

d·i·e

Deutsches Institut für
Entwicklungspolitik



German Development
Institute

Discussion Paper

18/2020

Monitoring in German Bilateral Development Cooperation

A Case Study of Agricultural, Rural Development and Food Security Projects

*Sarah Holzapfel
Cornelia Römling*

Monitoring in German bilateral development cooperation

A case study of agricultural, rural development and
food security projects

Sarah Holzapfel

Cornelia Römling

Bonn 2020

Discussion Paper / Deutsches Institut für Entwicklungspolitik
ISSN (Print) 1860-0441
ISSN (Online) 2512-8698



Except as otherwise noted this publication is licensed under Creative Commons Attribution (CC BY 4.0). You are free to copy, communicate and adapt this work, as long as you attribute the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE) and the authors.

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.d-nb.de>.

ISBN 978-3-96021-129-7 (printed edition)

DOI:10.23661/dp18.2020

Printed on eco-friendly, certified paper

Dr Sarah Holzapfel coordinates the municipal development policy for the city of Dortmund. She wrote this paper during her time as a researcher in the research programme “Inter- and Transnational Cooperation with the Global South” at the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).

Email: sarah.holzapfel@gmx.net

Dr Cornelia Römling is an evaluator at the Competence Centre for Methods at the German Institute for Development Evaluation (Deval). This paper was part of her work during her time as a researcher at the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE) in the research programme “Inter- and Transnational Cooperation with the Global South”.

Email: Cornelia.Roemling@deval.org

Published with financial support from the Federal Ministry for Economic Cooperation and Development (BMZ).

© Deutsches Institut für Entwicklungspolitik gGmbH
Tulpenfeld 6, 53113 Bonn
☎ +49 (0)228 94927-0
☎ +49 (0)228 94927-130
Email: die@die-gdi.de
<http://www.die-gdi.de>



Acknowledgements

This discussion paper is part of the research project “Promoting food security in rural Sub-Saharan Africa: The role of private sector investments and results-oriented approaches” of the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE) and is funded by the special initiative “One World, No Hunger” (SEWOH) of the Federal Ministry for Economic Cooperation and Development (BMZ).

We would like to thank Manfred Metz, who conducted several of the interviews. Furthermore, we would like to thank DIE colleagues Michael Brüntrup, Christine Hackenesch, Raoul Herrmann, Heiner Janus, Stephan Klingebiel and Hannes Öhler for helpful comments and suggestions on prior versions of this discussion paper and this version. Last but not least, we thank all interview partners, who contributed to the paper.

June 2020

Sarah Holzapfel

Cornelia Römling

Contents

Acknowledgements

Abbreviations

Executive summary	1
1 Introduction	5
2 Background on RBM	7
2.1 History of the RBM debate	7
2.2 General setup of an RBM system	8
2.3 International quality criteria of RBM	9
2.4 RBM in German development cooperation	13
3 Methodology	15
3.1 Sample	15
3.2 Data collection and analysis	16
3.3 Limitations	17
4 Measuring results in German bilateral development cooperation	18
4.1 Step 1: Agreeing on objectives and developing a theory of change	18
4.2 Step 2: Selecting indicators to monitor objectives	20
4.2.1 Development process of indicators	20
4.2.2 Specific formulation of indicators	21
4.2.3 Relevance of indicators	23
4.3 Step 3: Finding data source and determining data collection methods	25
4.4 Step 4: Collecting baselines	27
4.5 Step 5: Setting targets	28
4.6 Step 6: Collecting monitoring data and using evaluation information	28
4.7 Step 7: Reporting and using findings	30
5 Conclusions and recommendations	31
References	37
Figures	
Figure 1: Setting up an RBM system	9
Figure 2: Primary and secondary data sources on different results levels (n=140)	25

Figure 3: Primary data sources (n=140)	26
Figure 4: Secondary data sources (n=140)	26
Figure 5: Usefulness of results (n=196)	31

Tables

Table 1: German “ <i>Wirkungsmatrix</i> ” (results matrix)	14
Table 2: Sample of projects	16
Table 3: Number of objectives per project (n=13)	19
Table 4: Example of a goal on a wrong results level	19
Table 5: Number of indicators per module	20
Table 6: Indicator formulation	21
Table 7: Unspecific indicators	22
Table 8: Indicators with sub-indicators	23
Table 9: Example for incorrect results level of indicators	24
Table 10: Irrelevant indicators	24
Table 11: Indicators measuring national developments/developments within the target region	27
Table 12: Inconsistency between indicator and reported data	29

Abbreviations

BMZ	German Federal Ministry for Economic Cooperation and Development / Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung
DFID	Department for International Development
DIE	German Development Institute / Deutsches Institut für Entwicklungspolitik
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit
IFAD	International Fund for Agricultural Development
KfW	Kreditanstalt für Wiederaufbau
LFA	logical framework approach
MDG	Millennium Development Goal
M&E	monitoring & evaluation
OECD	Organisation for Economic Co-operation and Development
RBM	results-based management
SDG	Sustainable Development Goal
SEWOH	special initiative “One World, No Hunger” / Sonderinitiative “Eine Welt ohne Hunger”
WFP	World Food Programme

Executive summary

In the late 1990s, the discussion on the effectiveness of development cooperation gained importance within the scientific community and among aid providers. It reached a peak in 2005 with the adoption of the Paris Declaration on Aid Effectiveness. The Paris Declaration lays out five principles, among them the principle “managing for results”, which encourages donors and partner countries to manage and implement aid in a way that focusses on the desired results and to use information to improve decision-making (Organisation for Economic Co-operation and Development [OECD], 2018). Moreover, the establishment of the Millennium Development Goals (MDGs) in 2000 and of the Sustainable Development Goals (SDGs) in 2015 contributed to a results-oriented thinking of development actors. Many aid providers developed detailed strategies and guidelines to plan and implement projects to achieve predefined results. On the project level, results-based management (RBM) was introduced as a tool from the private sector to create a more efficient and effective way of implementing projects. RBM is a results-oriented management tool for planning, monitoring and evaluation (M&E). In parallel, impact-evaluation approaches spread from the health sector into the development community to assess projects’ effectiveness. Measuring results has thereby become a high priority in development cooperation.

This so-called results agenda also faces substantive criticism. Several practitioners see it as introducing a very high burden on reporting for partner countries as well as implementing agencies. Large results frameworks might overburden projects in terms of time resources and capacity. Furthermore, it is associated with diminishing ownership through top-down donor-imposed approaches. Critics fear that the results agenda is preventing local solutions and reducing flexibility in adapting projects to local contexts and unforeseen events. Many actors see providing accountability to donors and taxpayers as the main aim of RBM while neglecting the purpose of learning. The focus on accountability might lead to biased reporting. Project staff feel pressured to achieve quantitative targets, whereas discussions that are critical of continuous learning and adapting projects are not seen as equally important. Reviews of donors’ RBM systems show that setting up and implementing RBM is challenging, and that the complexity of measuring results is often underestimated. Common RBM challenges that have been identified are, for example, the attribution and aggregation of results, ensuring the ownership of partner countries and developing a learning culture. Despite the critiques, a pragmatic and methodologically sound RBM system can increase the effectiveness of development assistance and contribute to a learning culture. To do so, ownership and local needs must be top priorities in such systems. Lessons and learnings from other aid providers have been published which can help to create RBM systems that consider all of the common pitfalls (Organisation for Economic Co-operation and Development [OECD], 2019b).

German development cooperation has been criticised by the peer review of the OECD for not having a comprehensive concept for RBM (OECD, 2015). BMZ is undertaking efforts as part of a recent reform process to improve its impact orientation and its approach towards measuring results (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung [BMZ], 2020). Action plans on results and data are going to be developed to address existing challenges. The reform process is still at a very early stage, and details have not yet been determined. Whereas the OECD peer review analysed the general structure of German development cooperation at a meta level, this paper sheds light on the quality of

monitoring and measuring results in German bilateral development cooperation at the project level. By outlining the challenges of RBM in German development cooperation, we would like to contribute towards the reform process. Building on a sample of 13 projects of the two main implementing agencies – Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and Kreditanstalt für Wiederaufbau (KfW) – the quality of the projects' RBM and their results reports to the Federal Ministry for Economic Cooperation and Development (BMZ) are analysed in detail.

The findings of the analysis show that there are several methodological problems in the information provided by the implementing agencies. First, no comprehensive theory of change is developed before the start of the project. Second, indicators are too complex and resemble goals rather than measurement units. Furthermore, the indicators are often not appropriate for measuring the corresponding goals. Although the effort in formulating objectives and indicators might help the projects in the planning stages, the methodological drawbacks identified in this analysis suggest that the collected data is often of limited use for reporting. Third, there is a challenge in attributing the results of outcome and impact-level indicators to the projects. This is because methodologically appropriate methods are often not applied, and information from monitoring is used for indicators on higher results levels. However, monitoring cannot provide evidence about causality on the outcome and impact levels.

The analysis concludes that comprehensive guidelines for implementing agencies on how to measure and report on results are lacking. Existing guidelines are not very detailed and not publicly available. Implementing agencies have compiled their own more detailed RBM guidelines. These list BMZ requirements in addition to their own requirements, which in some cases slightly diverge from BMZ standards, for example in terms of definitions. Ownership by the partners and strengthening partner countries' RBM systems are not main characteristics of the RBM process, although some projects definitely outperform others in this regard. As a result, the RBM system appears to be relatively complex and incoherent, and it does not answer the most important questions, for example with regard to attribution.

A reformed RBM system would hold important potential in contributing towards making German development assistance more effective. Overall, ways need to be found so that learning between different donors and partners on RBM is intensified. The guiding principles from the OECD on measuring development results could support a reform of the RBM system, even though these guidelines are not very concrete regarding the implementation of RBM systems (OECD, 2018, 2019a).

This analysis offers several recommendations to increase the quality of RBM systems, and therefore also the level of development cooperation's effectiveness.

First, it is of utmost importance to develop a more comprehensive RBM system that defines the setup, the responsibilities of all stakeholders and the usage of the collected data. The system should be created in a participatory manner and include all relevant stakeholders. Learning needs to be one of the main purposes of the system – additionally to accountability – to contribute towards increasing aid effectiveness via an incentive to report high-quality data. The system should be as simple as possible and not lack methodological rigour. It should provide high-quality data and, at the same time, put emphasis on ownership and

flexibility to allow for local solutions. Processes to secure the systematic usage of the collected M&E data should be developed. Details on the RBM system should be summarised in a set of guidelines. Other donors' experiences can provide interesting lessons when reforming the German system (OECD, 2019b). Especially the approach of the International Fund for Agricultural Development (IFAD) should be considered, which simplified its RBM system after detecting challenges as described in an IFAD report (2017).

Second, additional resources and capacities on RBM in BMZ are needed to develop, implement and monitor a reformed RBM system. The BMZ 2030 reform process highlights the need for additional resources, as BMZ not only aims to reform its RBM system but to also link resource allocation to results (BMZ, 2020). Moreover, knowledge, expertise and the competence to set standards should be concentrated within one unit. Different units in BMZ have been responsible for setting RBM standards. For example, several sectoral divisions have developed standards or aggregation indicators to aggregate results across projects and programmes. Furthermore, some sector-specific RBM guidelines have been created, for example by GIZ. As a result, there are diverging approaches towards aggregating results as well as guidelines within German development cooperation. Since RBM is part of every project and programme, enough staff resources are needed to develop and implement an ambitious RBM system. Due to the complex nature of RBM, it is important that positions be filled by M&E experts. The strengthened unit should be responsible for developing an overarching RBM system with respective guidelines and setting up a data management system. In addition, it should be in charge of monitoring the implementation of RBM standards at all levels (BMZ, country, programme and project) and compiling results reports. It should work more closely with the evaluation and development research division. In addition, the consistent capacity-building of staff in all relevant institutions and within partner countries should be another main pillar to increase the quality of measuring results in German development cooperation. Strengthening capacities for RBM within BMZ does not mean taking ownership and implementation responsibility from the implementing agencies, but rather re-organising the handling of RBM within BMZ.

Third, this discussion paper identifies five specific challenges of the German RBM system that should be addressed. These are the involvement of partners (ownership), the usage of a theory of change, the quality of the indicators, efforts to strengthen and use partner countries' RBM systems, and the attribution of results to the project. The RBM system for a project is often developed with only limited involvement of the partners. As a result, the system and indicators are not "owned" by the partners. The results matrix and the current theory of change are highly standardised and too simple to represent complex projects. The theory of change should play a more prominent role, and the results matrix should become more flexible and adaptable. Indicator development is in need of a strong guidance and quality check to yield useful and relevant information. Secondary data of partner countries used by projects to monitor results is often of only limited quality, indicating that efforts need to be increased to strengthen partner countries' RBM systems. Lastly, the strategy to attribute results from projects on the outcome and impact levels needs to be more strongly connected to high-quality (qualitative and quantitative) impact-evaluation methods. A limited number of projects should be selected for evaluations, and only these should have to report on attributable quantitative changes at the medium-term outcome and impact levels. Evaluations should be conducted with very high methodological standards, which are mostly lacking at the moment. All other projects should mainly focus their monitoring

efforts on the output and short-term outcome level, where the link to a project's activities is reasonably clear. Medium-term outcome and impact indicators – ideally derived from existing sources in partner countries – can still be used as context indicators to describe the environment the project operates in and show whether development efforts of all actors involved are going in the right direction. However, due to the variety of external factors influencing results at these levels, they should not be used to assess a project's or programme's success.

In order to explore the potential medium-term outcomes and impacts, projects not selected for evaluations could rely more on qualitative methods. For example, Theory of Change verification workshops could be carried out with all involved stakeholders to discuss the (potential) outcomes and impacts as well as the respective challenges in achieving them. In addition, standardised perception-based surveys could be introduced to inform about results at higher levels within the results chain. Perception surveys ask project beneficiaries whether they have experienced a change (e.g. in diet, income) as a result of the project. The implementing agencies, in particular GIZ, already use a variety of qualitative methods (e.g. focus group discussions, semi-structured interviews) but do not systematically report the findings of qualitative analyses in their progress reports to BMZ.

1 Introduction

Governments have always been engaged in showing the effectiveness of aid to taxpayers. This effort has been reinforced since 2000, when several studies and public discussions questioned the effectiveness of development cooperation and the achievement of project-specific development goals. Scholars controversially discussed the impact of development aid on poverty reduction (Easterly, 2006; Sachs, 2005). At the same time, research proposed new approaches in the field of M&E to improve the measurement of results (Banerjee, 2011; Center for Global Development [CGD], 2006). At the political level, the Paris Declaration on Aid Effectiveness, adopted in 2005, highlighted results-orientation as one pillar to increase aid effectiveness. The subsequent Busan Declaration (2011) reconfirmed this focus on results with the establishment of the Global Partnership for Effective Development Cooperation (OECD, 2011). The MDGs – and from 2015 onwards, the SDGs – were another driver for donors to become more results-oriented.

These milestones contributed to the emergence of the so-called results agenda. Individual donors have continuously reviewed and restructured their planning and M&E to improve the focus on results and show the effectiveness of their investments. The formerly used approach for planning and managing projects – the logframe approach – was replaced by RBM. Whereas the logframe approach was more heavily focussed on planning, inputs and activities, the RBM approach shifted the focus more to the results on the outcome and impact levels.

On the one hand, RBM is seen as the panacea for effective development cooperation. On the other hand, the focus on results is controversially discussed. Critics argue that the results agenda is not promoting more effective development cooperation. Local, flexible and innovative approaches are necessary to cope with rather complex development challenges and specific local needs (Eyben, 2013; Sundberg, 2019). It is argued that the results agenda instead leads to a more top-down approach. Furthermore, the results agenda gives the impression that there is a relatively straightforward problem–solution nature of development cooperation projects, so that effective approaches in one location can be easily transferred to another setting.

Moreover, the aid effectiveness agenda not only emphasises the focus on results. It also demands an increase in ownership by partner countries, alignment with partner countries' strategies, harmonisation among donors and mutual accountability as important strategies to increase aid effectiveness (OECD, 2018). The focus on results may conflict with other aid effectiveness principles. Questions arise as to whether the results effort supports projects and donors to become more effective as desired, or whether it is used for domestic accountability purposes in donor countries without taking partner countries' perspectives into account (Chambers, 2017; Holzapfel, 2016).

Apart from the overall results agenda, also the usefulness of the applied tool – RBM – has been discussed in a critical manner because many challenges have been found with respect to the RBM systems of different countries and institutions (Mayne, 2007; Perrin, 2002; Vähämäki, Schmidt, & Molander, 2011; Vähämäki & Verger, 2019). Specifically, data quality remains a key problem. Also, the detailed reviews of the RBM systems of several aid providers, such as Norway (Balogun & Lloyd, 2017), Sweden (Shutt, 2016), the United Kingdom (Whitty & Valters, 2017), the World Bank (Independent Evaluation Group [IEG],

2017) and a synthesis by the OECD (2017) have come to the result that most RBM systems are facing challenges. Common areas for improvement identified by the OECD (2017) comprise (i) linking results to goals and building narrative, (ii) ensuring RBM approaches are fit for purpose, (iii) being realistic about attributing and aggregating results, (iv) enabling country ownership of results information, (v) linking results and performance to inform delivery and (vi) enhancing resources to build a learning culture.

Different stakeholders have developed and discussed the standards that guide the establishment of an RBM system (see e.g. IEG, 2012; Kusek & Rist, 2004; Vähämäki et al., 2011).

Building on common standards, as defined in the literature on RBM, this paper is analysing one case – RBM and, in particular, project monitoring in German bilateral cooperation, as reported to the ministry in a set of agricultural, rural development and food security projects. The paper builds on those RBM standards, as defined in the literature, that contribute to the overall goal of increasing aid effectiveness. That means that the standards selected not only lead to the best possible measurement of results, but also contribute to a high level of ownership, increase harmonisation, allow for an adaptation to local needs and provide flexibility to cope with highly complex setting.

Germany is the second-largest bilateral provider of official development assistance in absolute terms. The last OECD peer review came to the conclusion that Germany lacks a clear concept for RBM (OECD, 2015, p. 77). Although some changes have been introduced for M&E procedures recently, a comprehensive concept is still missing. This might have consequences for the quality of RBM. The overall quality of German RBM has so far only been analysed as part of the external quality control of GIZ, and in one study by GIZ in cooperation with KfW about results measurement in health projects and programmes (Koppenleitner et al., 2012). In addition, Schwegmann and Holzapfel (2018) refer to monitoring in Germany and the quality of M&E data in a briefing paper.

This paper analyses the current RBM system in German bilateral development cooperation. To do so, a sample of 13 agricultural, rural development and food security projects were analysed in detail. Two questions are at the centre of the discussion paper:

- Does the M&E information and reporting in German bilateral cooperation comply with international RBM standards?
- Which conclusions can be drawn from the analysis with respect to the usefulness of the collected information and the influence of the information on the effectiveness of the projects?

To shed light on these questions, results matrices, indicators, reported data and the approach to their development are checked against common RBM standards using project documents and in-depth interviews with project staff as data sources.

The aim of this analysis is to inform the results community about the status of the RBM system in Germany and to contribute towards ongoing reform debates in German development cooperation by suggesting possible improvements for the RBM system.

Even though challenges specific to RBM in the context of agricultural, rural development and food security projects exist, the challenges identified in this paper are largely

transferable to other sectors. The general relevance of our findings for German bilateral development cooperation was confirmed during a discussion of an earlier version of the paper within the BMZ working group “results matrix” in May 2019. This working group includes colleagues from the implementing agencies of German bilateral development cooperation working mainly in positions connected to M&E in their agencies. In addition, both authors have carried out research and evaluations on the German RBM system in other sectors, where the findings were similar.

The analysis is structured as follows: Section 2 presents background information on RBM systems and its methodological standards. Section 3 describes the methodology for our analysis. Section 4 provides the results of the analysis of the status of RBM in German bilateral development cooperation. Finally, Section 5 draws conclusions and elaborates recommendations for practitioners and policy-makers.

2 Background on RBM

2.1 History of the RBM debate

The accountability of actors for their expenditures has always been important in development policy. As new public management came up in the 1980s and 1990s, state institutions in general had to account for the efficiency and effectiveness of public spending more than before. As a consequence, many countries integrated stronger accountability obligations into their guidelines and adopted private-sector tools to measure success in different sectors (Hood, 1995). The global financial crisis that began in 2007 led to greater austerity and an increased need to account for results.

The development cooperation sector began discussing the achievements of development aid in a particularly critical manner at the beginning of the 2000s. Within the academic community, some argued that development cooperation is effective only under certain conditions, for example when there is a minimal degree of institutional development (Burnside & Dollar, 2000). Others argued that the poor quality of measured results impedes reliable conclusions about whether aid is effective or not (Banerjee, 2011; CGD, 2006; Easterly, 2006). The increasing number of rigorous impact evaluations in the field of development cooperation created new opportunities. Policy-makers appreciated the knowledge that these evaluations generated about effectiveness and in relation to evidence-based policy-making. One of the most prominent examples of the rise in impact evaluations for policy-making is the Progres program in Mexico (Skoufias & McClafferty, 2001; Skoufias, Parker, Behrman, & Pessino, 2001). Within the international aid system, results-orientation was agreed upon as one goal by aid providers in the Paris Declaration to contribute towards increasing aid effectiveness (Abdel-Malek, 2015; OECD, 2018). In the following years, both the Accra Agenda for Action and the Busan Declaration continued to stress this focus on results (OECD, 2008).

These developments in the international aid system and inputs from academic research led to strong pressure being put on aid providers, implementing agencies and developing countries to demonstrate the effectiveness of their development measures. Although M&E has been part of development cooperation for a long time, all stakeholders started investing

more resources into developing instruments and systems to measure their results (Chambers, 2017). This was the start of the so-called results agenda.

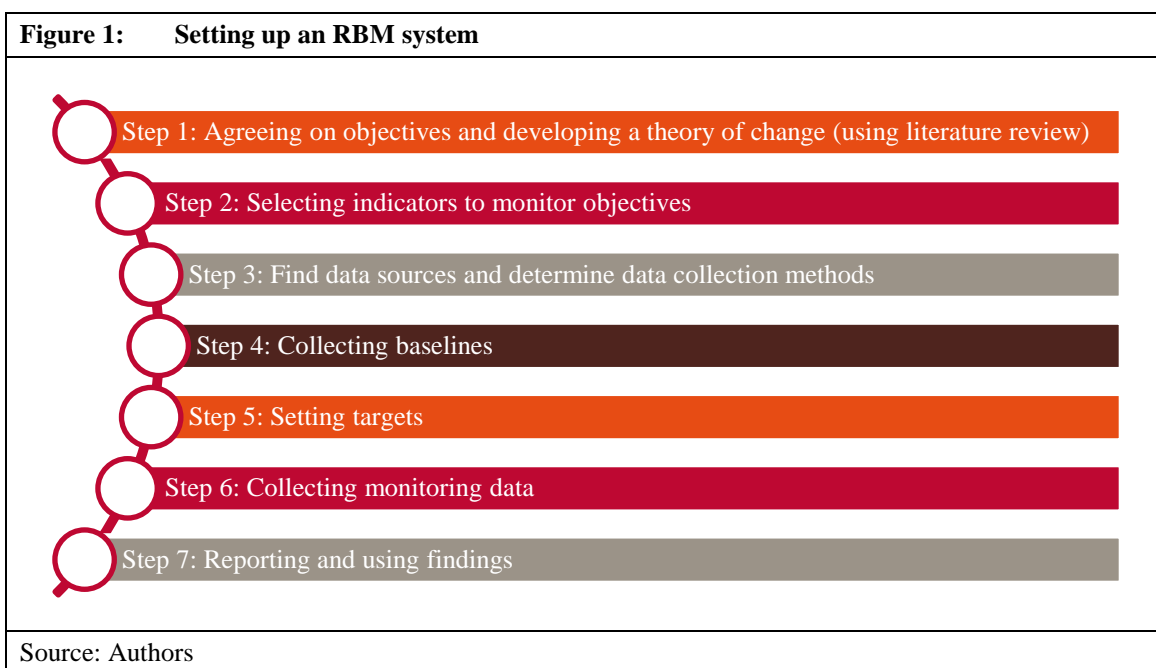
This increased focus on results led to strong criticisms. Shutt (2016) and Eyben (2013) describe in their works how the focus on results may work against local and innovative approaches, and thus may limit the effectiveness of development cooperation. They highlight that it is largely impossible to identify problems and solutions in development cooperation as clearly as the results agenda suggests, and that solutions from one context cannot be transferred to another. Alternative approaches such as *Doing Development Differently* and *Problem-Driven Iterative Adaption* have been developed to cope with this shortcoming (Andrews, Pritchett, Samji, & Woolcock, 2015; *Doing Development Differently*, 2014). In the view of Eyben (2013) and Shutt (2016), these are more apt to deal with the complexities of partner countries' problems, to ensure a strong rooting of the projects in the local environment and to provide a highly flexible approach for adequate solutions. Furthermore, there have also been criticisms about the focus on quantitative results and the proliferation of bureaucracy associated with the results agenda. It is argued that M&E efforts limit development results, as too many resources are spent to measure results, and the focus is shifted away from the intervention (Chambers, 2017). The resources (time, financial and capacity) needed for setting up and implementing an RBM system can indeed be high and are often underestimated.

However, it is possible to address the critics' concerns and to integrate ownership and flexibility in an RBM system. Moreover, systems can be created that are not focussed on classical quantitative figures and do not overburden staff. For this to happen, common standards for an RBM system that is derived from the literature are presented in the following section.

2.2 General setup of an RBM system

RBM is a strategy for measuring results that was adopted from business management in which performance and the achievement of output, outcome and impact are central components (OECD, 2010). It is one example of a business tool being integrated in public management in line with the thinking of the new public management approach. It stands for a complex procedure within an institution to measure results and use them to take management decisions. There are two main purposes: using information for learning and management, and reporting on results to fulfil accountability demands (Binnendijk, 2000).

RBM was developed from the previously used logical framework approach (LFA), and is relatively similar when looking at its main characteristics (Örtengren, 2016). However, whereas LFA – or, in general, earlier M&E approaches – focussed more on implementation, now there is a stronger emphasis on results, that is, results at the outcome level. The following seven steps are commonly part of an RBM system (IEG, 2012; Kusek & Rist, 2004; United Nations Development Group [UNDG], 2011):



The management of the system, that is, using results information for continuous learning and for adapting projects, is the main pillar, which makes it useful and is the reason why it is used in the end (Örtengren, 2016). Unfortunately, as the United Nations Development Group (2011) points out, the term “management” is often neglected in RBM.

Many donors provide information on their specific concepts for measuring results (e.g. Department for International Development [DFID], s.a.; IEG, 2012; Örtengren, 2016; UNDG, 2011). These concepts are influenced by the general priorities of the institution/country, for example a more quantitative focus by DFID and the World Bank, and a greater focus on ownership by the Swedish International Development Cooperation Agency (Keijzer, Klingebiel, Örnemark, & Scholtes, 2018). BMZ does not yet have a comprehensive concept showing its approach towards measuring results (see Section 2.4 for further details).

2.3 International quality criteria of RBM

The quality criteria for RBM systems can diverge, depending on the specific objectives and priorities. When the purpose of RBM is focussed purely on measuring results as accurately as possible, other goals such as ownership, flexibility and harmonisation will be neglected. Qualitatively measurable aspects or the cost–benefit ratio of collecting the data would also not be important priorities. This paper assumes that all different goals (i.e. measuring results accurately, promoting ownership, maintaining flexibility and harmonising M&E efforts) need to be considered in order to achieve more effective projects. Therefore, quality criteria are highlighted that contribute to this broader concept of aid effectiveness.

The process of building an RBM system needs to be very participatory, reflexive and flexible (IEG, 2012; IFAD, 2002). This ensures that the RBM system is understood and accepted by all stakeholders, and that there is a greater incentive to contribute. Additionally, the usage of M&E structures in the partner countries or the development of new RBM

systems by the partner countries should be supported. This improves the capabilities of the partner countries to become self-reliant, take over the assessment of the aid projects and monitor the development of their country. General guiding principles have been developed by the results team of the OECD in stakeholder consultations (OECD, 2019a). Several aspects are highlighted to cope with the challenges found in OECD reviews of RBM systems: a learning culture, ownership, transparency and a simple but reliable system.

Step 1: Agreeing on objectives and developing a theory of change

In step 1 of setting up an RBM system, an overall objective is formulated, together with the partner, that is derived from a needs assessment, for example. In general, an objective can be formulated more broadly than an indicator, which has to be very concrete and specific in terms of the information it requires (IEG, 2012). This overall objective is then the basis for developing a theory of change. The overall objective is located at the impact level, and then underlying objectives are formulated that can be ascribed to the outcome and output level.

In order to decide on the activity to achieve these objectives, reporting from other projects and additional impact-evaluation literature should be used to inform decisions with regard to selecting the most promising intervention in terms of effectiveness, and the most useful intervention in terms of the needs of the partners (IEG, 2012). This illustrates the life cycle approach (UNDG, 2011). Reporting is not only done for the project itself; it should also feed into decision-making on new development projects.

This process needs to be carried out together with, and according to, the needs of the partner (IEG, 2012; UNDG, 2011). The theory of change can then be used to fill in the results framework (IEG, 2012).

Step 2: Selecting indicators to monitor objectives

The selection of indicators is challenging but very important, as they are the crucial tool for measurement. An indicator aims to measure an objective as accurately as possible in order to provide information on the achievement or progress of the goal. Although several indicators will often be necessary to cover the content of an objective comprehensively, the number of indicators should be limited to two to three in order not to overburden management and reduce costs in data collection and evaluation (IEG, 2012; Kusek & Rist, 2004; UNDG, 2011).

Quality criteria for indicators are manifold.¹ But any set of quality criteria will require that the indicator should be as clear as possible as well as easily understood and unambiguously formulated; this is commonly described under the criterion “specific” (IEG, 2012; Kusek & Rist, 2004; UNDG, 2011). Therefore, no words should be used that lead to subjective interpretations. A third party should be able to collect data without the need for interpretation, for example when staff changes occur. An indicator should answer the questions of “who” should achieve “what” and “where” and be as clear as possible (Global Affairs Canada, 2016).

1 E.g. the concept of “SMART” indicators (specific, measurable, achievable, relevant and time bound) presented by the World Bank’s IEG (2012), or the slightly different concept for indicators “CREAM” (clear, relevant, economic, adequate and monitorable) (Kusek & Rist, 2004).

Another important criterion, “measurability”, refers to whether the indicator can actually be measured – and for reasonable cost and effort (UNDG, 2011). An indicator is measurable if either adequate secondary data exists or there are sufficient financial resources and capacities to collect primary data. When agreeing on an indicator, the data collection process should already be defined and thought through.

The criterion relevance of an indicator implies that the indicator is – according to expert opinion – a good indicator for measuring the object, and that it corresponds to the objective in terms of content (IEG, 2012). That means at the same time that the indicator is on the same results level as the objective. Reviews found difficulties especially in the definition of relevant indicators at higher results levels (Vähämäki et al., 2011). A common mistake is that indicators at these levels are identified on a lower level, for example the output level, as data can be collected more easily and results are attributable to the intervention.

Qualitative and quantitative indicators should be part of the RBM system to cover all aspects of the intervention. Many qualitative indicators are quantified in the end, for example “% of users satisfied with the service”. Apart from perception and opinion indicators, all indicators are defined as qualitative in cases that need further criteria to make the assessment objective and where counting is not straightforward or there is no way of presenting the information clearly. The results agenda is instead associated with quantitative indicators, and some of the aid providers recommend using quantitative indicators or implicitly promoting the usage of quantitative indicators (IEG, 2012).

In terms of harmonisation, but also to avoid common mistakes, it is helpful to search for existing indicators from other sources (Kusek & Rist, 2004). One useful source in this regard are the SDG indicators, which are mostly located at higher results levels (Inter-Agency and Expert Group on SDG Indicators, 2019). Some institutions also set indicators that the projects should contribute towards or provide exemplary indicators as guidelines (IFAD, 2017; World Bank, 2013a).

Step 3: Finding data sources and determining data collection methods

The term “sources of verification” – often used in logical frameworks – refers to data sources and data collection methods. These are to be identified in step 3. Primary as well as secondary data can be used. In terms of ownership but also survey fatigue and efficiency, secondary data should be given priority (IEG, 2012). However, availability and data quality are the main challenges of secondary data (Kusek & Rist, 2004). Primary data collection is, on the other hand, accompanied by significant costs and great effort. Moreover, experts need to be involved to collect high-quality data.

In step 3, it is also important to note down if the information can be made available through routine monitoring, or if an evaluation has to be carried out. Monitoring focusses at the input, activity and output levels (e.g. resources spent, number of workshops conducted and number of people receiving a food security package) (Riely, Mock, Cogill, Bailey, & Kenefick, 1999). Such information can be collected relatively easily on an ongoing basis and is generally directly attributable to a project. Evaluations aim to answer more complex questions, which can oftentimes not be answered directly using monitoring data (e.g. impact questions); in these cases, more sophisticated methods of data collection and analysis need to be applied (Riely et al., 1999; UNDG, 2011). Evaluations are important with regard to the possible attribution of the results to the project. Short-term outcomes (e.g. technologies

adopted by farmers with assistance from the project, use of outputs and services provided by the project) are influenced by external factors, but a link to the project is usually reasonably clear. At higher results levels (e.g. household income, yields or food security), external factors such as weather patterns and general economic developments may play such a large role that, without an evaluation, it is not possible to draw any conclusion about whether the project has had any influence on the results measured.

It is one of the main problems of RBM in development cooperation that attribution can often not be demonstrated at higher results levels due to the lack of applying methods that help in drawing conclusions on causality (IFAD, 2017; OECD, 2017). For outcome- and impact-level indicators, only research designs of a quantitative nature using comparison groups (e.g. randomised controlled trials), before–after comparisons with comparison groups or qualitative methods that systematically test the contribution of the intervention to the results obtained can provide evidence. With respect to the less-known qualitative methods, this can take the form of a so-called contribution analysis (Mayne, 2008), a realist evaluation (Westhorp, 2014), a performance story (Roughly & Dart, 2009) or process tracking (Collier, 2011). All of these methods can only be applied in evaluations because of the amount of time they require.

Step 4: Collecting baselines

Baseline values should be collected before the implementation of the project in order to have a reference point against which success of the project can be assessed (UNDG, 2011). Baseline values at the output level will often be zero, as these results stem directly from the project (roads built, people trained, etc.). Most donors oblige implementing agencies to collect baseline data, as it provides a reference point with regard to the measured results (Kusek & Rist, 2004).

Step 5: Setting targets

Indicators are a means to set exact targets, and thus increase motivation to achieve certain benchmarks (IEG, 2012). One criterion with respect to targets is that they should be achievable, that is, targets are set realistically (IEG, 2012). The baseline value provides important information for defining the target value. Furthermore, local knowledge and knowledge of similar projects in different settings is necessary to develop a realistic and ambitious target value. At the same time, external factors (e.g. weather events, economic downturns and diseases) can influence the achievement of the targets (positively and negatively) and need to be investigated in detail when assessing the success of the objective. A final aspect mentioned in all quality standards with regard to defining a target relates to the definition of a time horizon (time-bound). The RBM system needs to determine by when the objective should be reached (Asian Development Bank, 2006; IEG, 2012; World Bank, 2013b).

Step 6: Collecting monitoring data

Now the RBM system can go into action and data can be collected. Monitoring data will be collected continuously, especially if the source is own administrative data such as training data. Furthermore, own primary data can be collected on the agreed schedule. Secondary data can be collected either whenever it is available or on agreed points in time. Data collection needs to take place as consistently as possible. Therefore, the baseline data

collection sets the framework for the following data collection processes. That means that the data sources and data collection instruments used should be the same at different points in time (e.g. at baseline, midterm review and project completion). Otherwise, the data is not comparable. A further quality criterion in this regard is to collect data regularly without gaps. Nevertheless, measuring resource-intensive outcome- and impact-level data annually might be a waste of resources, because the effects will come about only after some years. Monitoring outcome- and impact-level data without conducting an evaluation does not provide informative data, as the information cannot be attributed clearly to the project. In analysing the data, it is important to use all the information available. This means that monitoring as well as evaluation findings should be consulted to triangulate findings as much as possible.

Step 7: Reporting and using findings

A clear system needs to be elaborated for how the collected data is to be used on the different levels, and for what and by whom. The pure monitoring of data feeds directly into the steering process during the implementation of the project. Formats need to be determined in which the data is reported to the relevant audiences.

These elaborations on the different steps of an RBM system show how many different aspects need to be considered when setting up and implementing an RBM system. This overview on the main characteristics and quality criteria for RBM systems will be the basis for our analysis in Section 4.

2.4 RBM in German development cooperation

Measuring results and presenting impact is important in German development cooperation. Still, a comprehensive and coordinated RBM system has not yet been established by BMZ. This was also the result of the OECD peer review on development cooperation in 2015 (OECD, 2015). BMZ refers on its website to aid effectiveness and emphasises the importance of ownership by the partner countries as well as alignment with the partner countries as priorities for project implementation (BMZ, 2009, 2010-2019b, 2011). A direct reference to improving the measurement of results or focussing more on results – as foreseen also in the Paris Declaration – is not made.

With regard to M&E, BMZ emphasises the role of evaluation for German development cooperation (BMZ, 2010-2019a). Some information on M&E can be found in the general guidelines for German bilateral cooperation (BMZ, 2008). These guidelines refer, among other things, to aspects of regular reporting in the form of progress reports and final reports after the projects end.

More precise information on measuring the results of development projects in bilateral cooperation is given in internal manuals (BMZ, 2012). Still, this information is mixed in with general project planning procedures and does not provide comprehensive information. Currently, BMZ is involved in a comprehensive reform process, in which further changes that touch upon RBM in its development cooperation strategy can also be expected (BMZ, 2020).

The main tool of reporting in German development cooperation is the results matrix (*Wirkungsmatrix*²), which is obligatory for bilateral projects. It entails three results levels: output, module and programme objective, corresponding to the structure of German bilateral development cooperation (Table 1). Single projects are called modules, and thematically close modules in one country are combined within a programme to increase harmonisation and coordination. However, it is noted that these three BMZ results levels should at the same time correspond to the usually defined results levels (i.e. output = output, module = outcome, programme = impact) (BMZ, 2012). Only at the programme level is there some flexibility, and the objective can also be defined at the outcome level.

No further differentiation within the different levels (e.g. several outcome levels building on each other) is foreseen. The results matrix sets a relatively rigid three-level results model for reporting.

Guidelines specify that there should be one objective for each programme and one objective for each module. Only in complex technical cooperation projects with several fields of activity is it possible to define more than one module objective. At the output level, several objectives can be formulated. The number of indicators at the programme and module levels should be limited to three to five, whereas there is no limit given for the output level. Indicators are required to be specific, precise, relevant, realistic, measurable and time-bound. However, exact definitions of these quality criteria are not provided in the documents of the ministry.

Table 1: German “Wirkungsmatrix” (results matrix)			
Summary	Success indicators	Sources of verification	Assumptions/risks
Programme objective = Impact level (Long-term effects, including positive and negative, primary and secondary, produced by a development intervention, directly or indirectly, intended or unintended)	Baseline Target value Reported value 1: Reported value 2: ...		
Module objective = Outcome level (Direct (short-term and medium-term), positive and negative, intended and unintended effects that result from the use of outputs for the target group/public goods)	Baseline Target value Reported value 1: Reported value 2: ...		

2 With regard to harmonisation, it is important to note, that *Wirkungsmatrix* is not an official German translation, according to the OECD glossary for results matrix or logical framework, but a completely new term. The glossary proposes using “*Ergebnismatrix*” or the English term “results framework” for similar products.

Table 1 (cont.): German “ <i>Wirkungsmatrix</i> ” (results matrix)			
Output (Technical capacities, personal competence or knowledge (outputs) that result from the use of resources and the implementation of activities). The products, capital goods and services that result from a development intervention)	Baseline		
	Target value		
	Reported value 1:		
	Reported value 2:		
	...		
Source: Own translation and adaption from BMZ (2012)			

Within BMZ, different units are in charge of M&E. There is a division for evaluation and development research, while the division “effectiveness, transparency, quality standards” is responsible for the results matrices used at the project and programme levels, as well as for developing a system to aggregate results at BMZ. Both of the implementing agencies, GIZ and KfW, have their own divisions covering M&E matters and are equipped with more resources for M&E than BMZ, which has very limited human resources in that area. They also provide their own guidelines laying out BMZ’s requirements for RBM as well as additional requirements. For example, GIZ requires that a theory of change (*Wirkungsmodell*) is formulated in all projects. This theory is to be used as the basis for drafting the results matrix. The theory of change is, however, not part of the reporting to BMZ.

The main responsibility for developing results matrices lies with the implementing agencies. BMZ regional divisions provide input and approve the results matrix. Although regional divisions can seek advice from the effectiveness division, there is no centralised quality control of results matrices within BMZ. Quality control is mainly done internally in the implementing agencies. According to the OECD peer review (OECD, 2015), BMZ has to ensure it has the adequate capacities and incentives in place, and it needs more steering from the top to establish a stronger results culture. Currently, a reform process is ongoing in BMZ, with the aim of reforming the RBM system.

3 Methodology

3.1 Sample

This discussion paper is based on an internal paper for the BMZ special initiative “One World, No Hunger” (SEWOH). The internal paper aimed at providing recommendations on how to improve RBM in agricultural, rural development and food security projects. This paper examines data from a sample of projects of bilateral cooperation in those three thematic areas.³ Even if the projects were selected from a limited number of sectors, lessons can also be drawn that go beyond these sectors (see Section 3.3). We highlight only those

³ Related internal work was financed by the responsible department at BMZ. These are all non-SEWOH projects.

challenges that are transferable to other sectors and have been encountered in previous research or evaluation practice by the authors. Important selection criteria for projects were:

- bilateral technical and financial cooperation projects
- activity in the thematic area of the special initiative SEWOH⁴
- projects already in progress (started before 2016)
- necessary project documentation available

The selected projects all started before the procedural reform (*Gemeinsame Verfahrensreform*) in German development cooperation that took place in 2017. As M&E was not a main focus of the procedural reform, no substantial changes in measuring results occurred. Therefore, the results from the analysis of the projects for this paper are still relevant today. Thirteen projects were selected for an in-depth analysis. Eleven projects are implemented in Sub-Saharan Africa. Two projects are located in Asia. Six projects are financial cooperation projects of KfW, and seven projects are technical cooperation projects implemented by GIZ.

Region	Type of cooperation	Number	Total
Sub-Saharan Africa	Technical cooperation	6	11
	Financial cooperation	5	
Asia	Technical cooperation	1	2
	Financial cooperation	1	
Total		13	13

Source: Authors

3.2 Data collection and analysis

As a basis for this discussion paper, documents relating to the 13 projects were analysed, and 17 interviews with project staff were conducted. We received project proposals, amendments, yearly reports and final reports from BMZ and the implementing agencies. In addition, project staff were asked whether there was additional information that could be shared. In several projects, staff made baseline studies, midterm evaluations and final/impact evaluations available. The document analysis covered all project documents that were made available. In all projects, project proposals and at least one – but mostly several – report were made available. Final reports had already been made available for 6

4 The sampled projects are drawn from the six thematic fields that SEWOH focusses on: (1) food security, (2) resilience and food security in crises and conflicts, (3) innovation in the agricultural and food sector, (4) ecological-social structural change in rural areas, (5) sustainable use of natural resources in rural areas, (6) responsible land and land use rights. All projects are financed by BMZ's bilateral programme and not by SEWOH itself.

of the 13 projects. Apart from that, six baseline studies, one midterm study, five impact evaluations and three evaluations fed into the analysis. Document availability is as of 2017. In total, 73 documents were analysed, mainly proposals (22) and annual reports (30). An analysis table was developed to undertake the systematic document review. In this table, information on the different quality standards was collected per indicator. Answer categories were determined for each of the quality standards and were either binary (yes/no), a fixed set of categories (high, medium, low) or free text (e.g. for data source: national statistics).

Another source of data were insights from semi-structured interviews, which were conducted for the 13 projects with the contract managers and/or M&E managers of the implementing agencies. Seventeen persons were interviewed. The interviews covered the following topics:

- selection of indicators, definition of impact logic and objectives
- data collection and monitoring process
- use of the data for decision-making and management

Whereas the document review covered reported information, the interviews provided mainly information on processes. Furthermore, the interviews shed light on the reasons for missing information in the reporting and provided background information on the process of data collection and usage.

The analysis is structured according to the different steps of setting up the RBM system:

1. agreeing on objectives and developing a theory of change
2. selecting indicators
3. finding data sources
4. collecting baselines
5. setting targets
6. collecting monitoring data
7. reporting and using findings

The quality criteria for RBM – and indicators, in particular – described in Section 2 guide the analysis. Conclusions and recommendations are derived from the analysis.

3.3 Limitations

The reports to the ministry that were analysed for this paper entail comprehensive information. Still, implementing agencies suggested that internal reporting may be more detailed than the documents for BMZ. Therefore, the challenges encountered might not necessarily apply equally within the implementing agency's internal reporting systems.

However, the interviews suggest that procedures vary considerably between different projects, so that not all projects have much more detailed information available. In addition, project staff had the opportunity to provide additional information on their RBM system in case their efforts went beyond BMZ requirements. Nevertheless, a general quality criterion is that all information is easily understandable by third parties, that is, BMZ, without consulting further data sources.

Regarding the use of reported information for learning, this discussion paper focusses on the use of information by BMZ. Conclusions do not refer to the implementing agencies' internal learning and evaluation systems.

The sample of bilateral projects is relatively small and only agricultural, food security and rural development projects were analysed. Whereas there might be some characteristics specific to this field in M&E, such as the types of indicators used, this paper analysed general quality issues, such as the ways of formulating indicators and reporting. The guidelines are the same for all sectors. Some additional working aids may exist in some areas, but in general the rules to be applied are the same. In the following sections, the paper analyses the implementation of German RBM by means of the standards presented in Section 2.

4 Measuring results in German bilateral development cooperation

4.1 Step 1: Agreeing on objectives and developing a theory of change

Overall, the interviews reveal that the procedures for preparing project proposals and results matrices vary and are not standardised across projects. Whereas in many projects staff cooperate closely with the partners, other project staff develop the results matrix mostly on their own and discuss it in-depth only at a late stage of planning with their partners. The frequency or depth of the exchange with the BMZ during the process varies as well.

Furthermore, the analysed information shows that projects which are implemented in partnership with multilateral donors cannot always adhere to the German results matrix, as other donors have different guidelines. This illustrates that the development of the results matrix not only depends on the implementing agencies and their planning, but that it is also shaped by BMZ or other donors. Although the results matrix and indicators are discussed and agreed with partners during the appraisal, partners often do not play a major role in the selection of indicators.

Project staff highlighted in interviews that results matrices benefited in particular from already elaborated matrices from prior projects in cases of follow-up projects. This reveals the complexity of developing a results matrix.

With regard to the different objectives in the results matrix, the number of objectives in the projects varies from project to project and does not depend on the size of the project (according to the total budget or budget of technical/financial cooperation). The projects have between one and nine objectives (see Table 3).

Objective	Min.	Max.
Programme	0	1
Module	1	4
Output	0	7
Total	1	9

Source: Project documents and own elaboration

At the programme and module levels, most projects determine only one objective. At the output level, on the other hand, an average of three to four objectives are formulated per project. Some projects do not formulate an objective at the programme level and/or output level. Even if there are case-specific reasons why no programme objective is formulated (e.g. lack of affiliation to a programme), it leaves the results matrix incomplete. This limits the usefulness of the tool, as it further reduces the theory of change, and therefore the understanding of the mechanisms of the project.

When analysing the consistency of the level of the goal with the predefined level in the matrix, it becomes evident that projects do not always interpret the BMZ levels of the results matrix as output, outcome and impact. In 8 of the 13 projects, one or more goals are not formulated at the correct level, using the BMZ definition (BMZ, 2012) as a basis. In total this refers to 15 of the 71 goals. The main challenge is the distinction between module and output levels, where this mismatch happens at a higher rate.

Results level (according to the project)	Objective	Comment
Output	Farmers improve the marketing of their products.	This refers to how farmers change their behaviour (effect that results from the use of outputs) and measures a short-term outcome.
Module (outcome)	The capacities of pastoralists to cultivate sustainably natural resources in region x are strengthened.	This goal describes an output goal (capacities, which result from the implementation of activities).

Source: Authors

With regard to the formulation of the objectives, often several aspects are combined in one objective. Thus, several indicators are necessary to cover the contents of the objective. The tendency to formulate relatively complex objectives might be the result of the requirement to generally define only one module and one programme objective.

Overall, the limited number of levels in the results matrix is a main problem in the German monitoring system, leading to very complex goal formulations and to missing results levels.

4.2 Step 2: Selecting indicators to monitor objectives

4.2.1 Development process of indicators

After project objectives are defined, the next step is to identify suitable measurement units in the form of indicators to assess whether objectives have been achieved. The interviews do not reveal an exact procedure for defining indicators. Most of the projects either rely on indicators from previous projects and/or rely on information and discussions with national partners. Programme-based approaches are relatively prone to high levels of ownership by the partners in this regard. Projects with different donors involved suffer from their distinct guidelines. As a consequence, indicators are either adopted despite not fitting with the German system (e.g. activity indicators) or new indicators are developed in parallel to satisfy the needs of BMZ. Many projects involve the partners when selecting indicators. However, some see the development of the indicators rather as an internal procedure within German development cooperation, with little involvement by the partners. According to the interviewees, this depends on the involvement and the ownership of the partners in the project in general.

The total number of indicators per project varies widely. The results in Table 5 show that projects use between 4 and 22 indicators to measure results. However, only a few projects determine less than 10 indicators (3 out of 13). The smallest number of indicators is set at the programme level, and the largest number of indicators at the output level – in line with the different number of objectives at the different results levels.

Objective	Mean	Min.	Max.
Programme	2.85	0	6
Module	4.7	2	10
Output	7.53	0	14
Total	15.08	4	22

Source: Project documents and own elaboration

Four projects do not define indicators at the output or programme level. Several indicators are not reported on further after introducing them in the proposal or amendment. This applies to half of the projects (7 of 13; 46 of 196 indicators). In their reporting, the staff of some projects give a reason for eliminating the indicator later on. They refer, for example, to data collection problems or the excessive costs of data collection. These are both indications of problems of measurability, which become evident when data is collected for the first time, for example for baseline values. Other reasons for the omission of indicators refer to changes in the results matrix, for example shifts in objectives and indicators. There is also an incidence of 34 out of 196 indicators in three projects that are no longer reported, and no further explanation has been given. In nine projects, changes in the indicators of the results matrix took place over time. Either indicators were newly added or the formulation of the indicator was changed for similar reasons as mentioned above.

According to the definition in Section 2, 110 of 196 indicators are of a quantitative nature, and 86 (quantified) are qualitative indicators. As expected, all projects contain quantitative indicators; four projects do not define qualitative indicators. Two projects had to take over other donors' indicators, which are purely quantitative. This might also be connected to the nature of the projects, as projects with governance activities refer more strongly to qualitative indicators.

4.2.2 Specific formulation of indicators

When assessing the formulation of indicators, it seems that certain formulations prevail in reporting in German (bilateral) development cooperation. A full sentence including the target is formulated, so that it oftentimes resembles goal formulations. This leads to some confusion on the reporting level and makes the indicator less specific. Table 6 shows an example of a "German" indicator vs. a methodologically correct definition of an indicator, according to international standards in RBM (see e.g. Inter-Agency and Expert Group on SDG Indicators, 2019; Kusek & Rist, 2004). DFID highlights in its guidelines that indicators should not be formulated as objectives, indicating that the problem might exist with other donors as well (DFID, 2011).

Table 6: Indicator formulation	
Typical "German" formulation	Correct formulation
xx formally registered small enterprises are still on the market after one year.	Number of formally registered small enterprises Baseline: xx Target: xx Until: xxxx
Note: To monitor the development over time ("after one year"), the regular data collection and tracing of enterprises is sufficient. It does not need to be mentioned in the indicator. Source: Project documents and own elaboration	

Moreover, some formulations make indicators unspecific. "To strengthen the ..." or subordinate clauses such as "to properly manage ..." or "increase in ..." or "rise in ..." are not meant to be part of an indicator. Instead, only the pure measurement unit should be mentioned. The target value will then determine in which direction the value of the measurement unit should change. A consequential error of this formulation is that it creates confusion with regard to the baseline values. For example, when "increase in" or "rise in" is part of the indicator, project staff determine the baseline value often as zero, even though the starting point of the actual measurement unit should be in the centre.

Other terminologies hamper the measurability of an indicator. Formulations such as "decision-oriented", "climate change-relevant/-oriented" and "food security-relevant/oriented" are terms that can be interpreted and assessed very differently. Particularly in qualitative indicators, formulations are often not specific (e.g. indicators in which the improvement of institutions or the contribution of target groups is to be measured). If unclear terminology is used, a clear definition needs to be provided.

An additional frequent problem is the incomplete specification. The questions what, where and who should be answered (Table 7). The different problems due to an unspecific formulation apply to almost half of the indicators (86 of 196).

Indicator	Problem	Suggestion
The improved institutional framework contributes to strengthening the drought resilience of the (agro-) pastoral population.	The formulation of an objective rather than an indicator. What exactly does “improved institutional framework” mean? How is “strengthening of drought resilience” determined? How can this be measured? “Where” is not answered.	Percentage of households applying for food aid in the rural programme area during the last drought; number of xy strategies implemented in the programme area.
The equal participation of women and the protection of their interests are guaranteed within the framework of cultural (and economic) opportunities.	Two aspects included: participation and protection of interests.	Two single indicators should be developed. More information is needed to decide on exact indicators. The term “framework” is unclear.

Source: Project documents and own elaboration

A related problem is that, from the indicator alone, it often does not become clear among which group or within which geographical region data is collected. For example, the indicator “share of rural food-insecure households” does not provide information about whether the indicator measures food security at the national level, at the regional level or specifically among project beneficiaries and non-beneficiaries within the target region. Even after consulting project reports, it is often not possible to establish clarity about the target group of the indicator.

Overall, most of the analysed projects suffer from unspecific indicators. Often, the text outside of the results matrix and/or further reporting needs to be consulted in order to understand exactly what should be measured.

Furthermore, indicators often combine two or more measurable aspects. These indicators are often already indicated as being unspecific under the aforementioned criteria. The inclusion of sub-indicators leads to a high level of complexity and a high probability that measurable aspects will be neglected in data collection and reporting. Moreover, the formulation of indicators with different sub-indicators can be problematic if electronic databases are used with only one possible value per indicator. The following examples illustrate how these combined indicators could be separated (Table 8).

Table 8: Indicators with sub-indicators	
Indicator with sub-indicators	Included single indicators
xx formally registered small enterprises founded by young people supported by the project, are still on the market after one year and make a profit of at least xy/year.	Number of formally registered small enterprises by young people supported by the project, target: xx until xxxx
	Number of formally registered enterprises founded by young people supported by the project that are still on the market after one year, target: xx until xxxx
	Profit in xx/year by small, formally registered enterprises set up by young people supported by the project, target: xx mill. x until xxxx
An interministerial working group consisting of ministries of x, y and z has been set up and adopts three frameworks relevant to employment.	An interministerial working group consisting of ministry of x, y and z has been set up. (Possible criteria: members of all ministries, meeting at least twice a year)
	This interministerial working group adopts three frameworks relevant to employment. (Who decides what is relevant? Specify certain criteria)
Source: Project documents and own elaboration	

4.2.3 Relevance of indicators

When examining the relevance of the indicators for measuring the objective, several challenges can be detected. There is a greater inconsistency in the assignment of the indicators to the specific results level in comparison with the assignment of the objectives to the predefined levels. In every project, indicators can be found that are assigned to the wrong results level. It is noticeable that it is mostly indicators for the programme level objective that are formulated at too low of a level, often even at the output level. As at least one indicator can be found in each project that is formulated at the wrong level, this indicates a general challenge. The result is that, although programme objectives are formulated, no medium-term outcomes or impacts can actually be measured, as indicators are measuring only short-term effects.

Objective	Indicator	Problem
Programme level: The drought resilience of the (agro-)pastoral population in the x and x regions is strengthened.	Number of households diversifying their food and feed supply with the help of small-scale agro-pastoral irrigation systems.	Objective and indicator not on the same results level: Indicator rather on short-term outcome level.
	Matching indicators for objective (depending on the local context): <ol style="list-style-type: none"> 1. Percentage of households that did not have to sell livestock/assets in the last drought. 2. Number of households that did not have to borrow in the last drought. 3. Number of households that did not suffer hunger in the last drought. 4. Number of households that have made use of an existing emergency fund. 	However, these indicators are not very realistic for monitoring within a three-year ⁵ technical cooperation project, as they can only be measured if a drought occurs. Therefore, the goal formulation should be rethought, as the goal is not measurable in a three-year project. Broader thematic evaluations could provide information in this regard, but not regular monitoring.
Matching objective for indicator: The (agro-)pastoral population has security mechanisms available in case of drought.		

Source: Project documents and own elaboration

A further aspect on relevance is that, with respect to their content, 36 out of 196 indicators are not directly related to the objective, and 34 indicators represent the content of the objective only to a limited extent. This ranges from only slight deviations in content to completely different contents. It applies to all projects analysed.

Objective	Indicator	Challenge
The population in the programme area improves its food supply and increases its income from agricultural production.	At least 53% of national rice consumption will be covered by national production in 2015.	Coverage of national rice consumption does not allow for any conclusions about general food supply or income, even if "production" means "rice production".
Villages in the target regions use existing land use plans, including clear usage rights for communal land, for the management and development of their land resources and as their contribution to district action planning.	The equal participation of women and the protection of their interests are guaranteed within the framework of cultural (and economic) possibilities.	The objective is at the village level, the indicator is at the individual level, the content of the participation of women is independent of the use of land-use plans by the villages.
The regional coverage of school meals is in line with the work plan of the programme.	Number of schools supported by WFP.	Number of schools has no significance for regional coverage. More correct would be: Number of schools supported by WFP per region / district / community.

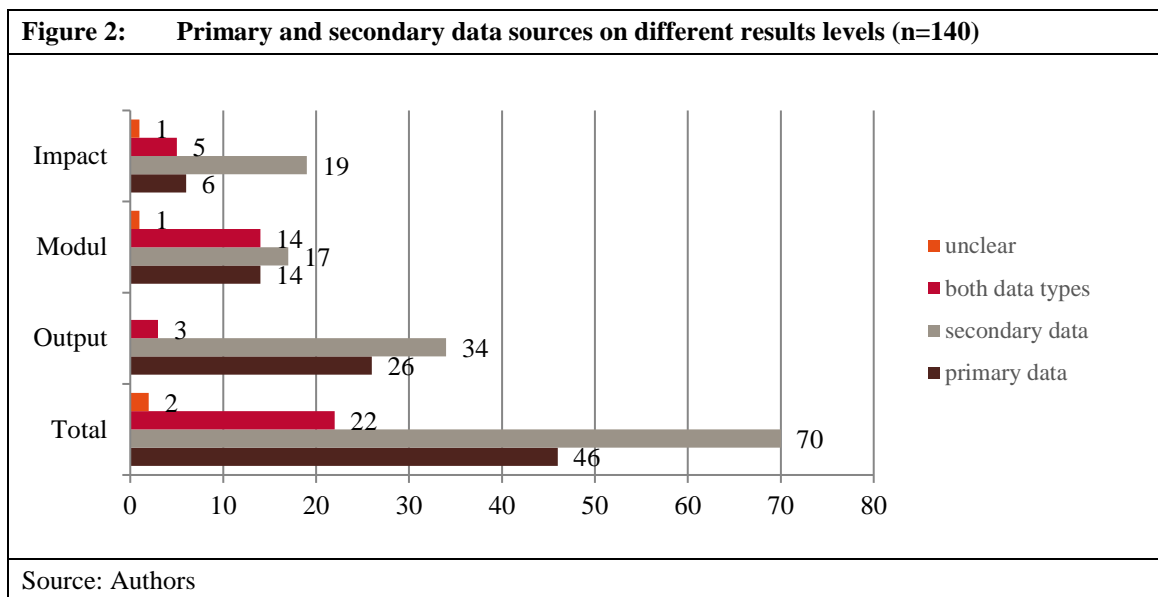
Source: Project documents and own elaboration

⁵ Programmes can have a different time horizon than single projects (modules). However, it is not possible to adapt programme objectives in case a module has a shorter timeframe than the programme.

Furthermore, the analysis shows that for 14 objectives, the set of indicators defined does not cover one or more aspect(s) of the objective. Only in one project do indicators cover all contents of the objective sufficiently. In the remaining 12 projects, at least one gap is detected.

4.3 Step 3: Finding data source and determining data collection methods

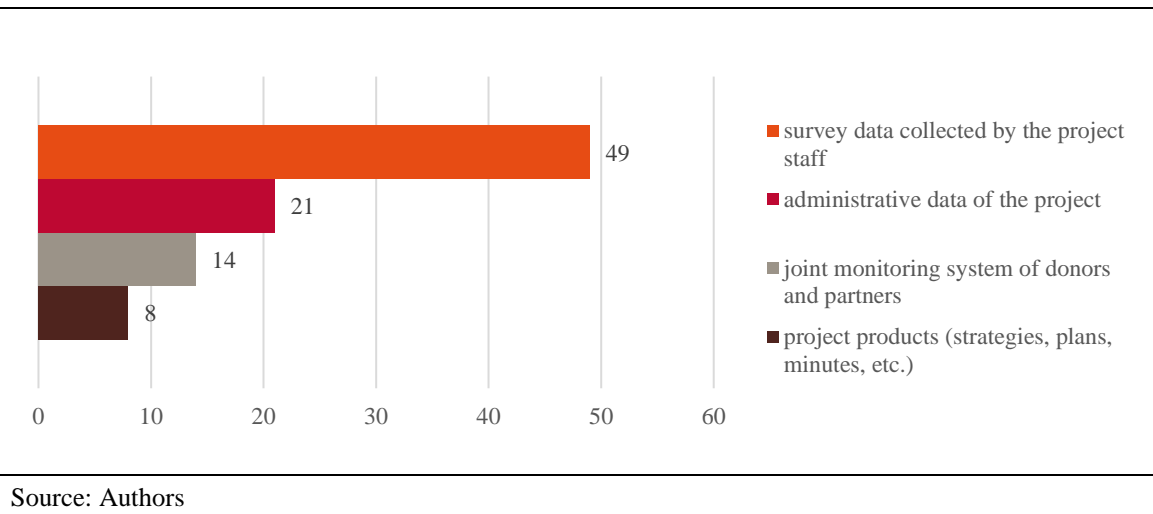
The project documents comprise data sources and data collection methods for the indicators. There are 140 out of 150 indicators that show their sources of verification in the results matrix. In general terms, information is more complete and conclusive on data sources than



information on data collection methods, which are often not clearly identified. For example, the term “own survey” is ambiguous in German. It can mean semi-structured interviews as well as a standardised survey. But also terms such as “own collection” are used, which do not allow conclusions to be made about the kind of data collection method applied.

More indicators make use of secondary data than primary data (70 vs. 40). Twenty-two indicators reveal data sources that include both primary and secondary data. Overall, no project relies only on primary or secondary data, when sources of verification of all indicators are analysed (Figure 2). Furthermore, there is a tendency that, on the impact level, secondary data is used more often. Also on the output level, this percentage is relatively high. This is due to the fact that, because of the project-specific content, a higher rate of primary data collection can be expected.

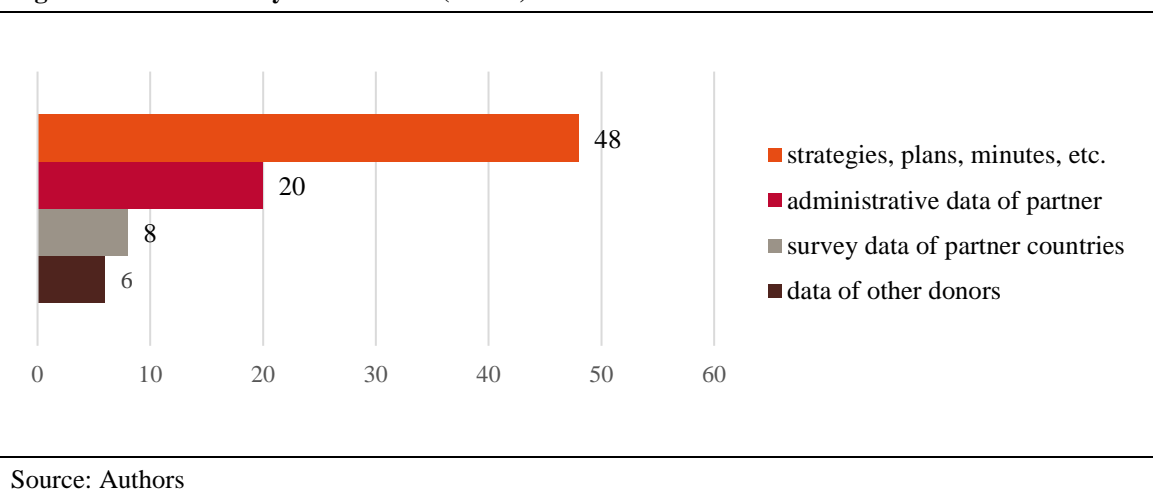
With respect to primary data, the data collection methods refer mainly to surveys or monitoring data. The data source indicated most frequently is an independent survey, whereby it is not always clear whether a quantitative survey or a qualitative interview series is implied. Overall, clear information on data collection methods is most likely to be found in evaluations or baseline studies, but not in the reporting. In addition, for 14 indicators, projects refer to a joint monitoring system with other donors and/or the partner government.

Figure 3: Primary data sources (n=140)

The results matrices of our sample instead do not refer to qualitative data collection methods. However, interviewees refer to qualitative data collection, for example focus group discussions, but do not highlight them in the reporting.

A more detailed analysis of data collection methods is often not possible, as the information provided is often limited, and the interviewees rarely provide any further information. As far as the information available is concerned, comparison groups for allocating the results to the project were surveyed in four projects for eight indicators on the module and programme levels. However, there are methodological limitations, as in one case the before-and-after comparisons were not carried out for the comparison group. Overall, statistical procedures for establishing causality between results and measures are rarely applied.

Data sources for secondary data are mainly the partner countries and less so other donors. The most common sources are strategies, plans and protocols of institutions and organisations in the partner countries. Administrative data of the partner is also an important source for reporting. Survey data of partners is used less often.

Figure 4: Secondary data sources (n=140)

Various problems with the use of data from the partner countries are outlined in the reporting or in the interviews.⁶ For example, interviewees explained that the data that was planned to be used was presumably not free of political influences. Thus, the indicator was changed in the reporting, and a different data source was used. Another problem with secondary data sources in partner countries is whether the data can be provided on time. Because of these existing challenges, implementing agencies refrain at times from using partner data, as it creates a dependency, and data quality is a challenge. Interviewees mention that partner countries still face a heavy workload, as they provide different data to different donors.

Another problem with using secondary data is that it often does not reflect developments within the target regions and/or among beneficiaries. For example, several projects use indicators that measure results at the national level although the project is only active within one specific region. Other projects use data that is specific for the target region, but does not differentiate between project beneficiaries and non-beneficiaries. This not only creates problems with regard to the attribution of results, but it also has consequences with respect to the set of targets. If, for example, data is used that measures household income among project beneficiaries and non-beneficiaries, the target would need to be lowered because non-beneficiaries cannot be expected to have increased their incomes in the same way as beneficiaries (assuming that all else is equal).

Indicator	Data source	Problem
The share of rural food-insecure households decreases by 5% annually	National data (partner country)	The project focusses on one target region and only has a very limited influence at the national level.
The share of poor households in target regions decreases by 20% by 12/2015 (baseline: 2011)	Partner country's Expenditure and Consumption Survey	Data shows changes in poverty among the population within the target region but does not differentiate between programme beneficiaries and non-beneficiaries.

Source: Project documents and own elaboration

4.4 Step 4: Collecting baselines

The analysis continues now with a reduced sample of 150 indicators, because for 46 indicators no further reporting on baseline values, targets values and reported values is provided.

A first step in identifying a realistic target value is to collect baseline data and have a reference point when assessing the success of a project. There are 136 of 150 indicators, where further reporting exists, that explicitly indicate a baseline value. Indicators with several sub-indicators often provide only partial information because some aspects are neglected. Mostly baseline values are provided in the proposal, the amendment or the first

⁶ These problems refer mainly to outcome and impact data, as output data is rather project-specific and needs to be collected by the project staff.

reporting (117 of 150) guidelines. Baseline values before the first reporting are only provided in 79 cases. Projects do not report baseline values for 14 indicators.

Baseline values with zero are reported in 81 of 150 baseline values. Setting baseline values to zero is often criticised, but further analysis reveals that most of the zero values are connected to the contents, where this is plausible. For example, 56 are (quantified) qualitative indicators that deal with existence and non-existence, such as laws passed, strategies developed or documents produced. Thirteen other indicators refer to content that is new, for example farmers trained by the programme. In these two cases, starting at zero is plausible. Nevertheless, 12 baseline values are set zero, although the measurement does not necessarily have to start at zero. Particularly if reported values are expressed in percentages, zero is often chosen, even though for the calculation, the actual figure is needed as a reference point and should also be reported as a baseline value.

4.5 Step 5: Setting targets

Baseline values are then used to estimate target values. There are 141 of 150 indicators in this sample that indicate target values in their project documents; 112 indicators set a target value in the programme proposal or amended programme proposal. For 29 out of 150 indicators, the target values are set only after that, and nine indicators in six projects are not equipped with a target value at all. The implementing agencies specify either figures or percentages, which can be converted into a figure using the baseline value (baseline value: 200 households, target: 15 per cent increase → 230 households). In the case of qualitative indicators, descriptions of the target values are very diverse. They range from “a law is adopted” to “someone is in a position to...”. Difficult specifications are, for example, “empowered”, “established” or “is applied”, which need further explanation to facilitate the data collection and interpretation.

With regard to the assessment of whether realistic target values are chosen, the analysis clearly shows that target values are adjusted over time. Interviewees expressed that targets are regularly adjusted in order to be realistic. Some turn out to be overambitious, others were unexpectedly overachieved. In most cases, the target values are perceived as being realistic yet ambitious. Overall, external factors such as the Ebola crisis can make realistic target values unattainable, so there is always some uncertainty as to which value will be realistic.

As described before, in addition to a baseline and target value, indicators should be equipped with a time horizon. Only 51 out of 150 indicators provide such a time horizon. The high number of non-time-bound indicators is unexpected. Although it can be argued at the module and output levels that the achievement of objectives will be measured at the end of the project, this natural time horizon is not so clear at the programme level. In addition, the duration of modules and programmes is not always congruent.

4.6 Step 6: Collecting monitoring data and using evaluation information

Now the analysis focusses on the collected information and the reported values. There are 79 of 150 indicators that report a value for the indicator in the first reporting. In particular,

reporting takes place with regard to output indicators. In the second reporting, another 28 indicators (75 of 95 indicators), and in the last reporting, four indicators (57 of 66 indicators) receive a value. This shows that the response rate increases in the second and final reportings. There is a particularly high rate of reporting for outputs in the first reporting (already 75 per cent), and less than half for the outcome and impact indicators. This is reasonable, as the effects on higher results levels need time to develop, and surveys often cannot be carried out annually due to high costs.

Different aspects determine the reliability of the data collected, and thus the quality of the data. First, this analysis checked whether the data reflects what the indicator intends to measure; 95 of the 131 available values measure what the indicator claims to measure. For 10 indicators, some aspects were neglected, for example when they consisted of several sub-indicators. For 22 indicators, the content of the data differs from that foreseen by the indicator (Table 12). For six indicators, this aspect is not clear from the information provided.

Indicator	Reported value	Problem
The involvement of the rural population, especially women, in local planning and decision-making processes rises from xy% to xy% at the village, district and provincial levels in the selected provinces. The proportion of poor households and women participating in this process is at least equal to their share of the population.	Individual values for parts of the indicator are reported. Figures for the indicator as a whole are not available.	Possible cause: complex formulation of the indicator, with various sub-indicators.
xx young people supported by state and private service providers, x% of whom are women, generate additional income of x million x/year despite climate-related changes through self-employment or wage labour (inserted indicator for pure number).	Measures how many young people are reached and not how many young people have been able to increase their income.	Possible cause: indicator formulated with various sub-aspects so that parts are overlooked in the survey.
Source: Project documents and own elaboration		

Many interviewees were critical about whether the data collected is trustworthy and reflective of reality. On the one hand, they recognise that some questions are difficult for the surveyed persons to answer. In this case, only very rough estimates are made, such as in the case of recall data. On the other hand, reference is made to the possible tendency of being polite or supporting the continuation of the project, and therefore providing positive feedback to the interviewer.

A third aspect of data quality concerns whether the data is consistent and comparable over time. For the analysis in this paper, project documents were used to determine whether the same measuring method, the same measuring instrument (e.g. questionnaire) and the same sampling design were used for baseline, midline and end-line values. This is to ensure comparability that leads to a reliable interpretation of the performance of a measure. In total, 93 of 117 indicators that report at least two values (incl. baseline value) are rated as highly consistent, and only 3 as medium and 13 as low, in terms of consistency. An assessment cannot be made for eight indicators because sufficient information is lacking. The fact that the vast majority of data is consistent is a positive finding.

Lastly, it is important that attribution be discussed in a critical manner when values are reported, and with this the level of plausibility that the intervention contributes to the results.

The results show that 76 of 103 reported values in the first or second reporting of module indicators explain their values, but only 48 of them discuss matters of attribution. These discussions usually only take place if results have not been achieved. If the results are positive, the causal link to the project is assumed to be given. However, this link is usually extremely unclear in case of outcome- and impact-level data because impact-evaluation techniques are not applied. Usually data is reported without showing any comparative value (e.g. results for project beneficiaries compared to non-beneficiaries).

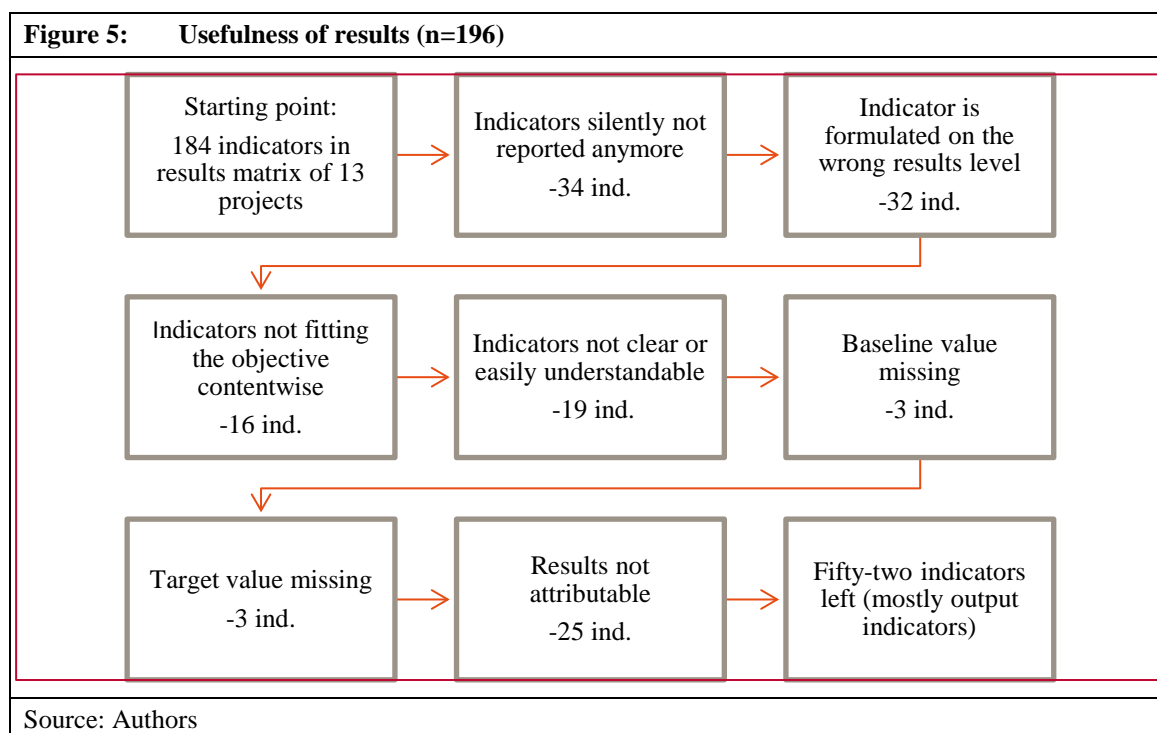
4.7 Step 7: Reporting and using findings

This discussion paper collected – only to a limited extent – data with regard to the process of reporting and using findings. This section therefore focusses more on an aggregate assessment of how useful reporting in German bilateral cooperation is and how findings are used.

The whole analysis shows that much of the information collected by the implementing agencies cannot be used without reservation. There are many methodologically flawed specifications in the RBM of German bilateral cooperation. Therefore, also the quality of the collected data can be questioned. For illustration purposes, all indicators are checked against the most important general quality criteria. The following quality criteria are cumulatively checked:

- The indicator is used throughout the whole reporting process.
- The indicator is formulated at the right results level.
- The content of the indicator matches the objective of the project.
- The indicator is specific.
- Baseline values are available.
- Target value is available.
- Results are attributable or attribution is discussed.

Very few of the 150 indicators meet the methodological requirements when examined for different quality criteria at the same time (Figure 5). Only slightly more than half of the indicators pass the hurdle of the first four stages on the quality criteria of indicators referring mainly to the content and formulation of the indicators. A few indicators do not provide any information about the baseline value, target value or data source. In contrast, the lack of discussion about attribution for indicators higher than the output level or applying impact-evaluation methods extremely diminishes the amount of usable data. This analysis sums up the quality challenges that exist in the RBM of German bilateral cooperation.



Implementing agencies use the data reported to the ministry but also collect additional data to have enough information for internal quality control and project adaption. However, there is no systematic use by BMZ of the information provided in the yearly reports of implementing agencies.

5 Conclusions and recommendations

The analysis of results measurement and RBM carried out in bilateral projects of German development cooperation shows that great efforts are being made to measure results. Projects often collect detailed primary data in order to document the achievement of the objectives in the reporting. At the same time, the analysis shows that the challenges of RBM are continuously underestimated. Capacities and expert knowledge are lacking within BMZ, the implementing agencies as well as in partner countries. There are large data gaps in the individual projects, and there is a strong potential for improvement in the quality of the indicators formulated and the data collected. Overall, the paper finds that international RBM standards are often not met, and that it is usually not possible to show a causal relationship between results measured at the outcome and impact levels and project activities. This is problematic because accountability of the implementing agencies towards BMZ focusses on these levels, whereas the output level is given less importance. Germany can learn from the experiences of other donors and reform its RBM system by taking into consideration the existing challenges (Vähämäki & Verger, 2019).

Subsequently, the main conclusions and recommendations derived from the analysis are presented.

Develop a comprehensive results-based management system

First and foremost, the analysis suggests that a **comprehensive RBM system** would need to be developed by BMZ. Several guidelines on RBM procedures are available. Yet, a comprehensive RBM framework and details for its implementation are missing. For collecting useful M&E information, it is important to set up a comprehensive RBM system that covers all levels: the agency level (BMZ), the country level, the programme level and the project (module) level. The system should aim at achieving high-quality data and, at the same time, put emphasis on ownership and – along with this – high degrees of flexibility and local solutions. It should be as simple as possible and not lack methodological rigour. BMZ should develop one comprehensive RBM guideline covering the purpose of, and details on, the management of the system as well as details on different tools. Other donors' experiences could be taken into consideration when setting up the RBM system. Especially the institutional approach of IFAD might be considered, which simplified its RBM system after detecting similar quality challenges to those found in this discussion paper (IFAD, 2017).

Our recommendations focus on the necessary improvements of RBM standards, as set by BMZ for German development cooperation. RBM within implementing agencies is not at the centre of the paper, but it remains equally important. Despite handing over the main responsibilities in planning and implementation to the implementing agencies, the ministry has the responsibility of setting and enforcing RBM standards as well as monitoring and reviewing the use of funding.⁷

The following features should be part of the RBM system. First, the RBM approach should be a flexible bottom-up instead of a top-down approach in order to reconcile the results principle of the aid effectiveness agenda with the ownership principle. This requires that content adjustments by the BMZ in the planning process should always be introduced via discussions on the project's contents, and not solely by the adaptation of indicators. Second, learning, which is too often neglected in RBM systems, needs to be one of the main purposes – also from the BMZ perspective, and not only within implementing agencies – in order to achieve a higher level of project effectiveness and to contribute towards a higher level of data quality within the system. For this to happen, a stronger learning culture needs to be cultivated within BMZ in which it is acceptable that some measures might be ineffective as long as new ways are found to approach the problem. Third, clear procedures need to be developed on how the reported information is systematically used by BMZ. For example, syntheses of the report contents can play an important role in providing a more evidence-based way for BMZ's steering of development cooperation at the country and sector levels. The use of data can also be increased by sharing data in open formats and under open licences. Following the example of other donors, such as the United States Agency for International Development and DFID, Germany should increase transparency in development cooperation. Project documents (e.g. proposals, progress and final reports, evaluations) as well as primary data collected by implementing agencies and research partners should be made publicly available. Fourth, as part of an RBM system at the agency level, a list of standards indicators, which should be harmonised among implementing

⁷ See, for example, the technical cooperation and financial cooperation guidelines (BMZ, 2008) and the federal budget regulations on economic efficiency of development cooperation (Bundesministerium der Justiz und für Verbraucherschutz & Bundesamt für Justiz, 1969).

agencies, could be drawn up to allow for the aggregation of results across countries, implementing agencies and projects.⁸ BMZ is currently discussing this option within the BMZ 2030 reform process. Aggregate information on project results is valuable for reporting to the German Parliament and taxpayers. Still, caution is important whenever it comes to standardisation, as it might reduce ownership and local adaptation. Standards indicators should not be mandatory. Following a bottom-up approach of project monitoring, they should only be used by projects if they fit the project context very well and represent a relevant measure for the progress of the project from the perspective of the implementing agencies and local partners. Standards indicators should be part of the results matrices to ensure data quality.

A guideline for RBM as such will not be able to improve the quality of RBM. In addition, capacity-building for staff of ministries, implementing agencies and partner organisations as well as steady quality control of reported information are necessary to ensure data quality and the usefulness of the data.

Another recommendation is to concentrate the knowledge, expertise and standard-setting competence for RBM within one unit in BMZ, as it is spread right now over various divisions. The number of staff designated to RBM within the unit needs to increase significantly. This is also highlighted by the ongoing BMZ 2030 reform process, which aims *inter alia* to increase the results orientation of German development cooperation and to link resource allocation to results (BMZ, 2020). Its responsibilities should initially include developing the RBM system with respective guidelines and setting up a data management system. In further course, it should be responsible for monitoring the implementation of RBM standards at the agency, country, programme and project levels⁹ and act as an advisory unit for other BMZ units. In addition, the unit should compile regular results reports and make use of the information from these reports for learning. It should work closely with the evaluation and development research division as well as with the implementing agencies to develop a coherent approach.

Apart from the general establishment of a comprehensive RBM system, the following five aspects could be considered during the reform process and for further reforms.

-
- 8 Currently, BMZ uses a set of “aggregation indicators” to aggregate results across projects and countries. The indicators as well as the approach towards measuring aggregate results, however, are not harmonised among the two implementing agencies KfW and GIZ. A major difference is that GIZ reports on *ex post* results, whereas KfW reports on *ex ante* results.
- 9 According to the technical cooperation and financial cooperation guidelines BMZ (2008), implementing agencies are responsible for planning, implementing and monitoring development measures. They are also responsible for reporting to BMZ on progress made and for carrying out evaluations. Our recommendation to strengthen BMZ’s role in RBM does not imply that responsibilities should be taken from the implementing agencies. Rather, we suggest that BMZ should exercise its duty to monitor and review the use of funding more fully. This can be done by providing more comprehensive RBM standards, monitoring the implementation of these standards and using results reports as well as evaluation findings more systematically for learning and steering development cooperation measures.

Involvement of the partners and ownership

Information on the process of developing indicators shows that they are often developed on the basis of the partners' documents or adopted from previous projects. This points on the one hand to alignment with partners, and secondly to the usage of already established indicators. However, interviews have revealed that the selection of indicators is often done with only limited involvement of the partners. Partners often only become involved at a later stage of the process once the results measurement system has been finalised. This is not in line with the principle of ownership, where mainly the needs of the partners should be at the centre of attention.

Establishing a flexible and more complex theory of change

German reporting centres around the **results matrix**. With this tool, implementing agencies present all obligatory information on RBM to BMZ. A main problem of the tool is its limitation to three levels only and equating the structural levels of German bilateral cooperation – output, module and programme – with output, outcome and impact. Simplification is important, but at the same time, only having three levels will, in most cases, not come nearly close enough to representing the project.

In addition to the results matrix, BMZ should require a comprehensive theory of change from projects. GIZ has already introduced the theory of change as part of its monitoring system. Building a comprehensive theory of change helps project staff to think about how change will come about and which actors and external factors may influence results. It also encourages project staff to think critically about assumptions underlying their results model and to view development as a complex process. A theory of change is also the basis for later evaluating a project.

The results matrix can continued to be used as a simplification tool for reporting purposes. However, the results matrix as well as its new graphical version should become more flexible. The results matrix should be adaptable so that more than three results levels can be shown, and more complex projects can be reflected. Projects staff should develop the results matrix only after the theory of change has been completed. Key activities and results identified within the theory of change should be transferred to the results matrix.

Stronger guidance and quality check for indicators

Indicators are the heart of measuring the results and making each objective measurable. Detailed information by BMZ for the implementing agencies on how to develop indicators would need to be provided. Despite some internal quality controls within the implementing agencies, this discussion paper finds that there are methodological problems for many indicators.

BMZ should develop detailed guidelines on how to define an indicator correctly according to methodology. All indicators need to be checked for quality. A system should be designed in which – while implementing agencies continue to be mainly responsible for quality checks against clear criteria – BMZ also randomly checks a share of projects. Capacity-building within BMZ and consistency of approaches among implementing agencies is important to change the general way of formulating indicators. Checklists and reference indicators offer further guidance and should not only – as is currently done – be developed

for certain sectors in a decentralised manner, but also provide coherent guidance across sectors on formulating indicators. Although there is a tendency to use classical quantitative indicators in German bilateral cooperation, the coverage of qualitative aspects should be further promoted. Qualitative indicators should be more precise and need to entail exact assessment criteria.

Increase efforts to strengthen partner countries' RBM systems

With regard to data collection, interviews show that using the secondary data of the partner countries as data sources presents a major challenge. Data availability and data quality are often low. In addition, there is only limited cooperation and coordination with partner countries and other development actors when primary data is collected. Furthermore, parallel RBM systems by donors and a lack of harmonisation among development actors impede improvements in data quality and availability in many developing countries. Programme-based approaches show that using partner countries' systems can work. Therefore, efforts to use and strengthen the partner countries' RBM systems need to be reinforced. One option would be to mainstream support of national RBM systems and to include a project component that supports partners' RBM systems into each technical and financial cooperation project. As part of such a component, funds could be provided to strengthen or expand the capacities of the local project partners. The long-term goal should be to use partner countries' RBM systems and to reduce the need for primary data collection in projects.

Increase the focus on producing attributable results

Another large challenge found in this analysis is the lack of attribution of results at the outcome and impact levels to the development cooperation measure. This is an important concern of the BMZ and the international aid community, but action has been relatively limited with no systematic approach on impact evaluations in place in German development cooperation to date (Bruder, Faust, & Krämer, 2019).

The current practice of using monitoring data to report about the outcomes and impacts of development projects should be discontinued because a causal link between projects' activities and results measured cannot be established. Instead, this discussion paper proposes the following.

First, in cases where evaluations are conducted, more impact evaluations with high methodological standards should be carried out as part of project- and programme-level RBM to establish causality. For transparency and awareness-raising, reported data on higher impact levels (medium-term outcomes and impacts) should always provide information about the impact-evaluation design used to receive the information.

Second, projects not selected for an evaluation should no longer have to report attributable quantitative changes at the medium-term outcome and impact levels. One option is to focus at the output and short-term outcome levels when reporting project results. Short-term outcomes show the direct effects for the target group/ public goods that result from the use of outputs. Although they are also influenced by external factors, the link to the projects' activities is much clearer compared to changes at the medium-term outcome and impact levels (household income, yields, etc.). For example, a project providing training on sustainable land measurement practices could measure the number of practices adopted by

beneficiaries (in comparison to non-beneficiaries). Medium-term outcome and impact indicators – ideally derived from existing sources in partner countries – are still important as context indicators. Context indicators describe the setting a project operates in and give some indication about the success of the joint development efforts of all actors involved. They should not be used to assess a single project's or programme's effectiveness, however, because they are influenced by a variety of external factors.

Projects not selected for evaluations should rely more systematically on qualitative methods to explore (potential) medium-term outcomes and impacts. For example, Theory of Change verification workshops with all involved stakeholders could provide indications of which pathways in the theory of change are working as anticipated and which pathways face challenges. Adaptive programming needs to be applied to react to possible challenges with activities that support the theory of change. Alternatively, the approach used by IFAD (2017) could be followed. IFAD has introduced standardised perception surveys to measure outcomes. For example, instead of measuring quantitative changes in outcome indicators, IFAD asks project beneficiaries whether they have experienced a reprieve in their water shortage, increased income or improved quality in their diets as a result of the project. This significantly lowers the monitoring and reporting burden of projects. At the same time, it is likely to deliver a more accurate picture of the results of a project than presenting quantitative data on the outcome and impact levels when there is only very limited information available about the degree to which a project has contributed to the results measured. Other qualitative methods, such as focus group discussions and semi-structured interviews, are already being applied by the implementing agencies. However, findings from qualitative analyses are not yet being systematically reported in progress reports to BMZ.

References

- Abdel-Malek, T. (2015). *The Global Partnership for Effective Development Cooperation: Origins, actions and future prospects* (Study 88/2015). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).
- Andrews, M., Pritchett, L., Samji, S., & Woolcock, M. (2015). Building capability by delivering results: Putting Problem-Driven Iterative Adaptation (PDIA) principles into practice. In A. Whaites, E. Gonzalez, S. Fyson, & G. Teskey (Eds.), *A governance practitioner's notebook: Alternative ideas and approaches* (pp. 123-135). Paris: Organisation for Economic Co-operation and Development.
- Asian Development Bank. (2006). *An introduction to results management*. Mandaluyong: Author.
- Balogun, P., & Lloyd, R. (2017). *Evaluation of the Norwegian aid administration's practice of results-based management*. Retrieved from <https://norad.no/contentassets/8d8b2cbc48dd4a29872580a4845ed3d1/4.18-evaluation-of-the-norwegian-aid-administrations-practice-of-rbm.pdf>
- Banerjee, A. V. (2011). *Poor economics*. New York, NY: Public Affairs.
- Binnendijk, A. (2000). *Results-based management in the development co-operation agencies: A review of experience* (Background Report). Report presented at the meeting of the WP-EV, and the document was subsequently revised. Retrieved from <https://www.oecd.org/dac/evaluation/dcdndep/31950852.pdf>
- BMZ (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung). (2008). *Leitlinien für die bilaterale finanzielle und technische Zusammenarbeit mit Kooperationspartnern der deutschen Entwicklungszusammenarbeit*. Bonn: Author.
- BMZ. (2009). *Operationsplan zur Umsetzung der Pariser Erklärung 2005 und des Accra Aktionsplans 2008 zur Steigerung der Wirksamkeit von Entwicklungszusammenarbeit*. Bonn: Author.
- BMZ. (2010-2019a). *Ergebnisse: Evaluierungsberichte, Berichte zu Reviews und Wirkungsmessungen*. Retrieved from <https://www.bmz.de/de/ministerium/evaluierung/ergebnisse/index.html>
- BMZ. (2010-2019b). *Ziele: Die Aid-Effectiveness-Agenda – Wirksamkeit der Zusammenarbeit steigern*. Retrieved from <https://www.bmz.de/de/ministerium/ziele/parisagenda/index.html>
- BMZ. (2011). *Der Operationsplan des BMZ zur Umsetzung der Pariser Erklärung 2005 und des Accra Aktionsplans 2008 zur Steigerung der Wirksamkeit von Entwicklungszusammenarbeit: Umsetzungsbericht*. Bonn: Author.
- BMZ. (2012). *Handreichung und kommentierte Gliederung für Programmvorschläge für gemeinsame EZ-Programme*. Mimeo: Author.
- BMZ. (2020). *Gemeinsam weiter / Zukunft denken - BMZ 2030*. Retrieved from http://www.bmz.de/de/mediathek/publikationen/reihen/infobroschueren_flyer/infobroschueren/sMateriale485_bmz_2030.pdf
- Bruder, M., Faust, J., & Krämer, M. (2019). *Rigorous impact evaluation in German development cooperation*. (Policy Brief 5/2019). Bonn: German Institute for Development Cooperation (DEval).
- Bundesministerium der Justiz und für Verbraucherschutz & Bundesamt für Justiz. (1969). *Bundshaushaltsordnung*. Retrieved from <http://www.gesetze-im-internet.de/bho/BHO.pdf>
- Burnside, C., & Dollar, D. (2000). Aid, policies, and growth. *The American Economic Review*, 90(4), 847-868. Retrieved from <http://www.jstor.org/stable/117311>
- CGD (Center for Global Development). (2006). *When will we ever learn? Improving lives through impact evaluation*. Washington, DC: Author.
- Chambers, R. (2017). *Can we know better?* Rugby, UK: Practical Action Publishing.
- Collier, D. (2011). Understanding process tracing. *Political Science and Politics*, 44(4), 823-830. Retrieved from <http://dx.doi.org/10.1017/S1049096511001429>

- DFID (Department for International Development). (2011). *Guidance on using the revised Logical Framework*. London, UK: Author.
- DFID. (s.a.). *DFID's results framework: Managing and reporting DFID results*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/175945/DFID-external-results.pdf
- Doing Development Differently. (2014, October). Doing development differently: The DDD manifesto community [Blog post]. Retrieved from <https://buildingstatecapability.com/the-ddd-manifesto/>
- Easterly, W. (2006). *The white man's burden*. New York, NY: Penguin Press.
- Eyben, R. (2013). *Uncovering the politics of "evidence" and "results": A framing paper for development practitioners*. Brighton, UK: Institute of Development Studies.
- Global Affairs Canada. (2016). *Results-based management for international assistance programming at Global Affairs Canada: A how-to guide*. Ottawa: Author.
- Holzapfel, S. (2016). Boosting or hindering aid effectiveness? An assessment of systems for measuring donor agency results. *Public Administration and Development*, 36(1), 3-19. <http://dx.doi.org/10.1002/pad.1749>
- Hood, C. (1995). The "New Public Management" in the 1980s: Variations on a theme. *Accounting, Organizations and Society*, 20(2), 93-109. [http://dx.doi.org/10.1016/0361-3682\(93\)E0001-W](http://dx.doi.org/10.1016/0361-3682(93)E0001-W)
- IEG (Independent Evaluation Group). (2012). *Designing a results framework for achieving results: A how-to guide*. Washington, DC: Author and World Bank.
- IEG. (2017). *Results and performance of the World Bank Group 2016: An independent evaluation*. Washington, DC: Author and World Bank.
- IFAD (International Fund for Agricultural Development). (2002). *A guide for project M&E: Managing for impact in rural development*. Rome: Author and Office of Evaluation and Studies.
- IFAD. (2017). *Taking IFAD's results and impact management system (RIMS) to the next level*. Rome: Author.
- Inter-Agency and Expert Group on SDG Indicators. (2019). *Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development*. New York, NY: United Nations Statistical Commission.
- Keijzer, N., Klingebiel, S., Örnemark, C., & Scholtes, F. (2018). *Seeking balanced ownership in changing development cooperation relationships*. Stockholm: Expert Group for Aid Studies.
- Koppenleitner, R., Koob, E., Reade-Soh, N., Dirinpur, J., Lang, E., Schnellbach, K.-P., & Weinmann, S. (2012). *Wirkungsanalyse und Wirkungsmessung in Gesundheitsvorhaben der deutschen Entwicklungszusammenarbeit*. Bonn: Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH.
- Kusek, J. Z., & Rist, R. C. (2004). *Ten steps to a results-based monitoring and evaluation system*. Washington, DC: World Bank.
- Mayne, J. (2007). Challenges and lessons in implementing results-based management. *Evaluation*, 13(1), 87-109. <http://dx.doi.org/10.1177%2F1356389007073683>
- Mayne, J. (2008). *Contribution analysis: An approach to exploring cause and effect*. Montpellier: Consultative Group on International Agricultural Research.
- OECD (Organisation for Economic Co-operation and Development). (2008). *The Accra Agenda for Action*. Paris: Author.
- OECD. (2010). *Glossary of key terms in evaluation and results-based management*. Paris: Author.
- OECD. (2011). *Busan High-level Forum on Aid Effectiveness: Proceedings*. Paris: Author.
- OECD. (2015). *DAC-Prüfbericht über die Entwicklungszusammenarbeit: Deutschland 2015*. Paris: Author.
- OECD. (2017). *Strengthening the results chain: Synthesis of case studies of results-based management by providers*. Paris: Author.

- OECD. (2018). *Paris Declaration and Accra Agenda for Action*. Retrieved from <http://www.oecd.org/dac/effectiveness/parisdeclarationandaccraagendaforaction.htm>
- OECD. (2019a). *Guiding principles on managing for sustainable development results*. Paris: Author.
- OECD. (2019b). *Results-based approaches: How can a greater focus on results improve accountability, communication, direction and learning?* Retrieved from <https://www.oecd.org/dac/results-development/results-based-approaches/>
- Örtengren, K. (2016). *The LFA method: A guide to results-based management (RBM), efficient project planning with the aid of the Logical Framework Approach (LFA)*. Stockholm: Swedish International Development Cooperation Agency.
- Perrin, B. (2002). *Implementing the vision: Addressing challenges to results-focus management and budgeting*. Paris: Organisation for Economic Co-operation and Development.
- Riely, F., Mock, N., Cogill, B., Bailey, L., & Kenefick, E. (1999). *Food security indicators and framework for use in the monitoring and evaluation of food aid programs*. Washington, DC: Nutrition Technical Assistance Project (FANTA).
- Roughly, A., & Dart, J. (2009). *Developing a performance story report: User guide*. Canberra: Commonwealth of Australia.
- Sachs, J. (2005). *The end of poverty*. London: Penguin Books.
- Schwegmann, C., & Holzapfel, S. (2018). *Data for development: An agenda for German development cooperation* (Briefing Paper 12/2018). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).
- Shutt, C. (2016). *Towards an alternative development management paradigm*. Stockholm: Experten Group for Aid Studies, Sweden.
- Skoufias, E., & McClafferty, B. (2001). *Is PROGRESA working?* Washington, DC: International Food Policy Research Institute.
- Skoufias, E., Parker, S. W., Behrman, J. R., & Pessino, C. (2001). Conditional cash transfers and their impact on child work and schooling: Evidence from the PROGRESA program in Mexico. *Economía*, 2(1), 45-96. Retrieved from <http://www.jstor.org/stable/20065413>
- Sundberg, M. (2019). Donors dealing with “aid effectiveness” inconsistencies: National staff in foreign aid agencies in Tanzania. *Journal of Eastern African Studies*, 13(3), 445-464. Retrieved from <https://www.tandfonline.com/doi/pdf/10.1080/17531055.2019.1628384?needAccess=true>
- UNDG (United Nations Development Group). (2011). *Results-based management handbook: Harmonizing RBM concepts and approaches for improved development results at country level*. New York, NY: Author.
- Vähämäki, J., Schmidt, M., & Molander, J. (2011). *Review: Results-based management in development cooperation*. Stockholm: Riksbankens Jubileumsfond.
- Vähämäki, J., & Verger, C. (2019). *Learning from results-based management evaluations and reviews*. Paris: Organisation for Economic Co-operation and Development.
- Westhorp, G. (2014). *Realist evaluation*. London: Overseas Development Institute.
- Whitty, B., & Valters, C. (2017). *The politics of the results agenda in DFID: 1997-2017*. Retrieved from <https://www.odi.org/publications/10902-politics-results-agenda-dfid-1997-2017>
- World Bank. (2013a). *Core sector indicators and definitions*. Retrieved from <http://siteresources.worldbank.org/PROJECTS/Resources/40940-1367867968385/CoreSectorIndicatorsList.pdf>
- World Bank. (2013b). *Results framework and M&E*. Washington, DC: World Bank, Operational Policy and Quality Department.

Publications of the German Development Institute /
Deutsches Institut für Entwicklungspolitik (DIE)

Discussion Papers

- 17/2020 Laudage, Sabine. *Corporate tax revenue and foreign direct investment: potential trade-offs and how to address them* (46 pp.). ISBN 978-3-96021-128-0. DOI:10.23661/dp17.2020
- 16/2020 Cavatorta, Francesco, & Fabio Merone. *Never-ending reformism from above and dissatisfaction from below: The paradox of Moroccan post-Spring politics* (32 pp.). ISBN 978-3-96021-127-3. DOI:10.23661/dp16.2020.
- 15/2020 Eppler, Mirko, Stella Gaetani, Francy Köllner, Jana Kuhnt, Charles Martin-Shields, Nyat Mebrahtu, ...Carlotta Preiß. *Information and communication technology in the lives of forcibly displaced persons in Kenya* (31 pp.). ISBN 978-3-96021-126-6 DOI:10.23661/dp15.2020.
- 14/2020 Hackenesch, Christine, Julia Leininger, & Karina Mross. *What the EU should do for democracy support in Africa: Ten proposals for a new strategic initiative in times of polarisation* (45 pp.). ISBN 978-3-96021-125-9. DOI:10.23661/dp14.2020.
- 13/2020 Never, Babette, Jose Ramon Albert, Hanna Fuhrmann, Sebastian Gsell, Miguel Jaramillo, Sascha Kuhn, & Bernardin Senadza. *Carbon consumption patterns of emerging middle classes* (42 pp.). ISBN 978-3-96021-124-2. DOI:10.23661/dp13.2020.
- 12/2020 Wehrmann, Dorothea. *Transnational cooperation in times of rapid global changes: The Arctic Council as a success case?* (29 pp.). ISBN 978-3-96021-123-5. DOI:10.23661/dp12.2020
- 11/2020 Erforth, Benedikt. *The future of European development banking: What role and place for the European Investment Bank?* (36 pp.). ISBN 978-3-96021-122-8. DOI:10.23661/dp11.2020.
- 10/2020 Kaplan, Lennart. *Systemic challenges and opportunities of Franco-German development cooperation* (64 pp.). ISBN 978-3-96021-121-1. DOI:10.23661/dp10.2020.
- 9/2020 Stoffel, Tim. *Socially responsible public procurement (SRPP) in multi-level regulatory frameworks: Assessment report on policy space for SRPP regulation and implementation in Germany and Kenya* (55 pp.). ISBN 978-3-96021-120-4. DOI:10.23661/dp9.2020.
- 8/2020 Müngersdorff, Maximilian, & Tim Stoffel. *Strategies to strengthen socially responsible public procurement practices in German municipalities: A mapping exercise* (78 pp.). ISBN 978-3-96021-119-8. DOI:10.23661/dp8.2020.
- 7/2020 Stender, Frederik, Axel Berger, Clara Brandi, & Jakob Schwab. *The trade effects of the Economic Partnership Agreements between the European Union and the African, Caribbean and Pacific Group of States: Early empirical insights from panel data* (39 pp.). ISBN 978-3-96021-118-1. DOI:10.23661/dp7.2020.

[Price: EUR 6.00; publications may be ordered from the DIE or through bookshops.]

For a complete list of DIE publications:
www.die-gdi.de