KNOWING WHAT WORKS

Central project evaluation

Strategic Cooperation with International Organisations (SCIO), global project
Project number 2015.6258.6

Evaluation Report

On behalf of GIZ by Franziska Ehlert (Syspons GmbH) and Jana Wulff-Troeger (Syspons GmbH)

Published: January 2022

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Publication details

Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH is a federal enterprise that supports the German Federal Government in achieving its objectives in the fields of international education and international cooperation for sustainable development.

GIZ's Evaluation Unit reports directly to the Management Board. It is separate from GIZ's operational business. This organisational structure strengthens its independence. The unit is mandated to generate evidence-based results and recommendations for decision-making, to provide plausible verification of results and to increase the transparency of findings.

The Evaluation Unit commissioned external independent evaluators to conduct the evaluation. This evaluation report was written by these external evaluators. All opinions and assessments expressed in the report are those of the authors.

Evaluators:
Franziska Ehlert, Jana Wulff-Troeger (Syspons GmbH)

Authors of the evaluation report:
Franziska Ehlert, Jana Wulff-Troeger (Syspons GmbH)

Consulting:
Syspons GmbH
Prinzenstraße 84
10969 Berlin
T +49 30 69815800
info@syspons.com

Concept, coordination and management:
Claudia Kornahrens, GIZ, Head of Section
Ulrike Haffner, GIZ, Evaluation Manager
Central Project Evaluations Section
GIZ Corporate Unit Evaluation

Responsible:
Dr Albert Engels, GIZ, Director
GIZ Corporate Unit Evaluation

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

Editing:
International Correspondents in Education (ICE)

Registered offices:
Bonn and Eschborn

Friedrich-Ebert-Allee 32 + 36
53113 Bonn, Germany
T +49 228 4460-0
F +49 228 4460 - 1766

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

Design/layout:
DITHO Design GmbH, Cologne

Printing and distribution:
GIZ, Bonn

Printed on 100% recycled paper, certified to Forest Stewardship Council (FSC) standards.

Bonn, January 2022

This publication can be downloaded as a pdf from the GIZ website: www.giz.de/evaluierung.

Contents

List of figures and tables 4

Abbreviations 5

The project at a glance 7

1 Evaluation objectives and questions 8

1.1 Evaluation objectives 8

1.2 Evaluation questions 8

2 Object of the evaluation 9

2.1 Definition of the evaluation object 9

2.2 Results model including hypotheses 12

3 Evaluability and evaluation process 16

3.1 Evaluability: data availability and quality 16

3.2 Evaluation process 16

4 Assessment of the project according to OECD-DAC criteria 18

4.1 Long-term results of predecessor 18

4.2 Relevance 18

4.3 Effectiveness 25

4.4 Impact 36

4.5 Efficiency 40

4.6 Sustainability 45

4.7 Key results and overall rating 48

5 Conclusions and recommendations 51

5.1 Factors of success or failure 51

5.2 Conclusions and recommendations 51

List of references 54

Annex: Evaluation matrix 55

***Separate annexes**: Interview Coding List (with interview code only available to GIZ evaluation unit, password protected); and Evaluation matrix as an Excel file (in addition to annex within this document).*

List of figures and tables

Figure 1: Results model 15
Figure 2: Distribution of costs by outputs and overarching costs 42

Table 1: List of stakeholders of the evaluation and selected interviewee 17
Table 2: Rating of OECD/DAC criterion: relevance 24
Table 3: Alternatives hypotheses 26
Table 4: Assessment of module indicators 28
Table 5: Assessment of results hypotheses (reform-oriented measures) 33
Table 6: Assessment of alternative hypotheses (reform-oriented measures) 34
Table 7: Assessment of results hypotheses (topic-oriented measures) 35
Table 8: Rating of OECD/DAC criterion: effectiveness 36
Table 9: Rating of OECD/DAC criterion: impact 40
Table 10: Rating of OECD/DAC criterion: efficiency 45
Table 11: Table 4. Rating of OECD/DAC criterion: sustainability 48
Table 12: Overall rating of OECD/DAC criteria and assessment dimensions 50
Table 13: Rating and score scales 50

Abbreviations

BMZ	German Federal Ministry for Economic Cooperation and Development
CPE	Central Project Evaluations
DIE	German Development Institute
DG DEVCO	European Commission's Directorate-General for International Cooperation and Development
EC	European Commission
ECOSOC	United Nations Economic and Social Council
EFO	Externally Financed Output
ESF	Environmental and Social Framework
GEDO	German General Executive Director's Office
GDI	Global Delivery Initiative
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH
HLPF	High-Level Political Forum
ICR	International Cooperation with Regions for Sustainable Development
IFAD	International Fund for Agricultural Development
iLTE	International long-term expert
IO	International organisation
JICA	Japan International Cooperation Agency
KfW	Kreditanstalt für Wiederaufbau (Credit Institute for Reconstruction)
LNOB	Leave no one behind
OECD-DAC	Organisation for Economic Co-operation and Development–Development Assistance Committee
SDG	Sustainable Development Goal
SMART	Specific, measurable, achievable, realistic and time-bound
StäV	Permanent Mission of the Federal Republic of Germany to the United Nations
SCIO	Strategic Cooperation with International Organisations
TC	Technical cooperation
ToC	Theory of change
ToR	Terms of reference
UN	United Nations
UNDG	United Nations Development Group
UNDP	United Nations Development Programme
WB	World Bank

The project at a glance

Global project: Strategic Cooperation with International Organisations

Project number	2015.6258.6
Creditor reporting system code	43010 – multi-sectoral help
Project objective	With a view to implementing the 2030 Agenda, international organisations have further developed reform processes which BMZ defined as strategic
Project term	January 2016 to January 2020
Project volume	EUR 5,000,000
Commissioning body	German Federal Ministry for Economic Cooperation and Development (BMZ)
Lead executing agency	-
Implementing organisations (in the partner country)	United Nations (UN), United Nations Development Programme (UNDP), World Bank (WB), European Commission (EC), Global Delivery Initiative (GDI), International Fund for Agricultural Development (IFAD) – especially teams responsible for reform processes in the organisations
Other development organisations involved	N/A
Target group(s)	UN, UNDP, WB, EC, GDI, IFAD

1 Evaluation objectives and questions

1.1 Evaluation objectives

The evaluation of the global project 'Strategic Cooperation with International Organisations' forms part of the central project evaluations (CPEs) of the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. All projects commissioned by the German Federal Ministry for Economic Cooperation and Development (BMZ) with a commission value over EUR 3.0 million are included in the evaluation on a standard basis. The project under evaluation has been selected by means of a **random sample**.

Concerning the epistemological interest for the evaluation, the GIZ evaluation unit has voiced a twofold purpose: (1) the general goal is to achieve accountability; and (2) it aims to learn why and how different aspects of the project have or have not worked. As the project was concluded in January 2020 without a follow-up project, this can be defined as a **final evaluation**. However, a new project entitled 'Promoting the implementation of 2030 Agenda through more effective cooperation between multilateral actors' (PN: 2019.2035.4), which started in January 2020, incorporates some aspects of the evaluated project.[1] Hence, although set up with a different portfolio, the evaluation results should be useable in parts for the new project. Furthermore, the evaluation can also generate lessons learned for future projects with a similar set-up.

The **central stakeholders** of this evaluation are the GIZ evaluation unit, the project staff and the international partner organisations and networks. Furthermore, the German Federal Ministry for Economic Cooperation and Development (BMZ) is a central stakeholder. This includes, in particular, Division 400 (Policy issues of multilateral development policy, G7/G20) as the commissioning party of the project, as well as further divisions that drew on the project for support.[2] Further stakeholders are the Permanent Mission of the Federal Republic of Germany to the UN (StäV) and the German General Executive Director's Office (GEDO) at the World Bank. Regarding the project's partner structure, the international organisations whose reform processes the project supported are the central actors in this evaluation: United Nations (UN), United Nations Development Programme (UNDP), World Bank (WB), Global Delivery Initiative (GDI) and the European Commission.

1.2 Evaluation questions

The project is assessed on the basis of standardised evaluation criteria and questions to ensure comparability by GIZ. This is based on the OECD-DAC criteria for the evaluation of development cooperation and the evaluation criteria for German bilateral cooperation: relevance, efficiency, effectiveness, impact and sustainability. Aspects regarding the criterion 'coherence, complementarity and coordination' are included across the other criteria. Specific evaluation dimensions and analytical questions are derived from this given framework by GIZ. These evaluation dimensions and analytical questions are the basis for all CPEs in GIZ and can be found in the evaluation matrix (see Annex 1). In addition, the contributions to the 2030 Agenda and its principles (universality, integrative approach, leave no one behind, multi-stakeholder partnerships) as well as cross-cutting issues such as gender, the environment, conflict sensitivity and human rights are taken into account. Furthermore, aspects regarding the quality of implementation are included in all OECD-DAC criteria.

As a global project, 'Strategic cooperation with International Organisations' is part of BMZ 'International Cooperation with Regions for Sustainable Development' (ICR).[3] Additional evaluation questions are defined for

[1] This new project is not part of the ICR title but is part of the TC title.
[2] This includes divisions 110, 111, 400, 401, 403, 404, 405, 410.
[3] The ICR title is phasing out and the described new project (see 1.1) is now part of the TC title.

the ICR title, and can be found in the evaluation matrix (see Annex 1). Furthermore, during the inception phase, stakeholders voiced specific knowledge interests for this evaluation. While some of these questions, could be answered within the standardised evaluation questions,[4] the following were listed as additional questions (see Annex 1): (1) Under which circumstances is a secondment as a technical cooperation instrument particularly strategic and effective? (Int_12); (2) How does the project staff in Germany contribute to maximising the results achieved by the international long-term experts (iLTEs)? (Int_13); (3) What should a results matrix and monitoring system look like to best allow flexibility while still steering towards the project's strategic objectives? (defined during workshop on results model); (4) Considering that multilateral cooperation is continuously changing, how were the project's measures and instruments adjusted to remain relevant and effective throughout the project's lifetime? (Int_1). All questions were answered as cross-cutting issues of the relevance, effectiveness, impact and sustainability criteria.

2 Object of the evaluation

2.1 Definition of the evaluation object

Subject of this evaluation is the ICR-project '**Strategic cooperation with International Organisations**' (PN: 2015.6258.6), in the following named '(the) project'. The project's aim was to support specific international organisations' (IOs') reform processes that would make them 'fit' for implementing the 2030 Agenda. This was based on the project's analysis that structures and processes of IOs were not sufficiently geared towards adequately managing and implementing the Agenda. According to the project's analysis, such support was especially needed due to the complex, long-term and highly political nature of institutional change (GIZ 2017a). However, the project was not set up to support just any IO reform process, but specifically those that were also identified as strategic by BMZ being the project's commissioning party. This is reflected in the module's objective: 'With a view to implementing the 2030 Agenda, international organisations have further developed reform processes which BMZ defined as strategic'. The IOs were consequently defined as the project's **target group** and at the same time the partner for implementation. Furthermore, the project was intended to be a flexible instrument with which BMZ could respond to windows of opportunity for enhancing IO reform. Several BMZ sector – and institution – divisions could suggest IO processes to be supported by the project. For decision-making, BMZ and the project were guided by criteria such as the feasibility of the support, the maturity of the idea and the relevance of the topic (Int_5). Here, it was discussed whether a specific activity or change within an IO would also take place without the project's support, or whether other donors were willing and able to step in. While BMZ might not be considered a target group of the project in the strict sense of the term, it was intended to benefit politically from the project and can hence be considered a shareholder in the project's results.

As an **overall framework,** the project was first embedded in BMZ's 2013 Strategy for Multilateral Development Cooperation,[5] in which BMZ specifies how it aims to support IOs and multilateral processes (today updated to the multilateral strategy 2020, see 4.2). Due to its focus on supporting internal reform processes of IOs and networks, which have a global scope of work, the global project did not have a specific **geographical focus**. Originally, the project was set up to last for two years from January 2016 to December 2017, with a total budget of EUR 2,000,000 (project offer 2015). With the modification offer of May 2017, the project's duration was extended to June 2020 and the total budget increased to EUR 5,000,000. It was later shortened to January 2020,

[4] Interview partners of BMZ were particularly interested in the general learnings from this project which could be useful for future projects (Int_13). The GIZ project team is interested in receiving an assessment of whether the project was perceived as useful from outsiders, such as partner organisations (Int_5). Likewise, partner organisations were interested in what the central lessons learned from the project were from GIZ's point of view (Int_6).
[5] The title of the strategy 2013 is Eckpunktepapier für die multilaterale Entwicklungspolitik (BMZ 2013).

since all project goals had been achieved. With regard to the project's **financial delimitation**, no co-funding was received. However, the project received partner support through, for example, offices and work materials provided for experts seconded to IOs. All in all, the modification offer of 2017 estimated the partner support at EUR 1,500,000.

The project's **capacity development strategy** focused on IO reforms and centred on the organisational level in combination with capacity development activities at the individual level. As already mentioned, the project supported different partner organisations. Therefore, the project's activities are best understood by viewing them as **seven different and separate individual measures**. Instruments used in these measures were seconded international long-term experts (iLTEs) to IOs, financial contributions with clearly defined outputs, short-term consultancies, and a project staff based in Germany. The project's focus in terms of instruments was on iLTEs. In five out of seven measures, the project seconded an iLTE to a partner organisation (with personnel costs accounting for more than half of the project's budget).[6] Moreover, financing was used in addition to the respective secondments in the UN and the two WB measures (see details for each measure below). In the majority of cases, the decision to employ financing instruments was prepared and discussed with the help of the iLTEs, linking the two technical cooperation (TC) instruments. Only in one case (IFAD, see below) was a financing instrument employed without the framing by an iLTE.

All in all, the seven measures can be considered to fall into one of two categories: While some measures are (1) geared towards supporting internal reform processes in IOs, as described in the project's strategy (GIZ 2017a), (2) others have a rather thematical focus. The genesis of the project is the reason for these two categories of measures: while it was initially used as a 'tool' for secondments into rather topical processes that BMZ wanted to support, the modification offer of 2017 strongly reoriented the project towards overall reform processes and the 2030 Agenda (Int_15, 30). However, secondments initiated before the modification offer continued until the end of their agreed contract, so that a mix of 'topic-oriented measures' and 'reform-oriented measures' falls within the scope of this evaluation. The following paragraphs give an overview of each individual measure, as the following chapters will regularly refer to them. The three **topic-oriented measures** started between 2011 and 2015. They were financed by the project and its predecessor.

- **European Commission DG DEVCO – topic: resource mobilisation (Nov 2011 – Aug 2016)**: Funded by the project, an iLTE worked for the Budget Support and Public Financial Management Unit A4 of the European Commission's Directorate-General for International Cooperation and Development (DG DEVCO) from November 2011 to August 2016. The measure aimed to strengthen the technical and financial cooperation between German development cooperation and the EU in the areas of budget support and public finance, anchoring BMZ's Good Financial Governance approach in the EU (GIZ 2015).
- **United Nations Development Programme – topic: SDG 16 (Aug 2014 – Dec 2016)**: An iLTE was funded by the project to work in the position of policy advisor in the Bureau for Policy and Programme Support at the United Nations Development Programme (UNDP) from August 2014 to December 2016. The measure aimed to support and expand existing cooperation between BMZ/GIZ and UNDP and to support the exchange of knowledge, information and experiences between UNDP and GIZ/BMZ. The iLTE's central task was the creation of indicators towards SDG16 – promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable, and inclusive institutions at all levels.
- **European Commission DG DEVCO – topic: garment Initiative (May 2015 – Mar 2017)**: In this individual measure, an iLTE was funded for the Trade, Private Sector and Regional Integration Unit (C4) of DEVCO from May 2015 to March 2017. The iLTE's objective was to position and help shape the issue of sustainable global supply chains in the textile sector at the EU Commission and to coordinate it with the approaches pursued by BMZ.

In its **reform-oriented measures,** the project focused on the reform processes that had already been initiated by IOs but needed further support. Thereby, the project aimed to address the core problem identified in the 2017

[6] Exceptions were the GDI and the IFAD measures, see more detailed descriptions below.

modification offer, namely that IOs were not yet 'fit' enough to manage and implement the 2030 Agenda. Depending on the respective organisation, the specific needs of support and the nature of the reform process differed, although the measures were generally implemented in those units within the IOs that were responsible for addressing key challenges of the reform processes. Contrary to the topic-oriented measures, the project also used financing in addition to iLTEs.

- The project shortly supported the financial reform process of the **International Fund for Agricultural Development (IFAD)** by commissioning an external consultant from February to April 2019. The consultant worked closely with the team of IFAD's Risk Management Unit and contributed to the modernisation of the organisation's financial architecture by assisting in drafting the new Capital Adequacy Framework.
- The project supported two measures with and within the **World Bank**. In 2013 under WB Group President Kim, the organisation started to look at reforms under the heading Science of Delivery (SoD) – the science of successful implementation of development policy projects. In this context, WB initiated the **Global Delivery Initiative (GDI)** together with GIZ and other institutions (donors, universities, think-tanks). The Initiative was founded as a partnership with the objective of enhancing the quality of implementation. In 2016, the GIZ project manager of the project under evaluation took over the role as co-chair together with WB. During his term until mid-2019, the project supported the GDI with personnel support of the staff in Germany including the project manager as well as financing. In terms of personnel support, the project provided strategic advice to the GDI Secretariat and helped to organise the GDI Annual Conferences, among other operational support. In terms of financing, the project supported the implementation of Delivery Labs – opportunities for peer exchange on delivery challenges.
- The second measure within WB supported the implementation process for the new **Environmental and Social Framework (ESF)**. The GIZ project supported the implementation team with an international long-term expert as senior advisor from September 2017 to July 2019. It also provided financing to the process. The iLTE focused on capacity building, supporting the training of Bank staff and partners, the preparation of a middle and long-term capacity-building strategy and the proposal for a multi-donor trust fund. Additionally, the iLTE helped develop a tool for project-level capacity needs assessments. The project also financed the pilots to test this capacity needs assessment in practice.
- Lastly, the project supported the **UN fit for purpose process**, aimed to get the UN ready for the implementation of the 2030 Agenda by making the system more effective, relevant, capable of action, as well as being partner and results oriented. The GIZ project first financed support to the reform process's Independent Team of Advisors. Later, the project supported the reform by accompanying UNDP and the United Nations Development Group by providing financial support to the activities of the UN Senior Coordinator 'Fit for purpose and the 2030 Agenda'. Starting in March 2016, the project also seconded an iLTE to support the Senior Coordinator in his tasks, who continued the support through different stages in the reform process. In August 2018, the iLTE transferred to the UNDP Executive Office to support the reform process by providing UNDP with advice and input into United Nations Development Systems' reform-related topics.

2.2 Results model including hypotheses

The following section will present the results model of the project under examination. It depicts the interaction between the intended output, outcome and impact-level results of the project (see Figure 1). The project team based the results model on the project's process map of February 2017 which was then updated jointly by the project team and the evaluators during the inception phase in May 2020.[7] Hence, the version of the results model depicted below is a product of the inception phase.[8]

The results model is best understood by first considering the project's aspired outcome (module objective). According to the modification offer (05/2017), it reads: **With a view to implementing the 2030 Agenda, international organisations have further developed reform processes which BMZ defines as strategic.** This is to be achieved via three fields of action, as the results model depicts (see Figure 1): (a) the formation of ideas and alliances; (b) the design of reform processes; and (c) the implementation of reforms. These fields of action are derived from selected phases of an ideal typical reform cycle. Each field of action is represented by an output:

- Field of Action A aims to strengthen alliances to support German proposals in IO reforms for implementing the 2030 Agenda (**Output A**).
- Field of Action B has the objective to introduce methodological proposals which are useful for the design process of reforms in IOs for the implementation of the 2030 Agenda (**Output B**).
- Field of Action C aims to apply the reforms of IOs developed with a view to the 2030 Agenda on a pilot basis and as far as possible with German bilateral development cooperation (**Output C**).

As described in section 2.1, the project encompassed a range of different measures, which each aimed to support the achievement of the project's outcome individually within different IOs. Therefore, the results model identifies which measure targeted which output instead of differentiating between different activities (see Figure 1). The following overview outlines the overall results logic, the central hypotheses, risks, potential unintended results and the project's system boundary.

The hypothesis of **Output A** *is that reform processes necessary to enable the IO to implement the 2030 Agenda can be developed further by encouraging potential partners to participate and contribute to the process*. To reach this output, two intermediary steps have to be taken (**I5, I6**). First, the alliances need to be linked to processes in the respective IO that are relevant for German development cooperation (**I6)**. Second, trust also needs to be built among alliance partners to support reform proposals (**I5**). Hence, the project presupposes that its strategic communication and technical support can contribute to fostering stronger alliances and trust in the reform processes. Output A is strongly related to the activities of German political stakeholders, who drive the formation of alliances on a political level (**I8**) and also influence the creation of trust among alliance partners in support of reform proposals (**I5**).

Output B is based on the hypothesis *that useful methodological proposals can be used for further developing reform processes that are relevant to enable an implementation of the 2030 Agenda*. This hypothesis is based on the assumption that methodical contributions are necessary to advance reform processes and that windows of opportunity for these types of technical contributions emerge when stakeholders within the IO identify a methodical need which (1) if not served, impedes further progress and which (2) cannot be met from within the IO. Therefore, the underlying hypothesis is that if the project can offer methodological expertise at the opportune moment in a reform process, it can help enhance the quality of the entire process.

Output C refers to the implementation of reforms, the hypothesis being *that an application of reforms on a pilot basis, as far as possible with German bilateral development cooperation, further develops the respective reform*

[7] Throughout the project's implementation, the project team mainly worked with the aforementioned process map.
[8] The abbreviations used to relate to the results model in Figure 1 include: I for intermediary steps, Im for impacts and R for risks.

process towards 'fitness' for the 2030 Agenda. This hypothesis is based on the experience that reforms, even if officially adopted, need to be tested first (see modification offer of 2017). Through pilot projects in IOs, partners and member states can make their own experiences with the practical implementation. Following this logic, the underlying hypothesis is that if the project supports (1) the development of concepts for the implementation of reforms, (2) the coordination of concrete steps with stakeholders in different countries, and (3) the exchange of knowledge and experiences with regards to the reform implementation, it can help to ensure the rollout of reforms beyond their adoption. In this context, the project describes the interlocking with bilateral and multilateral development cooperation as a cross-cutting issue. The assumption is that bilateral experiences cannot only be beneficial for implementation (Output C), but also for internal discussions (Output A) and methodical and technical considerations (Output B) (GIZ 2017a).

As the results model shows, the diverse individual measures did not target every output. However, the underlying hypothesis which relates to **all three outputs** is that *if the support of reform processes is provided in a technically competent, flexible and integrated manner in the structures of the IO (I3), Output A, B and C can be achieved.* This is linked to further hypotheses concerning the Outputs:

- **Output A:** Through the technically competent and flexible support, which is integrated into the IOs structures (I3), alliances can be linked to processes in IOs that are relevant for German development cooperation (I6) and trust can be built among alliance partners for the support of reform proposals (I5).
- **Output B:** The type of support provided by the project (I3) allows that methodological proposals cover the needs of the respective IO (I4). In addition, this type of support (flexible, technically competent and integrated) partially creates ownership for methodological proposals within the IO and, thereby, enables Output B and C.

Apart from these central hypotheses, the results model also shows that the project's outcome is highly dependent on a range of external influencing factors. Here, two categories can be differentiated: **Risks** within the project's area of influence and those outside of it. One risk within the project's area of influence is that it favours forms of support to reform processes that are too selective and too short term (ending before its objectives can be achieved, considering the long and complex nature of reform processes) (**R5**); that IOs or stakeholders perceive the support for individual reforms as a one-sided representation of German interests (**R7**) and that the secondment does not provide the expected support or does not fulfil the role attributed to it (**R8**). These risks can interfere with most intermediary steps to Output A and C and directly interfere with Output B. In addition, a number of risks exist that lie outside the project's sphere of influence. First, it is possible that individual stakeholders or member states of the IOs hinder or block the reform process for political reasons (**R1**). This could potentially lead to an alteration of the process or its failure. Similarly, there is a risk that important stakeholders in the process change their priorities (**R2**), which could lead to alterations and frictional losses. Third, the speed and success of the reform process might be impeded through personnel changes (**R3**) within the IO. Due to different circumstances, such as new crises or priorities arising, a reform process might also lose momentum before the necessary changes are implemented (**R4**). Last, there is a risk that the international organisation does not use the project's support to its full potential, which might impact the individual measure's effectiveness (**R6**).

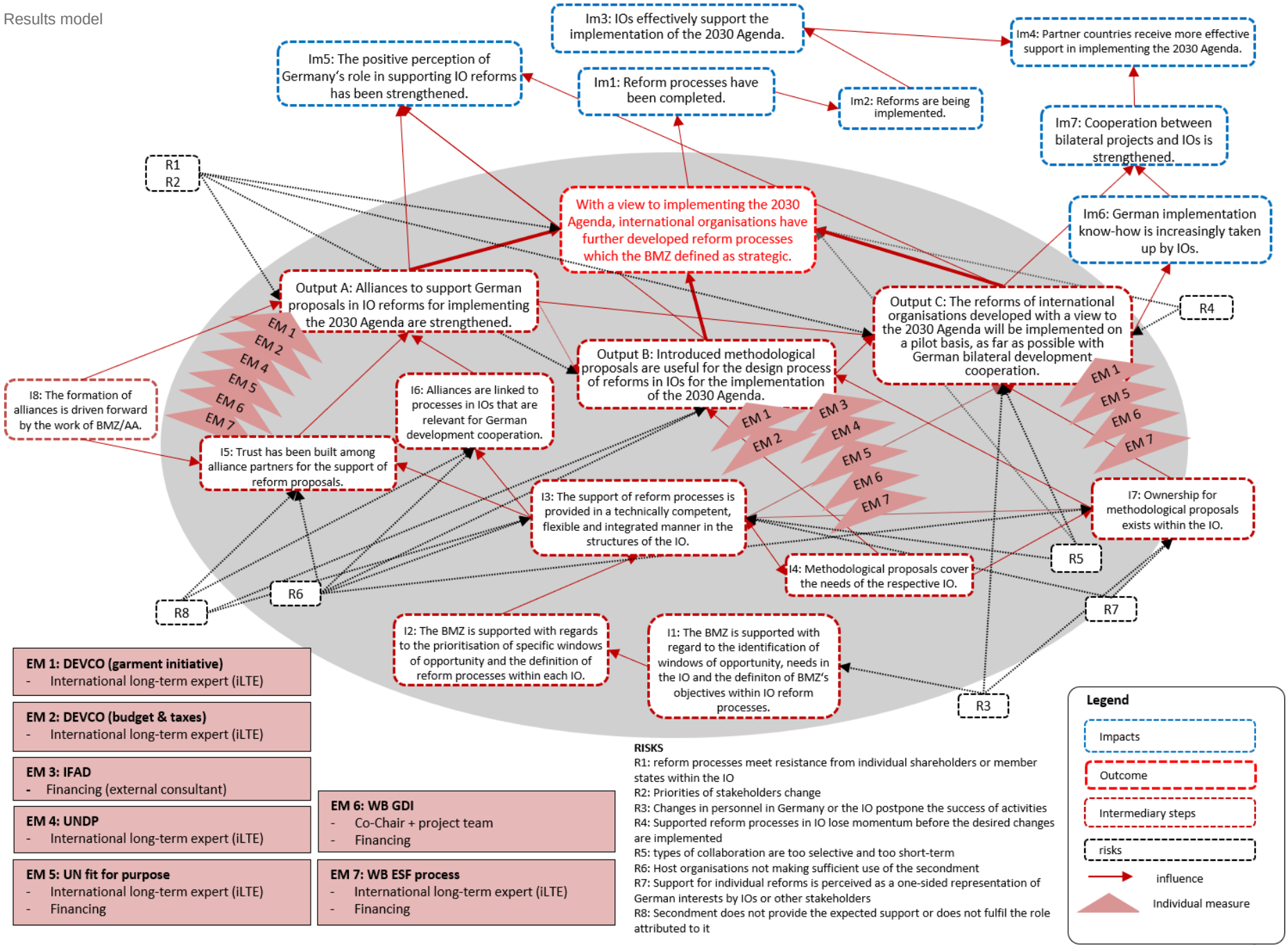
In a last step, the impacts and the system boundary of the project will be described. According to the project's results model, several **impacts** arise from achieving its outputs and outcomes. The project is based on the assumption that a further development of reform processes (outcome) ultimately leads to the completion and implementation of those reforms (**Im_1,2**), which then enables IOs to effectively support the implementation of the 2030 Agenda (**Im_3**). In general, due to the project's character, the objective to support the implementation of the 2030 Agenda follows a rather long results chain. Only if the IO actually implements the reform process and if the reforms indeed lead to a greater aptness to implement the 2030 Agenda, and if implementation is then actually followed through, would the intended long-term impacts be reached.

Besides the further development of reform processes, the introduction of methodological proposals (Output B)

and the strengthening of alliances to support German proposals in IOs (Output A), according to the results model lead to a more positive perception of Germany's role in supporting IO reforms (**Im_5**). Furthermore, the implementation of reforms on a pilot basis with German bilateral development cooperation (Output C) leads to an increased uptake of German implementation expertise (**Im_6**) and, in addition and as a direct result, further strengthens the cooperation between bilateral projects and IOs (**Im_7**). Both the strengthened cooperation between bilateral partners and IOs (**Im_7**) and the effective support to the 2030 Agenda by IOs (**Im_3**) in the long-term ensure that partner countries will receive more effective support in implementing the 2030 Agenda (**Im_4**).[9] These impacts lie outside of the project's **system boundary**. The system boundary, as presented in Figure 1, delineates which outputs and outcomes the project can directly influence, as opposed to those that lie within the responsibility of other actors (such as BMZ, the IOs or other stakeholders).

[9] In the spirit of the 2030 Agenda, impact evaluations usually consider potential interactions between social, economic and environmental results at the outcome and impact level. Due to the specific nature of this project, these dimensions are not directly applicable here, as the project's intended impacts are mostly at the level of BMZ and the IOs. For this reason, the results chain to implementation at partner country level is too long to be meaningfully assessed in this evaluation. However, the evaluation will focus on the extent to which the project worked towards integrating considerations on this issue area in their support of the 2030 Agenda implementation at IO level.

Figure 1: Results model



3 Evaluability and evaluation process

3.1 Evaluability: data availability and quality

The evaluators used a variety of data for the analysis including interview data (primary data), project documents such as the project's progress reports, as well as further secondary data provided by partner organisations. Concerning the available documents, the modification offer of 2017 was a major reference point, because it provided the basis for the project's results logic (see section 2.2). Taking into consideration that some of the measures implemented between 2016 and 2019 started prior to the modification, an assessment of the original project offer 2015 was also relevant. The evaluators further used the project's monitoring data, mostly in the form of its progress reports. The monitoring data itself largely consisted of reports submitted by the iLTEs to the monitoring staff in Germany. In these reports, the seconded experts described their tasks, current developments and the progress made in their respective reform process. Furthermore, monitoring was fed with the results of interviews conducted by a consultant for the progress report 2016. Therefore, the monitoring data used by the project was largely of qualitative nature.[10]

The evaluators used these data sources, of which especially the iLTE's activity reports turned out to be particularly useful. Furthermore, the interviews conducted in 2016 were helpful especially with the older measures, as some stakeholders were no longer available for interviews in this evaluation (DG DEVCO and UNDP). As for strategic reference documents, the BMZ strategy for multilateral development cooperation of 2013[11] provided the basis for assessment. Secondary data from partner organisations was helpful in parts to assess the overall context of the support.

Finally, there were no monitoring systems or data of partner organisations which could have been used for the project's monitoring by project staff or during the evaluation.

3.2 Evaluation process

The evaluation was implemented in the period from May to September 2020. Due to the Covid-19 pandemic, the evaluation had to take place remotely, using videoconferences to implement interviews. The evaluation team followed a participatory approach to evaluation, which fosters ownership for evaluation results and provides the basis for learning that can be used in future projects. In addition, the evaluators were transparent on how evaluation results were derived from the data and gave GIZ project staff the opportunity to provide feedback on evaluation findings as a debriefing. The evaluation also considered questions that stakeholders on the partner side wanted to see addressed.

The stakeholders of the evaluation can be clustered into three main groups: (1) different actors within GIZ (the evaluation unit as the commissioning party of this assignment, the project team including the iLTEs, and the Sectoral Department, FMB); (2) BMZ as the commissioning party of the GIZ project, and more specifically the different divisions within BMZ that assigned the project with different individual measures (this includes the departments 110, 111, 400, 403, 404, 405, 410 and German representations at partner organisations (GEDO

[10] Monitoring staff did not use any separate lists, Kompass or other specific tools for gathering data and updating the target system and all baselines were chosen at 0.
[11] The strategy was published under the title Eckpunktepapier für multilaterale Entwicklungspolitik (BMZ 2013).

and StäV);[12] and (3) the international organisations that received support (DG DEVCO, UNDP, WB, UN and IFAD), as they were the project's direct target group.

Besides those three stakeholder groups, the evaluation team also included other measure-specific stakeholders. Regarding the project's support to GDI, this included other technical development actors. For universities and think-tanks, the German Development Institute (DIE) was also included as they accompanied the UN fit for purpose process. Considering that the project's logic and implementation was highly removed from the indirect target group, the wider population (and especially disadvantaged populations) of partner countries, it was neither feasible nor relevant for an assessment of the project to include them as stakeholders.

In general, interview partners were identified in close cooperation with the GIZ project team and iLTEs provided support for selecting stakeholders in partner organisations. During the inception phase, the GIZ project team and Syspons determined document and interview partner availability for all seven measures. As a result, the evaluation team decided to focus on UN reform, WB ESF and GDI measures. This focus is also reflected in the number of stakeholders interviewed for each measure. In total, the evaluation team conducted 28 interviews during the evaluation phase as shown in Table 1.[13] A range of personnel changes in BMZ and GIZ affecting the responsibilities of the project during the evaluated period (2016-2019) represented a challenge to the evaluation. Some relevant interview partners retired, left the involved organisations or changed their work unit. In addition, apart from the information available in documents, much of the steering and decision-making of the project took place in strategic workshops, visits, or phone calls between GIZ, BMZ and iLTEs. While all former GIZ project staff could be closely involved in the evaluation, neither BMZ officials responsible for the initial project phase nor officials from GEDO could be interviewed. This represents a limitation for the evaluation, which the evaluation team will make transparent in the assessment of the different criteria (see chapter 4). Researcher triangulation took place regularly by sharing and discussing interview results. Data triangulation also included the validation of preliminary insights with the GIZ project team during a virtual debriefing.

Table 1: List of stakeholders of the evaluation and selected interviewee

Organisation/company/target group	**Overall no. of persons involved in evaluation (including gender disaggregation)**	**No. of interview participants**	**No. of focus group participants**	**No. of workshop participants**	**No. of survey participants**
Donors	4 (1 female, 3	4	/	/	/
BMZ					
GIZ	14 (7 female, 7	14	/	/	/
GIZ project team/ iLTEs					
GIZ staff of other units					
Partner organisations (direct target group)	12 (5 female, 7 male)	12	/	/	/
European Commission					
UN (UNDP)					
World Bank					

[12] The commissioning department of BMZ (400), the GIZ project team, the GIZ Sectoral Department (FMB) and the GIZ Evaluation Unit also commented on the inception report.

[13] This is complemented by 13 explorative interviews with GIZ project staff, BMZ officials as well as selected partners of the project (WB, GDI, UN) conducted during the inception phase.

Organisation/company/target group	Overall no. of persons involved in evaluation (including gender disaggregation)	No. of interview participants	No. of focus group participants	No. of workshop participants	No. of survey participants
IFAD					
Other stakeholders (public actors, other development projects, etc.)	4 (1 female, 3 male)	4	/	/	/
GDI members (JICA, BRAC USA, UNDP)					
Permanent Mission of the Federal Republic of Germany to the UN (StäV)					
Universities and think-tanks	1 (female)	1	/	/	/
German Institute for Development Politics (DIE)					

4 Assessment of the project according to OECD-DAC criteria

4.1 Long-term results of predecessor

The project under evaluation was preceded by a similar project (PN 2012.6251.8) running from 2012 until 2015. After careful consideration during the inception phase, the evaluation team decided that the predecessor would not be considered in this evaluation. Although similar to the follow-up project (PN 2015.6258.6) in its basic orientation, it was more broadly used as an instrument for secondments into IOs and less strictly oriented towards supporting selected reform processes (Int_15). Consequently, the secondments followed individual objectives that are not directly related to the project logic developed for the phase under evaluation (2016 through 2019). Furthermore, very limited data is available from this first project phase. Taking into consideration these different factors, the evaluation team decided that the predecessor will not be fully assessed.

4.2 Relevance

The relevance criterion analyses the extent to which the objectives of a development project are consistent with beneficiaries' requirements, regional needs, global priorities as well as partners' and donors' policies.

Evaluation basis and design for assessing relevance

The assessment of the relevance criterion is structured in four dimensions: In the **first relevance dimension**, the evaluation team analysed how far the project's strategy was aligned with relevant strategic reference frameworks. In this regard, the evaluation team mainly took into account the BMZ strategy for multilateral development cooperation of 2013 and the report to the High-Level Political Forum on Sustainable Development 2016. Interview partners in BMZ also pointed out that the multilateral strategy 2020 approach, although

published after the project term, already affected BMZ's approach to the project in spirit (Int_13). In addition, qualitative interviews with BMZ officials and the Permanent Mission of the Federal Republic of Germany to the UN (StäV) complemented the document analysis for this first dimension. In addition, the evaluators assessed whether the project was subsidiary to efforts of other partners. In the context of this project, the evaluation team defined these 'efforts of other partners' as 'support by other countries or members to the reform processes'. The team also assessed whether the project complemented other regional, sectoral or global projects. Furthermore, as a global project under the ICR title, the evaluation matrix for this dimension included additional questions: the evaluation team analysed to which Sustainable Development Goals (SDGs) the project contributed, and to what extent the project was geared towards solving a global challenge and closed gaps with regard to this global challenge. To answer these questions, the evaluation team conducted a document analysis of project documents and interviewed GIZ stakeholders and partner organisations.

The **second relevance dimension** deals with the suitability of the project's strategy to match core problems and needs of the target group. According to the project's progress reports, the project's direct target group was defined as IOs and networks which were supported by individual measures. Each measure addressed different needs within diverse (reform) processes. Therefore, to answer the questions listed in the evaluation matrix, Syspons differentiated between the individual measures to clarify the specific contexts. Qualitative interviews with the target group and iLTEs were essential for answering the questions in the second dimension.[14]

The **third and fourth relevance dimension** analysed the extent to which the project design is suitable for achieving the chosen project objective, and to what extent the conceptual design was adapted to contextual changes as required. For assessing these dimensions, the results model and the underlying hypotheses, the project proposal 2015 and the modification offer of 2017 served as the central basis. In addition to the document analysis, the evaluation team also conducted interviews with the GIZ project team and stakeholders in partner organisations.

Analysis regarding relevance

The analysis of relevance dimension 1 focused on the project's alignment with the relevant strategic reference frameworks. Considering first the project documents, especially the initial project proposal (2015) and the 2017 modification offer, a strong connection to the BMZ multilateral strategy 2013 and German report to the High-Level Political Forum (HLPF) 2016 can be identified. The initial project proposal from 2015 in fact embedded the three key points of the multilateral strategy 2013 in its results model: (1) effective cooperation and reform agendas; (2) systematic and goal-oriented agenda-setting; and (3) links between bilateral and multilateral cooperation. In turn, the 2013 strategy even explicitly referred to the predecessor of the project under the ICR title as a means by which BMZ operated a systematic and goal-oriented agenda-setting at the international level (BMZ 2013: 10), showing the close conceptual interconnection between the overall strategy and the project. Moreover, interview partners at BMZ, as well as the project team indicated that they saw the project's design as evolving over time, alongside the evolution of BMZ's priorities (Int_13, 28). BMZ eventually started emphasising the 2030 Agenda implementation more strongly as a priority, for example HLPF 2016, where Germany's report expressed that 'the goal to improving the international framework for sustainable development around the globe [and] gearing multilateral organisations to the promotion of sustainable development around the globe' (German Federal Government 2016: 16). The project followed suit, introducing its modification offer in 2017 with a very clear focus on IO reform processes that were directly related to 2030 Agenda implementation, shifting away from the initial proposal's focus on agenda-setting. This is illustrated by comparing the three measures at UNDP and the EU initiated before 2016, with the logic of those measures supported since 2016 and after the modification offer of 2017. Within the former measures, the focus was thematical and the iLTEs did not support reform processes of the respective organisations. Contrary to the EU

[14] The evaluators also assessed the additional aspects included in dimension 2: gender equity and the **leave no one behind (LNOB) principle** of the 2030 Agenda (e.g. whether the project was designed to reach particularly disadvantaged). As both aspects turned out not to be relevant for the project, which had IOs as the target group, the evaluation team did not refer to the LNOB or **gender equity** in the analysis.

and UNDP measures, the support to UN fit for purpose, WB ESF process, GDI and the support to IFAD did not have this type of thematical focus, but emphasised the overall aim to make multilateral organisations 'fit' for the 2030 Agenda management and implementation. While an update of BMZ's 2013 multilateral strategy was only published in 2020, interview partners indicated that the project kept close to the strategy development and 'picked up' on the new strategy's spirit (Int_9, 13). This shows how the project adapted its concept in an ongoing manner to stay relevant with a view to the strategic reference frameworks.

With its strong focus on the 2030 Agenda, the project related to the **Sustainable Development Goals** in two ways. First, the project corresponds strongly with SDG 17 – Strengthen the means of implementation and revitalise the global partnership for sustainable development. This is best exemplified through the support to the UN fit for purpose process, WB ESF process, the GDI and IFAD, which were directly supposed to enhance implementation capacity. Furthermore, some individual measures implemented by the project also show a thematic fit to the 2030 Agenda, e.g. SDG 12 (work on the garment initiative) or SDG 16 (iLTE at UNDP).[15] Second, a support to reaching SDGs can be understood as 'in-built' into the project's design (even if the chain of results from the project's support to fostering the SDGs is long, as sections 4.3 and 4.4 will show), since it sought to enable IOs to become better managers of the 2030 Agenda. The interviews of the evaluation show that there was one crucial prerequisite to ensuring this 'in-built' relevance in practice: choosing reform processes for support that were in fact related to the 2030 Agenda implementation (Int_9, 15, 18). According to interview partners in the respective IOs, this link could effectively be confirmed, even if the argument often reads relatively technical:

- In the case of the **UN fit for purpose** process, the reform according to interviewees has potential to directly lead to a better implementation of the 2030 Agenda (Int_10, 17, 22). That is, the reform of the Resident Coordinator system addresses the UN's core problem of fragmentation, which leads to a variety of projects with limited resources that only created reduced impact and can therefore not sufficiently contribute to a transformation in terms of the 2030 Agenda. The reform, by strengthening the role of the Resident Coordinator, set the course for a more common and collective programming of UN agencies at country level (Int_34).
- For **WB ESF process**, the potential with regard to the 2030 Agenda lies in the application of the new ESF for all new WB investment project financing. The assumption is that better managing social and environmental risks will ensure implementation that is more strongly in line with and even contributes to fulfilling the 2030 Agenda (Int_2, 8, 11).

In this vein, both the analysis of the project's documents and the interviews with relevant stakeholders show that the project was geared towards **solving a global challenge**, as stipulated under the ICR title through which the project was funded. Taking an analysis of the multilateral system's weaknesses as its basis, the initial project proposal and modification offer argue that the project could make selected technical-methodological contributions to relevant agendas (2015 proposal) and processes (2017 proposal) in IOs. This sees them as the crucial actors for supporting countries in their sustainable development, promoting the coordination of various donors, and creating the forum for international, cross-sectoral debates (GIZ 2017a: 5). In the interviews conducted during the evaluation, all stakeholders confirmed this argument, underlining that moving reforms forward inside an IO might be a political, but at least as much a technical exercise, in which input from technical experts can have important value added (Int_3, 4, 5, 15). Therefore, a technical cooperation measure such as the one under examination here was generally seen as uniquely positioned to impact the multilateral system, even if – naturally – in small and selective dosage (Int_3, 15).

Given the project's relatively unique set-up as directly working inside IOs, there were no other **regional, sectoral or global GIZ projects** that the project complemented in a direct manner. However, the interviews showed that all iLTEs exchanged information with stakeholders in other GIZ projects, such as the GIZ–World Bank Sector project (Int_8, 10, 18, 19, 20, 29, 32).

[15] See European Commission (2016): Collect More – Spend Better: Achieving development in an inclusive and sustainable way.

Seeing as the project was designed to strictly support agendas or reform processes that already existed within its partner organisations, it was directly **subsidiary to its partners' efforts**. The project documents show that an existing IO agenda or framework was in fact treated as prerequisite to the project's support, as its instruments could not have worked without an existing reform process (see, for example, modification offer of 2017). This also relates directly to the project's target group orientation (**relevance dimension 2**). As explained in the introductory sections of this report, the IOs supported by the project had in fact a dual role as partners and target group, at the same time. Ensuring partner orientation in this project, therefore, also meant ensuring target group orientation. On a general conceptual level, the stepping-stone for this was the selection of agendas and processes for support that already existed within the IOs and were owned by them (as argued in the project documents). The interviews with different stakeholders in the evaluation show that the decisive moment for ensuring target group orientation then lay in the selection and conceptualisation of individual measures, including the concrete terms of reference for the TC instruments employed by the project. Here, both the partner IO, as well as BMZ and the project agreed on the right fit of the instrument selected (secondment of an iLTE or financing instrument) (Int_1, 2 4, 9, 11, 12, 13). The interviews with stakeholders in the supported teams and units showed that the project adequately used this moment (Int_2, 6, 11, 12, 16, 17, 21, 22, 23, 24, 25, 27, 31). The following examples further illustrate that the project was able to meet the different needs of the target group:

- In the **UN fit for purpose** reform process, the UN Senior Coordinator 'Fit for purpose and the 2030 Agenda' needed quick personnel support in 2016 to prepare the Quadrennial Comprehensive Policy Review. Interviewees expressed that this position required an extensive skill set, including a comprehensive understanding of the UN system and different agencies (Int_12, 16). From the partner perspective, the iLTE deployed by the project was the right and timely fit, and the quick secondment enabled the continuous work of the supported team (Int_12, 16).
- In 2016, the **World Bank** created a team for implementing the new Environmental and Social Framework (ESF). It became apparent in the interviews that the team faced an enormous workload at the time and required personnel support (Int_2, 11, 23). Stakeholders further expressed the need for capacity-building knowledge as the team itself did not possess sufficient expertise (Int_2, 11, 23). Thus, from a partner perspective, the secondment of an iLTE in capacity building was the right fit. Likewise, the second instrument used in the ESF measure – the financing of pilots to test the project-level capacity needs assessment – was rated extremely useful for WB, as it created practical experiences that otherwise would not have been financed at the time (Int_11, 23).
- In the case of the **Global Delivery Initiative**, the analysis showed that, after initial WB financing, the Initiative was still at the development stage of aiming to become a global partnership with diverse members. The interviews showed that the takeover of the co-chair position and the operational support provided by the project team (e.g. support for the organisation of conferences, joint strategical planning with the GDI Secretariat) combined with financing were addressing the relevant 'weak spots' for the Initiative to finally reach a global level (Int_6, 18, 27, 33).

Apart from this illustrated match with the target group's needs, the evaluation team also analysed the strategic relevance of the measures for **BMZ as a shareholder in the project's results**. The analysis showed that the supported processes were generally relevant from the Ministry's point of view. This was assured via two mechanisms. First, a series of strategy workshops conducted with BMZ were used to select those individual support measures that had the most potential from the Ministry's political point of view. Second, interview partners also explained that 'windows of opportunity' for support were jointly monitored by BMZ and the project, and a generally close exchange and relationship assured the individual measure's relevance going forward (Int_3, 4, 5, 15). On this basis, especially the ESF and the UN fit for purpose measure were cited as highly relevant by BMZ interviewees (Int_1, 4, 13). Questions were only raised in relation to the GDI measure. The project's support to the Initiative, from some stakeholders' points of view, was not entirely coherent with the logic and political relevance of the other processes (Int_1, 15). The GDI is a platform which aims to enhance project implementation by its diverse members. Thus, by design, it is closer to implementation than to a political/IO reform process, and therefore further away from making IOs 'fit' for Agenda 2030 implementation

(focusing on delivery of support for direct target groups). Generally, perspectives on the GDI voiced during the evaluation varied strongly. With commissioning the project, BMZ initially affirmed the GDI measure and supported GIZ in taking on the co-chair role.[16] During the evaluation, however, interviewees at the Ministry reflected critically on whether the support to the Initiative was indeed beneficial for reaching BMZ's strategic objectives. This is because the Initiative in its set-up was closer to implementation than to IO reform and policy-making. At the same time, interviewees at the Ministry recognised especially the strategic relevance of the co-chair position within the GDI. Within GIZ as well, perspectives on the GDI varied. While some interview partners at GIZ remarked that the Initiative was a logical element in the project's set-up, since supporting reform implementation was part of its purpose, others saw it more as an opportunity for GIZ to position itself vis-à-vis other implementing organisations and IOs (Int_15, 18). These diverse perspectives on the GDI support will also be discussed in the analysis of the efficiency and the sustainability criterion (see 4.3 and 4.6).

Turning next to **relevance dimension 3**, it became apparent in the analysis that the project was mostly adequately designed for the chosen project objective. The analysis showed that the assumptions behind the theory of change (ToC) were mostly correct. The description of the results model in 2.2 already showed that the project was based on three aspects required for reforms – the formation of alliances (field of action A), the preparation of technical proposals (field of action B), and the implementation of reforms (field of action C). As mentioned already, it became apparent in the interviews that the target group, staff in IOs, did indeed need support in all three fields of action (see also 4.3). However, the analysis also showed that the original ToC had deficiencies. As it turned out during implementation, it was too ambitious regarding the interlocking with bilateral and multilateral development cooperation. The results model had foreseen the implementation of reforms on a pilot basis, as far as possible with German bilateral development cooperation (Output C), which should also have led to an increased uptake of German implementation expertise (Impact 6) (Int_3, 5). This turned out to be unrealistic due to systemic challenges, for example, the different programming cycles of multilateral partners and the bilateral level (Int_3) or the reluctance of bilateral partners to share information on delivery challenges (Int_5). Consequently, a link to German development cooperation could only be established as part of the project's topic-oriented measure at the EU level, which did not support reform processes and, therefore, did not directly correspond to the project's ToC (see 2.2). In reaction, the GIZ project team focused more strongly on the piloting and implementation preparation aspects of Output C. This will be critically examined in section 4.5 on efficiency.

Another deficiency in the project's design relates to Output A, which originally aimed to support alliances for German proposals in IO reforms for implementing the 2030 Agenda (GIZ 2017a). The analysis showed that this focus on *German* proposals in the project design did not correspond to the reality of the iLTEs' work and was criticised during the evaluation (Int_10, 19, 20, 32). It became apparent that, while the project supported (reform) processes which BMZ defined as strategic, proposals themselves were not *German*. Instead, what was needed and supported by the project were alliances for the technical proposals, which were developed within partner organisation's teams. Consequently, the results model was true to the spirit of the project objective, although not totally aligned to the objective as written in Output A (see also sections 4.3 and 4.4).

A third weakness became apparent when considering the adequacy of the project's target system as its basis for results-oriented steering.[17] A close examination of the module and output indicators showed that they contained very limited information about each individual measure. Both the module indicators and the output indicators aggregate results across the different contexts, giving no indication of the progress of the individual reform process, or the project's contribution to it (for example: 'four new methods for process design of reforms […] have been applied by international organisations' (module indicator 3)). This is further intensified by the fact that the module indicators are output-heavy and count the delivery of products rather than results at the level of reform processes – that is, module indicator 3 above tells us that methods have been applied, but not

[16] As explained in section 3.2, the evaluators could not interview BMZ officials responsible for the initial project phase.
[17] The target systems of all GIZ projects follow the requirements of the commissioning party (BMZ).

what their effect was. This leads to a weakness of the target system in depicting the progress of each individual measure and hence, as the basis for results-oriented steering.

As was discussed with the different stakeholders of the evaluation in the in-depth interviews, this weakness of the target system is caused by the complexity of the project on the one hand, and the need for flexibility (as well as demand for flexibility by BMZ) on the other (Int_3, 7). The project bridges extremely different implementation contexts, unlike many other TC measures, which made it very challenging to design a meaningful target system (Int_3, 7). Not only does it span fundamentally different processes; there is also no overarching relationship between the processes that could be monitored (comparable to, for example, a programme level). For this reason, the designers of the project initially discussed the option of setting it up as a fund. This would have meant to explicitly factor in the diversity of implementation contexts by allowing beneficiaries (in this case IOs) to access a given amount of funding for their activities, while designing a target system specifically for their project objectives (under the pre-condition of it fitting the overall project objectives). Due to this set-up, fund systems are generally more flexible than classic TC measures; not least because they can run for a longer time (up to 10 years) than a TC-measure. However, the option of a fund-design did not receive sufficient support and was in the end discarded (Int_3). The resulting target system's weaknesses, as described above, strongly limit its usefulness both as a steering instrument for achieving the project's objectives, and as a monitoring tool to assess any given reform process' progress.

Lastly, looking at the risks identified in the modification offer (2017), the evaluators generally found them to be plausible and appropriately considered in the project's conceptualisation.

With a view to **relevance dimension 4**, the analysis showed that the project adapted well to **changes occurring in its implementation context**. First, the adaption of the project's concept in the modification offer of 2017 was a reaction to a changed focus in the commissioning unit of BMZ (Int_5, 13). Together with the GIZ project team, the objective was to steer the project towards the support of strategic reform processes rather than to serve a series of thematical needs by partners (Int_5, 15, 30).[18] Second, it became apparent in the analysis that the project's entire design was based on the constant adaption to changes in the contexts of the implementation partners (the IOs/network). As the reform processes lay in the partner organisations' hands, GIZ project staff had extremely limited influence on changing context factors. It further faced diverse and changing institutional contexts in each measure. The findings of the analysis showed that its broad and flexible design was the project's main **coping strategy** to deal with these challenges. To illustrate this, to enable the iLTEs to adjust to dynamics within the partner organisation, the project did not define the exact results, which needed to be achieved in advance (Int_3, 10, 19, 20). Furthermore, the project flexibly used the instrument of financing to support products or pilots, which the project team and iLTE's identified as relevant (Int_3, 8). These examples show that in practice, the iLTEs continuously adapted to the institutional contexts to ensure the relevance of their support. Consequently, the selection of competent secondee's for the respective positions was a critical tool for the project to influence its relevance. The following examples exemplify the needed adaption by an iLTE: in the case of ESF support, the iLTE was initially supposed to work on the capacity development strategy for the new ESF (Int_8). However, the political process to agree and finally adapt the ESF Guidance Note for borrowers took longer than expected. As a result, his foreseen task was put on hold. In the meantime, the iLTE identified another task closely related to the original assignment: the development of a tool for project-level capacity needs assessment (Int_8, GIZ 2019e). It became apparent in the interviews that the iLTE, thereby, could ensure the relevance of the support (Int_8, 23, 31, see also 4.3). The iLTE's change in tasks, however, did not require the project design (its target system) to be adapted as the task – the development of a strategy for capacity building – had not been defined in the original project design (GIZ 2017c, 2018, 2019a). This further illustrates the above-mentioned finding, that the project adapted to the complex implementation contexts by leaving its objectives relatively broad.

[18] Prior to the modification offer of 2017, the project had been primarily used as an instrument to second iLTEs to different IOs for different sector projects and BMZ units (see 2.1).

Assessment regarding relevance

As the analysis showed, the project was designed in line with the relevant strategic reference frameworks and pursued the ambitious goal of closing gaps in the solution of global problems, as was intended for projects conceptualised under the ICR title. It was also successful in matching the target group's needs. This is reflected in the high satisfaction of stakeholders with the project's support. BMZ, as an additional shareholder of the project's results, also confirmed the strategic relevance of the supported processes. However, some stakeholders at BMZ also assessed the GDI support measure as not fully in line with the project's logic, being closer to implementation than to a political/IO reform process, and therefore further away from making IOs 'fit' for Agenda 2030 implementation. Nevertheless, the evaluators did not assess this aspect negatively as the analysis suggests that perspectives on this measure vary beyond verifiability for the evaluators. Furthermore, the commissioning party must have initially supported the measure for it to be implemented. This results in a rating of 30 out of 30 points in dimension 1 and 30 out of 30 points in dimension 2. Regarding the project's design, the analysis revealed limitations. The planned link to German development cooperation (Output C) and the support of 'German' proposals[19] (Output A) turned out to be neither relevant nor feasible for reaching the project's objective. However, the overall assumption that reforms needed support to develop concepts for implementation and follow through with pilots remained true. Moreover, the target system was found not to provide an adequate basis for fully depicting the project's results, and hence for results-oriented steering. Apart from these three limitations, the results model and hypotheses were adequately designed for the chosen project objective, resulting in a rating of 10 out of 20 points in relevance dimension 3. The analysis further revealed that to deal with the variety of contexts in which the measures were implemented, the project's design was flexible and broad and adapted well to changing context conditions. Having chosen competent experts who reacted appropriately to changing context conditions, the project managed to stay relevant even through turbulences within the IOs. It is therefore awarded the full score of 20 out of 20 points in dimension 4.

Table 2: Rating of OECD/DAC criterion: relevance

Criterion	Assessment dimension	Score and rating
Relevance	The project design[20] is in line with the relevant strategic reference frameworks.	30 out of 30 points
	The project design matches the needs of the target group(s).	30 out of 30 points
	The project is adequately designed to achieve the chosen project objective.	10 out of 20 points
	The project design was adapted to changes in line with requirements and readapted where applicable.	20 out of 20 points
Overall score and rating		Score: **90 out of 100 points** Rating: Level 2: successful

[19] As opposed to any proposal that was in the reform processes' interest, no matter which member state introduced them.

[20] The project design encompasses project objective and theory of change (ToC = GIZ results model = graphic illustration and narrative results hypotheses) with outputs, activities, instruments and results hypotheses as well as the implementation strategy (e.g. methodological approach, CD strategy, results hypotheses).

4.3 Effectiveness

The criterion effectiveness assesses whether the project has achieved its intended results. The evaluation matrix lists all evaluation questions assessed under the criterion of effectiveness (Annex 1).

Evaluation basis and design for assessing effectiveness

The assessment of effectiveness is structured along three evaluation dimensions. Effectiveness dimension 1 deals with the question whether the project has achieved its objective on time and in accordance with the module indicators (project objective indicators) agreed upon in the offer:

- **Module indicator 1:** four new German proposals have been incorporated into strategy documents for beginning or further developing reforms to strengthen the implementation of the 2030 Agenda in international organisations.
- **Module indicator 2:** eight new stakeholders support reform processes to implement the 2030 Agenda in international organisations (e.g. GDI, UN Reform, World Bank Social and Environmental Standards).
- **Module indicator 3:** four new methods for process design of reforms (e.g. procedures, moderation, coordination, monitoring) have been applied by IOs.
- **Module indicator 4:** two reform measures by IOs (e.g. within the framework of the GDI, World Bank Social and Environmental Standards) have been applied in an exemplary manner (one of which interlock with German bilateral cooperation).

In addition, the evaluation team addressed the shared knowledge interest by the evaluation's stakeholders: How can a results matrix in this type of project generally be helpful for steering towards the project's strategic objectives.

In **effectiveness dimension 2**, the evaluation team analysed how the activities and outputs of the project contributed to the attainment of the project objective. To answer this question, the evaluation team chose a *generative/mechanism approach* as the design for this evaluation. It is best suited under the given circumstances, as it answers the guiding question of why and how the observed impacts were achieved. For the implementation of this approach, the evaluation team conducted a **contribution analysis**, which analyses the extent to which observed (positive or negative) impacts can be related to the project (Mayne 2001).

A contribution analysis seeks to identify alternative explanations that may explain observed impacts and hence does not seek to prove that one factor 'caused' the observed results. In practice, for a contribution analysis, data from various sources is gathered to analyse the causal hypotheses between inputs, outputs, outcomes and impacts formulated in the ToC. The analysis then seeks to construct a credible '**performance story**' to show whether the intervention was a relevant factor, possibly together with other factors, to lead to change (Mayne 2001).

For this evaluation, the results model depicts the project's ToC and builds the basis for analysing the causal hypotheses between inputs, outputs, outcomes and impacts. To test the results logic and to eventually create a 'performance story', the evaluation team distinguished between the project's individual measures. Hence, as part of the inception phase, the data collection possibilities for each measure were discussed and categorised. Furthermore, the project staff defined the project's support to the ESF process, the UN fit for purpose process and the GDI being particularly important (Int_3). The inception mission also showed that data availability was higher for the UN fit for purpose, WB ESF, GDI and IFAD measures, than for the DEVCO and UNDP measures. Therefore, these measures will be emphasised most in the following analysis, while the analysis of the topic-oriented measures is only briefly presented. The following results hypotheses built the basis for the analysis of the reform-oriented measures:

1) The project's support to the reform process contributed to fostering stronger alliances for and trust in the reform process. (Output A)

2) Support to the reform process offered methodological expertise at an opportune moment in the reform process and hence helped enhance the quality of the entire process (Output B)
3) If in the reform process, the project supports (a) the development of concepts for the implementation of reforms, (b) the coordination of concrete steps with all relevant stakeholders, and (c) the exchange of knowledge and experiences with regard to the reform implementation, it can help to ensure the rollout of reforms beyond their adoption. (Output C)
4) If the support to the organisation is provided in a technically competent and flexible manner, which is integrated into the structure of the IOs, the project can create a link to German bilateral development cooperation. (Output C)
5) If the support to the reform process is provided in a technically competent and flexible manner, which is integrated into the IOs' structures, it is able to provide the needed support in Output areas A, B and C and partially ensures ownership by the IOs/the network for the project's contributions. (Outputs A, B & C)

These results hypotheses were adapted to analyse the two DEVCO measures and the UNDP measure (topic-oriented measures) (see Table 5), as the hypotheses presented above are strongly derived from the reform-oriented measures.

In addition to the contribution analysis, the evaluation team complemented the method with **process tracing**. Process tracing focuses on the processes or mechanisms which led to an observed result. It explicitly considers different hypotheses that may explain an observed outcome and, thereby, incentivises an examination of alternative contributing factors. Even though contribution analysis also seeks to identify alternative explanations that may explain observed results and impacts, this aspect in practice often recedes into the background. Hence, the evaluation team used process tracing to ensure that alternative explanations would be sufficiently considered. The following alternative hypotheses were considered in the evaluation of the ESF, UN fit for purpose and GDI measures:

Table 3: Alternatives hypotheses

Output A	1) Not the project's support to the reform process, but the performance of the IO's reform team were critical to fostering stronger alliances for and trust in the reform process. 2) Not the project's support to the reform process, but member states and other relevant stakeholder's ownership and will to reform were critical in fostering stronger alliances for and trust in the reform process.
Output B	3) Not the methodological expertise brought in by the project, but the IO's in-house expertise were critical to enhance the processes' quality. 4) Technical support measures by other member states, rather than those of the project, provided support in a technically competent and flexible manner, leading to methodological proposals covering the needs of the respective IO.
Output C	5) Not the project's support measures, but the respective partner countries' and stakeholder's will and ownership were critical to ensure the rollout of reforms beyond their adoption. 6) Not the project's support measures, but the push and ownership from within the IO were critical to ensure the rollout of reforms beyond their adoption.

Regarding the topic-focused measures, due to the limited scope of the evaluation's report and the limited data availability, the evaluation team decided to not conduct process tracing with adapted alternative hypotheses for the UNDP measure and the two DEVCO measures.

Lastly, **effectiveness dimension 3** addresses unintended positive and negative results. As part of the contribution analysis, the evaluation team assessed whether the project has produced any positive or negative unintended results at outcome or output level and, if so, why. All in all, the evaluation team used the following **data collection methods** to assess the criterion of effectiveness in all three dimensions: (a) document

analysis of project documents (e.g. progress reports, modification offer), other relevant documents (e.g. the reports by iLTEs)[21]; (b) semi-structured interviews (e.g. project staff, staff in international partner organisations and in BMZ divisions). Contribution analysis increases the risk of a bias on behalf of the researcher, as the estimation effect and its causality depend in a greater manner on qualitative considerations. Thus, to counterbalance any limitations in the effectiveness analysis, the evaluation team included several means of **triangulation** in the evaluation design. First, data sources were triangulated where possible. This also included the validation of preliminary insights with GIZ project staff and the target group (IOs) (e.g. through interviews and a debriefing session). Second, the evaluation team ensured researcher triangulation by reflecting results throughout the evaluation phase.

Analysis regarding effectiveness

Effectiveness dimension 1 examines whether the project has achieved its objective on time and in accordance with the indicators agreed upon in the modification offer of 2017. As all indicators were achieved by the end of the project's term, the evaluation team re-examined whether the project had indeed achieved its objective, but focused more strongly on the question whether the way the project measured indicators and counted target values was plausible. During the analysis, the evaluators remarked that they learned very little about each reform process by looking at the target system. As described in section 4.2, this is because the module indicators aggregate results across the different contexts, are output-heavy and count the delivery of products rather than results at the level of reform processes.

This is exemplified by the following analysis, in which the evaluators conducted the standard procedure for analysing effectiveness dimension 1, namely by examining whether the indicators have been achieved as agreed in the project offer. For working with the indicators, the evaluators had already assessed whether they fulfil the Specific, Measurable, Achievable, Realistic and Time-bound (SMART) quality criteria as part of the inception phase (see Table 4). According to this assessment during the inception phase, two of the outcome indicators required adaptions. These two adaptions, which will be presented alongside with all other module indicators in the following analysis, seem to lower the project's level of ambition. However, in the project's reality, the adaptions represent the actual results (see assessment according to SMART criteria in Table 2). In the following examination, it becomes clear that the analysis amounts to 'counting' products rather than examining results.

The **first** (adapted) **indicator (M1)** targeted by the project was: **four new German proposals have been incorporated into strategy documents for setting or further developing reforms to strengthen the implementation of Agenda 2030 in international organisations**. The last progress report shows that the target value was overachieved with five proposals included in strategy documents. This included two results in the UN fit for purpose measure, two results in the WB ESF measure and one result in the DG DEVCO garment initiative measure. In the UN reform process, the project's iLTE supported a pooled funding strategy for UNDP (GIZ 2019a) and preparations for the UN General Assembly Resolution 72/279 (GIZ 2018). At WB, the ESF Guidance Notes for Borrowers were passed in June 2018 (GIZ 2018) and a multi-donor trust fund was designed for capacity development for ESF implementation (GIZ 2019a). At DG DEVCO the iLTE prepared the Commission Staff Working Document, 'Sustainable garment value chains through EU development action' (GIZ 2017c). Therefore, the module indicator as formulated in the modification offer of 2017 was met/overachieved. The **second** (adapted) **indicator (M2)** targeted by the project was that **eight new stakeholders express support and/or agreement to a proposal supported by the project within the framework of the reform processes for implementing the 2030 Agenda in international organisations (e.g. GDI, UN Reform, WB Social and Environmental Standards).** The last progress report shows that the target value for the original indicator was overachieved with nine new stakeholders (GIZ 2019a). Of these nine new stakeholders, according to the project's counting, eight were for GDI (GIZ 2017c, 2018, 2019a) and one support expressed to the DG DEVCO garment initiative, with the OECD adopting guidelines and France adopting a law both in

[21] The main document sources were the project's modification offer of 2017 and the project's progress reports (GIZ 2016a, 2017c, 2018, 2019a).

relation to sustainable supply chains. The analysis showed that most new GDI members not only joined the Initiative, but also provided substantial input to the GDI, for example, at conferences (Int_3). Following this logic of the GDI measure, new members and their content contributions can plausibly be counted as a support and/or contribution to the GDI's objective. Hence, the evaluation team found the measuring of M2 to be plausible.

The **third indicator (M3)** the project aimed to achieve was that **four new methods for process design of reforms (e.g. procedures, moderation, coordination, monitoring) have been applied by IOs**. The last progress report shows that the target value was overachieved with a total of five new methods (GIZ 2019). The counted results included two results in the UN fit for purpose measure, two results in the WB ESF measure and one result in the DG DEVCO garment initiative measure. At the UN, the counted results included an analysis of existing functions and capacities (GIZ 2017c) and the systematic approach for 'SDG Integration' in the programme development of UN agencies (GIZ 2019a). Delivery Labs were implemented as part of the GDI measure (GIZ 2018) and established as a permanent, quarterly offer at WB (GIZ 2019a). In addition, another method included an analysis of weaknesses conducted with the New Partnership for Africa's Development (GIZ 2018). At DG DEVCO, the new method referred to a comprehensive overview of DEVCO's commitment to sustainable value chains in the textile and clothing sector. In conclusion, the evaluation team found the counting of target values for M3 to be plausible.

The **fourth indicator (M4)** targeted by the project was that **two reform measures by IOs (e.g. within the framework of the GDI, WB Social and Environmental Standards) have been applied in an exemplary manner (one of which interlocks with German bilateral cooperation)**. According to the last progress report 2019, the target value was overachieved with a total of four applied reforms. In 2019, the counted results for M4 included (1) the first steps to implementation of UN fit for purpose at UNDP, including, for example, pooled funding; and (2) the piloting of the Capacity Needs Assessment Tool at WB (GIZ 2019a). In 2017 the counted results included (3) an EU project on sustainable garment supply chains in cooperation with GIZ related to the DG DEVCO garment initiative measure; and (4) the implementation of reforms at the UN including a partnership with WB, a flagship initiative on 'gender parity' and a reform of the Departments of Social and Economic Affairs (GIZ 2017c). The evaluation team found the counting of the target values for M4 to be plausible. All in all, the analysis showed that the project plausibly achieved all module indicators.

Table 4: Assessment of module indicators

Module indicators according to the latest modification offer	Achieved / not achieved	Assessment according to SMART criteria/ assessment	Adapted module indicators	Achieved / not achieved
new German proposals have been incorporated into strategy documents for beginning or further developing reforms to strengthen the implementation of the 2030 Agenda in international organisations. Baseline: 0 target value: 4 actual value: 5	achieved	**Relevant:** The indicator is not relevant as it measures the number of German proposals and not the technical proposals supported by the project. Hence, the indicator does not measure the project's contributions clearly enough and suggests that these could also be German political contributions.	technical proposals supported by the project have been incorporated into strategy documents for beginning or further developing reforms to strengthen the implementation of the 2030 Agenda in international organisations. Baseline: 0 target value: 4 actual value: 5	achieved
8 new stakeholders support reform processes to implement the 2030 Agenda in international organisations (e.g. GDI, UN Reform, World Bank Social	achieved	**Measurable:** It is unclear how support for a reform process as a whole can be measured.	8 new stakeholders express support and/or agreement to a proposal supported by the project within the framework of the reform processes for implementing the 2030	achieved

and Environmental Standards). Baseline: 0 target value: 8 actual value: 9			Agenda in international organisations (e.g. GDI, UN Reform, World Bank Social and Environmental Standards). Baseline: 0 target value: 8 (data sources: e.g. interviews with stakeholders in IOs, reports by iLTEs) actual value: 9	
new methods for process design of reforms (e.g. procedures, moderation, coordination, monitoring) have been applied by international organisations. Baseline: 0 target value: 4 actual value: 5	achieved	The indicator is SMART	/	/
2 reform measures by international organisations (e.g. within the framework of the GDI, World Bank Social and Environmental Standards) have been applied in an exemplary manner (1 of which interlock with German bilateral cooperation). Baseline: 0 target value: 2 (1 of which interlock with German bilateral cooperation) actual value: 5	achieved	The indicator is SMART	/	achieved

The results for effectiveness dimension 1, show that all module indicators were achieved. However, the analysis also showed that the achievement of the indicators provided little information on the actual success of the project's support (see also 4.2). As was exemplified by the analysis above, the results counted under each indicator sometimes contain products of a very different nature with different meanings in their respective reform context, therefore making it arbitrary what was counted or not. Indeed, some indicators were reached on the basis of the results achieved in almost one measure alone. Indicator M2's achievement, for example, was mostly based on new GDI members, while other indicators were reached by counting results in several measures. As a result, focusing on the indicators provides little information on the effectiveness of the project. In contrast, the analysis under effectiveness dimension 2 below will reveal the important results created by the project. The target system, however, largely fails to account for those. This weakness is in-built in the target system, as was elaborated in section 4.2.

Two consequences arise from the target system's design. First, as mentioned above, the information acquired from studying the indicators is of limited value. Second – and as a direct result of the first observation – the target system did not allow for strategic steering on its basis but was geared towards ex post monitoring only. The interviews with the project team show that when taking any strategic decision, such as launching a new support instrument in favour of a given process (e.g. the additional grants given to WB ESF), the target system was of no relevance. Rather, the expert opinions of the iLTEs, the assessment of the window of opportunity by the project team and/or BMZ, and close exchange with the host organisation built the foundation for such decisions (Int_3, 15).

The following section analyses how the activities and outputs of the project contributed to the attainment of the project objective, looking beyond the target system (**effectiveness dimension 2**). For this purpose, the

evaluation team discussed the hypotheses along the three fields of action. However, the analysis below will differentiate between the reform processes, telling the individual measures performance stories. To analyse the project's contributions along the fields of action, the evaluators differentiated between reform-oriented and topic-oriented measures. Figure 1 in section 2.2 provides an overview of which measure targeted which output.

Analysis of reform-oriented measures

In **field of action A**, the project aimed to strengthen alliances to support German proposals in IO reforms for implementing the 2030 Agenda. For this purpose, the project tried to support reform processes by contributing to foster alliances for and trust in the reform processes.

Looking at the results in field of action A, it became obvious from the interviews and the document analysis that different kinds of successful alliances were formed. While in the UN fit for purpose process the project contributed to alliances for the support to technical reform proposals, in the case of GDI, building alliances meant gaining new members to strengthen the Initiative as a reform platform. In both the UN and the GDI measures, the building of alliances/gaining new members was essential for moving the reform/the reform platform forward (outcome). As the analysis showed, the project was successful in supporting the UN fit for purpose process and the GDI in this field of action. For the Initiative, insights suggest that already the decision of GIZ to become co-chair built trust in GDI, as GIZ was considered a rather progressive and thoughtful development actor (Int_33). The project could further recruit new members through organising events to address potential partners, for example, the Africa Road Show (GIZ 2017c). In the case of UN, successful support to foster alliances meant providing the needed technical expertise to convince stakeholders of the reform proposals. To provide an example, after the Secretary-General's reform reports[22] were published, the iLTE (as member of the (reform) coordinating team) was in the position to explain (and defend) the proposals. In this context, an interviewee described that the iLTE played a significant role in responding to member states' questions with explanatory notes (Int_12). Contrary to the UN and the GDI measure, the project's support to the ESF process in terms of alliances was not directly related to the supported reform process within the organisation (WB), but focused on an uptake of the ESF *outside* of WB by other institutions (GIZ 2019e). Thus, the results in field of action A in the ESF measure were less essential for further developing the reform (outcome). Instead, the project provided information for similar ESF reforms in other institutions. For example, with a four-day workshop series, the iLTE explained the ESF to partners, such as KfW, who had in parts already decided to use the ESF in their own organisations (GIZ 2019e).

Taking these different types of alliances into account, insights from the interviews further show that the project did not support alliances around *German* proposals (as this was neither aimed for nor within the project's sphere of influence in either of the three contexts) but technical proposals developed within the IOs' teams (Int_4, 10, 28). However, regarding UN fit for purpose, interviews suggest that the iLTE also contributed to the German political support to the reform by providing non-confidential insights into the reform's progress (GIZ 2016–19, Int_3, 4, 10, 28).

All in all, the examination of field of action A showed that the project in fact delivered important results for moving reform processes forward. These results included the recruitment of new members for the GDI, the technical expertise convincing different stakeholders of the UN fit for purpose reform proposals and the creation of partnerships for the ESF. Thus, the hypothesis in field of action A could also be confirmed (see Table 5).

In **field of action B**, the project aimed to provide methodological expertise at the opportune moment in the reform process and hence help to enhance the quality of the entire process.

Regarding the results in field of action B, the analysis showed that the project did indeed provide methodological expertise which helped to further develop the reforms. This became obvious in interviews with

[22] The titles of these two reports are 'Repositioning the United Nations development system to deliver on the 2030 Agenda: ensuring a better future for all' in July 2017 (UN 2017a) and 'Repositioning the United Nations development system to deliver on the 2030 Agenda: our promise for dignity, prosperity and peace on a healthy planet' in December 2017 (UN 2017b).

partners who confirmed that the secondees were able to successfully use their skills to produce products which helped to move the respective reform processes forward (Int_2, 11, 12, 16, 23). It was further underlined by the finding that partner organisations used the project's products and documents and integrated them into the IO's own documents – showing that the provided proposals were indeed useful. In case of UN fit for purpose, the iLTE, for example, directly contributed to the writing of the Secretary-General's reports to the United Nations Economic and Social Council (ECOSOC).[23] Moreover, the secondee developed a technical note on Resident Coordinator funding, which was included as an annex to the General Assembly's resolution 72/270 (GIZ 2018, GIZ 2016–19). Both, the Secretary-General's reports, and the resolution were crucial milestones on the way to further develop the fit for purpose reform. For the ESF measure, the new ESF was already adopted and the iLTE worked as part of the implementation team. In this role, the iLTE, for example, supported the initiation of e-learning material and face-to-face trainings, which are still being used by WB for capacity building (GIZ 2019e, Int_23). These products generally helped to take the new framework from theory into practice. As a new standard is only effective if stakeholders are aware of it and act accordingly, the project's support to capacity building was a central component for contributing to an effective implementation of the reform.

In terms of timing, the interviews confirmed that the project's support came at the opportune moment regarding the UN, the GDI and IFAD measures. For the UN fit for purpose process and IFAD, the project's support came at a critical moment, allowing continuity in the organisation's work on the reforms (Int_12, 16, 24).[24] Similarly, in the case of GDI, interviews suggest that the support of the project for the rather new Initiative was critical at the time (Int_18, 27). In the ESF case, the situation was different. Due to non-foreseeable delays at the political level, the iLTE could not directly start to work on the planned main task, which could, consequently, not be finalised until the end of the project's support (GIZ 2019e, Int_8). Though this changed context condition suggests that the moment for the project support was less opportune, the interviews showed that the iLTE was able to adapt the support accordingly: stakeholders in WB perceived the iLTE's technical support as extremely useful (Int_2, 11, 23, 31) and further expressed that the ESF implementation team needed personnel support and expertise in capacity building at the time (Int_12, 11). This shows that despite the changed context conditions, the iLTE was able to provide contributions which effectively supported the ESF implementation team.

The examination of field of action B revealed that the project succeeded in providing needed methodological expertise to move the supported reforms forward. The UN fit for purpose process, the ESF process and the GDI made significant progress during each measure's support and the project successfully contributed to enhancing the methodological quality of reform documents. Moreover, the project adequately chose the timing of its support. Only in the ESF case, external context conditions made the project's moment of support less opportune. Nonetheless, the project was able to steer accordingly so that the project's contributions still helped to further develop the ESF reform. Thus, all in all, the analysis could also confirm hypothesis 2 (see Table 3).

In **field of action C**, the project aimed to implement reforms on a pilot basis, if possible, with a link to German bilateral development cooperation.

Looking at field of action C, the analysis showed that the project was able to generate effective results to support reform implementation. In the ESF process, the iLTE worked directly with the ESF implementation unit and the project-financed pilots in countries[25] to test the capacity needs assessment tool developed by the iLTE (GIZ 2019e, WB 2020c). Thus, with the project's help, one reform component could directly be taken to implementation. In addition to this piloting of reform steps, the project produced different types of products that supported the partners in moving towards reform implementation: for the UN, the analysis showed that the reform is still ongoing and the negotiation of and the design for the reform proposals (Output A and B) took up

[23] Ibid.
[24] In case of UN fit for purpose this refers to the preparation for the Quadrennial Comprehensive Policy Review. In case of IFAD this refers to the preparation of IFAD's risk management system.
[25] The pilots were implemented in Ruanda, Sierra Leone, Somalia, Tajikistan, Peru, Marokko, Bangladesh and Nepal (GIZ 2019a).

most of the secondee's time (GIZ 2016–19). However, as advisor to UNDP,[26] the iLTE supported first preparations of rollouts on a technical level. The secondee, for example, prepared the drafting of a pooled funding strategy for the agency and developed briefing kits for country offices for its rollout (GIZ 2016–19, GIZ 2018). An interview partner confirmed that this support was helpful to prepare implementation at UNDP (Int_22).

Reform implementation does not apply in the same sense to GDI. The project did not support a reform process, but a 'reform platform'. Hence, no rollout of 'reforms' beyond their adoption was required. Rather, in this measure, implementation referred to the piloting and use of methods and tools developed by and with GDI members. This included, for example, the financing of 11 Delivery Labs (WB 2020a) or the piloting of Delivery Labs at a conference in Addis Ababa (GIZ 2018).[27] All in all, due to the limited contributions in terms of rollouts regarding UN and GDI measures, the analysis only partially confirmed hypothesis 3 (see Table 5). The analysis could confirm aspects 1 and 2 of the hypothesis underlying field of action C, which stated that if in the reform process, the project supports (1) the development of concepts for the implementation of reforms; and (2) the coordination of concrete steps with all relevant stakeholders. However, aspect (3) – the exchange of knowledge and experiences with regard to the reform implementation to ensure the rollout of reforms beyond their adoption – could not be fully confirmed. Furthermore, due to the context of the UN and GDI measure, implementation of reforms was less relevant in these measures at the time.

The second hypothesis underlying field of action C (hypothesis 4) related to the project's objective to create a link to German bilateral development cooperation when implementing reforms on a pilot basis. As already stated in section 4.2, this link turned out to be a challenging goal for the project and could not be achieved in the reform measures,[28] and the project team, therefore, focused on general implementation support (Int_5).

Last, across all three fields of action, the project aimed to provide its support in a technically competent and flexible manner integrated into the IOs' structures. These objectives were preconditions for creating effective results in all fields of action: the project directly supported IO teams in very dynamic institutional and political contexts. Thus, the iLTEs' technical competence, their flexibility to react adequately to changing contexts and evolving needs as well as their integration into the IO teams were necessary for enabling direct support to the development of reforms.

Regarding the iLTEs' technical competence, it became obvious in the interviews with partners that the secondee's had the required expertise to contribute to the respective process (Int_2, 11, 12, 16, 23). The findings further showed the secondees' flexibility as they continuously adapted to the requirements of changing contexts and of their teams. As described in field of action B, for the ESF measure, the iLTE adapted to unexpected political delays and identified a different product than was initially intended (a project-level capacity needs assessment tool), which interviewees at WB rated as very helpful (Int_23, 31). Furthermore, the achieved results in the UN measure (e.g. technical note for resident coordinator funding) suggest that the iLTE was able to react flexibly to changing tasks as the reform process evolved between 2016 and 2019. Furthermore, the analysis also showed that the secondees were well integrated into the IO teams. In fact, the interviewees in WB and UN described the iLTEs as full members of their teams (Int_2, 11, 12, 22, 23). Moreover, the analysis showed an additional factor which ensured ownership for the project's contributions: from the interviewees' perspective, the adoption and acceptance of the iLTEs' products and contributions was possible because of their 'neutrality' – from the target group's point of view the secondees did not push for a *German* political agenda, but provided their support under technical aspects only (Int_2, 11, 12). All in all, the project managed to meet the three preconditions, which further explains the achieved results for the different fields of action. Thus, the analysis could also confirm hypothesis 5 (see Table 3).

[26] The iLTE was advisor at UNDP from August 2018 to January 2020.
[27] Interviews further revealed that the project initially aimed to initiate a broader event series of Delivery Labs (Int_5). According to the project team, the failing of this idea was also linked to the aversion of several stakeholders to address their delivery challenges publicly (Int_5).
[28] Only the topic-oriented measure at DG DEVCO (garment initiative) established a link.

Table 5: Assessment of results hypotheses (reform-oriented measures)

#	Hypotheses	hypotheses for reform-oriented measures	Data sources
1	The project's support to the reform process contributed to fostering stronger alliances for and trust in the reform process. (Output A)	confirmed	modification offer, progress reports, reports of iLTEs, interviews
2	Support to the reform process offered methodological expertise at the opportune moment in the reform process and hence helped enhance the quality of the entire process. (Output B)	confirmed	modification offer, progress reports, reports of iLTEs, interviews
3	If in the reform process, the project supports (a) the development of concepts for the implementation of reforms, (b) the coordination of concrete steps with all relevant stakeholders and (c) the exchange of knowledge and experiences with regard to the reform implementation, it can help to ensure the rollout of reforms beyond their adoption. (Output C)	partially confirmed (reason: only partially confirmed for UN fit for purpose and GDI measure)	modification offer, progress reports, reports of iLTEs, WB completion report of EFOs, interviews
4	If the support to the organisation is provided in a technically competent and flexible manner, which is integrated into the IOs structures, the project can create a link to German bilateral development cooperation. (Output C)	not confirmed (reason: link to German bilateral cooperation could not be established)	modification offer, progress reports, reports of iLTEs, interviews
5	If the support to the reform process is provided in a technically competent and flexible manner, which is integrated into the IOs structures, it is able to provide the needed support in Output areas A, B and C by also ensuring ownership by the IOs/the network for the project's contributions. (Output A, B & C)	confirmed	modification offer, progress reports, reports of iLTEs, WB completion report of EFOs interviews

In addition to this contribution analysis, the evaluation team also included process tracing to assess hypotheses of alternative explanations for the results achieved in each field of action. The evaluators considered these hypotheses by creating counterfactual situations with the interview partners, thinking through what might have happened without the project's support. The interviewees' descriptions suggest that the project's support was critical for achieving the results in each measure and could not have been replaced with a different instrument (e.g. only financing). For the UN measure, according to one stakeholder, if the project had not seconded an iLTE, it would not have been possible to find a special advisor with the right skill set on time (Int_16). For the GDI measure, all interview partners emphasised the critical role of the project's support in strengthening the Initiative (Int_6, 26, 27, 33). In the case of the ESF measure, interview partners said that although WB might have developed a multi-donor trust fund and a tool for capacity needs assessment at some point, the project's support led to a faster achievement of the results (Int_8, 23). One interview partner further highlighted that the team could not have done it without the iLTE (Int_11). Similarly, for the IFAD measure, the partner organisation stated that even though there might have been eventual financing from a different source, the project's support avoided further prolonging the development (Int_24). Nonetheless, considering the project's design, it is apparent that both the project's support and the IOs teams' (and the GDI Secretariat's) ownership and performance were critical for any reform progress (see Table 4). The analysis further showed that no other member state directly supported the respective tasks in the reform teams. Thus, hypothesis 4 could not be confirmed (see Table 4). All in all, the insights suggest that each measure was a critical support to the partner organisations.

Table 6: Assessment of alternative hypotheses (reform-oriented measures)

#	Alternative hypotheses	UN measure	ESF measure	GDI measure
1	Not the project's support to the reform process, but the performance of the IO's reform team were critical to fostering stronger alliances for and trust in the reform process	both: project and IO staff	confirmed	both: project support and GDI Secretariat
2	Not the project's support to the reform process, but member states and other relevant stakeholder's ownership and will to reform were critical in fostering stronger alliances for and trust in the reform process	both: project and member states	confirmed	not applicable
3	Not the methodological expertise brought in by the project, but the IO's in-house expertise were critical to enhance the processes' quality	both: project and IO	both: project support and IO staff	both: project support and GDI Secretariat
4	Technical support measures by other member states, rather than those of the project, provided support in a technically competent and flexible manner, leading to methodological proposals covering the needs of the respective IO	not confirmed	not confirmed	not applicable
5	Not the project's support measures, but the respective will and ownership of partner countries and stakeholders were critical to ensure the rollout of reforms beyond their adoption	not applicable	both: project support and IO	not applicable
6	Not the project's support measures, but the push and ownership from within the IO were critical to ensure the rollout of reforms beyond their adoption	confirmed	both: project support and IO	not applicable

Analysis of topic-oriented measures

The following section analyses the three topic-oriented measures at UNDP and DG DEVCO. As described in section 2.1, these measures did not follow the results logic of the modification offer of 2017. Nonetheless, the three measures can be understood along the fields of action A, B and C. The results hypotheses were slightly adapted (see Table 5).[29] As the evaluation put a focus on the UN fit for purpose, the ESF and the GDI measures, the following analysis is only briefly presented for each measure. The evaluation team also focused on the period under evaluation (2016-2019) and, hence, did not consider the support of the iLTE prior to 2016.

- **UNDP – topic SDG 16.** At the time of the project's support, UNDP was the lead agency on evaluating governance (Int_17). From the partner's perspective, the measuring of SDG 16, the main task of the iLTE, was, thus, of strategic importance for the agency (Int_17). The interviewee in UNDP concluded that the secondee's work was of high importance to the organisation (Int_17). All in all, the analysis shows that the project successfully supported UNDP in the fields of action A and B.[30]
- **DG DEVCO – garment initiative.** Overall, the analysis revealed that the project's iLTE generated effective results in all three fields of action. In field of action A, the iLTE contributed to fostering alliances for the topic with stakeholders within the EU and member states (Int_25, GIZ 2017b). In terms of fields of action B and C, Unit C4 described that the iLTE's support to personnel and expertise covered the needs of the team at the time (Int_25). The findings of the analysis also suggest that the technically competent and flexible support integrated into the IO's structures ensured IO ownership of the developed documents (Int_20, 25).
- **DG DEVCO budget and taxes.** The limited insights[31] in this measure suggest that the iLTE provided useful contributions in fields of action A and B. With the contribution to 'Collect more spend better', the iLTE shaped a strategic document on the topic (GIZ 2017d, Int_19). With regard to fostering alliances, the

[29] The parts focusing on reforms were replaced by the term topic and the hypothesis which referred to reform implementation was excluded.
[30] The documents available for the evaluation included the reports by the consultant who also conducted interviews with three stakeholders form UNDP and two stakeholders from StäV on the project's support to UNDP for the progress report 2016.
[31] As the evaluation team could not conduct interviews with stakeholders in the partner organisation, the analysis of the support is based on an interview with the iLTE and the document analysis including the assessment of the iLTE's results by the project's consultant.

iLTE supported close cooperation between actors such as DEVCO and BMZ (Int_19). All in all, the available data also suggests that the project's support helped to anchor BMZ's Good Financial Governance approach within DEVCO (GIZ 2017d, Int_19).

Table 5 shows the four hypotheses for the topic-oriented measures, which the evaluation team assessed as part of the contribution analysis.[32]

Table 7: Assessment of results hypotheses (topic-oriented measures)

#	Adapted hypotheses	UNDP-SDG 16	DG DEVCO garment initiative	DG DEVCO budget and taxes
1	The project's support contributed to fostering stronger alliances for the topic. (Output A)	confirmed	confirmed	confirmed
2	Support to the IO offered methodological expertise at the opportune moment and hence helped enhance the quality of the entire process. (Output B)	confirmed	confirmed	confirmed
3	If the support to IO is provided in a technically competent and flexible manner, which is integrated into the IOs structures, the project can provide the support needed and create a link to German bilateral development cooperation. (Output C)	not confirmed	confirmed	not sufficient evidence
4	If the support is provided in a technically competent and flexible manner, which is integrated into the IOs structures, it is able to provide the needed support by also ensuring ownership by the IO for the project's contributions. (Output A, B and C)	confirmed	confirmed	not sufficient evidence

In **dimension 3** of the effectiveness criterion, the evaluation team analysed whether the project produced any **unintended positive or negative results**. Generally speaking, the project's design was so broad that unintended results could hardly occur. The entire project design aimed to provide flexible support. Accordingly, the evaluation did not reveal any unintended positive results and interview partners did not report any negative consequences for the different measures. Last, the evaluation's findings also suggest that risks were appropriately taken into account by the project team. Most risks did not materialise or were adequately dealt with.

Assessment regarding effectiveness

In terms of effectiveness, the evaluation team assessed the project as successful (85 of 100 points). It achieved all outcome-level indicators and most output-level indicators by the end of the project. However, the evaluation showed that the project's target system was not very useful for assessing the effectiveness of individual reform processes. Hence, the target system was not practical either as a tool for strategically steering the project. All in all, the analysis revealed that measures of the nature discussed here need a different target system design that allows insights into the diverse supported processes. The evaluators had to take this into account when rating effectiveness dimension 1, and attributed 25 out of 40 points.

The findings of the analysis for dimension 2 show that the project's support was generally highly effective. The individual measures all made important contributions to their respective reform processes that were rated as essential by the IOs (and the GDI network). The counterfactual situations discussed as part of the process tracing also showed that the project's contributions could not easily have been replaced by other support measures. All in all, the effective dimension 2 of the project resulted in 30 out of 30 points being awarded. Due to the project's flexible design, no unintended positive results occurred. Further, the evaluation did not identify any negative results. Considering that the project also took potential risks into account, the evaluators attribute 30 out of 30 point in effectiveness dimension 3.

[32] The evaluation team used the following data sources for the analysis: interviews with the iLTEs and stakeholders in the partner organisation, the iLTEs reports, the project progress reports and the report of the consultant for the progress report 2016.

Table 8: Rating of OECD/DAC criterion: effectiveness

Criterion	Assessment dimension	Score and rating
Effectiveness	The project achieved the objective (outcome) on time in accordance with the project objective indicators.	25 out of 40 points
	The activities and outputs of the project contributed substantially to the project objective achievement (outcome).	30 out of 30 points
	No project-related (unintended) negative results have occurred – and if any negative results occurred, the project responded adequately. The occurrence of additional (not formally agreed) positive results has been monitored and additional opportunities for further positive results have been seized.	30 out of 30 points
Overall score and rating		Score: **85 out of 100 points** Rating: Level 2: successful)

4.4 Impact

The criterion impact measures the extent to which objectives of a project contributed to reaching overarching development results.

Evaluation basis and design for assessing impact

In **impact dimension 1**, the evaluation team analysed to what extent the intended overarching development results have occurred or are expected to occur. The project's main objective was that IOs would, in the long run, complete their reforms (Im_1) and implement them (Im_2). The assumption was that as a consequence of this support, IOs would effectively manage the implementation of the 2030 Agenda (Im_3), which would then result in partner countries receiving more effective support in implementing it (Im_4). Besides these main impacts, the project also aimed to generate an increased uptake of German implementation expertise (Im_6) and, as a direct result, meant to strengthen the cooperation between bilateral projects and IOs (Im_7). Moreover, the project's contributions also intended to lead to a more positive perception of Germany's roles in supporting IO reforms (Im_5).

As described in section 2.1, the project is a stand-alone measure. For this reason, there is no general programme objective against which the project can be measured. **Development Assistance Committee** (**DAC)/BMZ identifiers** were not assigned to this project and consequently cannot serve as reference points for assessing the criterion. The analysis therefore focused on identifying the extent to which the intended impacts illustrated in the results model, as well as additional impacts, have occurred.

As first findings during the inception phase already indicated, the supported reforms are partially still ongoing. Hence, to assess to what extent the intended impacts have been achieved, the evaluation team identified the status quo of each supported reform process. Due to the scope of the report, the analysis will provide examples to illustrate the main overall findings regarding all measures. The intended impacts also directly refer to the **SDGs**. The evaluation team already assessed the relevance of the supported processes and topics to the 2030 Agenda in section 4.2. The evaluators also presented the potential of the supported reforms for making IOs 'fit' for the 2030 Agenda. Thus, the following analysis of impact will not repeat those examples.

Additionally, impact dimension 1 dealt with the question how the ICR criteria contributed to strengthening overarching development results.

In **impact dimension 2**, the evaluation team assessed if and how the project contributed to the intended overarching development results. For this, the evaluators focused on the underlying hypotheses relating to the impact level in the project's ToC (see 2.2). These hypotheses were assessed with the help of a plausibility analysis on the project's contribution to achieving the intended impacts in the short and medium term. Additionally, as part of ICR-specific questions, the evaluation team also analysed to what extent the project contributed to innovations.

In **impact dimension 3**, the evaluation team also analysed if the project produced any positive or negative unintended results at the impact level, and how it dealt with these. As data collection methods for analysing the impact criterion, the evaluation team used (1) document analysis (e.g. progress reports, reports of iLTE) and (2) semi-structured interviews (e.g. with project staff, BMZ, staff of IOs, DIE, StäV, members of GDI).

Analysis regarding impact

First, the evaluation team analysed how the **ICR criteria** affected the project's achievements. GIZ projects under the ICR title had to follow a cross-regional, cross-sectoral and multi-stakeholder approach which aims to build strategic partnerships and to strengthen the international networking of relevant actors to solve global challenges. Based on interviews with the project team, the ICR criteria affected the project on several levels. Most importantly, the title enabled the overall project design as a measure which flexibly supported strategic reform processes (Int_3, 14). From an interview partner's perspective, this cross-sectoral strategic approach, which took a systemic perspective and supported reforms beyond 'topics', was a valuable design (Int_30, 29). Thus, it became apparent that this type of project design and the achieved results could not have been realised without the ICR title.

Second, the evaluators turned to analyse the results at impact level. To assess these results, the evaluators differentiated between the individual measures. Therefore, the following section provides an overview of the status quo of the supported processes (**impact dimension 1**):

- **UN fit for purpose process:** The UN reform is ongoing. Insights from interviews suggest that it is the most ambitious reform the UN has gone through up to this point (Int_12, 34). Most strands of the comprehensive reform have been agreed at the political level, while some strands were still being discussed at the time of the evaluation (Int_10, 34). Generally, from an interview partner's perspective, it is too early to judge the reform's outcome (Int_12, 22, 34). Nonetheless, stakeholders identified first positive results and described the high potential of the reform process (Int_4, 10, 34). As an example, one interviewee stated that with the decision for implementation, less fragmentation and more cooperation within the UN system can already be observed. This increased cooperation makes the implementation of the 2030 Agenda by the UN more likely (Int_34).
- **ESF process:** The supported implementation of the new Environmental and Social Framework is ongoing. It became apparent in the interviews that the capacity-building workstream supported by the project is a long-term exercise (Int_2, 31). The rollout continues with, for example, training sessions on the ESF for new staff (Int_23). The analysis further suggested that the progress on capacity building also depends on future financing – a multi-donor trust fund for financing capacity building is still being set up (Int_29).
- **GDI**: As described in section 4.3, the GDI measure did not directly support a reform process. It was rather a 'platform' for reforms which also aimed to spread methods of analysis and enhance the quality of implementation of its members. It became obvious in the interviews that the GDI has grown to a global partnership with a diversity of members (Int_6, 15). The analysis further showed that the Initiative is facing challenges regarding its future financing (Int_3, 18, 33, 35).
- **IFAD**: The Risk Management Unit continues to develop IFAD's risk management (Int_24).

- **UNDP:** In terms of SDG 16, UNDP had taken over the interim custodianship for three indicators for SDG 16. According to the interviewee at the partner organisation, the statistical methodology for the indicators was finalised and data collection methods were tested in pilot countries at the time of the evaluation. Additionally, the data warehouse was being built to collect data for each country for the Global SDG database (Int_17). These insights show that, overall, the objectives in terms of UNDP's role for strengthening SDG 16 could be reached.
- **DG DEVCO – garment initiative**: From the partner's perspective, since the project's support, the EU discussion on value chains has moved from a phase of analysis to a phase of regulation (Int_25). After the Commission Staff Working Paper, the DG Justice and Consumers initiated a study looking at a regulatory approach for different value chains. This study was still being finalised during the evaluation mission. From the partner Unit's point of view, the proposal of a regulatory approach on sustainable supply chains by the Commission is likely in the foreseeable future (Int_25).
- **DG DEVCO – budget & taxes:** Due to limited data availability, results at impact level could not be analysed in more detail for this measure.

This overview of the status quo shows that all assessable partner organisations made significant progress regarding the supported reforms and topics. In a next step, the evaluation team turned to analyse how the project contributed to the achieved progress (**impact dimension 2**). For this purpose, the plausibility analysis focused on the underlying hypotheses related to the impact level in the project's theory of change (see 2.2). According to the insights gained from interviews and document analysis, **Impact hypothesis 1** is generally plausible: **If reform processes are further developed (outcome), reforms will ultimately be completed (Im_1) and implemented (Im_2), which will then enable IOs to effectively support the implementation of the 2030 Agenda (Im_3)**. The analysis for the effectiveness criterion (see 4.3) showed that the project could provide significant support for further developing the reforms. The description of the status quo in impact dimension 1 further illustrated that these reforms made critical progress. Hence, it is plausible that the project's support in the long term could have contributed to the completion (Im_1) and implementation (Im_2) of reforms. However, regarding Im_3 of the hypothesis, it is too early to assess its plausibility. As the overview in impact dimension 1 (see also sections 2.2 and 4.6) showed, a variety of risks could still reverse the progress made.

Regarding **impact hypotheses 2**, the analysis showed that the hypothesis is only plausible in part. **The introduction of methodological proposals (Output B) and the strengthening of alliances to support German proposals in IOs (Output A) leads to a more positive perception of Germany's role in supporting IO reforms (Im_5).** Though reform processes are not yet finalised, it became apparent in interviews that supported unit's positively perceived role of Germany in supporting the respective processes and topics with technical contributions (Int_2, 11, 12, 16, 17, 22, 25). Contrary to the second part of the underlying hypothesis (Output A), the data from interviews also showed that the target group perceived Germany's role positively due to the 'neutrality' of its support, which did not aim to push for a German political agenda (Int_2, 12, 25). The positive perception, thus, relates to the finding that iLTEs were successful in becoming part of the organisations' teams and responded to partner needs with technical proposals (Output B) (see also 4.3). These insights suggest that the clear differentiation between the political and the technical levels (see 4.3) contributed to a positive perception of Germany's role.[33] Taking these findings into account, the limitation regarding the hypothesis is apparent – the project's support was positively perceived by interview partners who directly worked with the iLTEs. However, considering the finding that interviewees valued the 'neutrality' of the support, it seems plausible that the iLTEs were not perceived as a 'German' support outside of their direct units and teams. Due to this finding, the analysis showed that project could not directly contribute to a wider and broad, positive perception of 'Germany's' role in the process (Im_5). It further suggests that to reach a wider positive perception of Germany's role in supporting the processes, political action and commitment was needed. The project could, however, contribute to this impact by supporting the German political level with helpful technical information for the UN fit for purpose measure (see 4.3). The analysis in

[33] This deficiency of Output A – the focus on German proposals – has already been described in sections 4.2 and 4.3.

section 4.3 showed that this interplay of the political and the project levels was successful. However, an analysis of the wider perception of Germany's role in supporting the processes in WB and UN exceeded the scope of this evaluation.

Lastly, the evaluation team could not confirm impact hypothesis 3: The implementation of reforms on a pilot basis with, if possible, German bilateral development cooperation (Output C) leads to an increased uptake of German implementation expertise (Im_6). In addition, and as a direct result, it further strengthens the cooperation between bilateral projects and IOs (Im_7). As described in sections 4.2 and 4.3, the project faced a range of systemic challenges regarding the link to German bilateral development cooperation. The challenges included, for example, the different programming cycles of multilateral partners and the bilateral level (Int_3) or the reluctance of bilateral projects to share information on failures and challenges.[34] The project team, hence, focused on the implementation aspect of Output C (see also section 4.2). For this reason, the evaluation could not confirm the second part of the hypothesis (Im_7) and its connection to Output C – implementation of the reforms on a pilot basis was not realised with the help of German bilateral development cooperation.[35] Consequently it did not lead to a strengthened cooperation between bilateral projects and IOs. Likewise, as no reforms were implemented with bilateral development cooperation (Output C), this planned Output could not lead to an increased uptake of German implementation expertise.

Despite these findings, the analysis showed that the project did contribute to the uptake of German implementation expertise (Imp_7) with the following two contributions: (1) The implementation of Delivery Labs as one method introduced in the GDI measure (Int_5, 15). (2) The approach to capacity building introduced as part of the support to the ESF process (Int_8).

The plausibility analysis of all the three intended impact hypotheses suggests the following: The project contributed to the achieved progress directly connected to the completion of reforms (hypothesis 1). With a view to the project's objective, this hypothesis is also the most important for the project's overall success. The other intended impacts, regarding the positive perception of Germany's role and the link to German bilateral projects, link to Output A and C. The evaluation already discussed the limitations regarding the project's activities for these outputs in 4.2 and 4.3. In the plausibility analysis of this section, the same limitations became apparent – the missing link to the bilateral level and the limited contribution of the project for 'German' proposals and, hence, 'Germany's perception'. It became obvious in the analysis (sections 4.2, 4.3, 4.5) that both these limitations were in practice considered by the project. For this reason, the two impact hypotheses, which were only partially or not confirmed by the analysis, corresponded less to the reality of the specific measures.

As part of impact dimension 2, the evaluation team also analysed the ICR-specific question to what extent the project contributed to **innovations**. Generally, as described in impact dimension 1 regarding ICR criteria, it became apparent that the project's approach in itself was innovative (Int_3). Moreover, regarding the individual measures, the analysis showed that especially GDI introduced innovative approaches. This included the introduction of Delivery Labs, but also the overall approach of GDI, supported by the project, which focuses on delivery challenges and created a conversation around implementation among its members (Int_3, 14, 18). Finally, it should be mentioned that no project-related negative or positive results at impact level have been observed. As described in section 2.2, due to the project's design most risks were outside of the project's sphere of influence and were adequately considered by the project team.

[34] The reluctance to share failures publicly was one reason why a broad rollout of the GDI Delivery Labs connecting different stakeholder could not be realised (Int_5).

[35] The only link created to German bilateral development cooperation, which was monitored during the project's term, was created at EU level, which was not a reform-oriented measure. In the case of the EU DEVCO measure, the iLTE contributed to establishing a link to the EU development project for sustainable clothing chains. This new EU development project for sustainable clothing supply chains was adopted in December 2016. It supported the Vision Zero Fund which was an initiative of the German government during its G7 Presidency in 2015. The project document directly referred to the experiences of the German TC and proposed to include GIZ in the evaluation and monitoring of the Vision Zero Fund (GIZ 2017e, GIZ 2017b).

Assessment regarding impact

Regarding the **impact** criterion, the intended impacts were not yet achieved. This is mainly due to the long-term nature of institutional reforms. Reform processes are ongoing and implementation in different organisational settings has, in parts, just started. Nonetheless, the analysis showed that major progress was made in each of the measures and that the project's results plausibly contributed to the main intended impacts relating to a successful implementation of reforms for the 2030 Agenda in the long run (Im_1–3, hypothesis 1). Considering the positive perspective of Germany's role, the analysis showed that the project could contribute to this impact goal only to a limited extent. The intended impacts regarding the link to German bilateral development cooperation did not turn out to be feasible. Despite these described limitations, the evaluation showed that the project did contribute to the achieved progress towards reaching the main directly intended impact: the completion of reform processes. This results in a rating of 35 out of 40 points in impact dimension 1 and 25 out of 30 points in impact dimension 2. Negative results or unintended positive results at impact level have not occurred. Considering that the realisation of the impacts is long term and depends on partner organisations, it is plausible that the project team did not systematically monitor unintended results at impact level. For this reason, the evaluation team rated impact dimension 3 with 30 of 30 points.

Table 9: Rating of OECD/DAC criterion: impact

Criterion	Assessment dimension	Score and rating
Impact	The intended overarching development results have occurred or are foreseen (plausible reasons).	35 out of 40 points
	The outcome of the project contributed to the occurred or foreseen overarching development results.	25 out of 30 points
	No project-related (unintended) negative results at impact level have occurred – and if any negative results occurred, the project responded adequately. The occurrence of additional (not formally agreed) positive results at impact level has been monitored and additional opportunities for further positive results have been seized.	30 out of 30 points
Overall score and rating		Score: **90 out of 100 points** Rating: Level 2: successful)

4.5 Efficiency

The criterion efficiency measures the extent to which objectives of an intervention have been achieved cost-effectively. In the context of this evaluation, this is not to be understood in the sense of austerity – trying to spend as little budget as possible – but rather from a perspective of yield maximisation: an intervention is efficient when a maximum of results is achieved with the available financial resources. The evaluation matrix in Annex 1 details the evaluation questions assessed under the criterion of efficiency.

Evaluation basis and design for assessing efficiency

An efficiency analysis can generally be done at two levels: **production efficiency**, which measures the transformation of inputs to outputs, and **allocation efficiency**, which measures the transformation of inputs to outcomes or impacts.

When analysing the project's production efficiency (**efficiency dimension 1**), the principle of yield maximisation is applied. This principle analyses the extent to which the results achieved seem appropriate

given the resources invested. Furthermore, it asks how even more results could have been achieved with the same financial means. The objective is, thus, not to analyse whether the project could have cost less, but whether it used all opportunities to maximise results with the resources available. In analysing this dimension, the evaluation team applied the 'follow-the-money' approach, using the GIZ efficiency tool provided by their evaluation unit. Together with the project team, all costs were attributed to the project's three outputs with the help of this tool. However, it became apparent during this exercise that an allocation of costs to outputs can only give limited information about the efficiency of the project. Due to its unique set-up of working with individual measures, looking at the amount of resources invested in a given output allows for very limited conclusions as to the results achieved in a specific reform process. This is also due to the set-up of the target system and indicators, as will be elaborated in the following analysis (see also sections 4.2 and 4.3). Therefore, it was planned in the inception phase (see Inception Report) that costs would in addition be allocated to individual measures. However, the project team realised that this exercise would create a very high level of effort and workload, as a cost-to-individual-measure allocation was not readily available. Just as for the efficiency tool, every cost would have to be tracked 'backwards' to the individual measures. As it seemed unrealistic to complete this exercise in the given time frame, the project team decided in accordance with the evaluation team not to attribute costs to individual measures.

To counteract this shortcoming, the following analysis will complement the cost-output-analysis with some qualitative observations from section 4.3 (observed outputs specific to the reform process that were not measured by the module indicators), assessing them in relation to the resources invested. For this, one should bear in mind that no overview of costs per individual measure exists, as just explained. Therefore, the analysis will offer a reflection on the appropriateness of the instruments (iLTEs, financing) used rather than their cost. Importantly, for the type of project under examination here, there are no standards or benchmarks that allow a 'robust' assessment of whether alternative allocation decisions could have led to yield maximisation. Therefore, the assessment is primarily of a qualitative nature and draws on plausibility analyses.

Allocation efficiency (**efficiency dimension 2**) measures the transformation of inputs to outcomes or impacts. As already underlined in the Inception Report, the long causal chain depicted in the ToC from the inputs up to the impact level makes it difficult to analyse allocation efficiency, as many other intervening factors will likely have influenced observed outcomes and impacts. Therefore, the evaluation team will focus on the outcome level. However, the challenges mentioned under efficiency dimension 1 equally apply to efficiency dimension 2. When looking at the investment of resources in relation to the outcomes achieved by use of the module indicators, the evaluation team can draw no specific conclusions for the individual reform processes. To cope with this challenge, the analysis will use the same mechanism as for efficiency dimension 1 – drawing on some of the qualitative observations from section 4.3 (observed outcomes specific to the reform process that were not measured by the module indicators) to assess them in relation to the resources invested. Again, this will be restricted to the appropriateness of the instruments used. Furthermore, special emphasis was put on the project's internal steering processes to analyse whether they were appropriately geared towards the desired results at outcome level.

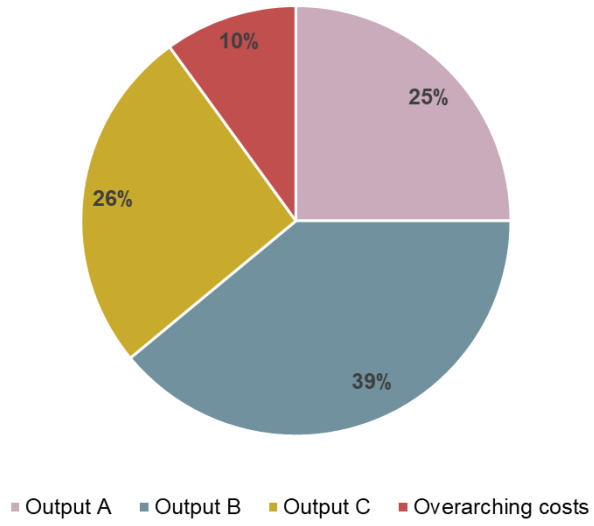
The evaluation of the efficiency criterion is based on the analysis of cost data by means of the GIZ efficiency tool, the intervention's progress reports, operational plans and steering structure as well as qualitative interviews with the project staff, partners and (external) stakeholders.

Analysis regarding efficiency

Beginning with **efficiency dimension 1**, the project's production efficiency, the analysis applied the follow-the-money approach using the efficiency tool for analysis. Taking into account all costs and attributing them to the respective outputs or overarching costs, the results of the analysis show that about 25% (approx. EUR 1.25 million) of the resources were invested in output A (alliances to support reform proposals), 39% (approx. EUR 1.95 million) of resources into output B (methodological proposals to support reform process) and 26%

(approx. EUR 1.3 million) in output C (implementation on pilot basis). Around 10% (approx. EUR 500,000) of the project's costs occurred as overarching costs (see Figure 2).[36] This distribution of costs is also roughly in line with the estimation of costs presented in the modification offer of 2017.

Figure 2: Distribution of costs by outputs and overarching costs



Looking at the latter category first, the investment of 10% of the project's total budget into **overarching costs** can be considered an average/normal relation to the outputs in the context of GIZ projects. The overarching costs of this project were mainly made up by administrative costs (especially time sheets), procurement of office supplies, as well as some travel and personal costs. For the latter, the analysis shows that only the project staff located in Germany (and not seconded to a reform process) assured the project's administrative procedures and hence incurred overarching costs, so that the seconded experts fully committed their time to working on outputs. In addition to handling the administrative procedures, the project team in Germany also contributed to the project's outputs.

Generally, the amount of overarching costs suggests that an appropriate amount of resources was invested in the outputs. Two factors may have benefited this good ratio. First, a good proportion of travel costs as well as some administrative costs were allocated to the outputs in the analysis, as they were considered necessary conditions to enable the iLTE's secondment or were otherwise related to the production of outputs (Int_3, 14). Second, the project benefited from partner contributions that also helped keep its overarching costs in check: office-related costs of iLTEs were naturally taken on by the IO that hosted the expert. Furthermore, for the GDI, the World Bank hosted the Secretariat and also contributed resources to the Initiative, increasing the amount of resources available for the project's outputs.

Looking next at the **costs invested into outputs**, these can be discussed from two perspectives. First, looking at the project's overall logic and ToC, the analysis can reflect whether the distribution of resources seems coherent from this point of view. Second, the analysis can take into account the results achieved in relation to the costs incurred. The following analysis will be structured along this logic.

Considering the ratio of costs invested in the three outputs, it immediately becomes apparent that output B (methodological proposals to support reform processes) received the highest investment of resources (39%), while outputs A (alliances to support reform proposals) and C (implementation on pilot basis) were allocated about a quarter of the project's budget each. Reflecting on the project's ToC, it seems coherent that the project invested most resources into output B. There are three reasons for this. (1) It can be said that the project had the greatest leverage to achieve results with respect to output B. For output A, the project could only go as far as its mandate allowed, as the more political aspects of lobbying for alliances had to be mostly covered by BMZ. As for output C, the piloting of reform processes had to be owned strongly by the respective IO and was not possible under all circumstances. (2) Considering the state of the reform processes that the project chose to engage in, giving methodological support through the German TC's expertise was identified as the most

[36] We understand overarching costs as costs that do not directly contribute to achieving an output but are spent to assure the general operation of the project (such as office equipment, costs related to administrative procedures etc.).

relevant strategy by the IOs (Int_3). (3) Considering the aspect from a mere ToC perspective, the leverage of directly influencing the reform process seemed greatest under output B (see also section 4.3).

Having reflected on the distribution of investments from the perspective of the project's logic, we can next consider the results achieved. As the analysis under section 4.3 showed, the outputs pursued by the project were largely achieved, with the exception of output C1 (cooperation agreements between IOs and German development cooperation projects).[37] As for this output, the discussion in sections 4.2. and 4.3 showed that some systemic barriers existed to creating successful cooperation between bilateral and multilateral development actors. After some attempts at linking German TC projects to multilateral initiatives, it soon became apparent that the targeted results could not be achieved. The project team reacted to this in two ways. First, it scaled down the ambition for indicator C1 by counting results such as contributions of different GIZ projects to GDI conferences and a publication. However, the wording of the indicator itself was never adjusted to reflect this decreased level of ambition, so that the project's reporting suggested the initial indicator was in fact fulfilled. Second, according to the project team, a stronger focus was put on achieving indicator C2,[38] which was assessed as more realistic to achieve (Int_5). From an efficiency point of view, it could be said that the project team reacted appropriately by investing less effort into a result that was unlikely to be achieved; however, this reflection should have been made transparent in the reporting. This would have enabled discussion on whether investment in C1 was still at all relevant, or whether the resources foreseen for the output were better invested somewhere else.

Looking at the resource investment into those output indicators that were completed makes apparent that the target values were in fact overachieved.[39] Superficially, this suggests that resources were invested to maximum use. However, this would build on the assumption that more outputs automatically lead to more longer-term results. During the interviews with the project team, it became apparent that this question was in fact not systematically reflected on during implementation (Int_3). From the evaluation team's perspective, this is not only a question of the project's financial management – it also underlines the weaknesses of the target system already discussed in section 4.3. Because the target system was set up to aggregate results of the different processes in an output-heavy format, it contained little information for the project's day-to-day steering and, hence, for investment decisions concerning specific reform processes. When making investment decisions, the project team consequently referred to other sources of information, such as the expert opinion of the iLTEs or discussions and decisions in the GDI Secretariat (Int_3).

For the same reason, it seems more meaningful to look beyond the mere cost-output analysis and consider the project's outputs more directly in relation to the reform processes it supported. This can mainly be done with a view to the appropriateness of the TC instruments used. Overall, the costs incurred in relation to the outputs were made up of personnel costs (which accounted for 55% of the project's total cost), complemented by financing instruments (grants) (accounting for 16.5% of total cost) and the administrative costs necessary to enable the deployment of experts to reform processes. As elaborated in section 4.3 when considering possible alternative strategies for the project to support reform processes (e.g. through larger amounts of financing), the focus on the secondment of experts was considered the best possible option to achieve the desired results by those IOs who received such support from the project (Int_11, 16). Considering the instrument of secondment from an efficiency point of view, one of its strengths becoming apparent in this evaluation is the number of different outputs it can generate while the financial investment of the project stays constant. Not only did the iLTEs produce outputs directly relevant to the reform process (e.g. briefing papers, strategies or workshops) but also served as expert advisors to German policy actors (BMZ, GEDO and StäV) and facilitators between different development actors (e.g. WB and KfW). As the interviews conducted during the evaluation showed,

[37] Indicator C1 states: four German development cooperation projects have agreed to cooperate with international organisations as part of German reform support.
[38] Indicator C2 states: In eight countries, bilateral and multilateral partners have prepared the implementation of the reform measures of international organisations.
[39] This concerns output indicator A2 (target: two new alliance partners support reform processes; achieved value: 11), B1 (target: six methodological contributions are rated as useful; achieved value: 8) and B2 (10 reports of IOs about the application of methods are available; achieved value: 11).

these strengths are contingent on the instrument being placed in the right position within the host organisation – a challenge tackled successfully by the project (Int_8, 10, 12, 15). According to several interview partners, this led to Germany having unique opportunities in shaping a process it might otherwise not have access to. Similar observations apply with respect to the instrument of financing/providing grants. In most cases, the project applied grants in combination with a secondment (UN fit for purpose, WB ESF), which allowed to draw on the iLTE's expertise when designing the financing instruments and helped create synergies between the TC instruments (Int_3, 8, 10).

Turning next to **efficiency dimension 2**, the transformation of inputs into outcomes, the insights elaborated above apply accordingly. Looking at the level of the module indicators, they were generally achieved by the project – in part even overachieved (see section 4.3). However, just as in relation to the outputs, they contain limited information on the efficient use of resources for every reform process, again underlining that financial steering was hardly possible on their basis. Looking at the results described in the effectiveness analysis that go beyond the assessment of indicators, the evaluation showed the significant outcomes produced by the project for each individual measure (e.g. the application of a new strategy for pooled funding that is now applied by UNDP). As section 4.6 on sustainability will underline, the results achieved are generally strongly anchored within the organisation, suggesting that they will be drawn on in the long run or have contributed to lastingly altered structures. From an efficiency point of view, this suggests that the resources initially invested by the project can create added value in the long run.

The only individual measure sparking greater discussion with respect to the outcomes achieved in light of the resources invested – especially against the backdrop of its longevity – was the support given to the Global Delivery Initiative (see section 4.6). While the project team could not provide the evaluators with a full overview of the resources invested in GDI, as explained above, a qualitative reconstruction based on the interviews with the project staff suggests that several staff members, especially the team working from Germany and the project manager, from time to time invested up to 50% or more of their time in the Initiative (Int_5, 21). Generally, it can be said that, during the co-chair term, the project accompanied GDI more closely than the other reform initiatives – this is due to the project manager's role as the Initiative's co-chair. For example, the project team provided policy and strategic advice to the GDI Secretariat and helped organise and accompany several of the GDI's conferences. Examining the strong involvement in the GDI in retrospect, the evaluation team gathered different perspectives on whether the investment of the project's resources in the Initiative was in fact justified. Some interviewees pointed to the unique opportunities created by the GDI, such as creating space for dialogue between development actors that would not otherwise have exchanged in depth – sometimes even creating joint ventures as a result (Int_3, 15). Others were more reluctant, saying that it soon became apparent that the Initiative may not have enough influence to persist in the long run (Int_18, 21). According to these voices, the Initiative always had a rather shaky standing in the World Bank and lost even more support when World Bank President Kim, who had emphasised that WB should focus more strongly on implementation, stepped down and some personnel changes took place in the GDI Secretariat. However, one needs to consider that once the project had invested into the GDI, it could not simply pull out – otherwise, any results already achieved might have been lost. (Int_3, 14). Today, the further financing of the Initiative is in question, making its future uncertain (Int_21, 26, 27,35).

Opinions also varied on how relevant the Initiative was from BMZ's point of view (section 4.2). Closer to implementation than to policy and logically some steps removed from IO reform, the evaluation team heard varying opinions on whether the Initiative was deserving of the resources it was given from the Ministry's political-strategic point of view (Int_1, 21).

Lastly, reflecting on the project's **financial steering processes** as a means to achieving efficiency overall, several mechanisms become apparent. With a view to the instrument of secondments, the most important steering moment was arguably when the expert's terms of reference (ToR) were negotiated with the IO, as the project then had the most leverage to define a strategic position for its iLTEs, working towards a maximum

'return on investment'. This was done successfully in all cases examined. With respect to the financing instruments, the project's steering style can be described as more ad hoc and opportunity-driven – a style that seems justified given the shared vision between BMZ and GIZ to react flexibly to windows of opportunity. The weakness of the system, however, lies in its limited traceability. Since the target system was not useful to guide financial steering decisions, it is hard to reconstruct in retrospect to what extent decisions were taken on the basis of systematic cost-benefit analyses, and whether moments to potentially change course were systematically considered (e.g. with respect to the GDI).

Assessment regarding efficiency

The analysis of **efficiency dimension 1** regarding the transformation of inputs into outputs showed that the project used the appropriate instruments to reach its strategic aims, and that the outputs brought about were in fact valuable contributions to the respective reform processes. Taking into account the distribution of costs to outputs from a conceptual/ToC point of view, the evaluators find it logical that the largest amount of resources was invested in output B. However, the evaluation has to reflect that the project's target system – and in consequence an analysis of the cost to output-relation – only yields limited information. More significantly, the target system was not useful for the financial steering of the project, which complicates an assessment of the mechanisms used for cost-benefit analyses. A financial monitoring along the lines of the individual measures did not exist. For this reason, the evaluators attribute 55 out of 70 points for efficiency dimension 1.
Looking at efficiency dimension 2, the evaluation shows that the project made significant contributions to the reform processes supported at the outcome level. However, the weaknesses of the target system identified under dimension 1 apply equally to the analysis of transformation efficiency. With a view to the outcome-resource relation in GDI, the evaluation team cannot give a conclusive answer on whether the investment was appropriate. Therefore, it seems that GDI offers an excellent example for institutional learning and could be studied in more depth to draw some lessons learned on investing in similar initiatives in the future (see Chapter 5 Recommendations). All in all, the evaluators attribute 28 out of 30 points for efficiency dimension 2.

Table 10: Rating of OECD/DAC criterion: efficiency

Criterion	Assessment dimension	Score and rating
Efficiency	The project's use of resources is appropriate with regard to the outputs achieved. [Production efficiency: Resources/outputs]	55 out of 70 points
	The project's use of resources is appropriate with regard to achieving the project's objective (outcome). [Allocation efficiency: Resources/outcome]	28 out of 30 points
Overall score and rating		Score: **83 out of 100 points** Rating: Level 2: successful

4.6 Sustainability

The criterion sustainability assesses the extent to which positive results of the project can be expected to continue after the end of the project's term. The evaluation matrix in Annex 1 details the evaluation questions assessed under the criterion of sustainability.

Evaluation basis and design for assessing sustainability

In **sustainability dimension 1**, the evaluation team analysed to what extent the project has made efforts to anchor the results in the structures of the international partner organisations. For this purpose, Syspons examined data with a twofold focus: (1) the project's strategy and design with regard to sustainability of its support within the different measures; and (2) the sustainability of the contributions made to the reform

processes. While the former aspect took into account potential exit strategies regarding the withdrawal of iLTEs (UN, WB) or the end of the co-chair term (GDI), the latter considered to what extent sustainability was considered in the iLTE's work and contributions to the reform processes. As the project forms part of the ICR title, it also had to live up to additional requirements regarding sustainability such as the building of networks, innovations (e.g. network approach) or the taking over by a classical technical cooperation project. The latter, due to the project's design, can be better understood as the taking over by partner organisations at the end of the project's term. When analysing the different measures, these additional requirements are discussed as cross-cutting.

Sustainability dimension 2 relates to a forecast of durability. The evaluation team examined to what extent the results of the project are permanent, stable, and long-term resilient (**plausibility analysis**). To answer this question, potential risks and other influencing contextual factors, as well as the project's mitigation strategies were investigated. For analysing this dimension of the sustainability criterion, the evaluation team used data collected on the status quo of the supported reform processes (section 4.4). Due to the scope of the evaluation, the evaluators focused mainly on the UN fit for purpose, the ESF and the GDI measures.[40]

Furthermore, the evaluation phase showed that it was neither feasible nor relevant for this type of project to consider the three dimensions of sustainability in its support. This is due to the project's design as a support to and within teams in IOs. Hence, this aspect was not assessed in the analysis. Due to the long results chain between the project's intended outcomes and impacts and the implementation of the 2030 Agenda in partner countries, as for the impact criterion, the analysis did not observe sustainability beyond the IO level.

As **data collection methods** to assess both dimensions of the sustainability criterion, the evaluation team used: (1) document analysis (e.g. project proposal, progress reports, reports of iLTEs); and (2) semi-structured interviews with project staff, staff of IOs and other stakeholder (e.g. German Development Institute, StäV).

Analysis regarding sustainability

In order to analyse to what extent the project has made efforts to anchor results in the structures of the partner organisations, the following section examines how sustainability was considered throughout the project term (**sustainability dimension 1**). The analysis showed that the project's influence on the sustainability of results within the IOs was limited. This is because the project's support was much shorter and selective than the nature of institutional reforms. It became apparent in the analysis that this limited influence was already considered in the project's modification offer (2017) and resulted in the following strategy: 'The project aimed to provide support tailored to the partners' needs and tied to existing teams or processes within the IOs, in order to increase chances for the sustainability of its results'. This strategy created ownership of the IOs for the produced contributions, as shown in section 4.3, and was supposed to limit the risks outside of the project's sphere of influence, like the loss of momentum for the reforms. It further defined the overall objectives and terms of support in ToRs with BMZ and the IOs. Apart from this strategy and the ToR, it became obvious in interviews that responsibility for the anchoring of results towards the end of the project's support depended strongly on the iLTE and, for the most part, on the partner organisations. It was, hence, the iLTE's responsibility to ensure a successful handover before the end the secondment (Int_3, 8, 19, 20, 32). This handover, of course, depended on the personnel resources which the IO made available to continue the tasks after the project term. The following examples in dimension 2 illustrate that the iLTEs were indeed able to adequately organise this handover. Nonetheless, the findings of the analysis suggest that a formalised exit strategy, for example included in the ToR, could have clarified and ensured the sustainability of results in advance. In this regard, the project could have made more use of its limited influence on sustainability.

Taking this overall context into account, the findings of the evaluation suggest that the project was successful

[40] As 4.4 has shown, the progress of the reform at IFAD is less complex than for the other reform-oriented measures and the topic-oriented measures did make significant progress.

in anchoring its results within partner organisations (and networks) in the short term. The following description of the individual measures (**impact dimension 2**) illustrates this finding:
In the **UN fit for purpose process** the reform continues. The iLTE now supports implementation at UNDP as part of the new project (PN.2019.2035.4). The analysis in section 4.3 as well as the continuation of the iLTE's support to UNDP show that the project has successfully helped to anchor the reform's first results in the short and midterm. This has been illustrated, for example, by the use of the iLTE's pooled funding strategy by UNDP. However, it also became apparent that the long-term success of the reform still faces a variety of risks. Besides the possible opposition from individual stakeholders within the UN, interview partners identified the overall pressure on multilateralism as the biggest risk to the reform's success (Int_3, 10): The UN depends on member states' support to follow through with the reform beyond the theoretical adoption of resolutions and proposals. This dependency becomes tangible in the following example: The reform aims to provide more core financing for the UN to strengthen effective, joint and coherent work of UN agencies at country level. This plan for financial reforms can only be effective if major donors, like Germany, act accordingly. If donors keep up strong UN agency specific funding, the financial reform goals will not have any effect (Int_34). Hence, the analysis showed that while the project used its influence to anchor results, long-term sustainability now depends on the political level. BMZ and StäV, as stakeholders of the evaluation, can, thus, actively contribute to the sustainability of results in the long term.

In the **ESF process**, the rollout of capacity building is an ongoing process. The continued implementation of the ESF mainly depends on WB and the responsible unit. In addition, the multi-donor trust fund, which was prepared with the project's support, could ensure long-term financing for capacity building. This finding shows that the project directly contributed to help ensure the sustainability of results. At the time of the evaluation, partners such as BMZ were still evaluating whether financing will be provided to the fund (Int_1, 29). Thus, as for the UN reform process, the sustainability of the achieved results also depends on member states and BMZ can play a role in ensuring its TC projects' long-term sustainability. Apart from this financial aspect, the analysis showed that personnel resources at the Bank were available. At the end of the iLTE's secondment the implementation team was dissolved and the iLTE's tasks were taken over by the newly created Special Unit, which continues to work on the topic and also followed up with the project-financed pilots for the capacity needs assessment (Int_11). Hence, while sustainability depends on the overall will within WB and provided financing by members and the Bank itself, the personnel and expertise built up within the IO for the topic, suggests that results can be sustained (Int_8, 11, 23, 31).

The case of the **GDI** shows that the project's responsibility concerning sustainability differed from the other individual measures. Compared to the supported institutional reforms at UN and WB, the project had more influence on shaping sustainability of the progress in the GDI. This was due to its close collaboration with the GDI Secretariat. During its co-chair term, project staff worked hand-in-hand with the Secretariat staff on developing strategies or preparing events (Int_6, 15, 21). In this context, the main responsibility for the project in terms of sustainability entailed the handover of the co-chair. Based on the conducted interviews, the project was able to support the recruitment of a successor – the Japan International Cooperation Agency (JICA) – and successfully organised the handover (Int_6, 35). From GDI members' perspective, the project could have done little more to positively influence the Initiative's future (Int_26, 27). The evaluation's results, hence, suggest that the project ended its support appropriately. However, in terms of the **durability** of the Initiative's progress the evaluation revealed great uncertainty. The major challenge for the GDI now is its financing (Int_21, 26, 27,35): at the time of the evaluation, the future of the GDI was unclear as WB decided to no longer finance the GDI Secretariat. Considering the possible scenario that the Initiative might lack sufficient resources in the future, interview partners differed in their assessment on whether any permanent results were, nonetheless, already achieved. Some mentioned the thought-provoking impulses at conferences and labs, which positively affect the approach to implementation of members in the long-term (Int_18, 33). Others underlined that the Initiative needs the continuous updating of their methods and the continuous exchange between members to create longer-term results (Int_26). With a view to these different perspectives and the assessment on the GDI in sections 4.3 and 4.5, the analysis could not generate a final estimation on the Initiative's long-term influence on

members. All in all, the analysis suggests that preventing the GDI's future financial challenges was beyond the project's influence.

Assessment regarding sustainability

Considering that an important responsibility for the durability of reform processes lies within IOs, the project used the opportunities at its disposal for influencing sustainability in a good manner. The analysis showed that the project's approach to anchoring results in partner organisations focused on considerations prior to the start of a measure. The project aimed to create ownership for its contributions by (1) providing *needed* support and (2) by seconding iLTE's to work directly with partner teams. This was supposed to secure sustainability. Looking at the findings, the assumption was correct. All partner organisations had resources available to take over the iLTE's tasks (e.g. Special Unit for ESF) and to further develop results in the short to midterm (e.g. DG DEVCO garment initiative). Necessary handovers were successful. As shown in 4.4, the results were also further used (e.g. capacity-building material in the ESF process, strategies for the UN reform process). The lack of a formalised exit strategy did not negatively affect the contributions' sustainability in this project. However, formalised exit strategies could have increased the project's influence on securing the sustainability of results. All in all, in sustainability dimension 1 – the strategy and activities for anchoring the results in the partner organisations – the evaluators attribute 45 of 50 points. While the relatively stable progress of the DG DEVCO garment initiative, the UNDP and the IFAD measures have been described in 4.4, this section showed the variety of risks that endanger the resilience and further development of the achieved progress in the reform-oriented measures. For the UN fit for purpose process and the ESF process risks prevail, while the continued existence of GDI is in fact unclear. The findings of the evaluation also revealed the responsibility of other stakeholders, mainly the IOs, but also BMZ and the German Ministry of Foreign Affairs (AA) (like other member states) to secure the sustainability of the progress. Finally, considering the described high risk with regard to the sustainability of GDI, the evaluators subtracted a few points. Though the project considered risks for the other measures in advance, the uncertainty regarding overall sustainability of the achieved progress leads to the rating of 40 out of 50 points.

Table 11: Table 4. Rating of OECD/DAC criterion: sustainability

Criterion	Assessment dimension	Score and rating
Sustainability	Prerequisite for ensuring the long-term success of the project: results are anchored in (partner) structures.	45 out of 50 points
	Forecast of durability: results of the project are permanent, stable and long-term resilient.	40 out of 50 points
Overall score and rating		Score: **85 out of 100 points** Rating: Level 2: successful

4.7 Key results and overall rating

The project aimed to strengthen IOs to further develop reform processes with a view to implementing the 2030 Agenda. In this context, it supported several partner organisations and a network and, hence, was implemented in diverse institutional and political contexts. The evaluation showed that the project was successful in achieving its objective. The following paragraphs summarise the key results for each evaluation criterion.

Regarding **relevance**, the evaluation showed that the project was closely aligned with the relevant strategic framework of BMZ as the commissioning party. The project also matched partner organisations' needs. As the project negotiated good positions for its iLTEs and selected competent experts, it was able to stay relevant despite complex and changing institutional contexts. Looking at the results model, the project was generally

well designed for the chosen project objective. However, the evaluation team noted several weaknesses in the project's concept: (1) The project was too ambitious regarding the link between bilateral and multilateral development cooperation in reform implementation. (2) The support to 'German' proposals turned out to be neither relevant nor feasible. (3) The target system was not suitable for adequately measuring the project's results. These findings led to a deduction of 10 points. Nevertheless, the project's main assumption, that German TC can provide helpful input for the implementation of IO reforms, remained true. All in all, the project was relevant in the different dimensions of the analysis. For this reason, the evaluators awarded **90 of 100 points**.

In terms of **effectiveness**, the evaluation showed that the project achieved all intended outcomes. Regarding the project's target system, however, the evaluation also revealed that it was not useful for assessing the project's effectiveness in the individual measures and was hence of limited use for strategic steering. Looking at the contribution of the individual measures to outputs and outcomes, a logical chain could be established. The analysis of the project's underlying hypotheses showed that most hypotheses could be confirmed. Overall, the evaluators attributed **85 out of 100 points**.

For the **impact criterion**, the evaluation's findings show that many of the intended impacts were not yet achieved due to the long-term nature of the supported reform processes. They also face a variety of risks that could still prevent the completion of reforms. In the case of the GDI, for example, it remains to be seen whether sufficient financing for the GDI Secretariat can ensure the continuance of the Initiative.

Nonetheless, the IOs made critical progress towards the main intended impacts – the completion of reform processes and their implementation. The analysis showed how the project's support contributed to this progress. However, the evaluation also revealed some limitations: The intended link to German bilateral development cooperation turned out not to be feasible. Additionally, the project's technical support was not perceived as 'German' contributions. Hence, the project could only to a limited extent contribute to reinforcing a positive perspective of Germany's role in the reform processes as stated in one impact hypothesis. All in all, the evaluation team attributed **90 out of 100 points**.

In terms of **efficiency**, the evaluation showed that the project used the appropriate instruments to reach its objectives. With a view to the distribution of costs and the ToC, the evaluation further showed that the project plausibly invested the largest amount in Output B, which provided the biggest leverage for the project. However, looking at the individual measures, it remains unclear whether the personnel and financial investment in GDI was appropriate considering the insecurities regarding its future. A further limitation in this criterion relates to the project's target system. As it was not useful for financial steering, the evaluators could not fully assess the mechanism used by the project for cost-benefit analysis. Due to these limitations, the evaluators attributed **83 out of 100 points**.

For the **sustainability** criterion, the evaluation's findings revealed that for the most part, the project used the opportunities at its disposal to anchor results in partner organisations. Within the individual measures the iLTEs – and (in GDI) the co-chair – successfully handed over their tasks and all partner organisations had resources available to ensure sustainability of results beyond the project's term. Nonetheless, formalised exit strategies for the individual measures could have clarified this process in advance, which now mostly depended on the iLTE's skills, networks and partner resources at the time. This could have increased the project's influence on the sustainability of results. Furthermore, a variety of risks still endanger the resilience and development of the achieved progress in the reform processes and require German political support. The evaluation showed that especially the GDI longevity faces immediate risks. Due to this uncertainty and the lack of formalised exit strategies, the evaluators attributed **87 out of 100 points**.

Table 12: Overall rating of OECD/DAC criteria and assessment dimensions

Criterion	Score (Max. 100)	Rating
Relevance	90 out of 100 points	Level 2: successful
Effectiveness	85 out of 100 points	Level 2: successful
Impact	90 out of 100 points	Level 2: successful
Efficiency	83 out of 100 points	Level 2: successful
Sustainability	85 out of 100 points	Level 2: successful
Overall score and rating for all criteria	**87 out of 100 points**	**Level 2: successful**

Table 13: Rating and score scales

100-point-scale (score)	6-level-scale (rating)
92–100	Level 1 = highly successful
81–91	Level 2 = successful
67–80	Level 3 = moderately successful
50–66	Level 4 = moderately unsuccessful
30–49	Level 5 = unsuccessful
0–29	Level 6 = highly unsuccessful

5 Conclusions and recommendations

5.1 Factors of success or failure

The evaluation revealed several factors of success in the project implementation. Considering the former, the project was able to identify **windows of opportunity** with BMZ and other political stakeholders which allowed the positioning of iLTEs in strategically relevant processes. This provided German technical cooperation with the unique opportunity to support processes in IOs which it otherwise would not have had close access to (Int_3, 15, 34). In this context, the **choice of instruments** – iLTEs in combination with financing – proved to be critical for the project's success. In the dynamic institutional and political settings, the iLTEs provided the needed technical support relevant for the respective processes and topics. Due to the **proximity to the supported processes** (e.g. WB ESF, GDI), the project was also able to identify relevant and effective financing options for reaching the project's objective. The project's target system, however, was not designed precisely enough for enabling the project team to steer towards results. Instead, for taking strategic steering decisions on the target system, the project relied on other sources of information, such as insights from the iLTEs. Therefore the **project's target system design** failed to provide information for steering. Whereas the evaluators conclude that the project's target system could have been improved, the finding raises the general question as to whether the requirements for TC measures are in fact a useful basis for this type of flexible and highly political project. Whether a different design might be useful in such cases could be discussed in the follow-up process to this evaluation.

In addition to the already described responsibilities of the iLTEs, they were also expert advisors for German political actors which helped to inform German political decisions, for example, in the case of the UN fit for purpose process (while respecting the limitation of their role in IOs). This role required **regular communication of iLTEs with political actors** and the up taking of the project's impulses by BMZ or StäV. Most importantly, considering this summary of the secondees' tasks, the project's success depended to a major degree on the **skill set of the iLTEs**. In this regard, the project successfully selected individuals which fit partner requirements: besides their technical expertise, the secondees disposed of the necessary experience and self-esteem to manoeuvre in complex institutional settings. They, further, possessed the necessary political sensitivity and awareness for their role, which allowed them to become part of the partner organisations' teams and units (Int_2, 11, 12, 16, 23, 25, 29, 31).

Finally, the project's success depended on **partner organisations' 'will' to reform** and each institutional context as the project could only achieve results in close cooperation with the supported teams and units. The resources provided by IOs after the project's term also ensured successful handovers and/or continuation of the iLTE's tasks.

At the same time, the partners' **institutional and political contexts** constituted the major **inhibiting factor** which affected the project during implementation and continues to influence the sustainability of the achieved results. While this factor lies outside the project's scope of influence, for the ESF process, it led to an adaption of the iLTE's tasks as political processes delayed original plans.

5.2 Conclusions and recommendations

This was the final evaluation of the project 'Strategic Cooperation with International Organisations', so that any conclusions and recommendations no longer feed into direct implementation. However, the new project 'Promoting the implementation of 2030 Agenda through more effective cooperation between multilateral

actors', which carries on parts of the project evaluated here, can learn from the results of this evaluation. Equally, any other future project set up to work at the level of IOs towards reform can benefit from them. Based on the results presented in this report, the evaluator's give the following recommendations:

- To GIZ project team and other relevant units: Analyse 'lessons learned' from the support to the Global Delivery Initiative as a learning instrument for potential future support to global initiatives with a similar character.

The evaluation results have shown that stakeholders of the project under examination, both at BMZ and GIZ, have divided impressions when reflecting on the support offered to the GDI. Some stakeholders argued that the Initiative did not fit the project's logic because it did not relate to a political reform process but offered a platform for exchange on so-called delivery challenges in practice. Others argued that the Initiative presented a unique opportunity for dialogue between stakeholders, including the political level, that otherwise would not have exchanged. With a view to the resources invested in the GDI, the analysis sparked questions as to the appropriateness, especially seen as major risks to the Initiative's long-term survival became apparent early on. On this basis, the evaluators reached no final conclusion as to the Initiative's 'appropriateness' in the project's portfolio, since any further examination would have been beyond the scope of this exercise.

Therefore, the evaluators recommend GIZ to initiate a lessons-learned process for the GDI, investing in more in-depth analyses with a view to the ambiguities that became apparent in this evaluation. From the evaluator's point of view, this could serve to inform any possible future engagement of BMZ and GIZ in comparable initiatives, both from BMZ's political-strategic, as well as from GIZ's implementation, point of view.

- To BMZ and future GIZ projects: When working in a project that spans multiple institutions and contexts, develop a target system which differentiates between different institutions or reform processes.

The analysis of effectiveness and relevance in this evaluation offered a critical angle on the set-up of the project's target system, which accumulated results across the individual reform processes in a strongly output-focused manner, even at the outcome level. The target system proved relatively unhelpful when trying to learn about progress and results at the reform processes' level and trying to determine the project's contributions to it. This was also reflected by the fact that the project team used the target system and results matrix for reporting purposes only and drew on other sources of information to inform their strategic decisions.
For the future set-up of GIZ projects that span such different institutions and contexts, we therefore recommend to set up target systems that are more strongly focused on the individual reform measure and allow actual steering of the process. This could take the form of structuring outputs along individual measures, which allow the output level to be process focused and defining module indicators along individual process lines. Alternatively, individual target systems could be developed for each measure. One option could be to design this type of project as a fund, to allow better steering across individual measures. The evaluation findings also raise the general question as to whether the current BMZ requirements for target systems are actually a useful basis for this type of flexible and highly political project. Whether a different design – such as a 'fund' – might be useful in such cases could be discussed in the follow-up process to this evaluation.

- To future GIZ projects: When working in a project that spans multiple institutions and contexts, use the differentiated target system for steering financial resources.

In parallel to the effectiveness analysis, the efficiency analysis also critically reflected upon the usefulness of the target system for financial steering. Accumulating costs along outputs, as practised in the follow-the-money approach, contained very little information on the transformation and allocation efficiency. While project staff indicated that during implementation, a financial monitoring took place at the level of individual measures, the respective documents could not be provided for the evaluators.
On this basis, we recommend setting up the financial steering processes of a project conceptualised like

‘Strategic Cooperation with International Organisations’ along the lines of the improved target system recommended above. Since there is no impact level that connects all the individual measures, and the project itself is in fact a sum of its parts, it should be steered accordingly to allow targeted cost-benefit analyses and investment decisions for each process. The KOMP procedures (cost-output-monitoring) that are currently being introduced at GIZ, are already a step in this direction.

- To future GIZ projects: When working with IOs on large reform processes, continue to employ a systematic, criteria-based selection process for TC engagement.

The analysis of relevance, efficiency and sustainability showed that the identification of potential support measures and their selection was a critical step for the project’s success. Potential windows of opportunity were closely monitored and communicated between BMZ, other political stakeholders and the project team. Furthermore, the project team together with BMZ reflected on the strategic relevance of potential support by considering, for example, whether a specific activity or change would take place without the project’s activity. The success of this criteria-based selection is illustrated by the effective contributions of the project across the diverse measures.

Therefore, the evaluators suggest to continue employing a systematic, criteria-based selection process for TC engagement in GIZ projects which support processes in partner organisations. From the evaluators’ point of view, though the success of large reform processes can never be foreseen, this detailed reflection could limit the risks of supporting a process which is either not relevant or fails in the future. This could include a detailed background analysis concerning the specific institutional challenges and risks with regard to the respective reform process. Considering the complex contexts of these processes, it could also be beneficial to include other stakeholders such as research institutes in the reflection process.

- To BMZ and future GIZ projects: When employing TC instruments in international organisations, continue working with ToR and make exit strategies more explicit.
-

The evaluation showed the challenges when steering a project which employs TC instruments in partner organisations. After the start of the deployment of an iLTE, the project’s influence on the progress in the supported processes and on the sustainability of its contributions was limited. In this context, the definition of expectations in ToR with BMZ and the IO in combination with the selection of skilled iLTEs proved to be a successful strategy. Nonetheless, the evaluation also showed that adequate handover of tasks and the follow-up by partners mostly depended on the iLTEs.

Based on the findings, the evaluators recommend to continue working with ToR and to explicitly include exit strategies for GIZ projects in a similar set-up. In this context, exit strategies refer to the definition of roles, responsibilities and processes towards the end of the project’s support. Often, IOs already have respective processes and procedures in place, to which the ToR could refer. From the evaluators’ perspective, making exit strategies explicit could clarify responsibilities in advance and increase the project’s influence on the durability of its contributions. Moreover, an additional insight underlines the importance of this recommendation: due to changed legal regulations, the maximum length of secondments is now limited. Taking into account the shorter duration of secondments, the clarification of expectations and available resources after the project’s support appears even more necessary.

Lastly, with a view to exit strategies and a follow-up of the project’s results, we recommend for BMZ to reflect on where it could best support the durability of results on the political level, e.g. by further pushing the respective issues in multilateral forums or applying agreed-upon principles in its own work (see example of supporting UN core funding mentioned in section 4.6 Sustainability).

List of references

BMZ (2013): Eckpunktepapier für die multilaterale Entwicklungspolitik.

BMZ (2020): BMZ-Strategie für eine starke europäische und multilaterale Entwicklungspolitik.

European Commission (2016): Collect More – Spend Better: Achieving development in an inclusive and sustainable way.

European Commission (2017): Commission Staff Working Document: Sustainable garment value chains through EU development action', SWD (2017) 147 final.

German Federal Government (2016): Report to the High-Level Political Forum.

GIZ (2015): Project proposal 'Strategic Cooperation in and with International Organisations II' (PN 2015.6258.6).

GIZ (2016a): Annual progress report of 'Strategic Cooperation in and with International Organisations' (PN 2015.6258.6).

GIZ (2016b): Reports of secondee to UNDP Governance and Peacebuilding Cluster.

GIZ (2016c): Final Report of secondee to DG-DEVCO Unit A4.

GIZ (2016–19): Reports of secondee to UN fit for purpose process.

GIZ (2017a): Modification offer of 'Strategic Cooperation with International Organisations' (PN 2015.6258.6).

GIZ (2017b): Reports of Consultant for the annual progress report 2016.

GIZ (2017c): Annual progress report of 'Strategic Cooperation in and with International Organisations' (PN 2015.6258.6).

GIZ (2017d): Final report of secondee to UNDP Governance and Peacebuilding Cluster.

GIZ (2018): Annual progress report of 'Strategic Cooperation in and with International Organisations' (PN 2015.6258.6).

GIZ (2019a): Annual progress report of 'Strategic Cooperation in and with International Organisations' (PN 2015.6258.6).

GIZ (2019b): Final report of 'Strategic Cooperation in and with International Organisations' (PN 2015.6258.6).

GIZ (2019c): Repetition offer 'Promoting the implementation of the 2030 Agenda through more effective cooperation with multilateral actors' (PN 2019.2035.4).

GIZ (2019d): Terms of Reference with specification of inputs for short-term expert assignment (to IFAD).

GIZ (2019e): Final Report of secondee to WB ESF process.

Mayne, J. (2001): Addressing attribution through contribution analysis: using performance measures sensibly. Canadian Journal of Programme Evaluation, 16, 1–24.

UN (2017a): Repositioning the United Nations development system to deliver on the 2030 Agenda: ensuring a better future for all: report of the Secretary-General. A/72/124-E/2018/3 (July 2017).

UN (2017b): Repositioning the United Nations development system to deliver on the 2030 Agenda: our promise for dignity, prosperity and peace on a healthy planet: report of the Secretary-General. A/72/684-E/2018/7 (Dec 2017).

UN (2018): General Assembly Resolution A/RES/72/279.

WB (2019): Completion Report: GDI EFO.

WB (2020a): Completion Report: GDI EFO.

WB (2020b): Completion Report: ESF EFO for Borrower Capacity Building.

Annex: Evaluation matrix

OECD-DAC criterion EFFECTIVENESS (max. 100 points)						
Assessment dimensions	**Filter – project type**	**Evaluation questions**	**Evaluation indicators**	**Data collection methods** (e.g. interviews, focus group discussions, documents, project/partner monitoring system, workshop, survey, etc.)	**Data sources** (list of relevant documents, interviews with specific stakeholder categories, specific monitoring data, specific workshop(s), etc.)	**Evidence strength** (moderate, good, strong)
The project achieved the objective (outcome) on time in accordance with the project objective indicators.(1) Max. 40 points	Standard	To what extent has the agreed project objective (outcome) been achieved (or will be achieved until end of project), measured against the objective indicators? Are additional indicators needed to reflect the project objective adequately?	Outcome Indicator 1: 4 new technical proposals supported by the project have been incorporated into strategy documents for setting or further developing reforms to strengthen the implementation of Agenda 2030 in international organisations	document analysis, interviews	1. documents of IOs, modification offer, reports of iLTEs, progress reports, results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good
			Outcome Indicator 2: 8 new stakeholders express support and/or agreement to a proposal supported by the project within the framework of the reform processes for implementing Agenda 2030 in international organisations (e.g. GDI, UN Reform, World Bank Social and Environmental Standards)	document analysis, interviews	1. modification offer, reports of SNEs, progress reports, additional documents (of e.g. GDI), results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good
			Outcome Indicator 3: 4 new methods for process design of reforms (e.g. procedures, moderation, coordination, monitoring) have been applied by international organisations	document analysis, interviews	1. documents of IOs, modification offer, progress reports, SNEs reports, results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good
			Outcome Indicator 4: 2 Reform measures by international organisations (e.g. within the framework of the GDI, World Bank Social and Environmental Standards) have been applied in an exemplary manner (1 of which interlocks with German bilateral cooperation)	document analysis, interviews	1. progress reports, modification offer, documents of IO, results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good
	Standard	To what extent is it foreseeable that unachieved aspects of the project objective will be achieved during the current project term?	The evaluation question will not be assessed, as all planned outcomes were achieved			
The activities and outputs of the project contributed substantially to the project objective achievement (outcome).(1) Max. 30 points	Standard	To what extent have the agreed project outputs been achieved (or will be achieved until the end of the project), measured against the output indicators? Are additional indicators needed to reflect the outputs adequately?	A1: 6 concrete agreements (e.g. on GDI, UN reform, WB Social and Environmental Standards) among bilateral and/or multilateral alliance partners have been reached during meetings for the discussion of reform proposals supported by the project	document analysis, interviews	1. progress reports, reports of SNEs, results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good
			A2: 2 new alliance partners have made substantive contributions (e.g. conducting a workshop, high-level appearance by a top executive) to events organised by international organisations on reforms to implement Agenda 2030 (e.g. in the context of GDI, UN reform, WB social and environmental standards) supported by the project	document analysis, interviews	1. progress reports, reports of SNEs, results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good

			B1: 6 methodological contributions to the design of reform processes (planned in the GDI, UN Reform, WB Social and Environmental Standards) made by the project have been assessed as useful for the reform processes by IO management staff	document analysis, interviews	1. progress reports, reports of SNEs, results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good
			B2: 10 field reports by staff members of international organisations on the application of methodological contributions to the design of reform processes as part of individual measures (envisaged in UN Reform, Mainstreaming Agenda 2030 in the OECD) describe the usefulness of the contributions for the respective process design, indicating that the proposals submitted are assessed as useful	document analysis, interviews	1. progress reports, reports of SNEs, results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good
			C1: 4 German development cooperation projects have agreed to cooperate with international organisations as part of German reform support (e.g. in GDI, World Bank Social and Environmental Standards)	document analysis, interviews	1. progress reports, reports of SNEs, results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good
			C2: In 8 cases, countries, bilateral and multilateral partners have dedicated resources (e.g. GDI, World Bank, Social and Environmental Safeguards) to the implementation of reform measures by IOs	document analysis, interviews	1. progress reports, SNEs reports, results matrix 2. Interviews with GIZ project staff and stakeholders in IOs	good
	Standard	How does the project contribute via activities, instruments and outputs to the achievement of the project objective (outcome)? (contribution analysis approach)	This question will be answered through the synthesis of results of the questions in dimension 1 and 2 of the effectiveness criterion			good
	Standard	Implementation strategy: Which factors in the implementation contribute successfully to or hinder the achievement of the project objective? (e.g. external factors, managerial set-up of project and company, cooperation management)	1. Success factors of the project cited by interview partners 2. Success factors cited in the project's documentation 3. Hindering factors of the project cited by interview partners 4. Hindering factors cited in the project's documentation	document analysis, interviews	1. results model, progress reports, modification offer, project proposal, reports of SNEs, capacity development strategy, assessment of individual secondments 2. Interviews with GIZ project staff, BMZ divisions, stakeholders in IO	good
	Standard	What other/alternative factors contributed to the fact that the project objective was achieved or not achieved?	1. Description of alternative factors cited in interviews 2. Description of alternative factors cited in progress reports	document analysis, interviews	1. project reports, reports of SNEs 2. Interviews with project staff, BMz divisions and stakeholders in IOs	good
	Standard	What would have happened without the project?	1. Description of hypotheses developed with interview partners concerning the outcome a) without the project's support b) with a different project strategy (e.g. no SNEs)	interviews	1. stakeholders in international partner organisations	good

No project-related (unintended) negative results have occurred – and if any negative results occurred the project responded adequately. The occurrence of additional (not formally agreed) positive results has been monitored and additional opportunities for further positive results have been seized. Max. 30 points	Standard	Which (unintended) negative or (formally not agreed) positive results does the project produce at output and outcome level and why?	1. Description of (unintended) negative or (formally not agreed) positive results according to interview partners 2. Description of identified (unintended) negative or (formally not agreed) positive results according to the progress reports and the project's monitoring	interviews	1. interviews with GIZ project staff, BMZ divisions and stakeholders in IOs 2. Progress reports *contribution analysis*	good
	Standard	How were risks and assumptions (see also GIZ Safeguards and Gender system) as well as (unintended) negative results at the output and outcome level assessed in the monitoring system (e.g. 'Kompass')? Were risks already known during the concept phase?	1. The degree to which (unintended) negative results were tracked in the project's monitoring system 2. The degree to which potential negative results were already reflected in the project proposal 3. Degree to which monitoring system was adjusted continuously as new risks were identified	document analysis, interviews	1. progress reports, modification offer, project proposal, results model 2. Interviews with GIZ project staff, BMZ divisions *contribution analysis*	good
	Standard	What measures have been taken by the project to counteract the risks and (if applicable) occurred negative results? To what extent were these measures adequate?	1. Description of mitigation strategies adopted by intervention towards risks 2. Identification of concrete examples for steering to avoid or counteract risks	document analysis, interviews	1. progress reports, reports of SNEs 2. Interviews with GIZ project staff *contribution analysis*	good
	Standard	To what extend were potential (not formally agreed) positive results at outcome level monitored and exploited?	1.Description of how unanticipated positive outcomes were identified in the project's monitoring 2. Description of exploitation of unintended positive results at outcome level according to different stakeholders a) Project staff b) BMZ c) IOs	document analysis, interviews	1. progress reports, reports of SNEs 2. Interviews with GIZ project staff, BMZ, stakeholders in IOs *contribution analysis*	good

(1) The first and the second evaluation dimensions are interrelated: if the contribution of the project to the objective achievement is low (2nd evaluation dimension) this must be considered for the assessment of the first evaluation dimension also.

(2) Risks in the context of conflict, fragility and violence: e.g. contextual (e.g. political instability, violence, economic crises, migration/refugee flows, drought, etc.), institutional (e.g. weak partner capacity, fiduciary risks, corruption, staff turnover, investment risks) and personnel (murder, robbery, kidnapping, medical care, etc.). For more details see: GIZ (2014): 'Context- and conflict-sensitive results-based monitoring system (RBM). Supplement to: The 'Guidelines on designing and using a results-based monitoring system (RBM) system', pp. 27–28.

OECD-DAC criterion RELEVANCE (max. 100 points)

Assessment dimensions	**Filter - Project Type**	**Evaluation questions**	**Evaluation indicators**	**Data collection methods** (e.g. interviews, focus group discussions, documents, project/partner monitoring system, workshop, survey, etc.)	**Data sources** (list of relevant documents, interviews with specific stakeholder categories, specific monitoring data, specific workshop(s), etc.)	**Evidence strength** (moderate, good, strong)

	The project concept (1) is in line with the relevant strategic reference frameworks. Max. 30 points	Standard	Which strategic reference frameworks exist for the project? (e.g. national strategies incl. national implementation strategy for 2030 Agenda, regional and international strategies, sectoral, cross-sectoral change strategies, if bilateral project especially partner strategies, internal analysis frameworks e.g. safeguards and gender (2))	1. Description of relevant policy frameworks mentioned in the project proposal and modification offer and referred to by stakeholders a) Eckpunktepapier für die multilaterale Entwicklungspolitik (2013) b) potential additional strategies identified during the evaluation phase 2. Comparison of the strategic principles /frameworks with the project's strategy	Document analysis, interviews	1. Modification offer of 2017, progress reports, SNEs reports, strategy for multilateral development cooperation (2013), strategies of reform processes supported by the project 4. Interviews with BMZ divisions, GIZ project staff and stakeholders in IOs	good	
		Standard	To what extent is the project concept in line with the relevant strategic reference frameworks?		Document analysis, interviews		good	
		Standard	To what extent are the interactions (synergies/trade-offs) of the intervention with other sectors reflected in the project concept – also regarding the sustainability dimensions (ecological, economic and social)?	1. Identification and qualitative reviews of the interactions of the intervention with other sectors in the project's implementation 2. Qualitative assessment of the interactions in relation to the 4 dimensions of sustainability	Document analysis, interviews	1. Modification offer of 2017, progress reports, strategies for multilateral development cooperation (BMZ: 2013) and government strategies on Agenda 2030 (e.g. report to (HLPF) 2016) 3. Interviews with BMZ divisions, stakeholders in IOs and GIZ project staff	good	
		Standard	To what extent is the project concept in line with the Development Cooperation (DC) programme (If applicable), the BMZ country strategy and BMZ sectoral concepts?	1. Description of relevant policy frameworks mentioned in the project proposal and modification offer and referred to by stakeholders. a) Eckpunktepapier für die multilaterale Entwicklungspolitik (2013) b) potential additional strategies identified during the evaluation phase 2. Comparison of the strategic principles /frameworks with the project's strategy	document analysis, interviews	1. project proposal, modification offer 2017, progress reports, SNEs reports, Eckpunktepapier für die multilaterale Entwicklungspolitik (2013), potential additional strategies identified during the evaluation phase 2. interviews with BMZ divisions, GIZ project staff	good	
		Standard	To what extend is the project concept in line with the (national) objectives of the 2030 Agenda? To which Sustainable Development Goals (SDG) is the project supposed to contribute?	1. Description of national objectives of the 2030 Agenda to which the project contributed. 2. Description of the IO reform processes supported by the project 3. Qualitative assessment of the degree to which the supported IO reform processes were relevant to Agenda 2030	Document analysis, interviews	1. project offer, modification offer of 2017, progress reports, reports of SNEs, government strategies on Agenda 2030 (e.g. report to HLPF 2016), documents of IOs 3. interviews with BMZ divisions, stakeholders in IOs and GIZ project staff	good	
		Standard	To what extend is the project concept subsidiary to partner efforts or efforts of other relevant organisations (subsidiarity and complementarity)?	1. Number of partners who confirm that the project is complementary and subsidiary to their efforts. 2. Qualitative assessment of the coherence between the partners' programmes and strategies and the project concept	document analysis, interviews	1. documents of IO on reform process, modification offer 2017, progress reports, reports of SNEs 2. interviews with stakeholders in international partner organisations	good	

		and SV/GV	To what extent does the project complement bilateral or regional projects? To what extent does it complement other global projects?	1. Identification of other bilateral or regional projects which the project complemented 2. Description of interview partner's assessment on how the project complemented bilateral or regional projects a) BMZ divisions b) GIZ project staff c) GIZ staff of other bilateral or regional projects d) stakeholders in IOs 3. Qualitative assessment of the degree to which the project complemented those other projects	document analysis, interviews	1. progress reports, modification offer, project proposal 2. interviews with GIZ project staff, BMZ divisions, stakeholders in IOs and (potentially) GIZ staff of other bilateral or regional projects	good
		and SV/GV	To what extent is the measure geared towards solving a global challenge that cannot only be effectively addressed bilaterally/ regionally?	This question will be answered through the synthesis of results			good
		and ICR	To what extent does the project complement bilateral or regional projects? To what extent does it complement other global projects?	This question appears twice in the matrix			
		and ICR	To what extent is the measure geared towards solving a global challenge that cannot only be effectively addressed bilaterally/ regionally?	This question appears twice in the matrix			
		and ICR	To what extent does the measure close gaps in the solution of global development problems where classical multilateralism reaches its limits?	This question will be answered through the synthesis of results			good
	The project concept (1) matches the needs of the target group(s). Max. 30 points	Standard	To what extent is the chosen project concept geared to the core problems and needs of the target group(s)?	1. Identification of needs of the target group through interviews 2. Comparison of the identified needs with the project's strategy	interviews, document analysis	1. interviews with stakeholders in IOs and GIZ project staff 2. project offer, modification offer, progress reports, SNEs reports	good
		Standard	How are the different perspectives, needs and concerns of women and men represented in the project concept?	1. Qualitative assessment of whether and how the project pushed for including the different perspectives, needs and concerns of women and men in the supported reform processes	document analysis, interviews	1. modification offer, progress reports, reports of SNEs, documents of IOs 2. interviews with GIZ project staff, BMZ and stakeholders in IOs	moderate
		Standard	To what extent was the project concept designed to reach particularly disadvantaged groups (Leave No One Behind (LNOB) principle, as foreseen in the Agenda 2030)? How were identified risks and potentials for human rights and gender aspects included into the project concept?	1. Qualitative assessment of whether and how the project pushed for including the LNOB principle in the reform processes it supported	document analysis, interviews	1. modification offer, progress reports, reports of SNEs, documents of IOs 2. interviews with GIZ project staff and stakeholders in IOs	moderate

	Standard	To what extent are the intended impacts regarding the target group(s) realistic from today's perspective and the given resources (time, financial, partner capacities)?	1. Qualitative assessment of the feasibility of reaching the intended impacts based on the perspectives of different stakeholders: a) GIZ project staff b) BMZ c) International organisations	document analysis, interviews	1. project offer, modification offer, progress reports, documents of IOs 2. interviews with GIZ project staff, BMZ and stakeholders in IOs	good
The project concept (1) is adequately designed to achieve the chosen project objective. Max. 20 points	Standard	Assessment of current results model and results hypotheses (theory of change, ToC) of actual project logic: - To what extent is the project objective realistic from today's perspective and the given resources (time, financial, partner capacities)? - To what extent are the activities, instruments and outputs adequately designed to achieve the project objective? - To what extent are the underlying results hypotheses of the project plausible? - To what extent is the chosen system boundary (sphere of responsibility) of the project (including partner) clearly defined and plausible? - Are potential influences of other donors/organisations outside of the project's sphere of responsibility adequately considered? - To what extent are the assumptions and risks for the project complete and plausible?	1. Qualitative assessment of the plausibility of causal hypotheses in the results models 2. Qualitative assessment of the plausibility of risks, assumptions and external factors named in the results model 3. Qualitative assessment of the implementation strategies 5. Qualitative assessment of the system boundaries according to different stakeholders a) Project staff b) BMZ c) IOs	document analysis, interviews	1. results model, modification offer, progress reports, documents on SNEs 2.interviews with GIZ project staff, BMZ division, stakeholders in IOs and other stakeholders (e.g. GEDO, StäV, members of GDI)	good
	Standard	To what extent does the strategic orientation of the project address potential changes in its framework conditions?	1. The extent to which changes in the framework conditions for the project are reflected in the project's progress reports (if applicable) 2. Qualitative assessment of whether and how the project's strategy was able to react to contingencies	document analysis, interviews	1. project proposal, modification offer, progress reports, results model 2. Interviews with GIZ project staff and BMZ	good
	Standard	How is/was the complexity of the framework conditions and guidelines handled? How is/was any possible overloading dealt with and strategically focused?	1. Degree to which the intervention can describe challenges regarding the framework conditions and guidelines as well as situations of overloading 2. Degree to which the intervention can describe coping strategies to deal with the named challenges	document analysis, interviews	1. project proposal, modification offer, progress reports, results model 2. Interviews with GIZ project staff and BMZ	good
The project concept (1) was adapted to changes in line with requirements and readapted where applicable. Max. 20 points	Standard	What changes have occurred during project implementation? (e.g. local, national, international, sectoral, including state-of-the-art sectoral know-how)?	1. Identification of changes/adaptions in the project strategy (if necessary) due to changed framework conditions 2. Qualitative assessment of the degree to which the project is capable of providing an overview of changes in the implementation that resulted	document analysis, interviews	1. project proposal, modification offer, progress reports, results model 2. Interviews with GIZ project staff and BMZ	good

	Standard	How were the changes dealt with regarding the project concept?	from changing framework conditions			good

(1) The 'project concept' encompasses project objective and theory of change (ToC, see 3) with activities, outputs, instruments and results hypotheses as well as the implementation strategy (e.g. methodological approach, CD-strategy, results hypotheses).

(2) In the GIZ Safeguards and Gender system risks are assessed before project start regarding following aspects: gender, conflict, human rights, environment and climate. For the topics gender and human rights not only risks but also potentials are assessed. Before introducing the new safeguard system in 2016 GIZ used to examine these aspects in separate checks.

(3) Theory of Change = GIZ results model = graphic illustration and narrative results hypotheses.

(4) Deescalating factors/ connectors: e.g. peace-promoting actors and institutions, structural changes, peace-promoting norms and behaviour. For more details on 'connectors' see: GIZ (2007): 'Peace and Conflict Assessment (PCA). Ein methodischer Rahmen zur konflikt- und friedensbezogenen Ausrichtung von EZ-Maßnahmen', p. 55/135135.

(5) Escalating factors/ dividers: e.g. destructive institutions, structures, norms and behaviour. For more details on 'dividers' see: GIZ (2007): 'Peace and Conflict Assessment (PCA). Ein methodischer Rahmen zur konflikt- und friedensbezogenen Ausrichtung von EZ-Maßnahmen', p. 135135.

(6) All projects in fragile contexts, projects with FS1 or FS2 markers and all transitional aid projects have to weaken escalating factors/dividers and have to mitigate risks in the context of conflict, fragility and violence. Projects with FS1 or FS2 markers should also consider how to strengthen deescalating factors/ connectors and how to address peace needs in its project objective/sub-objective?

OECD-DAC criterion IMPACT (max. 100 points)

Assessment dimensions	**Filter - Project type**	**Evaluation questions**	**Evaluation indicators**	**Data collection methods** (e.g. interviews, focus group discussions, documents, project/partner monitoring system, workshop, survey, etc.)	**Data sources** (list of relevant documents, interviews with specific stakeholder categories, specific monitoring data, specific workshop(s), etc.)	**Evidence strength** (moderate, good, strong)
The intended overarching development results have occurred or are foreseen (plausible reasons). (1) Max. 40 points	Standard	To which overarching development results is the project supposed to contribute (cf. module and programme proposal with indicators/ identifiers if applicable, national strategy for implementing 2030 Agenda, SDGs)? Which of these intended results at the impact level can be observed or are plausible to be achieved in the future?	1. Description of the status of the results at impact level of the results model 2.Qualitative assessment of the degree to which the supported reform processes plausibly contribute to reaching SDG 17	document analysis, interviews	1. progress reports, documents of IOs on reforms, results model 2. interviews with GIZ project staff, BMZ divisions and stakeholders in IOs	good
	and ICR	To what extent have the ICR criteria contributed to strengthening overarching development results?	1. Qualitative assessment of the contribution of ICR criteria to strengthening the results at impact level.	document analysis, interviews	1. progress reports, modification offer, project proposal 2. interviews with GIZ project staff, BMZ divisions	good
	Standard	Indirect target group and 'Leave No One Behind' (LNOB): Is there evidence of results achieved at indirect target group level/specific groups of population? To what extent have targeted marginalised groups (such as women, children, young people, elderly, people with disabilities, indigenous peoples, refugees, IDPs and migrants, people living with HIV/AIDS and the poorest of the poor) been reached?	Not applicable due to characteristics of the project			
The project objective (outcome) of the project contributed to the occurred or foreseen overarching development results (impact).(1) Max. 30 points	Standard	To what extent is it plausible that the results of the project on outcome level (project objective) contributed or will contribute to the overarching results? (contribution analysis approach)	1. Qualitative assessment of the plausibility of the results model 2. Qualitative assessment of the plausibility of the contribution of the project to the results at impact level of the results model	document analysis, interviews *contribution analysis*	1. progress reports, reports of SNEs, documents of IOs on reforms, results model, results matrix 2. interviews with GIZ project staff, BMZ divisions and stakeholders in IOs	strong

		Standard	What are the alternative explanations/factors for the overarching development results observed? (e.g. the activities of other stakeholders, other policies)	1. Qualitative assessment of alternative explanations/factors which explain the results at impact level, including: a) contributions by other members/shareholders of the IO b) work of IO itself (e.g. IO team working on reform process) c) additional explanations/factors identified during the evaluation	interviews *contribution analysis, process tracing*	1. interviews with stakeholders in IOs	moderate
		Standard	To what extent is the impact of the project positively or negatively influenced by framework conditions, other policy areas, strategies or interests (German ministries, bilateral and multilateral development partners)? How did the project react to this?		document analysis, interviews *contribution analysis*	1.progress reports 2. interviews with GIZ project staff, stakeholders in IOs, BMZ divisions and other stakeholders (e.g.StäV, GEDO)	strong
		Standard	What would have happened without the project?	1. Qualitative assessment of alternative developments of different stakeholders in the case of a) non-existence of the project b) alternative implementation strategies of the project	interviews *contribution analysis*	1. interviews with GIZ project staff, stakeholders in IOs and other stakeholders (e.g. StäV, GEDO, members of GDI)	moderate
		Standard	To what extent has the project made an active and systematic contribution to widespread impact and were scaling-up mechanisms applied (2)? If not, could there have been potential? Why was the potential not exploited? To what extent has the project made an innovative contribution (or a contribution to innovation)? Which innovations have been tested in different regional contexts? How are the innovations evaluated by which partners?	1. Qualitative assessment of the projects contribution to widespread impact with regards to: a) relevance b) quality c) quantity d) sustainability e) scaling-up approaches	document analysis, interviews contribution analysis, *process tracing*	1. progress reports 2. interviews with GIZ project staff, stakeholders in IOs and other stakeholders (e.g. partners of GDI, beneficiaries of ESF implementation pilots)	moderate
		and IZR	To what extent has the project made an innovative contribution (or a contribution to innovation)? Which innovations have been tested in different regional contexts? How are the innovations evaluated by which partners?	1. Qualitative assessment of the projects innovative contributions	document analysis, interviews *contribution analysis*	1. progress reports 2. interviews with GIZ project staff, stakeholders in IOs and other stakeholders (f.e. members of GDI, beneficiaries of ESF implementation pilots)	good
	No project-related (unintended) negative results at impact level have occurred – and if any negative results occurred the project responded adequately. The occurrence of additional (not formally agreed) positive results at impact level has been monitored and additional opportunities for further positive results have been seized. Max. 30 points	Standard	Which (unintended) negative or (formally not agreed) positive results at impact level can be observed? Are there negative trade-offs between the ecological, economic and social dimensions (according to the three dimensions of sustainability in the Agenda 2030)? Were positive synergies between the three dimensions exploited?	1. Description of unintended results at impact level a) Positive b) Negative	document analysis, interviews *contribution analysis*	1. Progress reports, reports of SNEs 2. interviews with GIZ project staff, stakeholders in IOs and other stakeholders (e.g. partners of GDI, beneficiaries of ESF implementation pilots)	moderate
		Standard	To what extent were risks of (unintended) results at the impact level assessed in the monitoring system (e.g. 'Kompass')? Were risks already known during the planning phase?	1. Risks of unintended negative results are included in the intervention's monitoring systems 2. Unintended negative results were already reflected in the project's proposal and modification offer	document analysis, interviews *contribution analysis*	1. Results model, progress reports, project proposal, modification offer 2. interviews with GIZ project staff, BMZ divisions, stakeholders in IOs	good
		Standard	What measures have been taken by the project to avoid and counteract the risks/negative results/trade-offs (3)?	1. Description of mitigation measures by the intervention towards a) risks b) negative results c) trade-offs	document analysis, interviews *contribution analysis*	1. progress reports, project proposal, modification offer 2. interviews with GIZ project staff, BMZ divisions and stakeholders in IOs	good

	Standard	To what extent have the framework conditions played a role in regard to the negative results? How did the project react to this?	1. Description of framework conditions that influence impacts a) changes in the political landscape b) institutional environment (BMZ) c) institutional environment in IOs (target group) d) strategies and activities of other German ministries e) activities by other IO member states / shareholders 2. Description of mitigation measures by the intervention towards risks	document analysis, interviews *contribution analysis*	1. progress reports, project offer, modification offer 2. interviews with GIZ project staff, BMZ divisions, stakeholders in IOs and other stakeholders (GEDO, StäV)	good
	Standard	To what extent were potential (not formally agreed) positive results and potential synergies between the ecological, economic and social dimensions monitored and exploited?	1. Description of exploitation of unintended positive results and synergies at impact level according to different stakeholders a) GIZ project staff b) BMZ c) international organisations 2. Documentation of unintended positive results at impact level in the monitoring system	document analysis, interviews *contribution analysis*	1. progress reports, results matrix 2. interviews with GIZ project staff, BMZ divisions and stakeholders in IOSs	good

(1) The first and the second evaluation dimensions are interrelated: if the contribution of the project outcome to the impact is low or not plausible (2nd evaluation dimension) this must be considered for the assessment of the first evaluation dimension also

(2) Broad impact (in German 'Breitenwirksamkeit') is defined by 4 dimensions: relevance, quality, quantity, sustainability. Scaling-up approaches can be categorised as vertical, horizontal, functional or combined. See GIZ (2014) 'Corporate strategy evaluation on scaling up and broad impact: The path: scaling up, the goal: broad impact' (https://www.giz.de/de/downloads/giz2015-en-scaling-up.pdf)

(3) Risks, negative results and trade-offs are separate aspects and are all to be considered

OECD-DAC criterion EFFICIENCY (max. 100 points)

Assessment dimensions	**Filter - project type**	**Evaluation questions**	**Evaluation indicators (pilot phase for indicators - only available in German so far)**	**Data collection methods** (e.g. interviews, focus group discussions, documents, project/partner monitoring system, workshop, survey, etc.)	**Data sources** (list of relevant documents, interviews with specific stakeholder categories, specific monitoring data, specific workshop(s), etc.)	**Evidence strength** (moderate, good, strong)
The project's use of resources is appropriate with regard to the outputs achieved. [Production efficiency: Resources/Outputs] Max. 70 points	Standard	To what extent are there deviations between the identified costs and the projected costs? What are the reasons for the identified deviation(s)?	Das Vorhaben steuert seine Ressourcen gemäß des geplanten Kostenplans (Kostenzeilen). Nur bei nachvollziehbarer Begründung erfolgen Abweichungen vom Kostenplan.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment*	1. cost-output allocation, efficiency tool 2. progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard	Focus: To what extent could the outputs have been maximised with the same amount of resources and under the same framework conditions and with the same or better quality (maximum principle)? (methodological minimum standard: Follow-the-money approach)	Das Vorhaben reflektiert, ob die vereinbarten Wirkungen mit den vorhandenen Mitteln erreicht werden können.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Das Vorhaben steuert seine Ressourcen gemäß der geplanten Kosten für die vereinbarten Leistungen (Outputs). Nur bei nachvollziehbarer Begründung erfolgen Abweichungen von den Kosten. Die übergreifenden Kosten des Vorhabens stehen in einem angemessen Verhältnis zu den Kosten für die Outputs. Die durch ZAS Aufschriebe erbrachten Leistungen haben einen nachvollziehbaren Mehrwert für die Erreichung der Outputs des Vorhabens.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good

Standard		Die übergreifenden Kosten des Vorhabens stehen in einem angemessen Verhältnis zu den Kosten für die Outputs.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good	
Standard		Die durch ZAS Aufschriebe erbrachten Leistungen haben einen nachvollziehbaren Mehrwert für die Erreichung der Outputs des Vorhabens.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good	
Standard	Focus: To what extent could outputs have been maximised by reallocating resources between the outputs? (methodological minimum standard: Follow-the-money approach)	Das Vorhaben steuert seine Ressourcen, um andere Outputs schneller/ besser zu erreichen, wenn Outputs erreicht wurden bzw. diese nicht erreicht werden können (Schlussevaluierung). Oder: Das Vorhaben steuert und plant seine Ressourcen, um andere Outputs schneller/ besser zu erreichen, wenn Outputs erreicht wurden bzw. diese nicht erreicht werden können (Zwischenevaluierung).	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good	
Standard	Were the output/resource ratio and alternatives carefully considered during the design and implementation process – and if so, how? (methodological minimum standard: Follow-the-money approach)	Das im Modulvorschlag vorgeschlagene Instrumentenkonzept konnte hinsichtlich der veranschlagten Kosten in Bezug auf die angestrebten Outputs des Vorhabens gut realisiert werden.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good	
Standard		Die im Modulvorschlag vorgeschlagene Partnerkonstellation und die damit verbundenen Interventionsebenen konnte hinsichtlich der veranschlagten Kosten in Bezug auf die angestrebten Outputs des Vorhaben gut realisiert werden.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good	
Standard		Der im Modulvorschlag vorgeschlagene thematische Zuschnitte für das Vorhaben konnte hinsichtlich der veranschlagten Kosten in Bezug auf die angestrebten Outputs des Vorhabens gut realisiert werden.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good	
Standard		Die im Modulvorschlag beschriebenen Risiken sind hinsichtlich der veranschlagten Kosten in Bezug auf die angestrebten Outputs des Vorhabens gut nachvollziehbar.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good	
Standard		Die im Modulvorschlag beschriebene Reichweite des Vorhabens (z.B. Regionen) konnte hinsichtlich der veranschlagten Kosten in Bezug auf die angestrebten Outputs des Vorhabens voll realisiert werden.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good	
Standard		Der im Modulvorschlag beschriebene Ansatz des Vorhabens hinsichtlich der zu erbringenden Outputs entspricht unter den gegebenen Rahmenbedingungen dem state-of-the-art.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good	

	Standard	For interim evaluations based on the analysis to date: To what extent are further planned expenditures meaningfully distributed among the targeted outputs?	This question will not be answered, because the evaluation is an ex post evaluation.			
The project's use of resources is appropriate with regard to achieving the projects objective (outcome). [Allocation efficiency: Resources/Outcome] Max. 30 points	Standard	To what extent could the outcome (project objective) have been maximised with the same amount of resources and the same or better quality (maximum principle)?	Das Vorhaben orientiert sich an internen oder externen Vergleichsgrößen, um seine Wirkungen kosteneffizient zu erreichen.	cost-output allocation, efficiency tool, interviews *Analysis based on plausibility analysis using the hypotheses formulated in the process tracing in the effectiveness analysis*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard	Were the outcome-resources ratio and alternatives carefully considered during the conception and implementation process – and if so, how? Were any scaling-up options considered?	Das Vorhaben steuert seine Ressourcen zwischen den Outputs, so dass die maximalen Wirkungen im Sinne des Modulziels erreicht werden. (Schlussevaluierung). Oder: Das Vorhaben steuert und plant seine Ressourcen zwischen den Outputs, so dass die maximalen Wirkungen im Sinne des Modulziels erreicht werden. (Zwischenevaluierung).	cost-output allocation, efficiency tool, interviews *Analysis based on plausibility analysis using the hypotheses formulated in the process tracing in the effectiveness analysis*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Das im Modulvorschlag vorgeschlagene Instrumentenkonzept konnte hinsichtlich der veranschlagten Kosten in Bezug auf das angestrebte Modulziel des Vorhabens gut realisiert werden.	cost-output allocation, efficiency tool, interviews *Analysis based on plausibility analysis using the hypotheses formulated in the process tracing in the effectiveness analysis*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Die im Modulvorschlag vorgeschlagene Partnerkonstellation und die damit verbundenen Interventionsebenen konnte hinsichtlich der veranschlagten Kosten in Bezug auf das angestrebte Modulziel des Vorhaben gut realisiert werden.	cost-output allocation, efficiency tool, interviews *Analysis based on plausibility analysis using the hypotheses formulated in the process tracing in the effectiveness analysis*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Der im Modulvorschlag vorgeschlagene thematische Zuschnitte für das Vorhaben konnte hinsichtlich der veranschlagten Kosten in Bezug auf das angestrebte Modulziel des Vorhabens gut realisiert werden.	cost-output allocation, efficiency tool, interviews *Analysis based on plausibility analysis using the hypotheses formulated in the process tracing in the effectiveness analysis*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Die im Modulvorschlag beschriebenen Risiken sind hinsichtlich der veranschlagten Kosten in Bezug auf das angestrebte Modulziel des Vorhabens gut nachvollziehbar.	cost-output allocation, efficiency tool, interviews *Analysis based on plausibility analysis using the hypotheses formulated in the process tracing in the effectiveness analysis*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Die im Modulvorschlag beschriebene Reichweite des Vorhabens (z.B. Regionen) konnte hinsichtlich der veranschlagten Kosten in Bezug auf das angestrebte Modulziel des Vorhabens voll realisiert werden.	cost-output allocation, efficiency tool, interviews *Analysis based on plausibility analysis using the hypotheses formulated in the process tracing in the effectiveness analysis*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Der im Modulvorschlag beschriebene Ansatz des Vorhabens hinsichtlich des zu erbringenden Modulziels entspricht unter den gegebenen Rahmenbedingungen dem state-of-the-art.	cost-output allocation, efficiency tool, interviews *Analysis based on plausibility analysis using the hypotheses formulated in the process tracing in the effectiveness analysis*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure	good

					3. interviews with GIZ project staff, stakeholders in IOs	
	Standard	To what extent were more results achieved through cooperation / synergies and/or leverage of more resources, with the help of other ministries, bilateral and multilateral donors and organisations (e.g. co-financing) and/or other GIZ projects? If so, was the relationship between costs and results appropriate or did it even improve efficiency?	Das Vorhaben unternimmt die notwendigen Schritte, um Synergien mit Interventionen anderer Geber auf der Wirkungsebene vollständig zu realisieren.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer,progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Wirtschaftlichkeitsverluste durch unzureichende Koordinierung und Komplementarität zu Interventionen anderer Geber werden ausreichend vermieden.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Das Vorhaben unternimmt die notwendigen Schritte, um Synergien innerhalb der deutschen EZ vollständig zu realisieren.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Wirtschaftlichkeitsverluste durch unzureichende Koordinierung und Komplementarität innerhalb der deutschen EZ werden ausreichend vermieden.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*	1. cost-output allocation, efficiency tool 2. project proposal, modification offer, progress reports, steering structure 3. interviews with GIZ project staff, stakeholders in IOs	good
	Standard		Die Kombifinanzierung hat zu einer signifikanten Ausweitung der Wirkungen geführt bzw. diese ist zu erwarten.	These questions will not be answered, because the project did not receive combi-financing		
	Standard		Durch die Kombifinanzierung sind die übergreifenden Kosten im Verhältnis zu den Gesamtkosten nicht überproportional gestiegen.			
	Standard		Die Partnerbeiträge stehen in einem angemessenen Verhältnis zu den Kosten für die Outputs des Vorhabens.	cost-output allocation, efficiency tool, interviews *Analysis based on the follow-the-money approach, Evaluators' assessment, possible subjectivity bias mitigated by researcher triangulation*		good

OECD-DAC criterion SUSTAINABILITY (max. 100 points)

Assessment dimensions	**Filter - project type**	**Evaluation questions**	**Evaluation indicators**	**Data collection methods** (e.g. interviews, focus group discussions, documents, project/partner monitoring system, workshop, survey, etc.)	**Data sources** (list of relevant documents, interviews with specific stakeholder categories, specific monitoring data, specific workshop(s), etc.)	**Evidence strength** (moderate, good, strong)
Prerequisite for ensuring the long-term success of the project: Results are anchored in (partner) structures.	Standard	What has the project done to ensure that the results can be sustained in the medium to long term by the partners themselves?	1. Description of the project's strategy to ensure sustainability 2. Description of the project's activities with regards to sustainability	document analysis, interviews	1. progress reports, reports of SNEs 2. interviews with GIZ project staff and stakeholders in IOs	good

Max. 50 points	Standard	In what way are advisory contents, approaches, methods or concepts of the project anchored/institutionalised in the (partner) system?	1. Description of contents developed within the project a. used by IO b. not used by IO	document analysis, interviews	1. progress reports, SNEs reports, IO documents 2. interviews with GIZ project staff and stakeholders of IOs	good
	Standard	To what extent are the results continuously used and/or further developed by the target group and/or implementing partners?				good
	Standard	To what extent are resources and capacities at the individual, organisational or societal/political level in the partner country available (long-term) to ensure the continuation of the results achieved?	1. Description of resources and capacities in IOs to continue processes supported by the project.	interviews	1. interviews with GIZ project staff, stakeholders in IOs, BMZ divisions and other stakeholders (e.g. d.i.e., GEDO, StäV, members of GDI)	good
	Standard	If no follow-on measure exists: What is the project's exit strategy? How are lessons learned for partners and GIZ prepared and documented?	1. Qualitative assessment of the project's exit strategy	document analysis, interviews	1. progress reports, reports of SNEs 2. interviews with GIZ project staff	good
	and Fragility	To what extent was the project able to ensure that escalating factors/dividers (1) in the context of conflict, fragility and violence have not been strengthened (indirectly) by the project in the long-term? To what extent was the project able to strengthen deescalating factors/connectors (2) in a sustainable way?				
Forecast of durability: Results of the project are permanent, stable and long-term resilient. Max. 50 points	Standard	To what extent are the results of the project durable, stable and resilient in the long term under the given conditions?	1. Description of risks potentially affecting sustainability 2. Assessment of extent to which project could influence risks 3. Description of mitigation strategies adopted by the project	document analysis, interviews	1. project offer, modification offer, progress reports, reports of SNEs 2. interviews with GIZ project staff, BMZ divisions, stakeholders in IOs and other stakeholders (e.g. GEDO, d.i.e, StäV, members of GDI)	moderate
	Standard	What risks and potentials are emerging for the durability of the results and how likely are these factors to occur? What has the project done to reduce these risks?				moderate

(1) Escalating factors/ dividers: e.g. destructive institutions, structures, norms and behaviour. For more details on 'dividers' see: GIZ (2007): 'Peace and Conflict Assessment (PCA). Ein methodischer Rahmen zur konflikt- und friedensbezogenen Ausrichtung von EZ-Maßnahmen', p. 135.
(2) Deescalating factors/ connectors: e.g. peace-promoting actors and institutions, structural changes, peace-promoting norms and behaviour. For more details on 'connectors' see: GIZ (2007): 'Peace and Conflict Assessment (PCA). Ein methodischer Rahmen zur konflikt- und friedensbezogenen Ausrichtung von EZ-Maßnahmen', p. 55/135.

Additional evaluation questions

Assessment dimensions	Evaluation questions	Evaluation indicators	Data collection methods (e.g. interviews, focus group discussions, documents, project/partner monitoring system, workshop, survey, etc.)	Data sources (list of relevant documents, interviews with specific stakeholder categories, specific monitoring data, specific workshop(s), etc.)	Evidence strength (moderate, good, strong)
Impact and sustainability (durability) of predecessor project(s)	Which of the intended impact of the predecessor project(s) can (still/now) be observed?	This question will not be answered as the predecessor will not be assessed.			
	Which of the achieved results (output, outcome) from predecessor project(s) can (still) be observed?	This question will not be answered as the predecessor will not be assessed.			
	To what extent are these results of the predecessor project(s) durable, stable and resilient in the long-term under the given conditions?	This question will not be answered as the predecessor will not be assessed.			
	In what way were results anchored/institutionalised in the (partner) system?	This question will not be answered as the predecessor will not be assessed.			
	How much does the current project build on the predecessor project(s)? Which aspects (including results) were used or integrated in the current project (phase)?	This question will not be answered as the predecessor will not be assessed.			
	How was dealt with changes in the project context (including transition phases between projects/phases)? Which important strategic decisions were made? What were the consequences?	This question will not be answered as the predecessor will not be assessed.			
	Which factors of success and failure can be identified for the predecessor project(s)?	This question will not be answered as the predecessor will not be assessed.			
Follow-on project (if applicable)	Based on the evaluations results: Are the results model including results hypotheses, the results-oriented monitoring system (WoM), and project indicators plausible and in line with current standards? If applicable, are there any recommendations for improvement?	This question will not be answered as no follow-on project exists.			
Additional evaluation questions	(1) Under which circumstances is a secondment as a technical cooperation instrument particularly strategic and effective?	1. Description of factors defining a strategic and effective secondment	document analysis, interviews	1. reports of SNEs 2. interviews with project staff and with stakeholders in IOs	good
	(2) How does the project staff in Germany contribute to maximising the results achieved by the iLTE?	synthesis of evaluation results and lessons learned identified with BMZ and project team			moderate
	(3) What should a results matrix and monitoring system look like to best allow flexibility while still steering towards the project's strategic objectives?	synthesis of evaluation results and lessons learned identified with BMZ and project team			moderate
	(4) Considering that multilateral cooperation is continuously changing, how were the project's measures and instruments adjusted to remain relevant and effective throughout the project's lifetime?	This question will be considered throughout the analysis as a cross-cutting scheme			moderate

(1) Please add additional questions of interests raised by the project including partner or target group during the inception phase that could not be included into the OECD-DAC criteria.

Photo credits and sources

© GIZ: Ranak Martin, Carlos Alba, Dirk Ostermeier, Ala Kheir

Disclaimer:

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

Maps:

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

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices:
Bonn und Eschborn

Friedrich-Ebert-Allee 32 + 36
53113 Bonn, Germany
T +49 228 44 60-0
F +49 228 44 60-17 66

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

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