KNOWING WHAT WORKS

# Evaluation of the average ratings of central project evaluations

Cross-sectional analysis – Main report

Conducted on behalf of GIZ by external evaluators

Deutsche Gesellschaft
für Internationale
Zusammenarbeit (GIZ) GmbH

# Imprint

As a service provider with worldwide operations in the fields of international cooperation for sustainable development and international education work, GIZ works with its partners to develop effective solutions that offer people better prospects and sustainably improve their living conditions. GIZ is a public-benefit federal enterprise and supports the German Government and a host of public and private sector clients in a wide variety of areas, including economic development and employment promotion, energy and the environment, and peace and security.

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


**Evaluators**
Lennart Raetzell, Matias Kraemer, Alexandra Hoppe and David Arens, Syspons GmbH

**Authors**
Lennart Raetzell, Matias Kraemer, Alexandra Hoppe and David Arens

**Consulting**:
Syspons GmbH
Prinzenstr. 84 Aufgang 1
10969 Berlin / Germany
T: +49 30 698 158 00
E: info@syspons.com
I: www.syspons.com



**Design, Coordination and Management**
Claudia Kornahrens, GIZ, Head of Section
Dr Vera Hundt, GIZ, Evaluation specialist
Benjamin Bräuer, GIZ, Evaluation specialist
Lucas Jacobs, GIZ, Evaluation specialist
Jana Moneke, GIZ, Evaluation specialist
Central Project Evaluation Section
GIZ Corporate Evaluation Unit

**Responsible**
Martha Gutierrez, GIZ, Director
GIZ Corporate Evaluation Unit

**Translation by**
Plunet - International Language Services by GIZ

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

**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
https://twitter.com/giz_gmbh

**Design**
now [nau], kommunikative & visuelle gestaltung,
Frankfurt

**Distribution**
GIZ, Bonn

Bonn 2024

# Contents

List of figures ..... 4

List of abbreviations ..... 5

1 Introduction ..... 6

Background and objectives of the cross-sectional analysis ..... 6

2 Methodological approach ..... 7

2.1 Object of the study ..... 7

2.2 Dataset and strength of evidence ..... 8

2.3 The analysis process ..... 9

Procedure during the preparatory phase ..... 9

Statistical processing and analysis of the CPE average ratings ..... 9

Content analysis of the evaluation reports ..... 10

Reporting and presentation ..... 11

3 Findings of the cross-sectional analysis ..... 11

3.1 Success rate of the CPEs ..... 11

3.2 Analysis of the CPEs and previous tools over time ..... 13

3.3 Analysis of the CPEs by evaluation criterion ..... 14

3.4 Analysis of the CPEs by project characteristic ..... 17

Project type ..... 17

Project size ..... 18

Project term ..... 19

3.5 Analysis of the CPEs by country and region ..... 20

Continents ..... 20

World regions ..... 21

Regional departments ..... 22

BMZ country categories ..... 24

Fragility ..... 25

3.6 Analysis of the CPEs by BMZ area and sector ..... 26

BMZ core areas ..... 27

BMZ areas of intervention ..... 27

3.7 Rating as a function of differences in evaluation ..... 29

Knock-out criteria ..... 29

The COVID-19 pandemic and its effect on CPEs ..... 30

The COVID-19 pandemic and its effects on project implementation ..... 32

Annex ..... 34

## List of figures

Figure 1: Population and sample ..... 8
Figure 2: Population and sample with success rate ..... 12
Figure 3: Success rate including 95% confidence interval ..... 13
Figure 4: Ratings over time, including previous evaluations (by reporting date) ..... 14
Figure 5: Distribution of ratings by evaluation criterion ..... 15
Figure 6: Average distribution of points ..... 16
Figure 7: Average ratings by over time (by reporting date) ..... 17
Figure 8: Ratings by project type *(N=207)* ..... 18
Figure 9: Ratings by project size (categories) *(N=207)* ..... 19
Figure 10: Ratings by project term (categories) *(N=207)* ..... 20
Figure 11: Average ratings by continent, taking sectors into account and BMZ core areas *(N=207)* ..... 21
Figure 12: Average ratings by world region, taking sectors into account and BMZ core areas *(N=207)* ..... 22
Figure 13: Ratings by operational department (regional departments of GIZ) *(N=207)* ..... 23
Figure 14: Ratings by BMZ country category *(N=138)* ..... 24
Figure 15: Ratings by fragility (as of reporting date) *(N=136)* ..... 26
Figure 16: Ratings by BMZ core area ..... 27
Figure 17: Ratings by BMZ area of intervention ..... 28
Figure 18: Ratings before and after introduction of the knock-out criteria *(N=207)* ..... 30
Figure 19: Average ratings by date of evaluation mission (effects of the COVID-19 pandemic) *(N=205)* ..... 31
Figure 20: Ratings by implementation modality of evaluation mission *(N=205)* ..... 32
Figure 21: Ratings by implementation modality of evaluation mission *(N=205)* ..... 33

## List of abbreviations

| | |
|---|---|
| BMZ | German Federal Ministry for Economic Cooperation and Development |
| CPE | Central project evaluation |
| CSA | Cross-sectional analysis |
| DAC | OECD Development Assistance Committee |
| EMC | Europe, Mediterranean, Central Asia |
| GIZ | Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH |
| GloBe | Sector and Global Programmes |
| ICR | International cooperation with regions programme |
| MV | Mean |
| N | Total number of the statistical population |
| n | Size of the sample |
| OECD | Organisation for Economic Co-operation and Development |
| ToC | Theory of change |
| ToRs | Terms of reference |

# 1 Introduction

## Background and objectives of the cross-sectional analysis

Syspons GmbH was commissioned by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH to conduct a **cross-sectional analysis** (CSA) of central project evaluations (CPEs) based on OECD DAC average and overall ratings (referred to as *CSA Insights*).[1]

CSAs prepare and summarise findings from evaluations by GIZ so that they can be used for company-wide learning. The *CSA Insights* aimed to gain comprehensive understanding from the **CPEs'** evaluation tool. GIZ has been using CPEs since 2017/2018 to regularly and systematically assess the results, cost-effectiveness and sustainability of projects that it implements together with partners on behalf of the German Federal Ministry for Economic Cooperation and Development (BMZ). CPEs for BMZ business address projects implemented by GIZ on behalf of BMZ in order to either support the partners' change projects or to work on sector and global issues. Evaluations are carried out on the basis of the six OECD DAC evaluation criteria: **relevance**, **coherence** (since the beginning of 2021), **effectiveness**, **efficiency**, **impact** and **sustainability** (BMZ, 2021).

GIZ's evaluations of projects commissioned by BMZ are essentially designed to serve three main purposes: to ensure transparency and accountability; to provide support for evidence-based decision-making; and to contribute to knowledge management and organisational learning. GIZ structures the planning, implementation and practical application of evaluations to optimise the contribution made by the evaluation process and the evaluation findings to these three basic functions.

Since 2020, the **COVID-19** pandemic has had a considerable effect on GIZ's work and the progress of the CPEs managed by the Evaluation Unit. Many projects have responded to the new situation by adapting or adding activities and changing outputs and objectives systems. At the same time, the pandemic presented the projects with a wide variety of challenges. For the CPEs, travel restrictions and measures to contain the pandemic meant that the usual fieldwork was replaced by remote missions. In accordance with the regulations in force in the respective country and the pandemic situation, either a semi-remote set-up or a purely remote set-up was applied.[2]

Against this background, as part of *CSA Insights* a cross-sectional analysis of CPEs completed to date was carried out based on ratings for the OECD DAC criteria. This analysis took the pandemic into account. The specific aim was to perform a **statistical analysis and presentation of the average and overall ratings for the OECD DAC criteria** for all CPEs available up to the deadline and determine possible explanatory factors. To this end, various questions and information requirements were defined together with the Evaluation Unit. These were incorporated into an **analysis grid**, in which they were further operationalised.

The report is structured as follows. Chapter 2 presents the methodological approach of the CSA. The chapter begins by describing the object of the study. It then discusses and assesses the dataset, the quality of the data and the strength of evidence. The evaluation process is subsequently explained. Chapter 3 presents and

[1] Besides analysing CPEs according to OECD DAC average and overall ratings, the cross-sectional analysis has two further components. These are: (1) determining the effects of the COVID-19 pandemic on projects and evaluations, and (2) analysing trends in selected thematic areas. These components were presented in separate GIZ reports.
[2] In the case of a semi-remote set-up, the local evaluators are able to move around in-country, whereas in a purely remote set-up they operate from home and the mission is carried out entirely virtually. In both scenarios, the international evaluators do not have to physically travel.

discusses the findings of the analysis. The findings follow two approaches: first, a statistical analysis of the CPEs was carried out using OECD DAC average and overall ratings for various criteria and thematic areas. Second, statistical trends (differences in average ratings) were described on this basis and explanatory factors were identified at selected points through a qualitative content analysis of the CPE reports. Chapter 4 contains the Annex.[3]

# 2 Methodological approach

This chapter presents the methodological approach of the CSA for the (statistical) assessment of the OECD DAC average and overall ratings of the CPEs. The first section begins by describing the object of the study. This is followed by an assessment of the dataset and the quality of the data for the analysis. The individual steps of the analysis are described in the final section on the analysis process.

## 2.1 Object of the study

The analysis covered **207 central project evaluations** that had been completed by the cut-off date (15 August 2022) and were available to GIZ. To select the projects to be evaluated in CPEs, GIZ works with a **representative sample** agreed on with BMZ. An annual (stratified) random sample of around 40% of the total population – all BMZ-funded projects with a commission value of at least EUR 3 million that expire in the following year – enables statistically representative conclusions to be drawn on the success rate of all projects completed over a **two-year period**. The CPE portfolio is supplemented by evaluations of projects that are proposed due to their particular strategic, political or methodological relevance.[4] The present analysis is limited to CPEs that were selected as part of the representative random sampling.

The 207 CPEs available were drawn from a **total of four annual samples**. These four CPE samples reflect **two representative selection procedures**. Figure 1 shows that, for the first two samples with a planned end of project in 2019/2020, a population of 425 projects was available. From this population, 165 projects were selected for evaluation by the CPEs. This corresponds to a coverage rate of 39%. Findings were already available for 150 of these evaluations; while for 15, the evaluation was not yet completed (91% completed). For the last two samples with a planned end of project in 2021/2022, 179 projects were selected from a population of 461. This also results in a coverage rate of 39%. As the samples were drawn later, the number of evaluations already carried out is lower. For these samples, 57 sets of findings were available, while 122 CPEs were still ongoing (32% completed).[5]

[3] In the overall context of the CSA, the present report is complemented by two further standalone reports for the other two defined objectives.
[4] GIZ, 2022: https://www.giz.de/de/downloads/giz2022-de-instrumentenpapier-zentrale-projektevaluierungen-bmz.pdf
[5] In the following section, for the sake of clarity a distinction is made only between the first sample in 2019/2020 and the second sample in 2021/2022.

Figure 1: Population and sample




## 2.2 Dataset and strength of evidence

The central dataset for the statistical evaluation comprised first of all a list of ratings and basic data for all commissioned CPEs that were made available to the Syspons team by GIZ at the beginning of the project. Second, the Syspons team was provided with the evaluation reports of all completed CPEs to identify possible explanations for the differences in the average ratings (statistical trends) as part of a content analysis.

Regarding the **strength of evidence of the statistical analysis of the success rate of the CPEs** for the success rate of the population, it should first be noted that the explanatory power still contains major statistical uncertainties. This is because, as described in the previous section, data for the samples has not yet been collected in full. The confidence intervals described in section 3.1 show the current strength of evidence of the conclusions on the success rate of the population. Regarding the **basic data** for the statistical analysis of the OECD DAC ratings, it should be noted that the basic data already collected and provided by GIZ contained some gaps that had to be filled in order to conduct a complete analysis. Accordingly, the Syspons team prepared the dataset and supplemented or updated it. For certain basic data (for example, the thematic classification of the evaluated projects according to BMZ's core areas and areas of intervention), as agreed the Syspons team made this determination themselves, rather than according to officially defined criteria. The team made a particular point of using information from the evaluation reports of the completed CPEs to minimise the risk of affecting the analysis. The gaps in the basic data were subsequently completely filled, so the dataset for the statistical analysis based on the OECD DAC ratings can be described as *very good*.

For the **content analysis of the factors** explaining the average differences in ratings (statistical trends), qualitative data were collected from 112 evaluation reports. The present content analysis is thus based on 66% of the available evaluation reports. The evaluation reports were selected systematically according to two criteria: first of all, on the basis of the preceding statistical analysis. Initially, differences in the ratings were identified (for example, average CPE ratings for bilateral projects as opposed to ratings for (supra)regional projects). Then, for each identified trend, the three best and the three worst rated CPEs were selected for content analysis. This selection of evaluation reports ensured that both particularly good approaches and particularly negatively rated approaches were included, which made it possible to identify explanations for each trend. Second, the selection includes the 90 most recent CPEs to identify patterns that are as up to date as possible and to be able to provide practical approaches.[6]

Despite this selection, it should be noted that the strength of evidence with regard to the explanatory factors depended on the informative value of the evaluation reports on the respective trend examined. For example, different CPE average ratings in relation to the project type can only be justified if direct, or at least indirect, reference was made to it in the evaluations. The dataset for the content analysis can be described as *good* at this point. Compared to the statistical data, however, this aspect constitutes a limitation of the CSA.

## 2.3 The analysis process

### Procedure during the preparatory phase

The aim of the preparatory phase was to define the focal points of the CSA and to obtain an overview of the object of the study and the available dataset. To this end, the Syspons team first conducted a detailed discussion with the client of the analysis – GIZ's Evaluation Unit.

Syspons then conducted in-depth interviews with selected representatives of GIZ's Evaluation Unit in order to inquire in detail about hypotheses, ideas and information requirements for the individual questions. The results of these interviews were incorporated into the further steps of the analysis and included in the concept note presented at the end of the preparatory phase, and the analysis grid.

### Statistical processing and analysis of the CPE average ratings

The aim of the statistical processing of the data was first to analyse and present the OECD DAC average ratings of the 207 available CPEs based on various criteria and thematic areas to obtain a broad overview of the ratings. In a second step, the initial findings were examined and conspicuous differences in ratings (statistical trends) were identified. These were discussed with the Evaluation Unit.

Concurrently with the preparatory phase, the Syspons team familiarised themselves with the dataset and prepared it for statistical analysis. The data provided on the ratings and basic data of all CPEs were cleaned in Excel and gaps were identified. These were filled in using information from the evaluation reports or through further research by the team. The data were then processed visually using the Tableau visualisation software. Based on the Terms of Reference (ToRs), the criteria and thematic areas for the analysis were agreed with the Evaluation Unit in the commission clarification meeting, and during in-depth interviews. In order to present the ratings based on these criteria and thematic areas, GIZ's ideas on certain types of analysis (for example, presentation of the sample and geographical maps) were first taken into account, and then an exploratory approach for further presentations was pursued. The product of the first analysis was made available to GIZ in PowerPoint format at the same time as the concept note (see above).

[6] In the course of further data collection by the CSA, all other evaluation reports were also examined and included in the analysis.

Based on these analyses, the Syspons datasets were examined for differences in the ratings. For this purpose, the findings were examined based on the selected criteria and thematic areas, and statistical trends (for example, different average CPE ratings in different regions) were considered.

The preliminary findings and the statistical trends were discussed in a joint meeting and in the follow-up. A decision was taken as to which evaluations should be included in the results report and which statistical trends should be examined qualitatively (see below). In the present report, the following **statistical trends** are examined in depth.

- **Differences in ratings by project type:** Why are bilateral projects rated differently on average than other project types? According to the CPE reports, how does the project type affect the CPE ratings?
- **Differences in ratings by project size:** Why do smaller projects have different average ratings compared to larger projects? What role does project size play in CPE rating?
- **Differences in ratings by region:** Why are projects more successful on average in some regions than in others? What factors can be identified from the CPEs?
- **Differences in ratings by BMZ country category:** Why are projects in different BMZ country categories (such as reform partners, transformation partners, nexus and peace partners, etc.) rated differently on average? What factors can be identified from the CPEs?
- **Differences in ratings by BMZ core area and area of intervention:** Why do the average ratings of CPEs differ for the various BMZ core areas and downstream areas of intervention? Which core areas and areas of intervention were rated particularly highly or particularly poorly, and what factors can be identified from the CPEs?
- **General rating trends:** Furthermore, general rating trends over time were considered and possible factors were discussed (e.g. the influence of the COVID-19 pandemic, remote evaluations and changes in the CPE rating scale).

The findings on these trend-related questions are presented in Chapter 3. The process of identifying explanations is described below in the following section on content analysis.

### Content analysis of the evaluation reports

The content analysis of the evaluation reports set out to identify explanations for the selected statistical trends or differences in ratings as part of a qualitative analysis. To this end, the (statistical) analysis was supplemented by a qualitative analysis of 112 evaluation reports (see section 2.2 on the selection of reports).

The central questions of the CSA, which were determined during the preparatory phase, were operationalised using a data collection interface, the definitive version of which was agreed on with GIZ. The following analytical categories were defined to identify explanatory factors: (i) What do the evaluations say about the **general conditions and challenges** (political, legal, social, etc.) of the projects? (ii) Which **project approaches**[7] were pursued by the projects, and how were these assessed by the evaluators? (iii) How were selected **results hypotheses** assessed in the evaluations, and why? (iv) Which **factors for success and failure** were listed in the conclusions of the CPEs?

In the **cross-section**, the team then examined how these factors are related to the respective **statistical trend**. Here, the team looked at what recurring factors might explain the differences in ratings. Some examples: are there certain general conditions and challenges, approaches or factors for success and failure that are clustered in certain regions? Do these clusters explain the differences in average ratings between regions? If so, which factors should be highlighted?

[7] In this context, we regard 'approaches' to mean the conceptual basis or the method adopted by the project to achieve its objectives and/or results in an effective way (e.g. training of trainers).

The data collection interface was programmed using the survey software SurveyXact. The interface contains both quantified questions with categories (yes/no/to some extent/no answer) and free-form text fields for the Syspons team's qualitative observations. Working with the survey software enabled the team to elicit the findings for each report on a standardised basis using the defined questions, and then analyse them efficiently. The datasets collected were analysed and synthesised for the analysis in terms of both quantity (frequency of observations in the report) and quality (type of observations).[8]

#### Reporting and presentation

The product of the analysis of the average ratings of CPEs is this report, which documents the key findings. It was prepared by the Syspons team and sent to the Evaluation Unit for comment. The feedback and comments were incorporated into the finalised version. The Syspons team also prepared and presented the content of the CSA in a PowerPoint presentation.

# 3 Findings of the cross-sectional analysis

This chapter presents the key findings of the CSA with respect to the average ratings of completed CPEs, together with possible explanatory factors. In the section below, the success rate of the CPEs is first of all presented at the aggregate level. Subsequently, the ratings for the CPEs are disaggregated by progress over time, general project characteristics, region, theme and sector, as well as evaluation characteristics. For this purpose, the selected statistical trends are described in more detail and underpinned with possible explanatory factors from the content analyses.

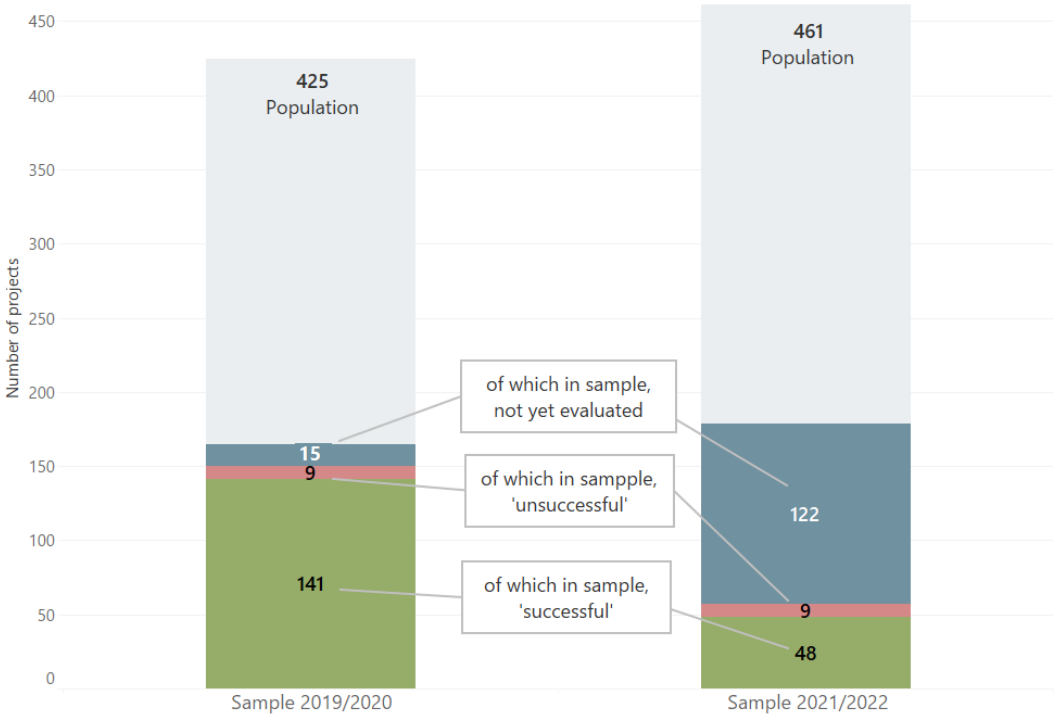
## 3.1 Success rate of the CPEs

In CPEs, BMZ-funded projects are rated on a scale from 1 ('highly successful') to 6 ('highly unsuccessful'), similar to the award of marks in German schools.[9] Based on this rating system, the success rate describes the proportion of projects with an overall rating of 1, 2 or 3. The success rate thus indicates the proportion of projects that were rated as successful by the external evaluation teams.

Figure 2 shows the samples, including the number of successful ratings, of the 207 completed CPEs (see section 2.1). Of the 150 CPEs from the first sample of projects scheduled to end during the period 2019/2020, 141 projects were rated as 'successful'. The **success rate** is thus 94%. In this sample, the findings for 15 CPEs are still pending, as the projects in question have been extended. In the later sample for the period 2021/2022, fewer CPEs have been completed so far: in 57 (out of a total of 179) completed CPEs, 48 projects were rated as successful. This corresponds to a success rate of 84%.

[8] Each report in *CSA Insights* is a standalone report. That said, the qualitative analysis process is also central to Part 2 and Part 3 of *CSA Insights*. The respective focus of the analysis with regard to the objectives is described in the relevant reports.
[9] The overall rating is comprised of the individual ratings for the various OECD DAC criteria (see section 3.3).

Figure 2: Population and sample with success rate

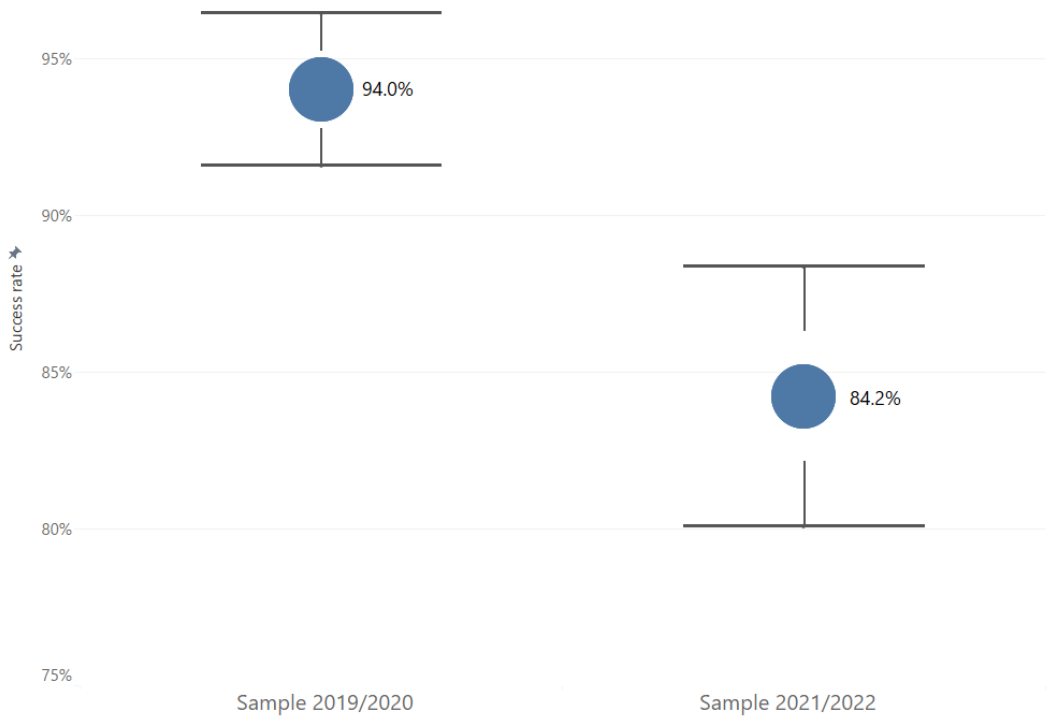



Due to the stratified random sample, the success rate is representative for all projects completed within a period of two years (population). This means that the success rate is meaningful beyond the sample for all BMZ-funded projects with a commission value of EUR 3 million or more. However, like any statistical estimate, the success rate is subject to a certain degree of uncertainty. This **uncertainty/meaningfulness of the success rate** initially depends on how many findings for the sample are already available. The more findings there are available, the more meaningful the success rate is for all projects in the population. Since CPEs are still outstanding for both the 2019/2020 sample and the 2021/2022 sample, the significance of the success rate is still limited in both cases – even if the uncertainty for the first two samples is significantly lower by comparison, given the low number of missing CPEs. However, even if all CPEs are available, a residual statistical uncertainty remains.

**To calculate the significance precisely it is possible to use the confidence interval, which explains the uncertainty and makes it tangible**. It specifies the range within which there is a 95% probability that the success rate for the corresponding population (all BMZ-funded projects) will lie. Figure 3 shows the current success rate for the two samples, and also indicates the confidence interval that can be used to describe the significance for all projects in the population.

**Applying the confidence interval, it already emerges for the 2019/2020 sample that the success rate of all projects in the population is well above 90%. The initial findings for the 2021/2022 sample show lower success rates of around 84%.** From the 2019/2020 sample findings, the success rate is expected to be between 96% and 92% for all projects in the population. As described above, there is greater uncertainty in the later 2021/2022 sample. Accordingly, the confidence interval in which the finding will lie with 95% probability is even wider for this sample. Here, the actual success rate can currently be estimated at 88% to 80%. The confidence interval will become smaller with the findings from further CPEs, which means that the significance of the success rate will increase accordingly.

Figure 3: Success rate including 95% confidence interval




## 3.2 Analysis of the CPEs and previous tools over time

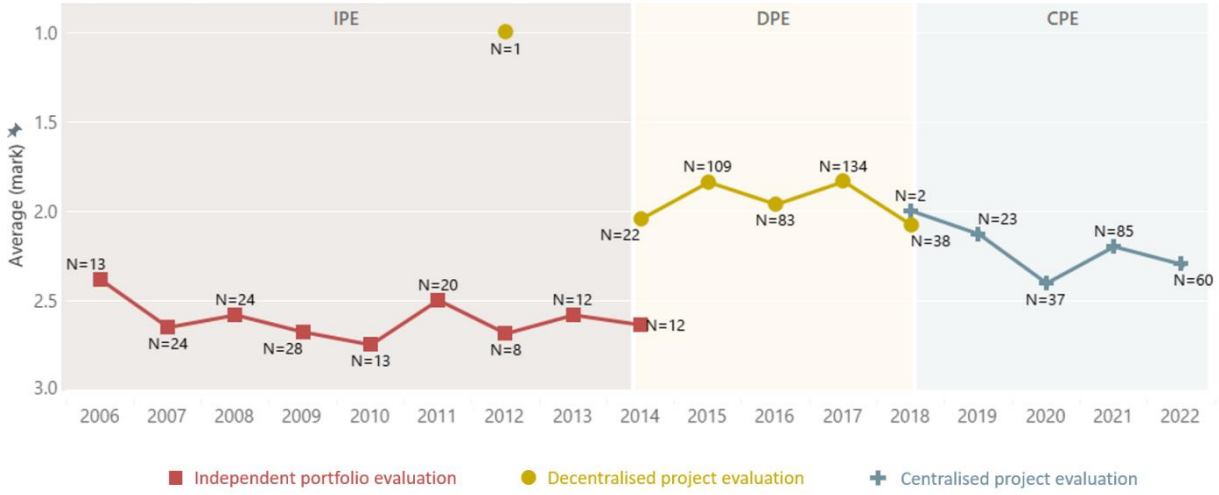
A look at the **GIZ evaluations and the associated ratings over time** allows the described findings on the success rate to be assessed. Here it is also worth comparing different evaluation tools that GIZ has established in its long-standing tradition of evaluation. The CPEs were introduced at the end of 2017 and have been a key evaluation tool for GIZ ever since. Besides the CPEs, other evaluation tools exist or have existed which, through their objectives, contribute or have contributed to the principle of 'knowing what works' applied in GIZ evaluations.[10] Two evaluation tools comparable to the CPE were the independent portfolio evaluations between 2006 and 2014 and the decentralised project evaluations between 2012 and 2018. Although these formats differed from CPEs in their respective implementation processes, they also rated projects using a six-point scale system and are therefore valid for (indirect) comparison with CPE findings.

**Looking at the timeline shown in Figure 4 indicates that there are clear differences when comparing the average ratings for the three evaluation tools: independent portfolio evaluations, decentralised project evaluations and CPEs.** First of all, it is striking that the lowest average ratings were awarded in the independent portfolio evaluations. In a total of 162 evaluations (average of 18 per year), an average rating of between 2.39 (2006) and 2.75 (2010) was awarded annually. By contrast, significantly higher average ratings of between 1.84 (2017) and 2.08 (2018) were awarded in the decentralised project evaluations. At the same time, significantly more decentralised project evaluations were carried out, the annual average being 97.[11] Ultimately, by comparison, the ratings of the 207 CPEs fall between the ratings of the two previous tools. Here, the highest average rating was awarded in 2018 (2.00), while the average rating in 2020 was the lowest at 2.41. So far, on average 41 CPEs have been conducted per year, with a gradual increase in the number of CPEs. While 23 CPEs were completed in 2019, in 2021 there were 85. However, the differences in the ratings do not permit directly comparable conclusions concerning the success or failure of the evaluated projects, as the differences

[10] GIZ, 2022: https://www.giz.de/en/aboutgiz/516.html.
[11] All but one of the decentralised project evaluations fall within the period 2014 to 2018. One decentralised project evaluation completed in 2012 can be identified as an '**outlier**' in Figure 4.

in rating could be due to different processes of the evaluation tools. For example, the independent portfolio evaluations were mostly conducted as *ex post* evaluations two to five years after project completion. By contrast, the decentralised and centralised project evaluations were or are predominantly commenced at an earlier point in time.

Figure 4: Ratings over time, including previous evaluations (by reporting date)[12]




**A closer look at the CPE ratings over time shows that the average CPE ratings decreased until 2020, but increased from 2021 onwards.** In the first year (2018), only two CPEs were completed. In each year that followed the figure was significantly higher. The highest number of ratings to date were awarded in 2021, when 85 CPEs were completed. In terms of average ratings, it can be seen that the ratings steadily decreased from 2018 (2.00) to 2020 (2.41). In 2021, the average rating improved, rising to 2.20. In 2022, the average rating was 2.30. These observations – and the possible link to the COVID-19 pandemic – are discussed in more detail in section 3.7.

In the findings presented so far on the success rate and the average ratings over time, all completed evaluations were considered in their entirety. At the same time, a closer analysis of the CPE ratings reveals differences that provide further insights into the evaluations and evaluated projects. These differences are discussed in the following sections.

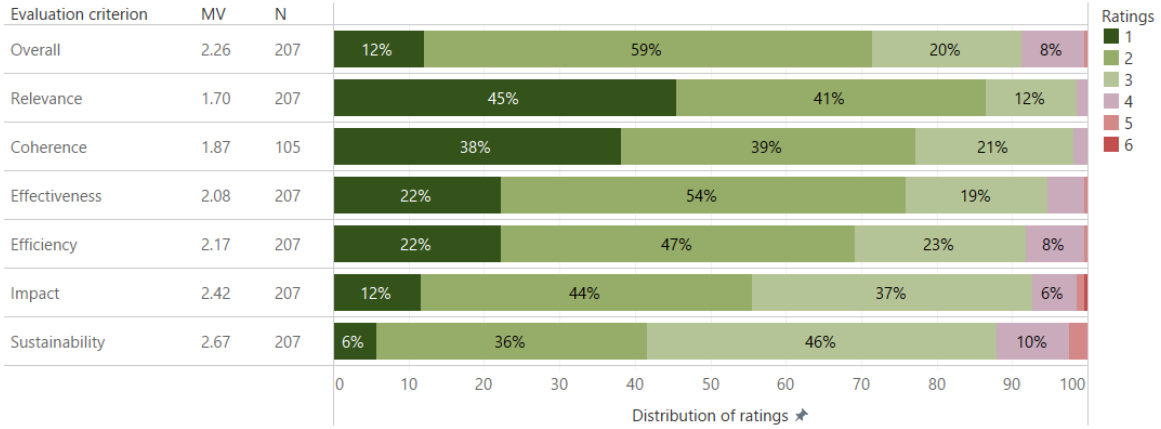
## 3.3 Analysis of the CPEs by evaluation criterion

Besides the overall ratings considered so far, points and ratings for individual OECD DAC criteria are awarded in each CPE in accordance with international evaluation standards.[13] Each project is rated on the basis of the six evaluation criteria: relevance, coherence (since the beginning of 2021, see section 3.7), effectiveness, efficiency, impact and sustainability. The rating based on these criteria forms a normative framework for determining the benefit of a project. Together, the criteria describe the desired characteristics of projects. Projects should be relevant to the context, be coherent with other interventions, achieve their objectives, generate results efficiently and have a positive impact that is sustainable.[14]

[12] For purposes of clarity, the chart shows only one section of the scale. The chart showing the complete scale can be found in the Annex.
[13] The **overall rating** is ultimately made up of the ratings for the individual criteria. The rating system allows evaluators to award up to 100 points per criterion. From the points awarded, the rating is calculated according to a fixed scale (rating 1: 100-92 points; rating 2: 91-81 points; rating 3: 80-67 points; rating 4: 66-50 points; rating 5: 49-30 points; rating 6: 29-0 points). The average of these ratings gives the overall rating for the project. There are exceptions due to the so-called knock-out criteria (see section 3.7).
[14] OECD, 2019.

A look at the **distribution of ratings in the overall rating and by evaluation criterion** (Figure 5) clearly shows where strengths and weaknesses were seen in the projects implemented by GIZ. First, closer examination reveals that the average overall rating of all 207 CPEs (N) is 2.26 (mean – MV). By far the largest share of all CPEs (59%) obtained the rating 2 (‘successful’). However, a much more differentiated picture emerges for the individual criteria.

Figure 5: Distribution of ratings by evaluation criterion




**The projects were rated particularly positively in respect of two evaluation criteria: relevance and coherence.** On average, the **relevance** of the projects was rated best by far (mean 1.70). In just under half of the CPEs, a rating of 1 (‘highly successful’) explicitly confirms that the projects are aligned with the needs, strategies and priorities of the partners or beneficiaries. According to the evaluators, a large proportion of the projects are designed to be relevant in their respective contexts and to pursue valuable objectives. In addition, the majority of projects were considered very **coherent** and thus compatible with other interventions (MV 1.87). Here, 77% of the projects achieved a rating/mark of 1 or 2. It should be noted that this criterion was not added until 2021, and ratings are available for 105 CPEs. Moreover, it is possible that introducing the coherence criterion affected the overall ratings of the CPEs, as this meant a change in how the mean value is calculated (see section 3.7). Nonetheless, it is clear that in the view of the evaluation teams, the projects do tend to support other interventions and generate synergies – both internally with other GIZ projects and with interventions by other actors. Accordingly, teams rated complementarity, harmonisation and coordination with other actors positively.

**By contrast, the ratings for the impact and sustainability criteria are lower in comparison.** The **impact** of the projects was rated on average 2.42, which is significantly lower than the overall average. The evaluators therefore still see potential for improvement in the longer-term (transformative) effects of the projects and the changes in systems or norms, as well as the possible effects on development goals. However, the **sustainability** criterion received the lowest average rating (MV 2.67). In many cases, the evaluators are sceptical about whether the impact of the projects will last. Nevertheless, it should be remembered that the ratings are within the range of successful, and only a fraction of the projects were rated as ‘unsuccessful’ (rating 4 or worse) for these evaluation criteria. At the same time, the comparatively poorer ratings for the criterion of sustainability align with the findings of a study by the German Institute for Development Evaluation (DEval), according to which the evaluation of sustainability in German development cooperation is highly inconsistent and must therefore be regarded as particularly challenging.[15]

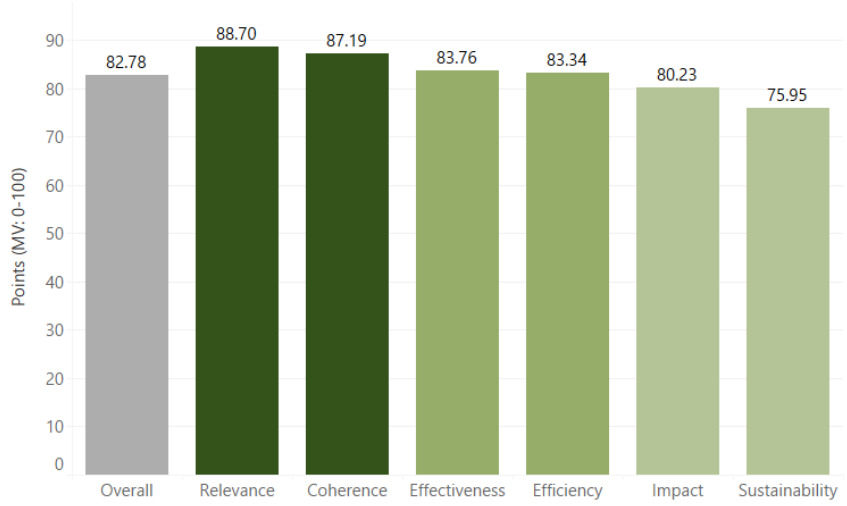
**The criteria of effectiveness and efficiency are within the overall average.** The **effectiveness** – the extent to which the projects have achieved or are expected to achieve their objectives – is on average ‘successful’ according to the evaluators (MV 2.08). The cost-effectiveness or **efficiency** of the projects is rated slightly

[15] DEval, 2018: https://www.deval.org/de/evaluierungen/laufende-und-abgeschlossene-evaluierungen/nachhaltigkeit-in-der-deutschen-entwicklungszusammenarbeit.

worse on average (MV 2.17). Accordingly, some projects that were more successful with regard to these criteria, and some that were less so were evaluated in the CPEs.

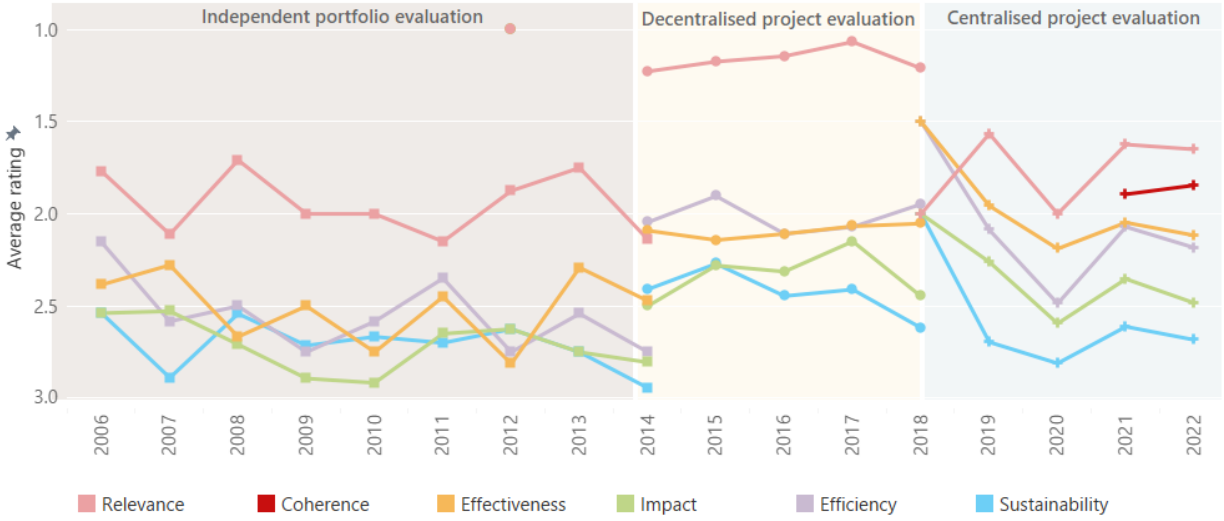
In an alternative presentation, Figure 6 shows the **average points for the overall rating and for the individual OECD DAC criteria on the 100-point scale that is also used at GIZ**. The differences described above are also brought to light in this presentation.

Figure 6: Average distribution of points




**These observations made in the CPEs for the individual evaluation criteria are comparable overall with the findings from the previous evaluation tools.** To highlight the commonalities, Figure 7 visualises the average ratings for the individual criteria over time for CPEs, decentralised project evaluations and independent portfolio evaluations (see section 3.2). This shows that, for example, the relevance criterion (shown in pink) also received by far the most positive ratings in the early evaluation tools. Further comparable trends can also be seen in the relatively low ratings for sustainability (shown in blue), especially when we compare CPEs and decentralised project evaluations. Moreover, the coherence criterion (shown in red) was included in the CPEs for the first time in 2021, and therefore is not to be found prior to that or in earlier evaluation tools.

Figure 7: Average ratings by over time (by reporting date)[16]




## 3.4 Analysis of the CPEs by project characteristic

With BMZ-funded projects, a distinction can be drawn between different project types. In addition, the evaluated projects can be classified according to further project characteristics. Considering the findings when disaggregated by these various (selected) project characteristics enables us to identify further differences in the CPEs.[17]

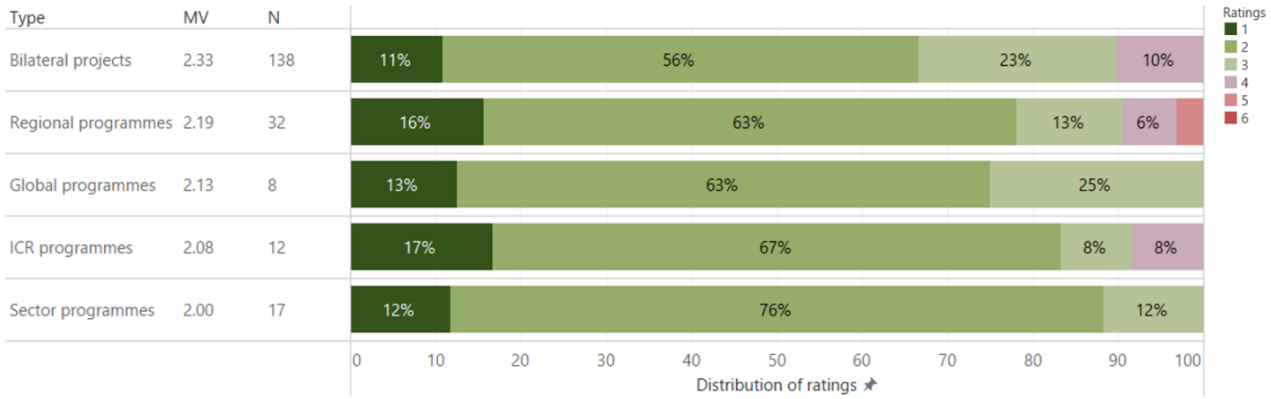
### Project type

One key characteristic of the evaluated projects is the **project type**. This is because the project type is a key factor affecting the entire design and implementation of a project, including the objectives and the target groups, as well as the respective cooperation partners. Figure 8 shows the distribution of ratings as a function of project type. First, it is striking that **bilateral projects** – i.e. projects that are implemented bilaterally in a partner country of German development cooperation in order to support partner-owned change processes – were evaluated in the majority of the CPEs (N=138 of 207, or 66%). A smaller proportion of the CPEs dealt with **other project or programme types** that address sectoral or global development issues through cross-country and cross-regional cooperation. These include regional programmes (N=32), global programmes (N=8), international cooperation with regions programmes (N=12) and sector programmes (N=17).

**A comparison between the project types shows that the average rating for bilateral projects is 2.33, which is lower than for all other project types.** Moreover, a comparatively large number of the 'unsuccessful projects' are bilateral projects. This observation raises the question of which factors associated with the project type were highlighted as particularly unhelpful or good in the CPEs. Trend box 1 lists some factors that may explain the differences in average ratings between bilateral projects and other project types. These factors were identified in the qualitative analysis of the CPE reports (see section 2.3). However, the differences could also be due to the evaluations themselves. For example, in the case of sector and global programmes, data collection was mainly carried out without on-site visits. In this context, section 3.7 shows that the type of data collection also affects the average rating.

[16] For purposes of clarity, the chart shows only one section of the scale. The chart showing the complete scale can be found in the Annex.
[17] The analysis presented here describes differences in CPE ratings in relation to selected project characteristics. The evaluated projects differ with respect to numerous other characteristics that could affect both design and implementation, and CPE ratings. A further disaggregation of the evaluated projects, e.g. by budget item as well as co-financing, is included in the Annex to this report.

Figure 8: Ratings by project type *(N=207)*




## Trend box 1: Effects of project type on CPE ratings

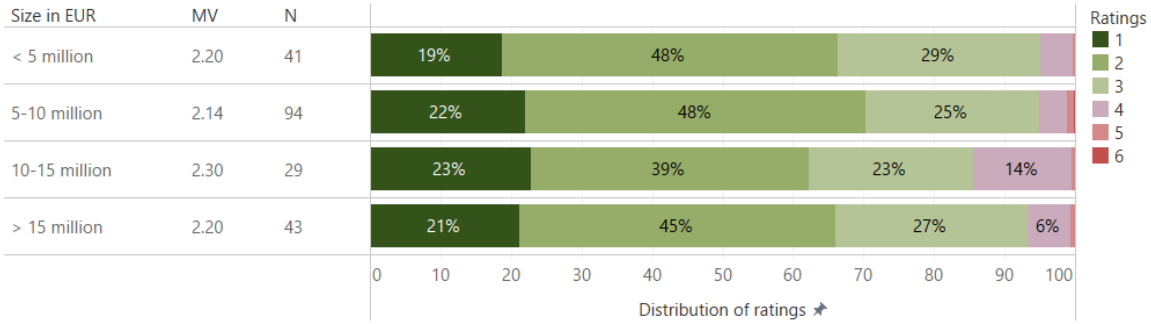
- In the case of **bilateral projects,** the CPE reports that were analysed qualitatively point more often than average to the **context** as a factor for failure. In particular, the CPE reports mention restrictions due to the COVID-19 pandemic. This is the case in 51% of the qualitatively evaluated reports of bilateral projects, compared to 29% of the other qualitatively analysed reports on non-bilateral projects. For example, due to the pandemic-related travel restrictions and hygiene regulations, face-to-face events could not take place or could only take place to a limited extent. This affects non-bilateral projects less frequently, as their activities more often include measures such as the promotion of networks or advisory services, which were less affected by the restrictions. Furthermore, a poor security situation (17% vs 40%) is also mentioned as a failure factor, for example due to armed conflicts that entail the cancellation of measures already planned.

- By contrast, cooperation with the stakeholders involved is highlighted as a success factor in **global programmes, sector programmes, international cooperation with regions programmes and regional programmes**, i.e. in 82% of the qualitatively evaluated CPE reports on these four project types, compared to 56% of all qualitatively considered CPE reports on bilateral projects. One example of successful cooperation is the use of participatory approaches. Inhibiting contextual factors that might have negatively influenced these **cooperation systems** were rarely mentioned here. **Internal factors** are also mentioned more often than average as success factors. It is true that there are more challenges in steering (41% vs 33%), for example due to steering structures that involve the clients and thus possibly limit the scope of the project. However, the expertise of the project teams (68% vs 48%) and the commitment of individual project staff members (14% vs 2%) are frequently highlighted.

### Project size

Besides the project type, there are other project characteristics that have a significant influence on the project design. For example, the **project size** determines a project's scope and implementation options. In Figure 9, the CPE ratings are broken down into four categories according to project size. Here we first of all see that so far both comparatively small projects (N=41 for projects with a budget <EUR 5 million), medium-sized projects (N=29 for projects with a budget of EUR 10–15 million) and particularly large projects (N=43 for projects with a budget >EUR 15 million) were part of the CPE sample and were rated. However, the majority of the projects rated had a budget between EUR 5 and 10 million (N=94).

Figure 9: Ratings by project size (categories) *(N=207)*




**The analysis based on project size does not reveal any statistically significant difference between the average ratings as a function of project size.** Projects with a budget of up to EUR 5–10 million achieved the highest average rating of 2.14, while projects with a budget of EUR 10–15 million or more only received a rating of 2.30. However, a correlation analysis does not confirm a statistical link between overall rating and project size. In other words, it is not possible to demonstrate statistically that the larger the project, the lower the overall rating awarded by a CPE. This observation contrasts with an earlier analysis of decentralised project evaluations from 2019, which found a *'significant difference in the relationship between overall project size and success'*.[18] Nonetheless, the qualitative analysis did identify several factors mentioned in the CPE that are linked to project size. These are shown in Trend box 2.

## Trend box 2: Effects of project size on CPE ratings

- **Projects with budgets of up to EUR 10 million** are slightly more successful on average. With regard to these projects, particularly strong emphasis is placed on close cooperation with the stakeholders involved and on the expertise of the project team. Moreover, project management is highlighted as a success factor. Examples include the monitoring system used, flexibility of project implementation and swift adaptation to new circumstances such as the COVID-19 pandemic.

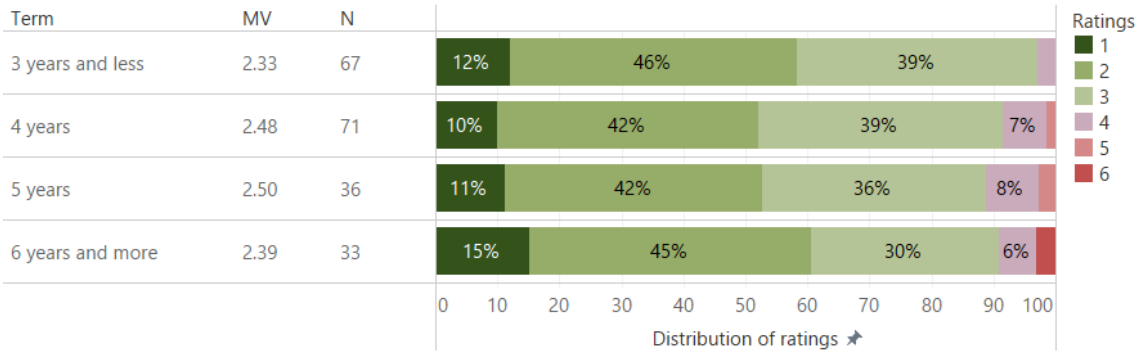
- By contrast, **projects with budgets of EUR 10 million or more** are somewhat less successful on average. In this case, however, cooperation with the stakeholders involved, which is highlighted as a positive in smaller projects, is named as a key factor for failure. In this context, reports point in particular to bureaucratic hurdles.

### Project term

The **project term** is also an important project characteristic. The project term specifies the time frame within which a project can achieve its objectives. Depending on this time frame, objectives can be defined and pursued. This characteristic therefore also has a significant effect on project design and implementation. In Figure 10, the CPE ratings are broken down into four categories according to project term: 138 evaluated projects had a term of up to four years and only 69 projects had a term of five years or longer. **No clear trend can be discerned in the average ratings as a function of the project term.**[19] This observation is consistent with the previously cited analysis of decentralised project evaluations from 2019, which also found no correlation between project term and project success.[20]

[18] GIZ, 2019
[19] A correlation analysis between the project term and the overall rating of the CPE also fails to identify a statistically significant link (p=0.19).
[20] GIZ, 2019.

Figure 10: Ratings by project term (categories) *(N=207)*




## 3.5 Analysis of the CPEs by country and region

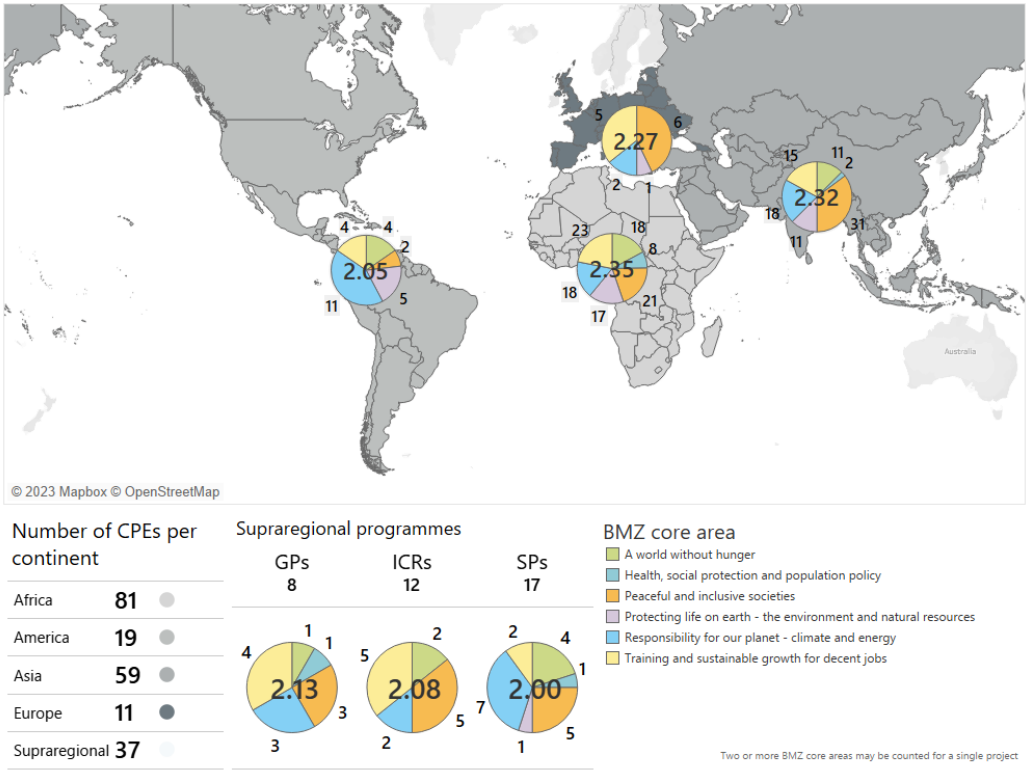
GIZ works with a large number of partner countries worldwide. The geographical conditions under which the projects are implemented are accordingly diverse. Moreover, the partner countries can be broken down into further categories. First of all, they are divided into various country categories in accordance with the BMZ 2030 reform strategy. Second, the fragility of partner countries is a factor to be highlighted for the feasibility of their implementation. A geographical analysis of CPE ratings for different continents, world regions and contexts (BMZ country categories and fragility of partner countries) can therefore provide insights into the importance of regional conditions for project success.[21]

### Continents

Figure 11 shows the average CPE ratings for different **continents** on a world map. In addition, the sectoral distribution of the evaluated projects (by BMZ core area) per continent is shown in a pie chart.[22] Supraregional projects that could not be clearly assigned to any continent are shown separately. This applies to global programmes, international cooperation with regions programmes (ICRs) and sector programmes.

[21] Due to the low number of CPEs at country level, no analysis per country was carried out. At the same time, the success/failure and thus the rating in general may also vary within a region. Moreover, the findings per region also depend on the criteria discussed in the other sections (e.g. the project characteristics).
[22] There are various ways of presenting a sectoral distribution of projects (see section 3.6). For the sake of clarity, the distribution by BMZ core area was chosen here.

Figure 11: Average ratings by continent, taking sectors into account and BMZ core areas *(N=207)*




**Most of the CPEs evaluated projects in Africa** (a total of 81 projects). A large proportion of the projects served to promote sustainable economic prospects (N=23, yellow) and peacebuilding (N=21, orange). Overall, however, the distribution by content in this context was very diverse: more than 10 projects were evaluated in each BMZ core area (with the exception of the sixth core area 'Health, social protection and population policy'; this, however, has only been defined as such since 2021 (see section 3.6)). The second-highest number of CPEs (51) were carried out in **Asia**. The projects evaluated had a clear focus on stabilisation and peacebuilding (N=31, orange). Significantly fewer projects were carried out on the **American continent** (19 CPEs) and in **Europe** (11 CPEs). While most projects in the Americas were evaluated in the area of climate and energy (N=11, blue), projects in Europe focused on the areas of stabilisation (N=6, orange) and sustainable economic development (N=5, yellow). The sectoral distribution of evaluations of supraregional programmes is also shown. For global and ICR programmes, sustainable economic development was among the most frequently evaluated areas (N=4 for global programmes and N=5 for ICRs, yellow). Sector programmes, on the other hand, were most frequently evaluated in the area of climate and energy (N=5, blue). Overall, however, the sectoral distribution of the CPEs is diverse here too.

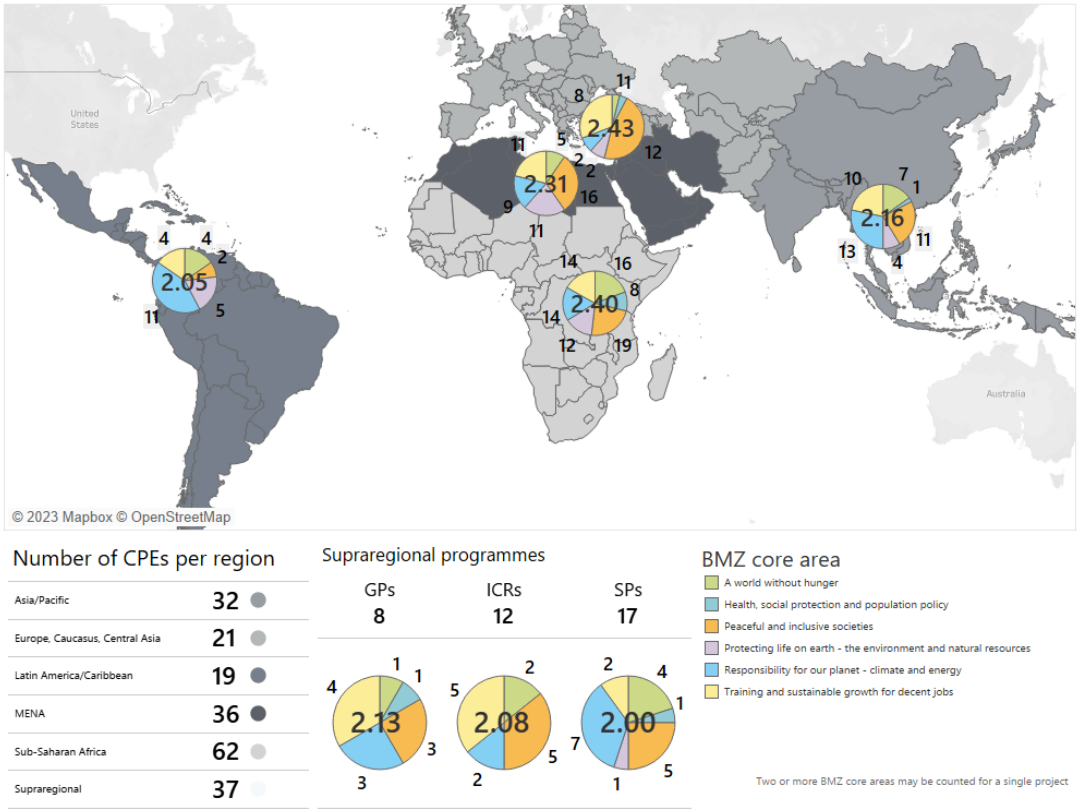
**The map also shows that the continents have different average ratings.** The lowest average rating is found in Africa, where 81 projects scored an average of 2.35. The continent with the best average rating is America (MV=2.05). Europe (MV=2.27) and Asia (MV=2.32) fall in between the two in terms of average ratings.

## World regions

Examining the findings by continent provides a good overview of the geographical distribution and ratings awarded by the CPEs. At the same time, the continents each subsume several regions. Thus, a more

differentiated analysis of **CPE ratings for various world regions**[23] provides a somewhat more accurate foundation. Figure 12 shows an alternative categorisation by the world regions Asia/Pacific (N=32), Europe, Caucasus, Central Asia (N=21), Latin America/Caribbean (N=19), Middle East and North Africa (N=36) and sub-Saharan Africa (N=62). As before, the sectoral distribution of the evaluated projects (according to BMZ core areas) is also presented.

Figure 12: Average ratings by world region, taking sectors into account and BMZ core areas *(N=207)*




**The findings in this map show parallels, but also new aspects compared to the previous map.** According to the analysis, the lower average values for Africa relate in particular to projects in sub-Saharan Africa. Here, the average value of 2.40 is below the average value for the continent as a whole (see Figure 11, MV=2.35). Projects in the Middle East and North Africa region – which is partly located in Africa – are rated better on average with 2.26. The ratings for projects in the Europe, Caucasus and Central Asia region are lowest overall (2.43). The inclusion of projects in Central Asia lowers the average rating – therefore, the new categorisation also has an impact on the average ratings of projects in the Asia/Pacific region, with ratings significantly better than the more comprehensive 'Asia' on the previous map. The findings for Latin America/Caribbean are the same as the previous findings for the continent of 'America, as German development cooperation does not operate in North America'.

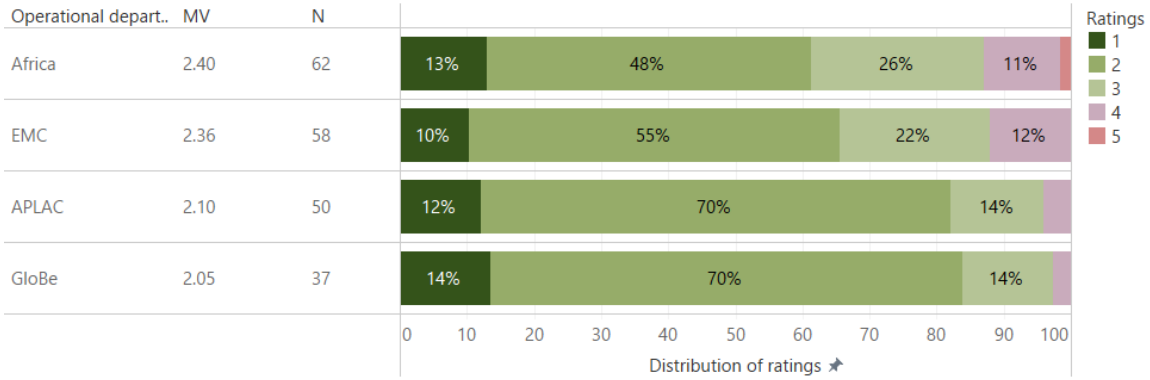
## Regional departments

A third geographical analysis of the CPEs was conducted for GIZ's internal **regional departments**.[24] Accordingly, this analysis is particularly useful for internal evaluations. Internally, GIZ is divided into four operational regional departments. As can be seen in Figure 13, 62 CPEs were implemented in the Africa area, 58 CPEs in

[23] The distribution of ratings by world region is shown as a bar chart in the Annex to the report. There is also another map with a more differentiated view of the world regions (Figure 29).
[24] It should be noted that the samples are proportionally stratified in line with the operational areas of GIZ, so that a representative selection was ensured (see section 2.2, GIZ, 2022: https://www.giz.de/de/downloads/giz2022-de-instrumentenpapier-zentrale-projektevaluierungen-bmz.pdf).

the Europe, Mediterranean, Central Asia (EMC) region and 50 CPEs in the Asia, Pacific, Latin America and Caribbean (APLAC) region. Furthermore, 37 projects were evaluated in the GloBe department, to which sector and global programmes are assigned. This confirms the trends shown in the previous maps (Africa and Europe/Central Asia with lower average ratings, especially compared to Asia and Latin America).

Figure 13: Ratings by operational department (regional departments of GIZ) *(N=207)*




Irrespective of the regional categorisation by continent, world region or GIZ regional department, the qualitative analysis of the CPE reports (see section 2.3) analysed factors that were mentioned in conjunction with the general geographical conditions in the CPEs. Trend box 3 lists the factors that may offer an explanation for the differences in ratings

.

## Trend box 3: Explanations for regional differences in CPE ratings

- Projects in **Asia/Pacific, Latin America/Caribbean** and **supraregional programmes** are on average more successful than others. It is striking that in the qualitatively examined CPE reports, cooperation with the stakeholders involved is listed both more often than average as a success factor and more often than average as a failure factor. This shows that the projects do experience difficulties in introducing a **permanent and/or a temporary cooperation system**. However, once that system has been introduced, it then proves to be a success factor. For example, for projects in Latin America/Caribbean, partner ownership is mentioned as a success factor more often than average: 33% of the qualitatively examined CPE reports from Latin America/Caribbean compared to 24% of all other reports. The **project team** component is also emphasised: the expertise of the project team and the commitment of individual project staff members are often cited as an internal success factor. This is reflected, for example, in the planning of projects in the Asia/Pacific region, for which sustainability strategies, among other things, are mentioned more often than average as a positive factor (25% vs 8%).

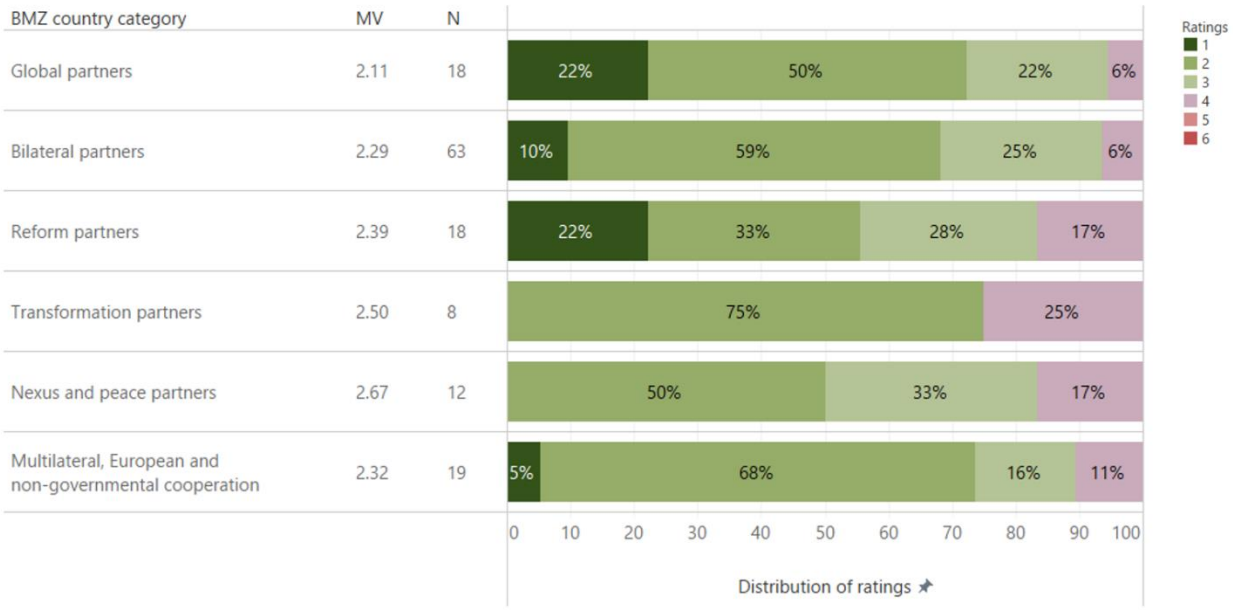
- By contrast, projects in the **Europe/Caucasus/Central Asia** and **Middle East/North Africa and sub-Saharan Africa** regions are less successful on average. For these projects, contextual factors and management challenges are listed as failure factors more often than average. With regard to the **context** of the projects, restrictions due to the COVID-19 pandemic, which led to measures having to be postponed or shifted to the digital space, and the security situation, were mentioned in particular. With regard to the **steering** of projects, the CPE reports examined qualitatively state this does not take place systematically in some cases, or does not involve relevant project stakeholders to a sufficient degree.

## BMZ country categories

The general conditions for projects implemented by GIZ also depend on the forms of cooperation that are politically managed by German development cooperation and the respective partner country or countries. In this context, BMZ has introduced new **partnership categories** with its BMZ 2030 reform strategy, which enable more strategic cooperation with partner countries. The aim is to steer the type of cooperation and to work on the priorities and areas in the partner countries as effectively as possible.[25] An analysis of the CPE ratings by these partnership categories shows how the practical implementation of the different types of cooperation was rated on average by the evaluators.[26]

**Overall, projects with global partners were rated highest in the CPEs completed to date. Projects with nexus and peace partners received the lowest average ratings** (Figure 14). On the one hand, projects with **global partners** work strategically on solving global future issues and protecting global goods (for example, environmental protection and climate action). In this regard, 18 projects were given an average rating of 2.11. On the other hand, **nexus and peace partners** in crisis and refugee regions are supported in stabilising and overcoming the causes of conflict. By comparison, the 12 CPEs that evaluated projects in these partner countries awarded the lowest average rating of 2.67.

Figure 14: Ratings by BMZ country category *(N=138)*




[25] BMZ, 2020: https://www.bmz.de/en/countries/reform-strategy-bmz-2030.
[26] The analysis only includes CPEs that could be assigned bilaterally to one partner country. This means that regional and supraregional projects (N=69) were not included in this section.

**For the projects in partner countries in other categories, the average CPE rating falls in between.** Joint development goals are pursued with bilateral partner countries in the long term. Here, besides the **'traditional' bilateral partnerships** (MV 2.29 for 63 CPEs), a distinction is drawn between two further forms: **reform partners** – countries that have proven to be particularly reform-oriented in recent years – receive special support (MV 2.39 for 18 CPEs). Second, targeted political and economic transformation processes in the EU's neighbourhood are supported in cooperation with **transformation partners** (MV 2.50 for 8 CPEs). There is no direct official cooperation with developing countries that do not fall into any of the above categories. However, development goals are also supported here through the promotion of civil society as well as by the EU and multilateral institutions. This type of cooperation received an average rating of 2.32 in 19 CPEs. Possible explanations for the differences in the ratings – especially for the comparatively well rated projects with global partners and the less well rated projects with nexus and peace partners – were collected in the qualitative analysis (see section 2.3). The findings are presented in Trend box 4.

## Trend box 4: Effects of country categories on CPE ratings

- Projects in the **global partners, bilateral partners and multilateral, European and non-governmental partners** categories are more successful on average. A qualitative analysis of the CPE reports for these projects reveals that cooperation with involved stakeholders is listed both as a success factor more often than average and at the same time as a failure factor more often than average. This also shows that a **permanent or temporary cooperation system**, once introduced, then becomes a success factor. With regard to the individual categories, other supporting factors stand out: projects with global partners seem to have a more appropriate provision of **resources** (38% of the qualitatively examined CPE reports for projects with global partners vs 17%), for example with regard to the integration of experienced long-term experts.

- Projects in the **reform partners, transformation partners and nexus and peace partners** categories are less successful on average. In projects with reform partners, it is striking that the **ownership by the stakeholders involved** (excluding partners) is mentioned in a negative light more often than average, for example as a consequence of frequent personnel changes. In projects with transformation partners, on the other hand, an above-average number of **challenges in managing** the project (44% vs 37%) were cited, which, according to the CPE evaluators, can be attributed inter alia to a lack of systematic monitoring and learning. In projects with nexus and peace partners, the **context** is often cited as an obstacle. Besides constraints due to the COVID-19 pandemic (60% vs 44%), this relates in particular to the security situation (80% vs 14%). Conflicts and fragile situations on the ground have relatively often led to delays and additional expense in these projects, partly due to inflation and rising fuel prices.

### Fragility

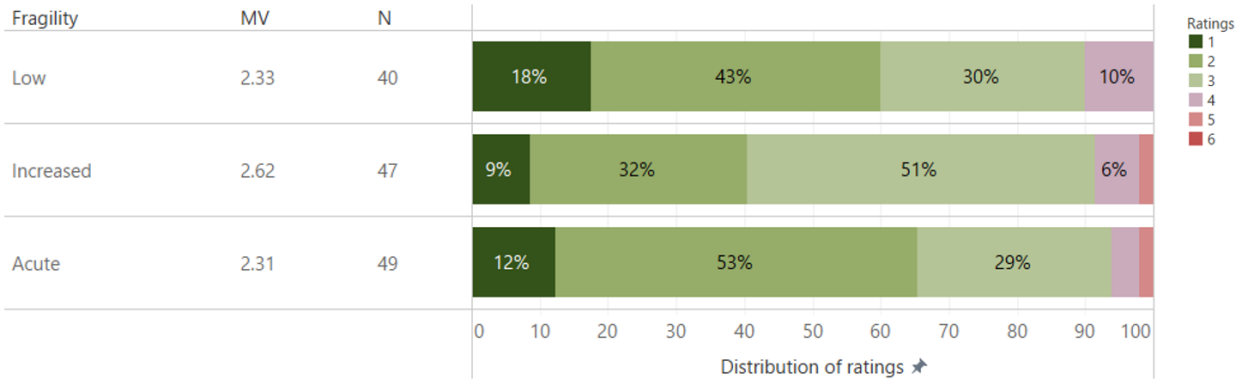
In the previous analyses, regional differences in rating and the different types of cooperation were addressed. Political and civil society stability is already an important factor here. Nonetheless, the issue of **fragility** should be highlighted once again; due to an increasing number of armed conflicts, around a quarter of the world's population now lives in fragile states, where security and development risks are high.[27] Fragility therefore has a major impact on the implementation of GIZ projects. This is because deliver effective support, extraordinary adaptability is required in terms of project planning and management, among other things.

[27] BMZ, 2020: https://www.bmz.de/en/countries/reform-strategy-bmz-2030

**Figure 15 shows how the CPE ratings of projects implemented in a more or less fragile context differ from other ratings.**[28] The categorisation was based on internal and confidential data using a security-based fragility analysis of the respective partner countries.[29] A distinction is made between countries with low fragility (low), partner countries with heightened fragility (increased) and partner countries with acute fragility, which includes all countries in violent conflict or country-wide crises (acute).

**It can be noted here that there is no clear correlation between the fragility of the partner country and the CPE ratings.** The average rating for projects in contexts with low fragility is 2.23 (N=40), which is slightly higher than the average rating for projects in contexts with increased fragility (MV=2.40; N=47). However, projects in contexts of acute fragility are rated more favourably than projects in the medium fragility category (MV=2.31; N=49). It is striking that the largest proportion of CPEs were conducted in contexts of acute fragility, which can be explained by the focus of many projects on peacebuilding and stability (see section 3.6). Overall, according to these data, the fragility of a partner country does not have a crucial effect on the CPE rating. A statistical correlation analysis also provided no evidence of a correlation.

Figure 15: Ratings by fragility (as of reporting date) *(N=136)*




To further validate the lack of correlation between the CPE ratings and fragility, a **second fragility index** was used and correlated with the CPE ratings. The **Fragile States Index** (FSI) assesses the fragility of countries according to a multidimensional index that takes other aspects into account, not just security-related aspects, and assigns each country an overall rating between 1 (not fragile) and 120 (very fragile).[30] To link this to the CPEs, the year of the evaluation mission was used rather than the reporting year. Although the analysis suggests a weak correlation between CPE ratings and fragility (the more fragile, the worse the rating), this is not statistically significant (see Figure 29 in the Annex).

## 3.6 Analysis of the CPEs by BMZ area and sector

Besides project characteristics and regional contextual factors, the CPEs also yield insights into the thematic orientation of the evaluated projects. The question of which thematic areas and why the evaluation experts consider GIZ projects to be particularly successful or less successful is central to GIZ's organisational learning.

Since the BMZ 2030 reform strategy, GIZ projects (and those of other implementing organisations in German development cooperation) have focused on five **core areas** (Peacebuilding; Food security; Training and sustainable growth; Climate and energy; Environment and natural resources).[31] In 2022, Health and social

[28] The analysis only includes CPEs that could be assigned bilaterally to one partner country. This means that in this section, regional and supraregional programmes (N=69) were not included. Moreover, two bilateral CPEs are not included, as no risk assessment was carried out for the country (Burundi) in the year of the evaluation report.
[29] An analysis based on the year of the evaluation mission (rather than the reporting year) did not reveal any significant differences.
[30] FSI 2022: Further information on the methodology of the index can be found on the website.
[31] BMZ, 2020: https://www.bmz.de/en/countries/reform-strategy-bmz-2030.

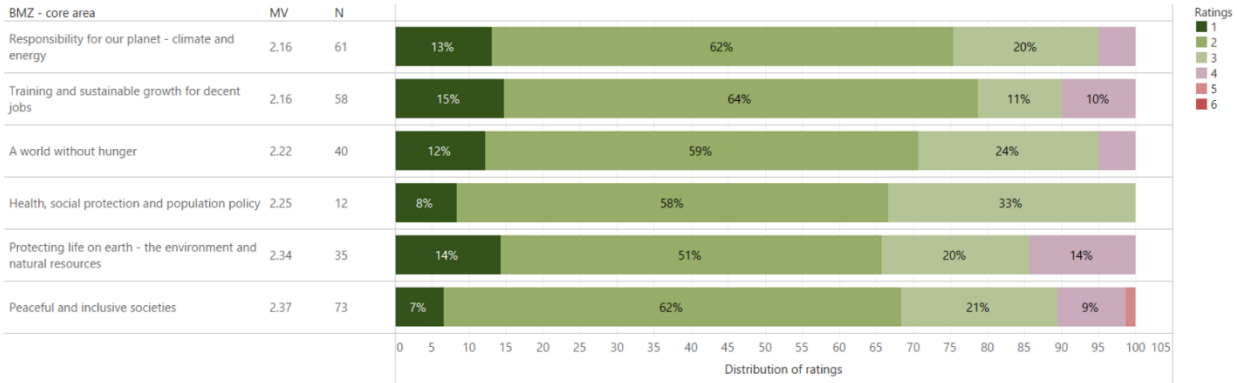
protection was added as a sixth core area.[32] These core areas define the priorities set for Germany's contribution towards implementing the Agenda 2030 and the 17 Sustainable Development Goals. For each of these core areas, three **areas of intervention** are also defined, which further differentiates the orientation of cooperation. In terms of their own content, the evaluated projects were assigned to *up to three* areas of intervention and core areas.

## BMZ core areas

**A look at the evaluations by BMZ core area shows that projects in the thematic areas of climate and energy and training and sustainable growth were rated most positively in the CPEs. By contrast, projects in the areas of peacebuilding and food security were viewed most negatively.** Figure 16 shows that projects in the core areas 'Responsibility for our planet – climate and energy' (N=61) and 'Training and sustainable growth for decent jobs' (N=58) scored on average 2.16. By contrast, the 73 evaluated projects in the core area 'Peaceful and inclusive societies' received a comparatively lower rating of 2.37. Moreover, the area of peacebuilding and stabilisation was where the majority of the CPEs were implemented, which supports the observation in the previous section concerning the large number of CPEs in fragile contexts. At the same time, the previous section did not identify a clear link between fragility and evaluation, which raises the question of why projects in the peacebuilding/stabilisation area are on average less successful than projects working in other core areas. Possible explanations for this phenomenon are taken up in the text box from the qualitative analysis at the end of this section (see section 2.3).

**The ratings in the other core areas fall in between the above.** Projects in the area of 'A world without hunger' are in third place in terms of average rating (MV 2.22 from 40 CPEs). This is followed by projects in the new core area 'Health, social protection and population policy', in which comparatively few projects have been evaluated so far in 12 CPEs (MV=2.20). Since the core area was not previously defined as such, this is understandable. In the environmental projects 'Protecting life on Earth – the environment and natural resources' (MV 2.34 from 35 CPEs), it is also striking that many projects were rated as 'highly successful' (14% with the best possible rating) as well as just under 15% rated as 'unsuccessful' (rating of 4).

Figure 16: Ratings by BMZ core area



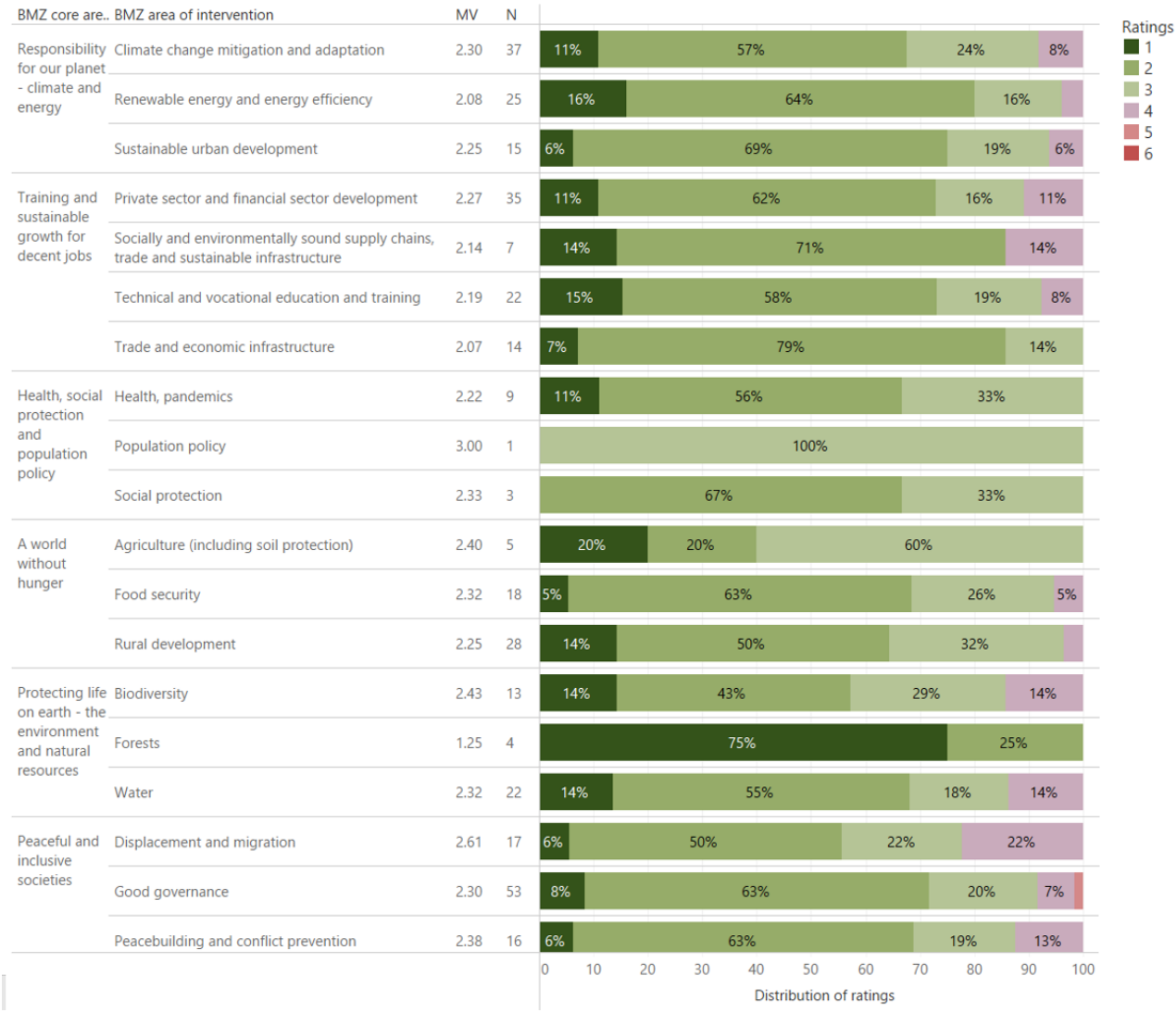

## BMZ areas of intervention

A more in-depth look at the ratings in different thematic areas is possible via the **areas of intervention**. Three areas of intervention are clearly defined for each BMZ core area. Figure 17 presents the ratings by area of

[32] BMZ, Model of areas 2022 (unpublished). In BMZ's model of areas, which was amended as of March 2022, the five core areas and their areas of intervention remain largely unchanged. One area of intervention in the core area 'Sustainable economic development, training and employment' has been redefined, and is now 'Socially and environmentally sound supply chains, trade and sustainable infrastructure', instead of 'Trade and economic infrastructure'. The new sixth core area and three areas of intervention have been added. The area of intervention entitled 'Health, pandemics and One Health' was previously defined as an initiative area.

intervention.[33]

Figure 17: Ratings by BMZ area of intervention




**If we look at the ratings in relation to the allocation of projects to areas of intervention, it becomes clear that the previously discussed average ratings for specific BMZ core areas can be explained by projects in several areas of intervention.** For example, the good rating in the thematic area of climate and energy can be explained in particular by projects in the area of intervention 'Renewable energy and energy efficiency', 80% of which were rated as 'highly successful' or 'successful' (N=25). At the other end of the rating spectrum, it can be seen that in the area of peacebuilding, projects in the area of intervention 'Displacement and migration' tend to be rated as less successful. The average rating here (with the exception of areas of intervention with <5 ratings) is the lowest by comparison with all areas (MV=2.61). It can also be seen that the previously described range of ratings in the thematic area of environment and natural resources can be attributed to the themes 'biodiversity' and 'water'. In both areas of intervention, just under 15% of the projects are rated as 'highly successful' (best possible rating) or 'unsuccessful' (rating of 4). As mentioned above, possible explanations from the qualitative analysis with regard to the rating differences as a function of thematic area are given in Trend box 5. [34]

[33] Areas of intervention to which no evaluated project has been assigned are not included in the chart. At the same time, projects can also be assigned to more than one area of intervention (or BMZ core areas).
[34] In addition to the explanations for the rating differences described in this report, a more in-depth content analysis based on 10 thematic areas was carried out in a separate report (third objective of *CSA Insights*).

## Trend box 5: Explanations for different CPE ratings as a function of thematic area

The BMZ core areas 'Peaceful and inclusive societies', 'Protecting life on Earth – the environment and natural resources' and 'A world without hunger' were less successful on average.

- A qualitative analysis of the CPE reports on the projects on the BMZ core areas '**Protecting life on Earth – the environment and natural resources'** and '**A world without hunger'** clearly shows that the context has adversely affected the projects. For example, an above-average number of references are made to the constraints caused by the **COVID-19 pandemic** and the **security situation** on the ground. In the case of the projects on the BMZ core area 'Protecting life on Earth – the environment and natural resources', this also had consequences for **ownership by the stakeholders involved** (excluding partners), whose priorities had in some cases changed considerably in the context of fighting the pandemic.

- In a qualitative analysis of the CPE reports on the projects in the BMZ core area '**Peaceful and inclusive societies'**, it is not the contextual factors that explain the different ratings. Instead, the CPE reports on these projects more often refer to **challenges in the management** of the projects, for example due to inadequate knowledge management or monitoring structures.

The 'core areas' and 'areas of intervention' defined by BMZ are not the only way to categorise project content. The projects can also be analysed in relation to the '**positioning themes**' defined internally by GIZ, for example, or internationally established creditor reporting system (**CRS) codes**, which donor countries use to report on their Official Development Assistance (ODA). The CPE ratings based on these two categories are attached as an annex to this report.

## 3.7 Rating as a function of differences in evaluation

In this final section on findings, we examine how changes in the general conditions for implementation of the CPEs themselves may have affected the evaluations. Accordingly, some CPEs also differ in their implementation processes and assessment criteria. Since the introduction of the CPEs, two factors should be emphasised. First, in 2021 the coherence criterion (see section 3.3) and the so-called **knock-out criteria** were added to the CPE evaluation criteria. Second, the COVID-19 pandemic is continuing to have a considerable impact on the course of CPEs.

### Knock-out criteria

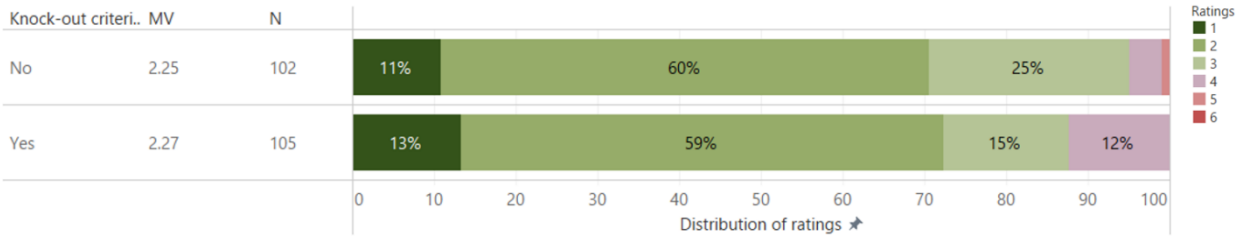
The introduction by BMZ of the **knock-out criteria** in 2021 ensured that projects that were rated as 'unsuccessful' in selected (central) evaluation criteria (see section 3.3) cannot be rated as 'successful' overall either. To be more precise: if a project receives a rating of 4 for one of the three criteria of effectiveness, impact or sustainability, it is classified as 'unsuccessful' overall.[35]

**Before the knock-out criteria were introduced, it was not clear what effect their introduction would have on the overall rating in CPEs.** On the one hand, it was possible that the average overall rating would deteriorate with the introduction of the knock-out criteria, since projects that had previously been rated as successful overall would now be rated 'unsuccessful' overall due to their poorer rating for individual criteria. On the other hand, it was possible that the introduction of knock-out criteria would influence the evaluation by the experts (even if this is explicitly not intended). The experts might rate a project slightly more favourably for these evaluation criteria, and thus still as 'successful', because they do not want to classify the project as 'unsuccessful'

[35] GIZ, 2022: https://www.giz.de/de/downloads/giz2022-de-instrumentenpapier-zentrale-projektevaluierungen-bmz.pdf.

despite the weaknesses in certain criteria.

Figure 18: Ratings before and after introduction of the knock-out criteria *(N=207)*




**A look at the distribution of ratings shows that, so far, the average overall rating has neither risen much nor fallen much with the introduction of the knock-out criteria** (Figure 18). While 102 projects received an average rating of 2.25 before the introduction of the knock-out criteria, the 105 projects evaluated according to the new rating system were awarded an average rating of 2.27. The hypothesis that the average rating would fall significantly with the introduction of the knock-out criteria cannot therefore be confirmed. A closer look at the ratings, however, provides indications that the second hypothesis may have played a role. On the one hand, the share of projects rated 4 or worse for the knock-out criteria effectiveness and sustainability has decreased (for the impact criterion, however, the number has slightly increased). Yet, above all, the share of CPEs that awarded ratings of just about 3 for all knock-out criteria has increased significantly.[36] This suggests that the evaluators are more frequently awarding a rating of 3 so that none of the knock-out criteria are triggered. Overall, the knock-out criteria have influenced the overall rating in nine CPEs to date. A comparison of the theoretical CPE ratings without the knock-out criteria shows that the average rating of 2.17 would have been significantly higher (see Figure 36 in the Annex). Since the rating would have been even better without the newly introduced criterion, it can be assumed that the knock-out criteria lowered the average overall rating and that other factors that led to a better rating on average were partially offset by this. It is important to note in this context that the rating is also influenced by other factors. The simultaneous introduction of the sixth evaluation criterion, coherence, might have an effect here, as this reduced the share of the usually less favourably rated criteria (above all impact and sustainability) in the overall rating. At the same time, the coherence criterion was previously rated higher than average, which raised the average rating (see section 3.3). Another factor is the temporal overlap between the introduction of the knock-out criteria (which began in 2021 for the CPEs) and the COVID-19 pandemic (March 2020 onwards); this is discussed in the next section.

### The COVID-19 pandemic and its effect on CPEs

The global **COVID-19** pandemic continues to have a significant impact on international development cooperation in general and the work of GIZ in particular. CPEs were was not spared either, and had to adapt to challenges such as travel restrictions and other measures to contain the pandemic.[37]

**Regarding the adjustments made to the CPEs during the pandemic, one point in particular should be emphasised:** following the outbreak of the pandemic, the evaluation missions took place without the usual field

[36] A rating is considered to be 'just about 3' if between 67 and 70 points were awarded for the respective criterion, so that the rating was just above 4 (from 66 points or less). Details of the analysis:

- For the **effectiveness** criterion, 4% of the CPEs have awarded the rating 4 or worse since the introduction of the knock-out criteria (the figure was 7% before introduction of the knock-out criteria). Moreover, 10% of the CPE-evaluated projects were rated just about 3 for the effectiveness criterion (the figure was 4% before introduction of the knock-out criteria).
- For the **impact** criterion, 8% of the CPEs have awarded the rating 4 or worse since the introduction of the knock-out criteria (the figure was 7% before introduction of the knock-out criteria). Moreover, 15% of the CPE-evaluated projects were rated just about 3 for the criterion impact (the figure was 9% before introduction of the knock-out criteria).
- For the **sustainability** criterion, 11% of the CPEs have awarded the rating 4 or worse since the introduction of the knock-out criteria (the figure was 14 % before introduction of the knock-out criteria). Moreover, 21% of the CPE-evaluated projects were rated just about 3 for the criterion sustainability (the figure was 15% before introduction of the knock-out criteria).

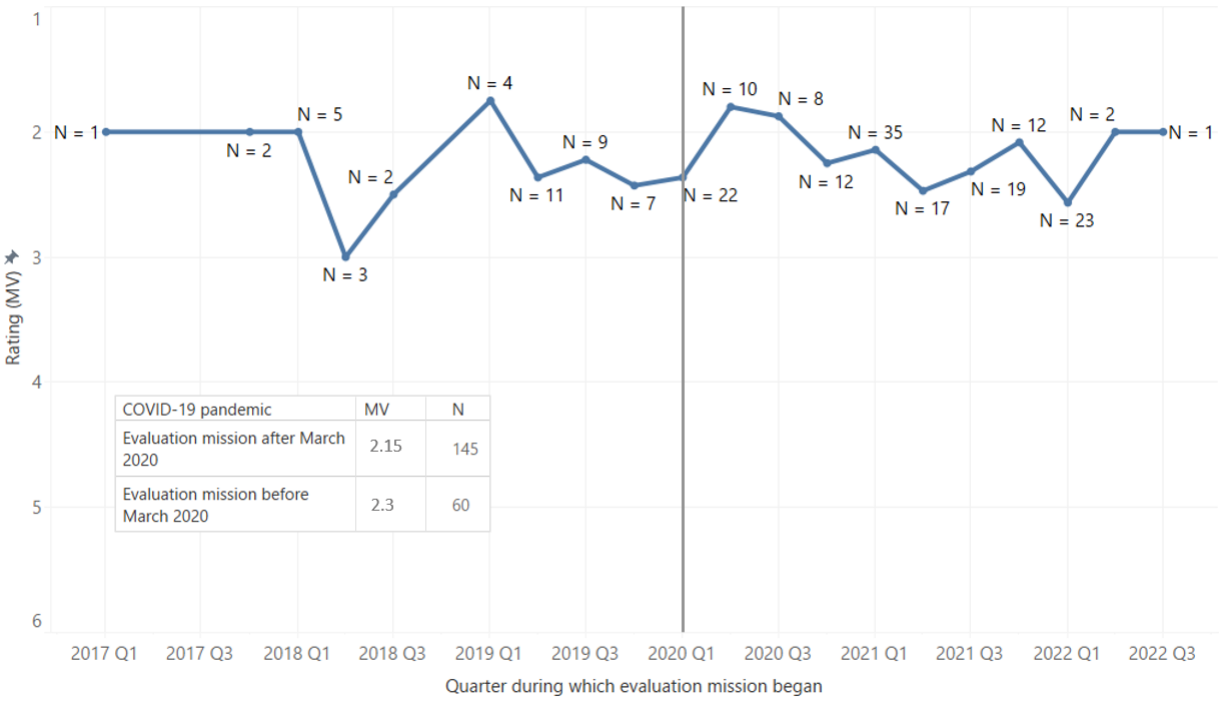
The datasets are shown in **Figures 34 and 35** in the Annex.

[37] The impact of the COVID-19 pandemic on the evaluated projects as well as on the implementation of CPEs will be addressed through an in-depth qualitative analysis of the CPE reports in a separate product (third objective of *CSA Insights*).

visits for data collection, and often replaced them with remote missions. Depending on the country-specific regulations and the respective pandemic situation, either a semi-remote solution (the local evaluators move around locally) or a purely remote solution (the mission is carried out entirely virtually) was chosen. Since these adjustments to the evaluation missions' implementation processes affected the evaluations' centralised data collection – and thus also the data quality – it cannot be ruled out that they affected the CPE ratings.

**Figure 19 shows the average CPE rating by date of evaluation mission, and provides an overview of the impact of the pandemic on CPE ratings.** Evaluation missions conducted before the outbreak of the pandemic are shown to the left of the dividing line. To the right of the dividing line are the CPEs whose evaluation missions were conducted after March 2020. Overall, it can be seen that the majority of the completed CPEs fall into the period after the pandemic (N=145). A smaller proportion of the missions were carried out before the pandemic (N=60). However, no clear trend in the graph can be discerned here, even though after the pandemic there was a somewhat more constant rating of just about 2. However, looking at the mean values when comparing the groups, it is striking that the average rating of CPEs with an evaluation mission after the pandemic is slightly higher (MV=2.15) than the average rating of CPEs with an evaluation mission before the pandemic (MV=2.30).

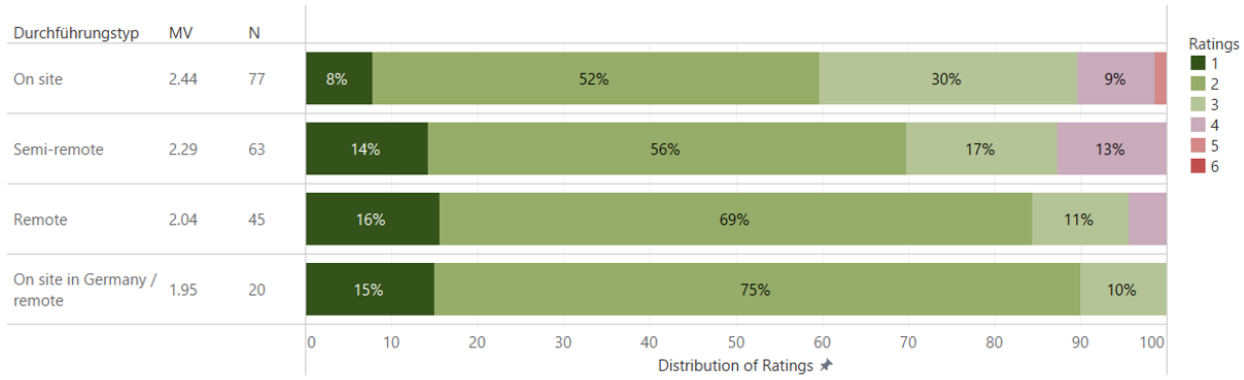
Figure 19: Average ratings by date of evaluation mission (effects of the COVID-19 pandemic) *(N=205)*[38]




A look at the different ratings as a function of the **implementation modalities of the evaluation mission** helps provide a more precise analysis. In particular, it was hypothesised in this context that the quality of the data in purely remote missions is not comparable to the quality of the data obtained during on-site visits. This is because in the remote variant, the evaluators do not have the opportunity to thoroughly examine the activities and results of the evaluated projects on-site. Rather, they have to 'rely' primarily on the statements made in telephone calls and video conferences by the interlocutors (project internals and externals). In semi-remote missions, on the other hand, on-site visits with local experts can be carried out. The data quality here is strongly dependent on the interaction between international experts (in Germany) and the local experts on-site.

[38] In Figures 19 and 20, two CPEs are not shown because they were replaced by evaluations of the co-financing partner, and no comparable data were available on the evaluation missions. Overall, the evaluations were conducted using the same rating scale.

Figure 20: Ratings by implementation modality of evaluation mission *(N=205)*




**The data show a clear correlation between the type of evaluation mission and the average rating.** Figure 20 shows that the ratings of CPEs with on-site missions are significantly less favourable (MV=44; N=77) than the ratings of CPEs with remote missions (MV=2.04; N=45). CPEs with semi-remote missions fall midway between the two in terms of average rating (MV=2.29; N=63). It is also striking here that almost all projects with remote missions were rated as 'successful' (ratings 1–3), while at least 10% of the projects evaluated through semi-remote missions and on-site missions were rated as 'unsuccessful' (ratings 4 or 5). Accordingly, the quality of the data in semi-remote and on-site missions seems to offer a deeper insight into the projects, so that negative points also become more apparent. A statistical analysis confirms these observations: CPE ratings correlate significantly with the implementation modality of the evaluation mission (the larger the remote component, the less negative the ratings). At the same time, however, there is no demonstrable statistical correlation when only CPEs with on-site and semi-remote evaluation missions are compared. Overall, these observations confirm the hypothesis stated at the beginning concerning the link between evaluation mission and data quality.

Finally, a further special case of evaluation missions is the implementation modality *on-site in Germany/remote*. Sectoral and/or global programmes are included here. These programmes cannot achieve the same added benefit with a field visit as bilateral projects. This is because their operations are widely dispersed both sectorally and geographically, and they often have no direct target groups in the partner countries.[39] The high average rating of these projects can also be explained by the fact that projects of this type were rated better overall, especially compared to bilateral projects (see section 3.4).

**The available data plausibly support the assumption that the previously described rise in ratings in 2021 (see section 3.3) was influenced by the high number of remote missions in 2020/2021.** In 2020, 17 data collection operations were carried out remotely for the first time; in 2021, 26 CPEs were implemented in this way. By contrast, the number of on-site missions decreased from 28 in 2019 to 14 in 2020 and 9 in 2021. This trend reversed again in 2022. As there were fewer pandemic-related restrictions, evaluation missions could again be increasingly conducted in the field (see Figure 35 in the Annex). This is consistent with the evaluation trend described for the years 2019 to 2022 (see section 3.3). However, as described in the previous sections, it can also be stated that especially the introduction of the coherence criterion from 2021 onwards had a positive effect on the overall rating, while the introduction of the knock-out criterion led to a decrease in the average rating. In other words, besides the implementation modality of the evaluation mission, there have been several factors that have affected CPE ratings since the beginning of the COVID-19 pandemic.

### The COVID-19 pandemic and its effects on project implementation

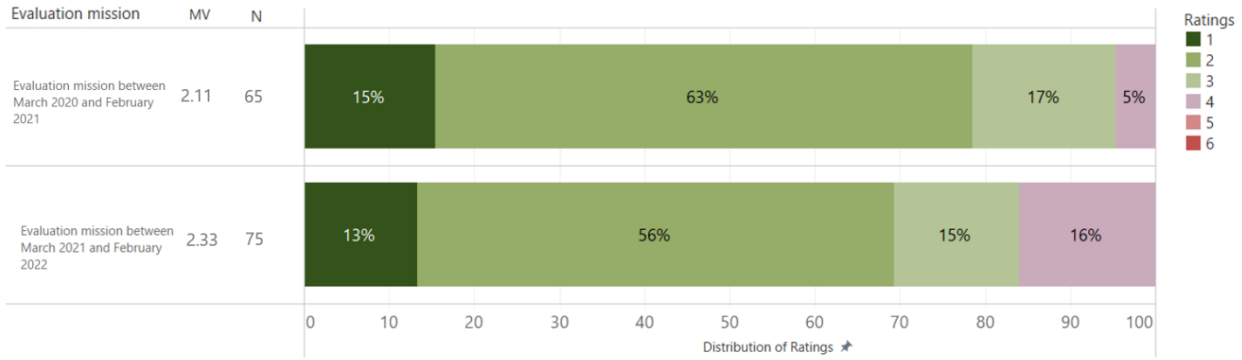
The COVID 19 pandemic also had a considerable effect on the **implementation of projects**. It can therefore be assumed that the pandemic also affected ratings, as the implementation of project measures was negatively

[39] After consultation with the Evaluation Unit, all sector programmes were included in this category. Furthermore, global programmes were categorised as either on-site, semi-remote or on-site in Germany/remote

impacted by the pandemic. The hypothesis is, therefore, the longer a project was implemented under pandemic conditions, the longer its implementation was adversely impacted, which in turn negatively affected ratings.[40]

**A comparison of the ratings between projects negatively affected by the pandemic over longer and shorter periods shows a correlation between rating and pandemic.** The analysis compared the CPE ratings of projects negatively affected by the pandemic over a shorter period (evaluation missions at the beginning of the pandemic between March 2020 and February 2021) and those that were negatively affected for significantly longer (evaluation missions between March 2021 and February 2022). This comparison confirms that the duration of the pandemic also affected average CPE ratings: projects that were affected by the pandemic for a shorter period are clearly rated better on average (MV=2.11) than those that were affected for a longer period (MV=2.33) (Figure 21). This finding aligns with other (qualitative) analyses. These observed, for example, that achievement of the benchmarks for the indicators was negatively affected by the pandemic, and in the CPEs the qualitative ratings for the criteria of effectiveness, impact, efficiency and sustainability were ultimately more negative.[41]

Figure 21: Ratings by implementation modality of evaluation mission *(N=205)*




[40] It should be noted that the ratings are a function of the effects described above.
[41] This conclusion is discussed in greater depth in the other components of the cross-sectional analysis.

# Annex

(1) Zusätzliche Abbildungen – unten aufgeführt

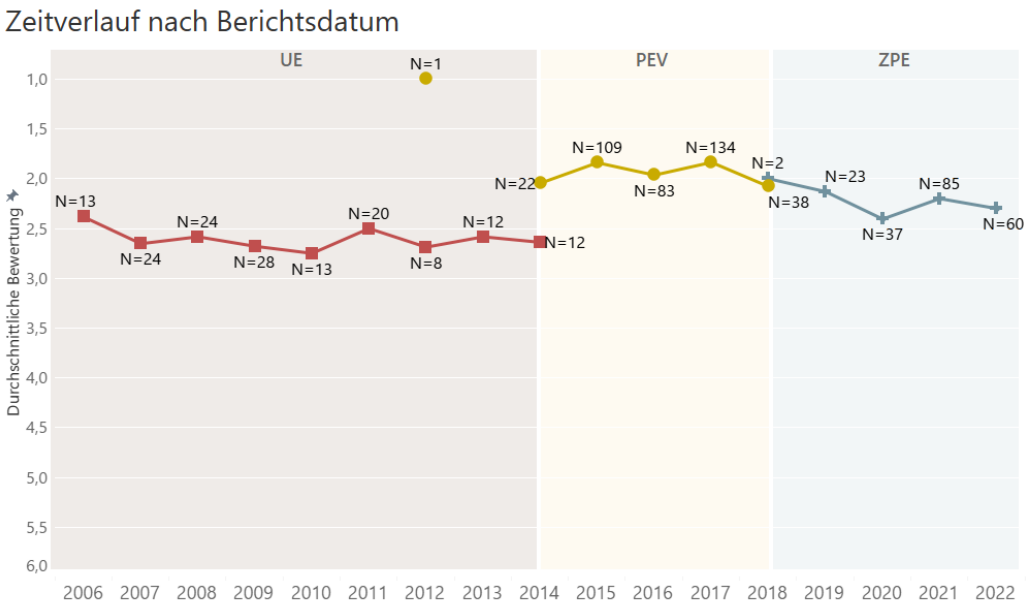
(2) Tabelle zu Erfolgs- und Misserfolgsfaktoren der qualitativ untersuchten Trends – unten aufgeführt

## Anhang 1 – Zusätzliche Abbildungen

### Weitere Abbildungen zu Kapitel 3.2: ZPE und frühere Instrumente

In **Abbildung 22** wird die in Kapitel 3.2 beschriebene durchschnittliche **Verteilung der Bewertungen im Zeitverlauf** unter Berücksichtigung der Evaluierungsinstrumente UE, PEV und ZPE dargestellt. Die Darstellung erfolgt auf einer **vollständigen Bewertungsskala** (1-6), um die Unterschiede zwischen den Evaluierungen unverzerrt abzubilden.

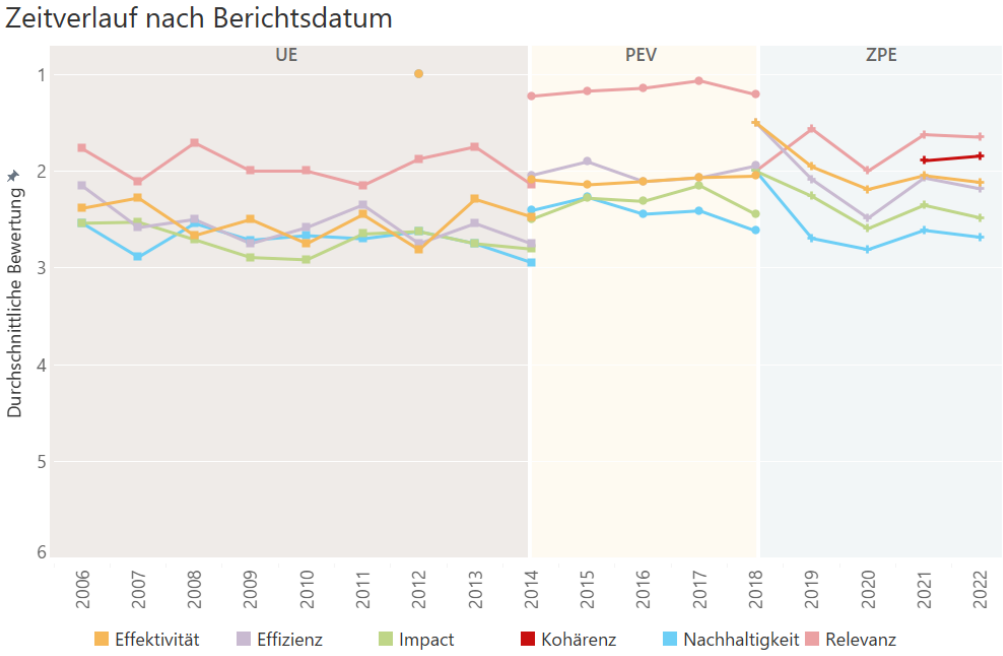
Abbildung 22: Bewertungen im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum) *(N=207)*



## Weitere Abbildungen zu Kapitel 3.3: Evaluierungskriterien

In **Abbildung 23** werden die in Kapitel 3.3 beschriebenen durchschnittlichen Bewertungen im Zeitverlauf **nach den Evaluierungskriterien** unter Berücksichtigung der Evaluierungsinstrumente UE, PEV und ZPE dargestellt. Die Darstellung erfolgt auf einer **vollständigen Bewertungsskala** (1-6), um die Unterschiede zwischen den Evaluierungen unverzerrt abzubilden.

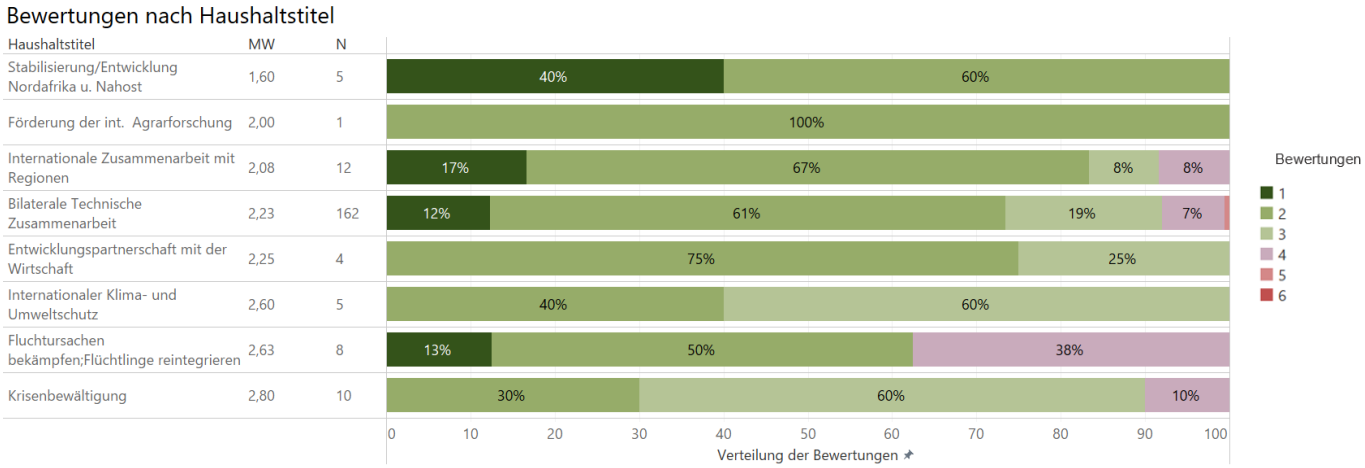
Abbildung 23: Bewertungen nach Evaluierungskriterien im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum) *(N=207)*




## Weitere Abbildungen zu Kapitel 3.4: Projektmerkmale

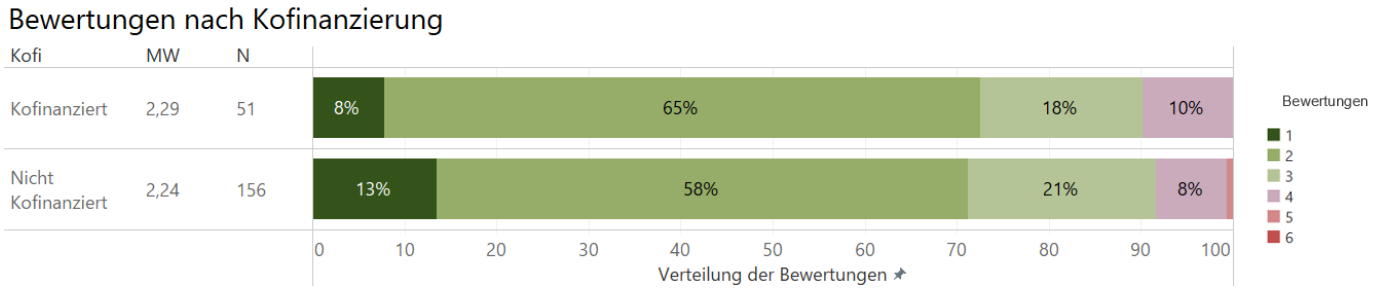
In **Abbildung 24** werden die ZPE-Bewertungen nach unterschiedlichen **Haushaltstiteln** dargestellt. Hier ist kein Trend zu erkennen, nicht zuletzt weil die Verteilung auf die Haushaltstitel sehr ungleichmäßig ausfällt.

Abbildung 24: Bewertungen nach Haushaltstitel *(N=207)*



In **Abbildung 25** werden die ZPE-Bewertungen nach **Kofinanzierung** (ja/nein) dargestellt. Die durchschnittliche Bewertung von nicht kofinanzierten Projekte ist etwas besser als die durchschnittliche Bewertung von kofinanzierten Projekten.

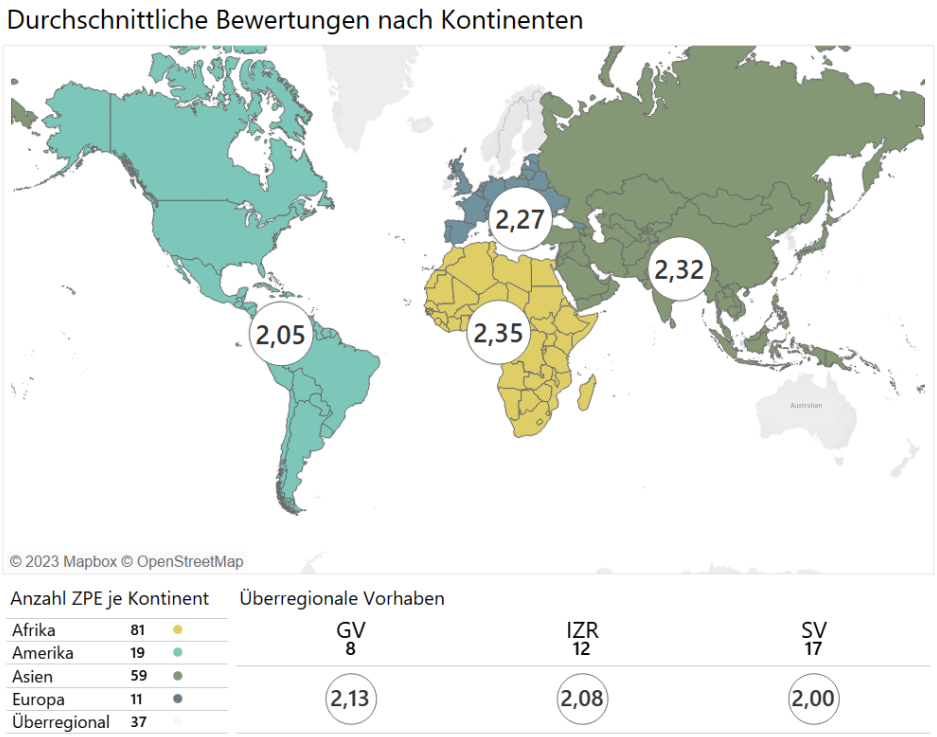
Abbildung 25: Bewertungen nach Kofinanzierung *(N=207)*




## Weitere Abbildungen zu Kapitel 3.5: Länder und Regionen

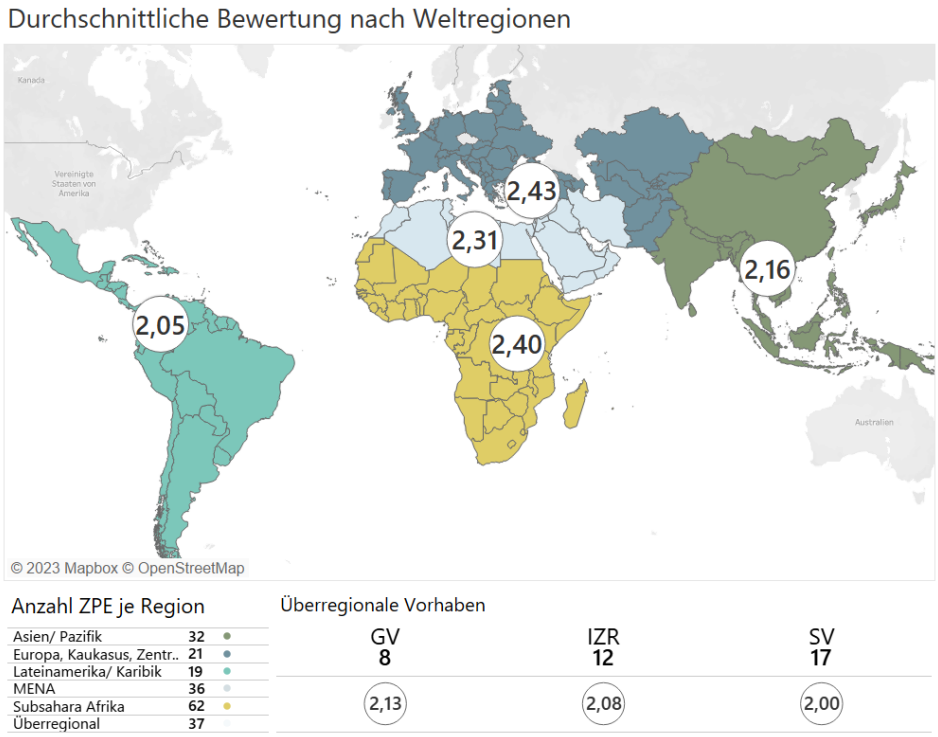
In **Abbildung 26** werden die in Kapitel 3.5 vorgestellten Daten für die einzelnen **Kontinente** (Abbildung 11) dargestellt – allerdings **ohne Berücksichtigung der sektoralen Verteilung der Projekte**.

Abbildung 26: Durchschnittliche Bewertungen nach Kontinenten ohne Berücksichtigung von Sektoren und BMZ-Kernthemen *(N=207)*



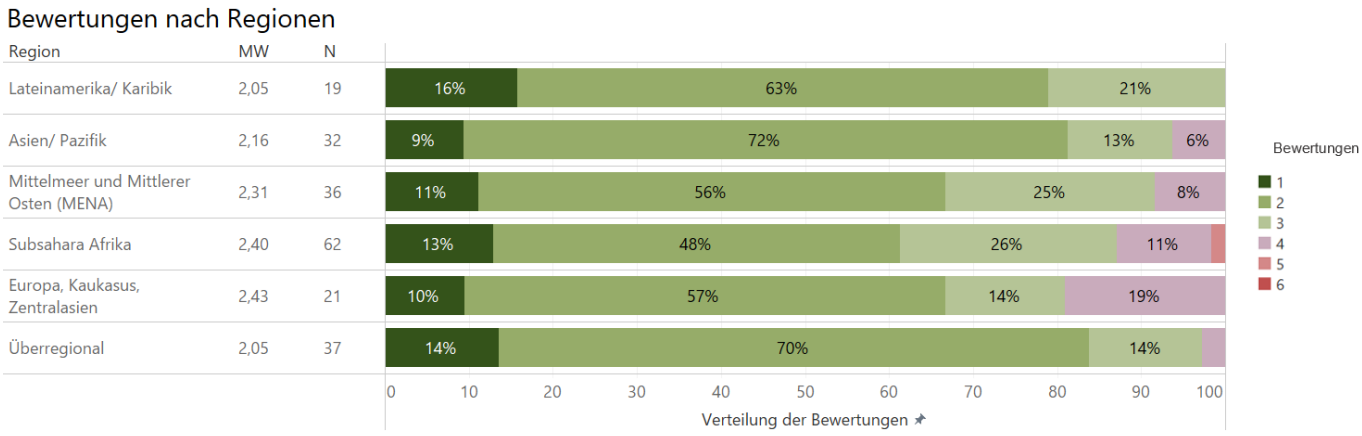
In **Abbildung 27** werden die in Kapitel 3.5 vorgestellten Daten für die verschiedenen **Weltregionen** (Abbildung 12) dargestellt – allerdings **ohne Berücksichtigung der sektoralen Verteilung der Projekte**.

Abbildung 27: Durchschnittliche Bewertungen nach Weltregionen ohne Berücksichtigung von Sektoren und BMZ Kernthemen *(N=207)*



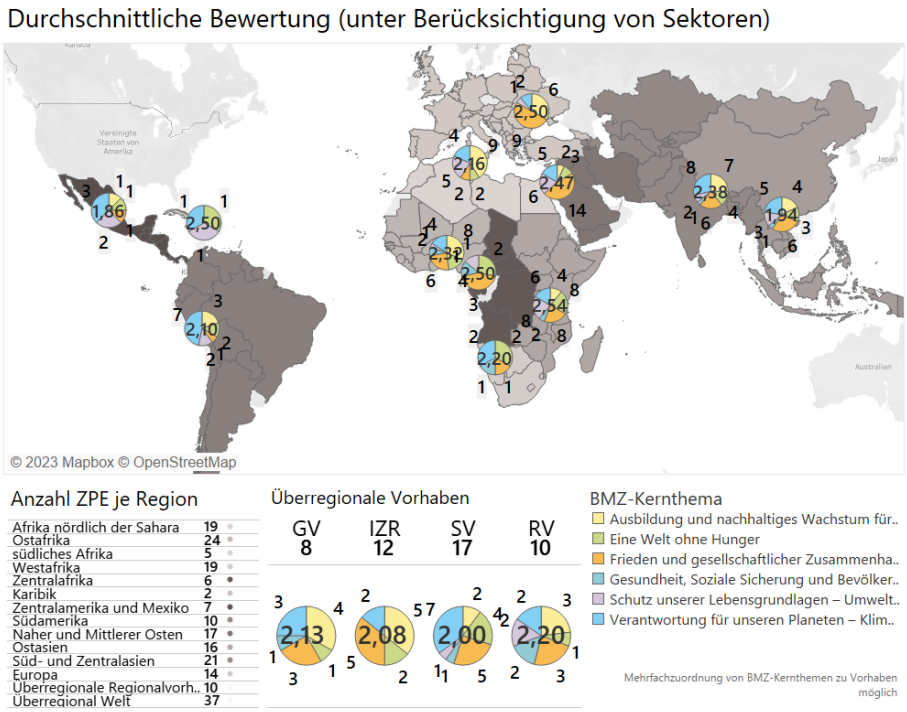

In **Abbildung 28** werden die in Kapitel 3.5 als Karte (Abbildung 12) vorgestellten Daten in Form eines **Balkendiagramms** dargestellt. Die Interpretation der Daten ist dem Bericht zu entnehmen.

Abbildung 28: Bewertungen nach Weltregionen *(N=207)*



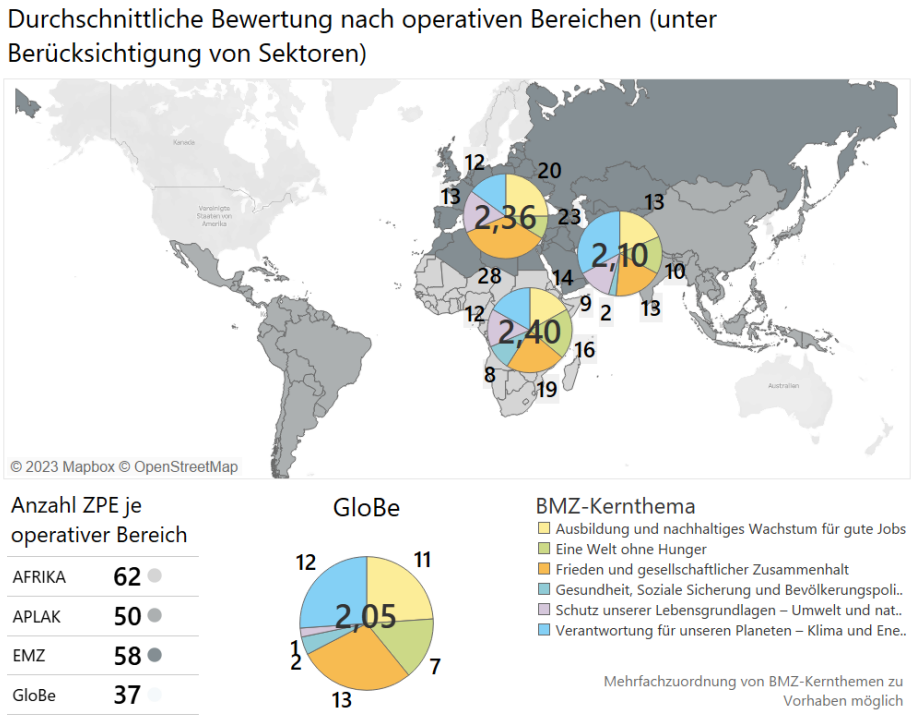

In **Abbildung 29** wird die Verteilung der durchschnittlichen Bewertungen nach **alternativen Weltregionen** dargestellt, wobei die Daten stärker differenziert werden. Die hier dargestellten Weltregionen wurden auf der Grundlage der DAC-Länderliste definiert.

Abbildung 29: Durchschnittliche Bewertungen nach differenzierten Weltregionen unter Berücksichtigung von Sektoren und BMZ-Kernthemen *(N=207)*



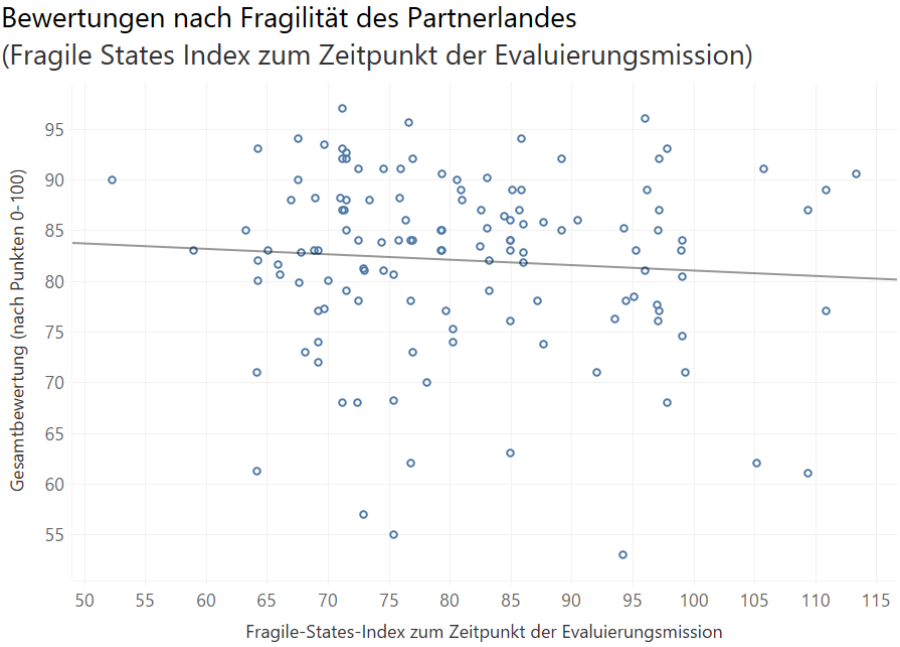

In **Abbildung 30** werden die in Kapitel 3.5 in Form eines Balkendiagramms (Abbildung 13) dargestellten Daten **in Form einer Karte** dargestellt. Die Interpretation der Daten ist dem Bericht zu entnehmen.

Abbildung 30: Durchschnittliche Bewertungen nach operativen Bereichen unter Berücksichtigung von Sektoren und BMZ-Kernthemen *(N=207)*



In **Abbildung 31** ist der Zusammenhang zwischen dem **Fragile State Index** und den ZPE-Bewertungen dargestellt. Zwar sind leichte Tendenzen erkennbar, doch kann durch die Analyse kein statistisch signifikanter Zusammenhang zwischen den ZPE-Bewertungen und der Fragilität nachgewiesen werden. Der zugehörige p-Wert der dargestellten Regressionslinie beträgt 0,390 (nicht signifikant). Für Kosovo, Simbabwe sowie die Palästinensischen Gebiete lagen keine Daten für den Index vor, so dass ZPE aus diesen Kontexten nicht in die Analyse eingeflossen sind.

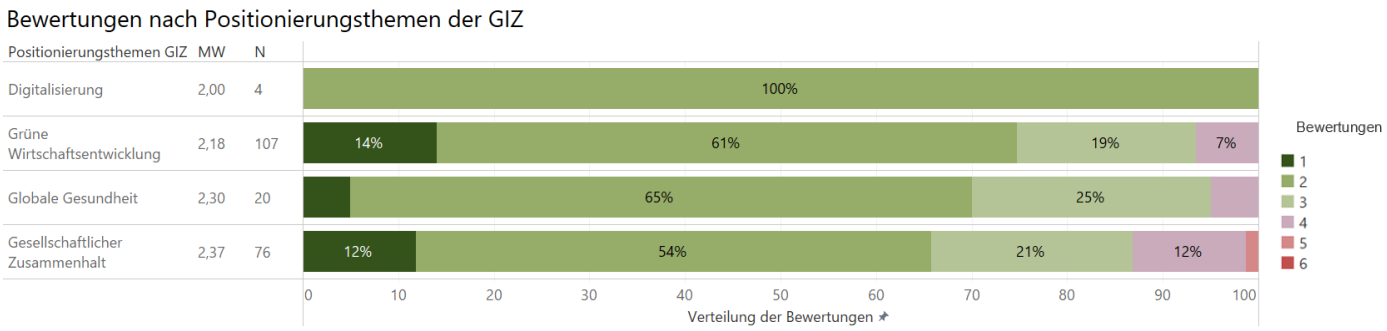
Abbildung 31: Bewertungen nach Fragilität II (Fragile State Index) *(N=132)*




## Weitere Abbildungen zu Kapitel 3.6: Themen und Sektoren

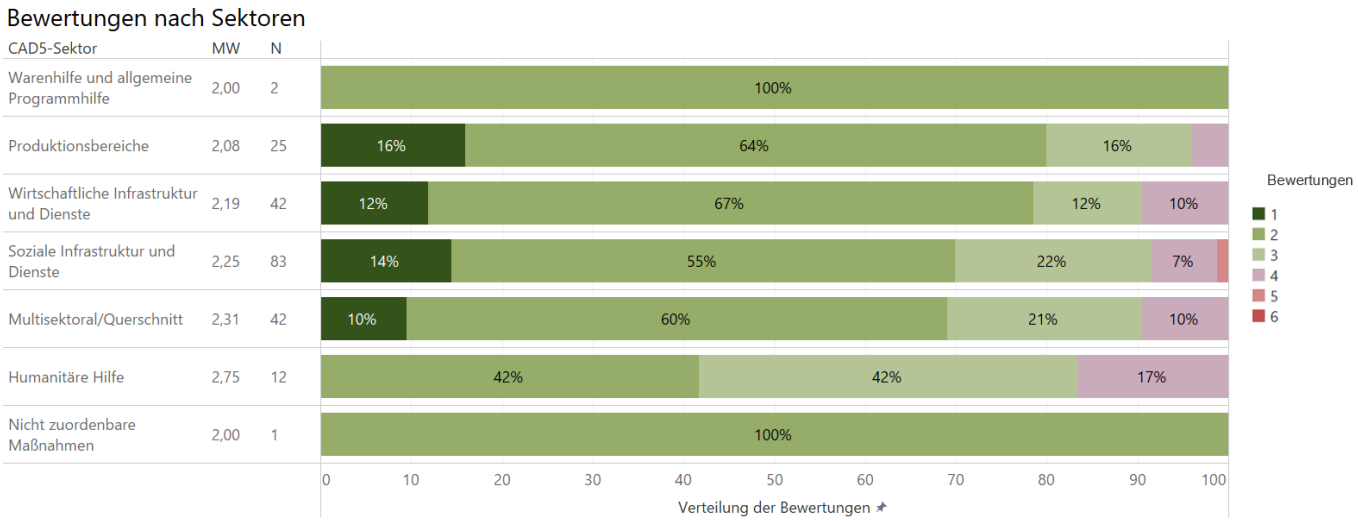
In **Abbildung 32** werden die ZPE-Bewertungen nach unterschiedlichen **Positionierungsthemen** dargestellt. Hier ist kein Trend zu erkennen, nicht zuletzt, weil die Verteilung sehr ungleichmäßig ausfällt.

Abbildung 32: Bewertungen nach GIZ-Positionierung *(N=207)*



In **Abbildung 33** werden die ZPE-Bewertungen nach unterschiedlichen **CRS Sektor Codes** dargestellt Hier ist kein Trend zu erkennen, nicht zuletzt weil die Verteilung sehr ungleichmäßig ausfällt.

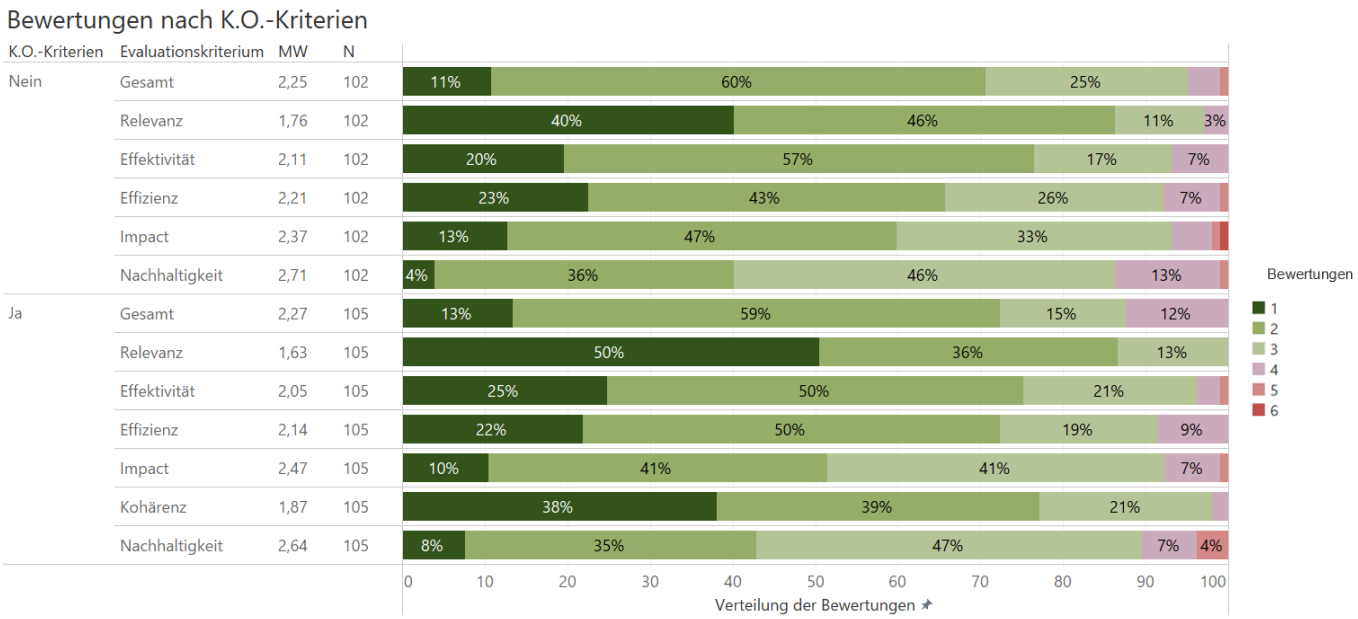
Abbildung 33: Bewertungen nach CRS Sektor Codes *(N=207)*




## Weitere Abbildungen zu Kapitel 3.7: Unterschiede in Evaluierungen

In **Abbildung 34** werden die ZPE-Bewertungen für die einzelnen Evaluierungskriterien vor und nach Einführung der K.O.-Kriterien dargestellt. Der Vergleich bezieht sich auf Abbildung 18 im Text.

Abbildung 34: Bewertungen für die einzelnen Evaluierungskriterien vor und nach Einführung der K.O.-Kriterien *(N=207)*



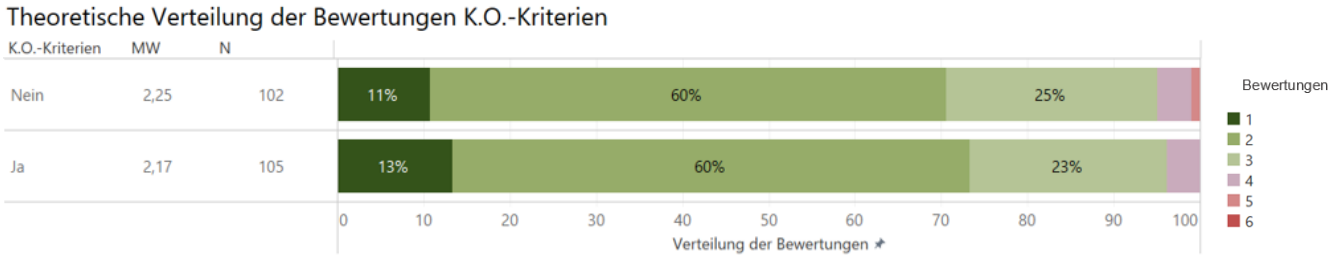
In **Abbildung 35** wird darüber hinaus der (gestiegene) Anteil an „schlechten“ Bewertungen der Note 3 für die K.O.-Kriterien Effektivität, Impact und Nachhaltigkeit dargestellt. Eine „schlechte“ Bewertung der Note 3 wurde als eine Punktzahl zwischen 67 und 70 für das jeweilige Kriterium definiert, d. h. knapp vor der Schwelle zur Bewertungsstufe 4 (ab 66 Punkte oder niedriger). Die Interpretation der Daten ist dem Kapitel zu entnehmen.

Abbildung 35: Anteil an „schlechten“ Bewertungen der Note 3 für die Evaluierungskriterien Effektivität, Impact und Nachhaltigkeit vor und nach Einführung der K.O.-Kriterien *(N=207)*

| | N | Effektivität | | | Impact | | | Nachhaltigkeit | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N (Gesamt) | N (Note 3) | davon "schlechte" Bewertung der Note 3 | | N (Note 3) | davon "schlechte" Bewertung der Note 3 | | N (Note 3) | davon "schlechte" Bewertung der Note 3 | |
| mit K.O.-Kriterien | 105 | 22 | **10** | **10%** | 43 | **16** | **15%** | 49 | **22** | **21%** |
| ohne K.O.-Kriterien | 102 | 17 | **4** | **4%** | 34 | **9** | **9%** | 47 | **15** | **15%** |

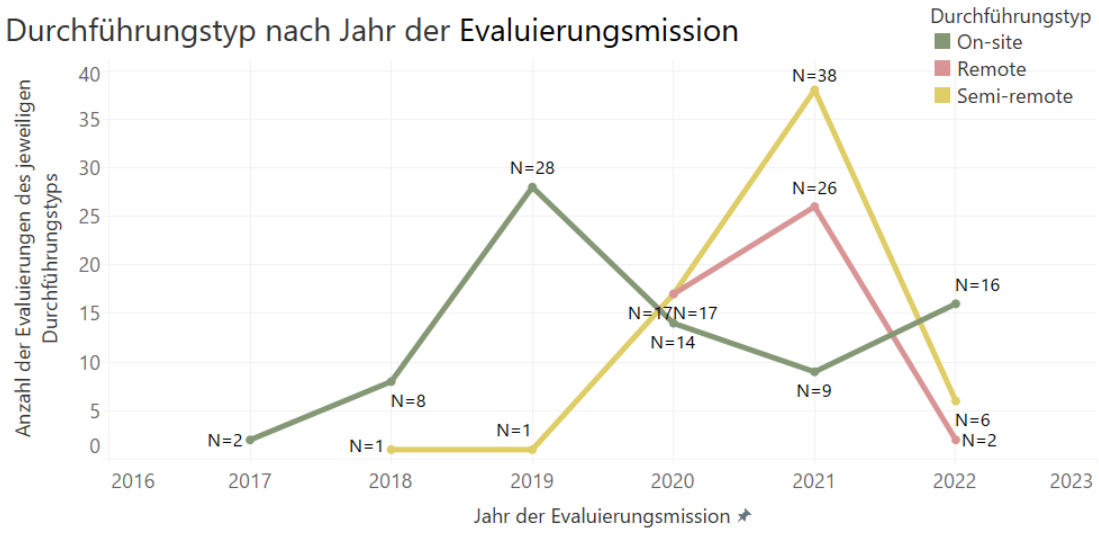
In **Abbildung 36** wird die theoretische Verteilung der ZPE-Bewertungen dargestellt, wenn die K.O.-Kriterien nicht eingeführt worden wären. Der Vergleich bezieht sich auf Abbildung 18 im Text. Die Interpretation der Daten ist dem Kapitel zu entnehmen.

Abbildung 36: Bewertungen ohne K.O.-Kriterien *(N=207)*




In **Abbildung 37** wird die Anzahl der Evaluierungsmissionen nach Durchführungstyp und Durchführungsjahr dargestellt. Vor-Ort-Missionen sind in Gelb dargestellt, *semi-remote*-Missionen in Grün, und *remote*-Missionen in Rot.

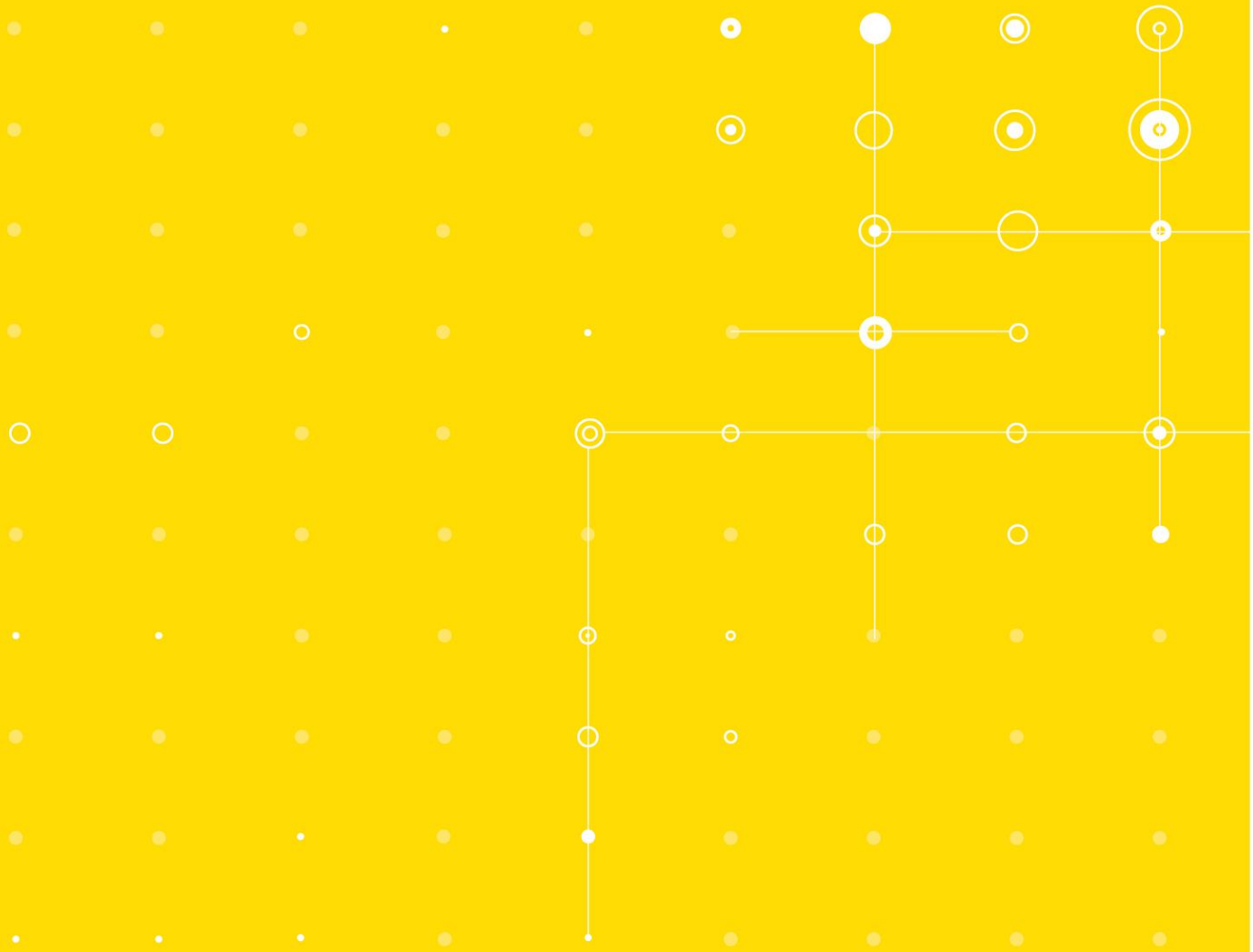
Abbildung 37: Anzahl der Evaluierungsmissionen nach Durchführungstyp und Durchführungsjahr *(N=185)*



# Anhang 2 – Tabelle zu Erfolgs- und Misserfolgsfaktoren für die qualitativ untersuchten Trends

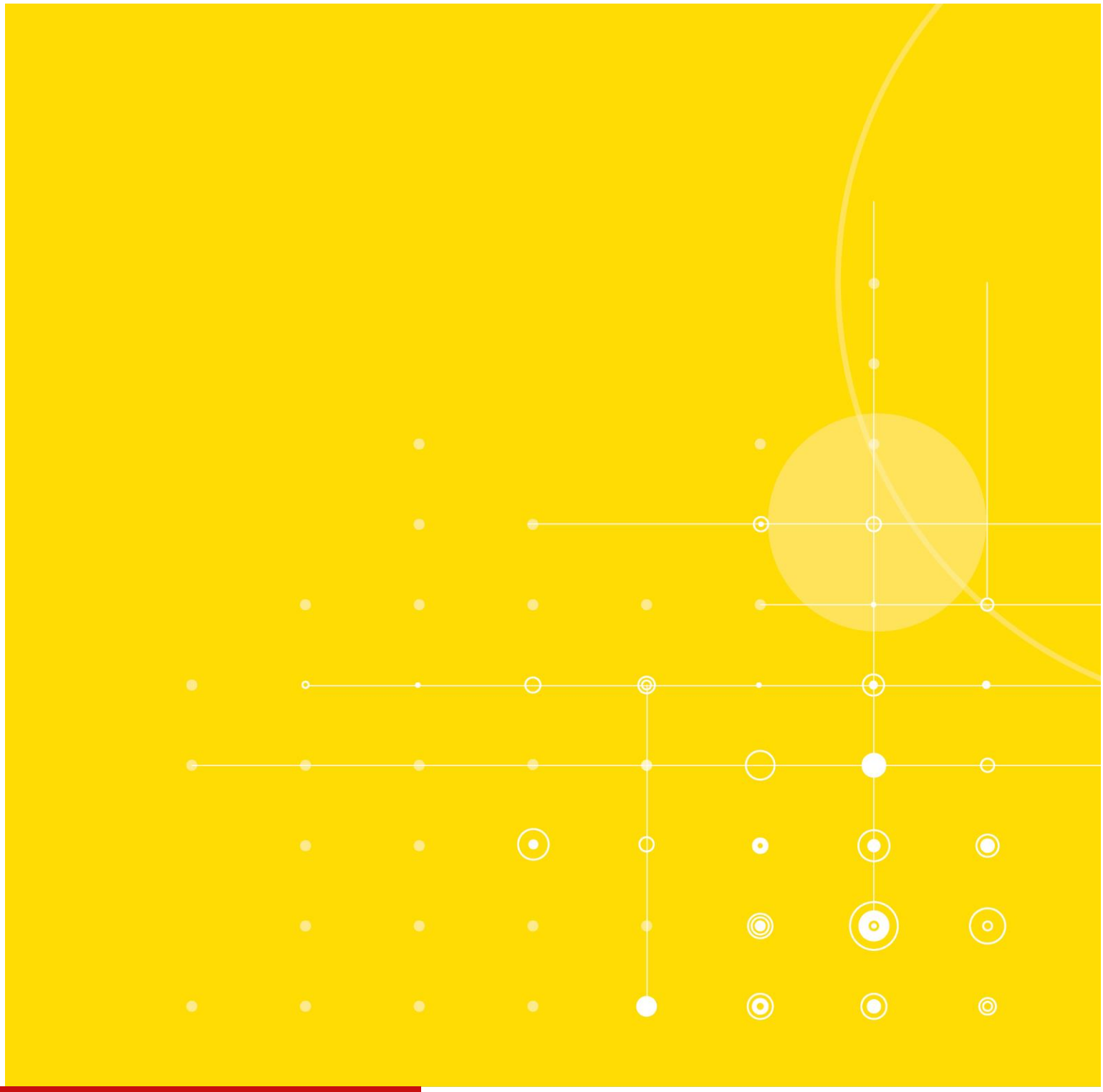
Tabelle 1: Erfolgs- und Misserfolgsfaktoren nach Projekttyp, Projektvolumen, operativem GIZ-Bereich und BMZ-Länderkategorie

| Trend | | Erfolgsfaktoren | Misserfolgsfaktoren |
|---|---|---|---|
| **Project type** | Bilaterale Vorhaben | - Ownership der Partner | - Kontext, insbesondere Einschränkungen durch die COVID-19-Pandemie sowie die Sicherheitslage |
| | Globalvorhaben, Sektorvorhaben, IZR-Vorhaben und Regionalvorhaben | - Zusammenarbeit mit beteiligten Stakeholdern<br>- Fachkompetenz des Projektteams<br>- Engagement einzelner Projektmitarbeiter *innen<br>- Synergien mit anderen Akteuren der deutschen EZ bzw. Akteuren der IZ | - Fehlende Ownership anderer Stakeholder (nicht Partner)<br>- Einschränkungen durch den Projekttyp |
| **Projektvolumen** | Vorhaben mit kleinerem Projektvolumen (bis zu 5 Mio. EUR) | - Zusammenarbeit mit beteiligten Stakeholdern<br>- Fachkompetenz des Projektteams<br>- Steuerung des Vorhabens, z. B. Monitoringsystem | - Zeitliche, finanzielle, personelle Ressourcen |
| | Vorhaben mit größerem Projektvolumen (15 Mio. EUR oder mehr) | - Ownership der Partner | - Zusammenarbeit mit beteiligten Stakeholdern, insbesondere bürokratische Hürden |
| **Operativer GIZ-Bereich** | Asien/Pazifik, Lateinamerika/Karibik und Überregional | - Zusammenarbeit mit beteiligten Stakeholdern<br>- Ownership der Partner<br>- Fachkompetenz des Projektteams<br>- Nachhaltigkeitsstrategien | - Zusammenarbeit mit beteiligten Stakeholdern |
| | Europa/Kaukasus/Zentralasien, Mittelmeer/Mittlerer Osten und Subsaharaafrika | - Ownership der Partner | - Kontext, insbesondere Einschränkungen durch die COVID-19-Pandemie sowie die Sicherheitslage<br>- Herausforderungen bei der Steuerung des Vorhabens |
| **BMZ-Länderkategorie** | Globale Partner und bilaterale Partner sowie multilaterale, europäische und nichtstaatliche Partner | - Zusammenarbeit mit den beteiligten Stakeholdern<br>- Ressourcenausstattung | - Zusammenarbeit mit beteiligten Stakeholdern |
| | Reformpartner, Transformationspartner sowie Nexus- und Friedenspartner | - Chancen im Partnerland | - Ownership der beteiligten Stakeholder<br>- Herausforderungen bei der Steuerung des Vorhabens<br>- Kontext, insbesondere Einschränkungen durch die COVID-19-Pandemie sowie die Sicherheitslage |



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

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



**Deutsche Gesellschaft für**
**Internationale Zusammenarbeit (GIZ) GmbH**

Registered offices
Bonn and Eschborn

Friedrich-Ebert-Allee 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