematic review on ethical challenges of 'field' research in low-income and middle-income countries: Respect, justice and beneficence for research staff?", *BMJ Global Health*, Vol. 6/7, BMJ Specialist Journals.

Stewart, R. et al. (2018), "Building capacity for evidence-informed decision making: An example from South Africa", *Evidence & Policy: A Journal of Research, Debate and Practice*, Vol. 14/2, pp. 241–258.

Stockmann, R. (2004), "Was ist eine gute Evaluation? Einführung zu Funktionen und Methoden von Evaluationsverfahren", *CEval-Arbeitspapier*, No. 9, Centrum für Evaluation (CEval); Universität des Saarlandes, Saarbrücken.

Syspons (2018), *Externe Qualitätskontrolle des GIZ-Ergebnisberichts 2017*, Syspons GmbH, Berlin.

United Nations General Assembly (2015), "Transforming Our World: The 2030 Agenda for Sustainable Development", *https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E* (accessed 25 October 2021).

USAID (2016), "Evaluation. Learning from Experience. USAID Evaluation Policy", United States Agency for International Development (USAID), Washington D.C.

Waddington, H. et al. (2012), "How to do a good systematic review of effects in international development: A tool kit", *Journal of Development Effectiveness*, Vol. 4/3, pp. 359–387.

White, H. (2009), "Theory-based impact evaluation: Principles and practice", *Journal of Development Effectiveness*, Vol. 1/3, pp. 271-284, doi:10.1080/19439340903114628.

White, H. (2019), "The twenty-first century experimenting society: The four waves of the evidence revolution", *Palgrave Communications*, Vol. 5/47, pp. 1–7.

White, H. and S. Sabarwal (2014), "Quasi-experimental design and methods", *Methodological Briefs, Impact Evaluation*, No. 8, , UNICEF, Florence, Italy.

Yanguas, P. (2018), *Why We Lie About Aid: Development and the Messy Politics of Change*, Zed Books, London.

German Institute for Development
Evaluation (DEval)
Fritz-Schäffer-Straße 26
53113 Bonn, Germany

Phone: +49 (0)228 33 69 07-0
E-Mail: *info@DEval.org*
www.DEval.org

With funding from the

Federal Ministry
for Economic Cooperation
and Development

DEval
GERMAN
INSTITUTE FOR
DEVELOPMENT
EVALUATION