

WISSEN WAS WIRKT

# Auswertung der durchschnittlichen Bewertungen von Zentralen Projektevaluierungen

Querschnittsauswertung – Hauptbericht

Im Auftrag der GIZ durchgeführt von externen Evaluator\*innen

## Impressum

Als weltweit tätiger Dienstleister der internationalen Zusammenarbeit für nachhaltige Entwicklung und internationale Bildungsarbeit entwickelt die GIZ mit ihren Partnern wirksame Lösungen, die Menschen Perspektiven bieten und deren Lebensbedingungen dauerhaft verbessern. Als gemeinnütziges Bundesunternehmen unterstützt sie die Bundesregierung und viele weitere öffentliche und private Auftraggeber in unterschiedlichsten Themenfeldern – von der Wirtschafts- und Beschäftigungsförderung über Energie und Umweltthemen bis hin zur Förderung von Frieden und Sicherheit.

Die Stabsstelle Evaluierung untersteht organisatorisch direkt dem Vorstand. Diese Organisationsstruktur stellt ihre Unabhängigkeit vom operativen Geschäft sicher. Die Stabsstelle Evaluierung generiert evidenzbasierte Ergebnisse und Empfehlungen. Sie sorgt für Transparenz der Erkenntnisse aus Evaluierungen und fördert deren Nutzung.

### Evaluator\*innen:

Lennart Raetzell, Matias Kraemer, Alexandra Hoppe, David Arens, Syspons GmbH

### Autor\*innen:

Lennart Raetzell, Matias Kraemer, Alexandra Hoppe, David Arens, Syspons GmbH

### Consulting:

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



### Konzeption, Koordination und Management

Claudia Kornahrens, GIZ, Gruppenleiterin  
Dr. Vera Hundt, GIZ, Spezialistin Evaluierung  
Benjamin Bräuer, GIZ, Spezialist Evaluierung  
Lucas Jacobs, GIZ, Spezialist Evaluierung  
Jana Moneke, GIZ, Spezialistin Evaluierung  
Gruppe Zentrale Projektevaluierung  
Stabsstelle Evaluierung der GIZ

### Verantwortlich:

Albert Engel, GIZ, Leiter Stabsstelle Evaluierung

### Lektorat:

Klaus Ahting

### Herausgeberin:

Deutsche Gesellschaft für  
Internationale Zusammenarbeit (GIZ) GmbH

### Sitz der Gesellschaft

Bonn und Eschborn

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

E [evaluierung@giz.de](mailto:evaluierung@giz.de)  
I [www.giz.de/wissenwaswirkt](http://www.giz.de/wissenwaswirkt)  
[www.youtube.com/user/GIZonlineTV](https://www.youtube.com/user/GIZonlineTV)  
[https://twitter.com/giz\\_gmbh](https://twitter.com/giz_gmbh)

### Design

now [nau], kommunikative & visuelle gestaltung,  
Frankfurt  
DITHO Design GmbH, Köln

### Vertrieb

GIZ, Bonn

Bonn 2023

# Inhalt

Tabellen- und Abbildungsverzeichnis .....	4
Abkürzungsverzeichnis .....	6
1 Einleitung .....	7
2 Methodisches Vorgehen .....	8
2.1 Untersuchungsgegenstand .....	8
2.2 Datengrundlage und Evidenzstärke .....	9
2.3 Auswertungsprozess .....	10
Vorgehen in der Vorbereitungsphase.....	10
Statistische Aufbereitung und Bewertung der ZPE Durchschnittsbewertungen.....	10
Inhaltliche Analyse der Evaluierungsberichte.....	11
Berichtslegung und Präsentation .....	12
3 Ergebnisse der Querschnittsauswertung .....	13
3.1 Erfolgsquote der ZPE .....	13
3.2 Auswertung der ZPE und früherer Instrumente im Zeitverlauf .....	15
3.3 Auswertung der ZPE nach Evaluierungskriterien.....	16
3.4 Auswertung der ZPE nach Projektmerkmalen .....	19
Projektvolumen.....	20
Projektlaufzeit.....	22
3.5 Auswertung der ZPE nach Ländern und Regionen.....	22
Kontinente .....	22
Weltregionen .....	24
Regionalbereiche.....	25
BMZ-Länderkategorien.....	26
Fragilität.....	28
3.6 Auswertung der ZPE nach Themen und Sektoren.....	29
BMZ-Kernthemen .....	29
BMZ-Aktionsfelder .....	30
3.7 Abhängigkeit der Bewertung von Unterschieden bei der Evaluierung .....	32
K.O.-Kriterien.....	32
Die COVID-19-Pandemie und ihr Einfluss auf die ZPE.....	33
Anhang.....	37

## Tabellen- und Abbildungsverzeichnis

Tabelle 1: Erfolgs- und Misserfolgskriterien nach Projekttyp, Projektvolumen, operativem GIZ-Bereich und BMZ-Länderkategorie .....	45
Abbildung 1: Grundgesamtheit und Stichprobe .....	9
Abbildung 2: Grundgesamtheit und Stichprobe mit Erfolgsquote .....	13
Abbildung 3: Erfolgsquote mit Konfidenzintervall.....	14
Abbildung 4: Bewertungen im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum) .....	15
Abbildung 5: Bewertungen nach Evaluierungskriterien .....	16
Abbildung 6: Durchschnittliche Punkteverteilung .....	18
Abbildung 7: Bewertungen nach Evaluierungskriterien im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum).....	18
Abbildung 8: Bewertungen nach Projekttyp (N=207).....	19
Abbildung 9: Bewertung Projektvolumen (Kategorien) (N=207).....	20
Abbildung 10: Bewertungen nach Projektlaufzeit (Kategorien) (N=207).....	22
Abbildung 11: Durchschnittliche Bewertungen nach Kontinent unter Berücksichtigung der Sektoren und BMZ-Kernthemen (N=207) .....	23
Abbildung 12: Durchschnittlichen Bewertungen nach Weltregionen unter Berücksichtigung von Sektoren und BMZ-Kernthemen (N=207).....	24
Abbildung 13: Bewertungen nach operativen Bereichen (Regionalbereiche der GIZ) (N=207) .....	25
Abbildung 14: Bewertungen nach BMZ-Länderkategorien (N=138) .....	27
Abbildung 15: Bewertungen nach Fragilität (N=136).....	28
Abbildung 16: Bewertungen nach BMZ-Kernthemen.....	30
Abbildung 17: Bewertungen nach BMZ-Aktionsfeldern .....	30
Abbildung 18: Bewertungen vor und nach Einführung der K.O.-Kriterien (N=207).....	32
Abbildung 19: Durchschnittliche Bewertungen nach dem Zeitpunkt der Evaluierungsmission (Einfluss der COVID-19-Pandemie) (N=205).....	34
Abbildung 20: Bewertungen nach Durchführungstyp der Evaluierungsmission (N=205) .....	34
Abbildung 21: Bewertungen nach Durchführungstyp der Evaluierungsmission (N=205) .....	36
Abbildung 22: Bewertungen im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum) (N=207).....	37
Abbildung 23: Bewertungen nach Evaluierungskriterien im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum) (N=207).....	38
Abbildung 24: Bewertungen nach Haushaltstitel (N=207) .....	38

Abbildung 25: Bewertungen nach Kofinanzierung (N=207).....	39
Abbildung 26: Durchschnittliche Bewertungen nach Kontinenten ohne Berücksichtigung von Sektoren und BMZ-Kernthemen (N=207).....	39
Abbildung 27: Durchschnittliche Bewertungen nach Weltregionen ohne Berücksichtigung von Sektoren und BMZ-Kernthemen (N=207).....	40
Abbildung 28: Bewertungen nach Weltregionen (N=207).....	40
Abbildung 29: Durchschnittliche Bewertungen nach differenzierten Weltregionen unter Berücksichtigung von Sektoren und BMZ-Kernthemen (N=207).....	41
Abbildung 30: Durchschnittliche Bewertungen nach operativen Bereichen unter Berücksichtigung von Sektoren und BMZ-Kernthemen (N=207).....	41
Abbildung 31: Bewertungen nach Fragilität II (Fragile State Index) (N=132).....	42
Abbildung 32: Bewertungen nach GIZ-Positionierung (N=207).....	42
Abbildung 33: Bewertungen nach CRS Sektor Codes (N=207).....	43
Abbildung 34: Bewertungen für die einzelnen Evaluierungskriterien vor und nach Einführung der K.O.-Kriterien (N=207).....	43
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).....	44
Abbildung 36: Bewertungen ohne K.O.-Kriterien (N=207).....	44
Abbildung 37: Anzahl der Evaluierungsmissionen nach Durchführungstyp und Durchführungsjahr (N=185).	44

## Abkürzungsverzeichnis

BMZ	Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung
DAC	Ausschuss für Entwicklungshilfe der OECD ( <i>OECD Development Assistance Committee</i> )
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH
MW	Mittelwert
N	Gesamtanzahl der Grundgesamtheit
n	Größe der Stichprobe
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung ( <i>Organisation for Economic Cooperation and Development</i> )
PEV	dezentrale Projektevaluierung
PN	Projektnummer
QSA	Querschnittsauswertung
SDG	<i>Sustainable Development Goals</i> [globale Nachhaltigkeitsziele]
UE	unabhängigen Portfolioevaluierung
ZPE	Zentrale Projektevaluierung

# 1 Einleitung

## Hintergrund und Ziele der Querschnittsauswertung

Die Syspons GmbH wurde von der Deutschen Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ) damit beauftragt, eine **Querschnittsauswertung** von Zentralen Projektevaluierungen (ZPE) nach OECD DAC-Durchschnitts- und Gesamtbewertungen durchzuführen (kurz: *QSA Insights*).<sup>1</sup>

In Querschnittsauswertungen werden Erkenntnisse aus Evaluierungen der GIZ analytisch aufbereitet und zusammenfassend dargestellt, um diese für das unternehmensweite Lernen nutzbar zu machen. Bei der Querschnittsauswertung *QSA Insights* bestand das Ziel darin, umfassende Erkenntnisse aus dem Evaluierungsinstrument der **Zentralen Projektevaluierungen** zu gewinnen. Mit den ZPE erfasst die GIZ seit 2017/2018 regelmäßig und systematisch die Wirkung, die Wirtschaftlichkeit und die Nachhaltigkeit von Vorhaben, die sie zusammen mit Partnern im Auftrag des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) umsetzt. Gegenstand der ZPE im BMZ-Geschäft sind Projekte, die die GIZ im Auftrag des BMZ durchführt, um entweder Veränderungsprojekte der Partner zu unterstützen oder Sektor- und Globalfragen zu bearbeiten. Die Bewertung erfolgt anhand der sechs OECD DAC-Evaluierungskriterien Relevanz, Kohärenz (seit Beginn 2021), Effektivität, Effizienz, Impact und Nachhaltigkeit (BMZ, 2021).

Grundsätzlich dienen Evaluierungen von im Auftrag des BMZ durchgeführten Projekten in der GIZ drei Grundfunktionen: Sie sollen Transparenz und Rechenschaftslegung gewährleisten, evidenzbasierte Entscheidungen unterstützen und einen Beitrag zu Wissensmanagement und organisationsbezogenem Lernen leisten. Planung, Durchführung und Nutzung von Evaluierungen gestaltet die GIZ daher so, dass der Beitrag des Evaluierungsprozesses und der Evaluierungsergebnisse zu diesen drei Grundfunktionen optimiert wird.

Seit 2020 hat die **COVID-19-Pandemie** erhebliche Auswirkungen auf die Arbeit der GIZ sowie den Ablauf der von der Stabsstelle Evaluierung gesteuerten Zentralen Projektevaluierungen. Viele Vorhaben haben den neuen Rahmenbedingungen durch angepasste bzw. zusätzliche Aktivitäten und Veränderungen der Leistungen und Zielsysteme Rechnung getragen. Gleichzeitig hat die Pandemie die Vorhaben vor unterschiedlichste Herausforderungen gestellt. Für die Zentralen Projektevaluierungen wurden sonst übliche Feldmissionen aufgrund der Reisebeschränkungen und Maßnahmen zur Eindämmung der Pandemie durch aus der Ferne durchgeführte *remote*-Missionen ersetzt. Gemäß den im jeweiligen Land geltenden Regelungen und der pandemischen Lage entsprechend wurde entweder ein *semi-remote*-Setup oder ein reines *remote*-Setup angewandt.<sup>2</sup>

Vor diesem Hintergrund wurde im Rahmen der *QSA Insights* eine Querschnittsauswertung von bisher abgeschlossenen ZPE nach OECD DAC-Bewertungen unter Berücksichtigung der Pandemie durchgeführt. Konkret wurde dabei eine **statistische Auswertung und Darstellung der OECD DAC-Durchschnitts- und Gesamtbewertungen** für alle bis zum Stichtag verfügbaren ZPE sowie die Ermittlung von möglichen erklärenden Faktoren angestrebt. Dazu wurden gemeinsam mit der Stabsstelle Evaluierung unterschiedliche Fragestellungen und Erkenntnisinteressen definiert. Diese haben Eingang in ein **Analyseraster** gefunden, in der diese weiter operationalisiert wurden

---

<sup>1</sup> Die Querschnittsauswertung hat neben der Untersuchung von ZPE nach OECD-DAC Durchschnitts- und Gesamtbewertungen zwei weitere Komponenten, nämlich (1) die Ermittlung des Einflusses der COVID-19-Pandemie auf Vorhaben und Evaluierungen sowie (2) eine Auswertung von Trends in ausgewählten Themenfeldern. Diese Komponenten wurden im Rahmen von gesonderten Berichten der GIZ vorgelegt.

<sup>2</sup> Im Falle eines *semi-remote*-Setups können sich die lokalen Evaluatoren\*innen vor Ort bewegen, wohingegen sie in einem reinen *remote*-Setup von zu Hause aus operieren und die Mission gänzlich virtuell durchgeführt wird. In beiden Konstellationen wird auf eine physische Anreise der internationalen Evaluatoren\*innen verzichtet.

Der Bericht ist wie folgt aufgebaut: In Kapitel 2 wird das methodische Vorgehen der Querschnittsauswertung dargestellt. In diesem Kapitel wird zunächst der Untersuchungsgegenstand beschrieben. Dann werden die Datengrundlage und die Datenqualität bzw. Evidenzstärke diskutiert und eingeordnet. Anschließend wird der Auswertungsprozess erläutert. Im Anschluss werden in Kapitel 3 die Ergebnisse der Auswertung und Analyse vorgestellt und diskutiert. Die Ergebnisse folgen zwei Ansätzen: Zum einen wurde eine statistische Auswertung der ZPE anhand von OECD DAC-Durchschnitts- und Gesamtbewertungen für verschiedene Kriterien und Themenbereiche durchgeführt. Zum anderen wurden auf dieser Grundlage statistische Trends (Unterschiede in den Durchschnittsbewertungen) beschrieben und an ausgewählten Stellen durch eine qualitative Inhaltsanalyse der ZPE-Berichte erklärende Faktoren ermittelt. Kapitel 4 enthält den Anhang.<sup>3</sup>

## 2 Methodisches Vorgehen

In diesem Kapitel wird das methodische Vorgehen der Querschnittsauswertung zur (statistischen) Auswertung und Analyse der OECD DAC-Durchschnitts- und Gesamtbewertungen der ZPE vorgestellt. Im folgenden Abschnitt wird zunächst der Untersuchungsgegenstand beschrieben. Anschließend werden die Datengrundlage und die Datenqualität in Bezug auf die Auswertung eingeordnet. Die einzelnen Auswertungsschritte werden im letzten Abschnitt über den Auswertungsprozess dargestellt.

### 2.1 Untersuchungsgegenstand

Der Untersuchungsgegenstand für die Auswertung und Analyse waren **207 zentrale Projektevaluierungen**, die bis zum Stichtag 15.08.2022 abgeschlossen waren und der GIZ zur Verfügung standen. Für die Auswahl der durch ZPE zu evaluierenden Vorhaben arbeitet die GIZ mit einer mit dem BMZ abgestimmten **repräsentativen Stichprobe**. Eine jährliche (geschichtete) Zufallsstichprobe von etwa 40 Prozent der Grundgesamtheit – alle BMZ-finanzierten Vorhaben mit einem Auftragswert von mindestens drei Millionen Euro, die im Folgejahr auslaufen – ermöglicht dabei statistisch repräsentative Aussagen über die Erfolgsquote aller in einem **Zeitraum von zwei Jahren** abgeschlossenen Vorhaben. Ergänzt wird das ZPE-Portfolio durch Evaluierungen von Vorhaben, die aufgrund einer besonderen strategischen, politischen oder auch methodischen Relevanz vorgeschlagen werden.<sup>4</sup> Die vorliegende Analyse beschränkt sich auf ZPE, die im Rahmen der repräsentativen Zufallsstichproben ausgewählt wurden.

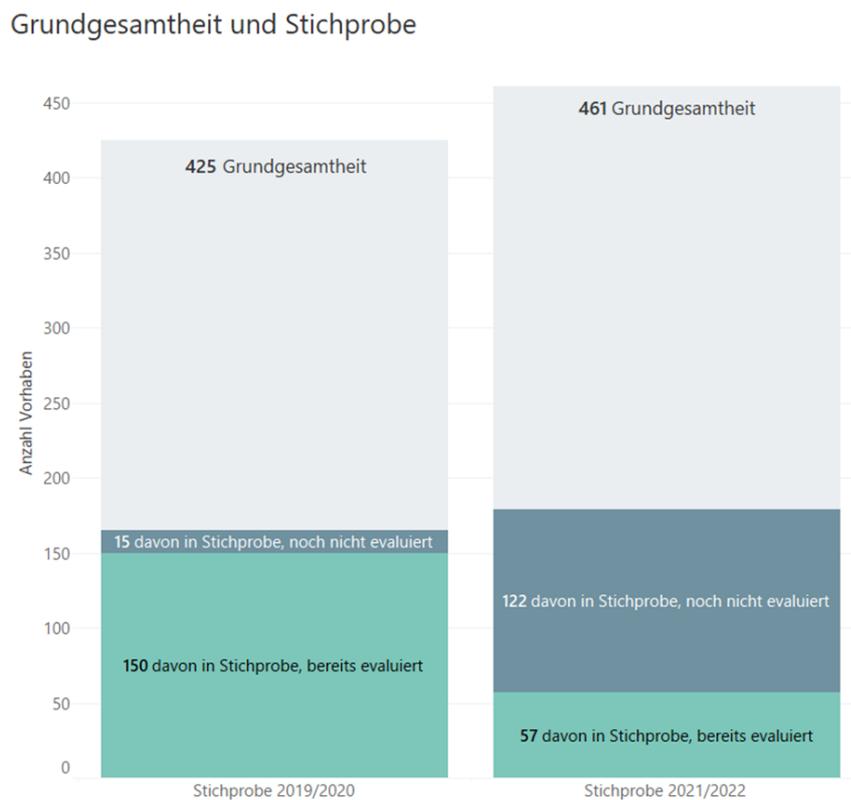
Die 207 vorliegenden ZPE wurden aus **insgesamt vier jährlichen Stichproben** gezogen. Diese vier ZPE-Stichproben stellen **zwei repräsentative Auswahlverfahren** dar. Abbildung 1 zeigt, dass für die ersten beiden Stichproben mit einem geplanten Projektende im Zeitraum 2019/2020 eine Grundgesamtheit von 425 Projekten vorlag. Aus dieser Grundgesamtheit wurden 165 Projekte für die Evaluierung durch die ZPE ausgewählt. Dies entspricht einem Deckungsgrad von 39 Prozent. Für 150 dieser Evaluierungen lagen bereits Ergebnisse vor, bei 15 war die Evaluierung noch nicht abgeschlossen (91 Prozent abgeschlossen). Für die letzten beiden Stichproben mit geplantem Projektende im Zeitraum 2021/2022 wurden 179 Projekte aus einer Grundgesamtheit von 461 ausgewählt. Dies ergibt ebenfalls einen Deckungsgrad von 39 Prozent. Da die Stichproben später gezogen wurden, ist die Zahl der bereits durchgeführten Evaluierungen geringer. Für diese Stichproben lagen 57 Ergebnisse vor, während 122 ZPE noch liefen (32 Prozent abgeschlossen).<sup>5</sup>

<sup>3</sup> Im Gesamtkontext der Querschnittsauswertung wird der vorliegende Bericht durch zwei weitere eigenständige Berichte für die beiden anderen definierten Zielsetzungen ergänzt.

<sup>4</sup> GIZ, 2022: <https://www.giz.de/de/downloads/giz2022-de-instrumentenpapier-zentrale-projektevaluierungen-bmz.pdf>

<sup>5</sup> Im folgenden Kapitel wird im Sinne der Verständlichkeit nur zwischen erster Stichprobe 2019/2020 und zweiter Stichprobe 2021/2022 unterschieden.

Abbildung 1: Grundgesamtheit und Stichprobe



## 2.2 Datengrundlage und Evidenzstärke

Die zentrale Datengrundlage für die statistische Auswertung war zum einen eine Liste mit Bewertungen und Rahmendaten aller beauftragten ZPE, die dem Auswertungsteam zu Projektbeginn von der GIZ zur Verfügung gestellt wurde. Zum anderen wurden dem Auswertungsteam die Evaluierungsberichte aller abgeschlossenen ZPE zur Verfügung gestellt, um im Rahmen einer inhaltlichen Analyse mögliche Erklärungen für die Unterschiede in den Durchschnittsbewertungen (statistische Trends) zu ermitteln.

Zur **Evidenzstärke der statistischen Auswertung der Erfolgsquote der ZPE** für die Erfolgsquote der Grundgesamtheit ist zunächst anzumerken, dass die Aussagekraft aufgrund der noch nicht vollständig erhobenen Stichproben, wie im vorangegangenen Kapitel beschrieben, noch größere statistische Unsicherheiten enthält. Dabei zeigen die in Kapitel 3.1 beschriebenen Konfidenzintervalle die aktuelle Evidenzstärke der Aussagen zur Erfolgsquote der Grundgesamtheit. Im Hinblick auf die **Rahmendaten** der statistischen Analyse der OEC DAC-Bewertungen ist anzumerken, dass die bereits erhobenen und von der GIZ bereitgestellten Rahmendaten einige Datenlücken enthielten, die geschlossen werden mussten, um eine vollständige Analyse durchführen zu können. Dementsprechend hat das Syspons-Team die Datengrundlage aufbereitet und an einigen Stellen ergänzt bzw. aktualisiert. Für bestimmte Rahmendaten (zum Beispiel die thematische Zuordnung der evaluierten Projekte nach den Kernthemen und Aktionsfeldern des BMZ) erfolgte die Zuordnung nach Absprache direkt durch Syspons und nicht nach offiziell festgelegten Kriterien. Dabei wurden insbesondere Informationen aus den Evaluierungsberichten der abgeschlossenen Zentralen Projektevaluierungen verwendet, wodurch das Risiko einer Beeinflussung der Analyse minimiert wurde. Die Datenlücken in den Rahmendaten konnten vollständig geschlossen werden, so dass die Datengrundlage für die statistische Analyse anhand der OECD DAC-Bewertungen als *sehr gut* bezeichnet werden kann.

Für die **inhaltliche Analyse der Faktoren**, die die durchschnittlichen Unterschiede in der Bewertung erklären (statistische Trends), wurden qualitative Daten aus 112 Evaluierungsberichten erhoben. Die vorliegende inhaltliche Analyse stützt sich somit auf 66 Prozent der verfügbaren Evaluierungsberichte. Die Auswahl der Evaluierungsberichte erfolgte systematisch nach zwei Kriterien: Erstens wurden die Evaluierungsberichte auf der Grundlage der vorangegangenen statistischen Auswertung ausgewählt. Zunächst wurden Unterschiede in den Bewertungen ermittelt (zum Beispiel durchschnittliche ZPE-Bewertungen von bilateralen Projekten im Gegensatz zu Bewertungen von (über-)regionalen Projekten). Dann wurden für jeden identifizierten Trend die drei am besten und die drei am schlechtesten bewerteten ZPE für die Inhaltsanalyse ausgewählt. Durch diese Auswahl der Evaluierungsberichte wurde sichergestellt, dass sowohl besonders gute Ansätze als auch besonders kritisch bewertete Ansätze berücksichtigt wurden und somit Erklärungsmuster pro Trend ermittelt werden konnten. Zweitens umfasst die Auswahl die 90 aktuellsten ZPE, um möglichst aktuelle Muster zu erkennen und praxisnahe Ansätze liefern zu können.<sup>6</sup>

Trotz dieser Auswahl ist zu beachten, dass die Evidenzstärke in Bezug auf die erklärenden Faktoren von der Aussagekraft der Evaluierungsberichte zum jeweils untersuchten Trend abhing. So können zum Beispiel unterschiedliche ZPE-Durchschnittsbewertungen in Bezug auf den Projekttyp nur dann begründet werden, wenn in den Evaluierungen direkt oder zumindest indirekt darauf Bezug genommen wurde. Die Datengrundlage für die Inhaltsanalyse kann an dieser Stelle als *gut* bezeichnet werden. Im Vergleich zu den statistischen Daten stellt dieser Aspekt jedoch eine Begrenzung der Querschnittsauswertung dar.

## 2.3 Auswertungsprozess

### Vorgehen in der Vorbereitungsphase

Ziel der Vorbereitungsphase war es, die Schwerpunkte der Querschnittsauswertung festzulegen und einen Überblick über den Untersuchungsgegenstand sowie die verfügbare Datengrundlage zu erhalten. Zu diesem Zweck hat das Syspons-Team zunächst ein ausführliches Auftragsklärungsgespräch mit der Stabsstelle Evaluierung der GIZ als Auftraggeberin der Auswertung geführt.

Im Anschluss hat Syspons vertiefte Interviews mit ausgewählten Vertreter\*innen der Stabsstelle Evaluierung der GIZ geführt, um Hypothesen, Vorstellungen und Erkenntnisinteressen für die einzelnen Fragestellungen vertieft zu erfragen. Die Ergebnisse dieser Interviews sind in die weiteren Analyseschritte eingeflossen und haben Eingang in die am Ende der Vorbereitungsphase vorgestellte Concept Note sowie das Analyseraster gefunden.

### Statistische Aufbereitung und Bewertung der ZPE Durchschnittsbewertungen

Das Ziel der statistischen Aufbereitung der Daten bestand einerseits darin, die OECD DAC-Durchschnittsbewertungen der 207 vorliegenden ZPE anhand unterschiedlicher Kriterien und Themenfelder auszuwerten und darzustellen, um einen breiten Überblick über die Bewertungen zu erlangen. In einem zweiten Schritt wurden die ersten Ergebnisse betrachtet und auffällige Bewertungsunterschiede (statistische Trends) ermittelt und mit der Stabsstelle abgestimmt.

Parallel zur Vorbereitungsphase hat sich das Evaluierungsteam mit der Datengrundlage vertraut gemacht und sie für die statistische Auswertung vorbereitet. Die bereitgestellten Daten zu den Bewertungen und Rahmendaten aller ZPE wurden in Excel bereinigt und Datenlücken identifiziert. Diese wurden durch Informationen aus den Evaluierungsberichten oder durch weitere Recherchen des Evaluierungsteams geschlossen. Die Daten wurden anschließend in der Visualisierungssoftware Tableau grafisch aufbereitet. Die Kriterien und Themenfelder für die Auswertung wurden auf Grundlage der Terms of Reference (ToR) mit der Stabsstelle Evaluierung

---

<sup>6</sup> Im Zuge der weiteren Datenerhebung der QSA wurden auch alle übrigen Evaluierungsberichte untersucht und in die Analyse einbezogen.

im Auftragsklärungsgespräch sowie in den vertiefenden Interviews abgestimmt. Zur Darstellung der Bewertungen nach diesen Kriterien und Themenfeldern wurden zum einen die Vorstellungen der GIZ zu bestimmten Auswertungstypen (zum Beispiel Darstellung der Stichprobe und geographische Karten) berücksichtigt, zum anderen wurde ein explorativer Ansatz für weitere Darstellungen verfolgt. Das Produkt der ersten Auswertung wurde der GIZ im PPT-Format zeitgleich mit der Concept Note (siehe oben) zur Verfügung gestellt.

Auf der Grundlage dieser Auswertungen wurden die Daten von Syspons auf Unterschiede in den Bewertungen untersucht. Zu diesem Zweck wurden die Ergebnisse anhand der ausgewählten Kriterien und Themenbereiche betrachtet und statistische Trends (zum Beispiel unterschiedliche durchschnittliche ZPE-Bewertungen in verschiedenen Regionen) untersucht.

Die vorläufigen Ergebnisse und die statistischen Trends wurden in einem gemeinsamen Termin und in der Nachbereitung diskutiert. Dabei wurde entschieden, welche Auswertungen in den Ergebnisbericht einfließen und welche statistischen Trends qualitativ untersucht werden sollten (siehe unten). In dem vorliegenden Bericht werden die folgenden **statistischen Trends** vertieft untersucht:

- **Bewertungsunterschiede nach Projekttyp:** Warum werden bilaterale Projekte im Durchschnitt anders bewertet als andere Projekttypen? Welchen Einfluss hat laut den ZPE-Berichten der Projekttyp auf die Bewertung der ZPE?
- **Bewertungsunterschiede nach Projektvolumen:** Warum weisen kleinere Vorhaben im Unterschied zu größeren Vorhaben unterschiedliche Durchschnittsbewertungen auf? Welche Rolle spielt die Projektgröße in der Bewertung der ZPE?
- **Bewertungsunterschiede nach Regionen:** Warum sind Projekte in einigen Regionen im Durchschnitt erfolgreicher als in anderen Regionen? Welche Faktoren können aus den ZPE abgeleitet werden?
- **Bewertungsunterschiede nach BMZ-Länderkategorien:** Warum werden Projekte in unterschiedlichen BMZ-Länderkategorien (zum Beispiel Reformpartner, Transformationspartner, Nexus- und Friedenspartner, etc.) im Schnitt unterschiedlich bewertet? Welche Faktoren können aus den ZPE abgeleitet werden?
- **Bewertungsunterschiede nach den Kernthemen und Aktionsfeldern des BMZ:** Warum fallen die durchschnittlichen Bewertungen von ZPE bei den unterschiedlichen Kernthemen und nachgelagerten Aktionsfeldern des BMZ unterschiedlich aus? Welche Kernthemen und Aktionsfelder wurden besonders gut bzw. besonders schlecht bewertet und welche Faktoren können aus den ZPE abgeleitet werden?
- **Generelle Bewertungstrends:** Darüber hinaus wurden allgemeine Bewertungstrends im Zeitverlauf betrachtet und mögliche Faktoren diskutiert (z. B. der Einfluss der COVID-19-Pandemie und *remote*-Evaluierungen sowie Änderungen des ZPE-Bewertungsmaßstabs).

Die Ergebnisse zu diesen trendbezogenen Fragestellungen werden in Kapitel 3 dargestellt. Der Vorgang zur Ermittlung der Erklärungsmuster wird im folgenden Schritt zur inhaltlichen Analyse beschrieben.

### **Inhaltliche Analyse der Evaluierungsberichte**

Ziel der inhaltlichen Analyse der Evaluierungsberichten war es, im Rahmen einer qualitativen Analyse Erklärungsmuster für die ausgewählten statistischen Trends bzw. Bewertungsunterschiede zu ermitteln. Dazu wurde die (statistische) Analyse durch eine qualitative Auswertung von 112 Evaluierungsberichten ergänzt (vgl. Kapitel 2.2 zur Auswahl der Berichte).

Die in der Vorbereitungsphase der Auswertung festgehaltenen zentralen Fragestellungen der Querschnittsauswertung wurden durch eine mit der GIZ final abgestimmte Erhebungsmaske operationalisiert. Dabei wurden

zur Ermittlung von erklärenden Faktoren die folgenden Auswertungskategorien festgelegt: (i) Welche Aussagen treffen die Evaluierungen über die **Rahmenbedingungen und Herausforderungen** (politisch, rechtlich, sozial usw.) der Vorhaben; (ii) Welche **Projektansätze**<sup>7</sup> verfolgten die Vorhaben und wie wurden diese von den Evaluator\*innen bewertet; (iii) Wie wurden im Rahmen der Evaluierungen ausgewählte **Wirkungshypothesen** bewertet und warum; und (iv) welche **Misserfolgs- und Erfolgsfaktoren** wurden im Fazit der ZPE aufgeführt?

Im **Querschnitt** wurde anschließend untersucht, wie diese Faktoren mit dem jeweils untersuchten **statistischen Trend** zusammenhängen. Dabei wurde ausgewertet, welche wiederkehrenden Faktoren es gibt, durch die die Bewertungsunterschiede erklärt werden könnten. Beispiele: Gibt es bestimmte Rahmenbedingungen und Herausforderungen, Ansätze oder Misserfolgs- und Erfolgsfaktoren, die sich in bestimmten Regionen häufen? Erklären diese Häufungen die unterschiedlichen Durchschnittsbewertungen zwischen den Regionen? Wenn ja, welche Faktoren sind dabei hervorzuheben?

Die Erhebungsmaske wurde für die Programmierung mit der Befragungssoftware SurveyXact erstellt und enthält sowohl quantifizierte Fragen mit Kategorien (Ja/Nein/Teilweise/Keine Angaben) als auch Freitextfelder für die qualitativen Beobachtungen des Auswertungsteams. Mit Hilfe der Befragungssoftware konnte das Auswertungsteam die Ergebnisse pro Bericht standardisiert anhand der definierten Fragen abfragen und anschließend effizient auswerten. Für die Auswertung wurden die gesammelten Daten sowohl quantitativ (Häufung von Beobachtungen in den Berichten) als auch qualitativ (Art der Beobachtungen) analysiert und synthetisiert.<sup>8</sup>

### **Berichtslegung und Präsentation**

Das Produkt der Auswertung der durchschnittlichen Bewertungen von Zentralen Projektevaluierungen ist der vorliegende Bericht, mit dem die zentralen Erkenntnisse dokumentiert werden. Dieser wurde durch das Auswertungsteam erstellt und der Stabsstelle Evaluierung zur Kommentierung übermittelt. Die Rückmeldungen und Kommentare wurden in die finalisierte Version eingearbeitet. Ferner hat das Auswertungsteam die Inhalte der Querschnittsauswertung in einer PowerPoint-Präsentation aufbereitet.

---

<sup>7</sup> Unter Ansätzen verstehen wir hierbei die konzeptionelle Grundlage bzw. Herangehensweise, das Vorhaben genutzt hat, um seine Ziele und/oder Wirkungen auf effektive Weise zu erreichen (zum Beispiel durch Training-of-Trainers-Maßnahmen).

<sup>8</sup> Jeder Bericht im Rahmen von QSA *Insights* steht für sich allein. Der qualitative Auswertungsprozess ist jedoch auch für Teil 2 und Teil 3 von QSA *Insights* von zentraler Bedeutung. Der jeweilige Fokus der Auswertung im Hinblick auf die Zielsetzungen wurde in den entsprechenden Berichten beschrieben.

# 3 Ergebnisse der Querschnittsauswertung

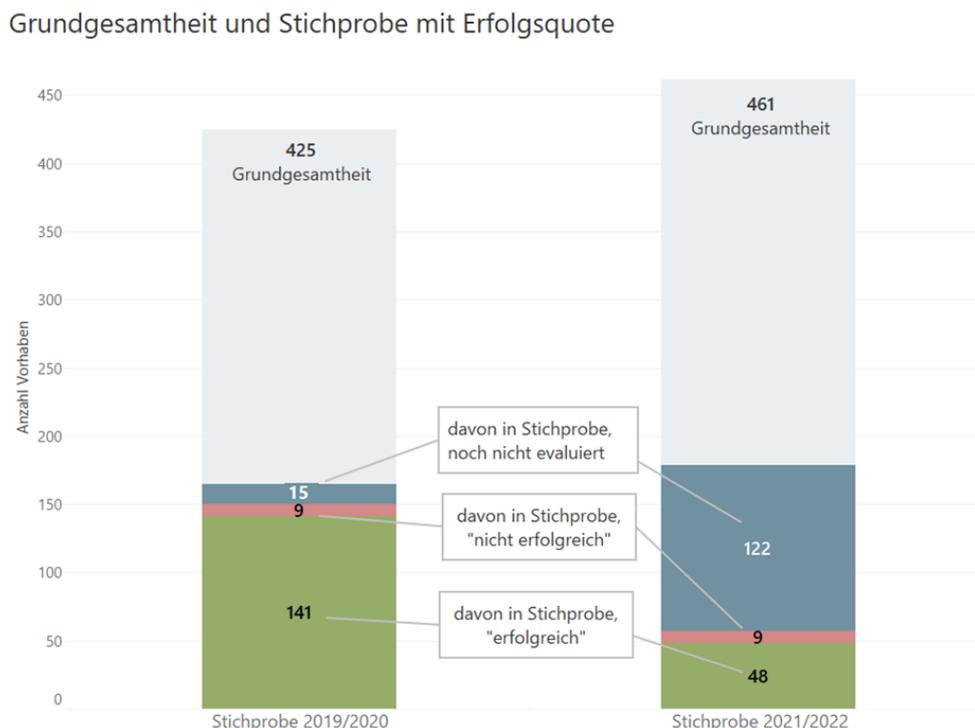
In diesem Kapitel werden die zentralen Ergebnisse der Querschnittsauswertung zur Auswertung und Analyse der Durchschnittsbewertungen von abgeschlossenen ZPE sowie mögliche erklärende Faktoren vorgestellt. Im folgenden Abschnitt wird zunächst die Erfolgsquote der ZPE übergreifend dargestellt. Anschließend werden die Bewertungen der ZPE nach Zeitverlauf, allgemeinen Projektmerkmalen, Regionen, Themen und Sektoren sowie Evaluierungseigenschaften differenziert. Dazu werden die ausgewählten statistischen Trends näher beschrieben und mit möglichen erklärenden Faktoren aus den Inhaltsanalysen unterlegt.

## 3.1 Erfolgsquote der ZPE

Vom BMZ finanzierte Vorhaben werden im Rahmen von zentralen Projektevaluierungen ähnlich wie bei Bewertungen in der Schule auf einer Skala von 1 („sehr erfolgreich“) bis 6 („gänzlich erfolglos“) bewertet.<sup>9</sup> Auf Basis dieses Bewertungssystems beschreibt die Erfolgsquote den Anteil der Projekte mit einer Gesamtbewertung von 1, 2 oder 3. Die Erfolgsquote gibt somit den Anteil der Projekte an, die von den externen Evaluierungsteams als erfolgreich bewertet wurden.

Abbildung 2 zeigt die Stichproben inklusive des Anteils erfolgreicher Bewertungen der abgeschlossenen 207 ZPE (vgl. Kapitel 2.1). Von den 150 ZPE aus der ersten Stichprobe von Projekten mit einem geplanten Ende im Zeitraum 2019/2020 wurden 141 Projekte als „erfolgreich“ eingestuft. Die **Erfolgsquote** liegt somit bei 94 Prozent. In dieser Stichprobe stehen noch die Ergebnisse für 15 ZPE aus, da die entsprechenden Projekte verlängert wurden. Bei der späteren Stichprobe für den Zeitraum 2021/2022 sind bisher weniger ZPE abgeschlossen worden: Bei 57 (von insgesamt 179) abgeschlossenen ZPE wurden 48 Projekte als erfolgreich bewertet. Das entspricht einer Erfolgsquote von 84 Prozent.

Abbildung 2: Grundgesamtheit und Stichprobe mit Erfolgsquote



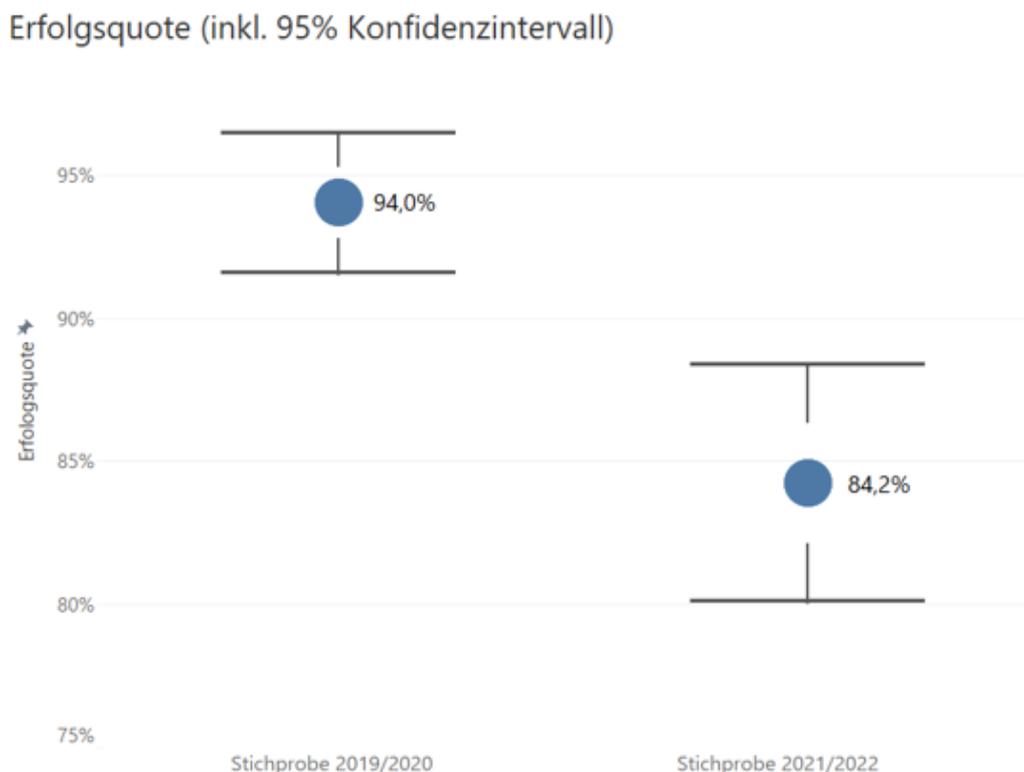
<sup>9</sup> Die Gesamtbewertung setzt sich dabei aus den Einzelbewertungen für die verschiedenen OECD-DAC-Kriterien zusammen (vgl. Kapitel 3.3).

Die Erfolgsquote ist durch die geschichtete Zufallsstichprobe für alle in einem Zeitraum von zwei Jahren beendeten Vorhaben (Grundgesamtheit) repräsentativ. Das bedeutet, dass die Erfolgsquote über die Stichprobe hinaus für alle BMZ-finanzierten Projekte mit einem Auftragswert ab drei Millionen Euro aussagekräftig ist. Allerdings unterliegt die Erfolgsquote wie jede statistische Schätzung einer gewissen Unsicherheit. Diese **Unsicherheit bzw. Aussagekraft der Erfolgsquote** hängt zum einen davon ab, wie viele Ergebnisse der Stichprobe bereits vorliegen. Je mehr Ergebnisse vorliegen, desto aussagekräftiger ist die Erfolgsquote für alle Projekte in der Grundgesamtheit. Da sowohl für die Stichprobe 2019/2020 als auch für die Stichprobe 2021/2022 noch ZPE ausstehen, ist die Aussagekraft der Erfolgsquote in beiden Fällen noch eingeschränkt – auch wenn die Unsicherheit für die ersten beiden Stichproben aufgrund der wenigen fehlenden ZPE im Vergleich deutlich geringer ist. Doch auch wenn alle ZPE vorliegen, bleibt eine statistische Restunsicherheit.

**Für die genaue Berechnung der Aussagekraft kann das Konfidenzintervall hinzugezogen werden, das die Unsicherheit erklärt und greifbar macht:** Es gibt den Bereich an, in dem mit 95-prozentiger Wahrscheinlichkeit die Erfolgsquote der dazugehörigen Grundgesamtheit (alle BMZ-finanzierten Projekte) läge. Abbildung 3 zeigt die aktuelle Erfolgsquote der beiden Stichproben und gibt zudem das Konfidenzintervall an, mit dem die Aussagekraft für alle Vorhaben in der Grundgesamtheit beschrieben werden kann.

**Mit Hilfe des Konfidenzintervalls ergibt sich für die Stichprobe 2019/2020 bereits, dass die Erfolgsquote aller Projekte in der Grundgesamtheit deutlich über 90 Prozent liegt. Die ersten Ergebnisse der Stichprobe 2021/2022 zeigen niedrigere Erfolgsquoten um 84 Prozent.** Aus den Ergebnissen der Stichprobe 2019/2020 wird eine Erfolgsquote zwischen 96 Prozent und 92 Prozent für alle Projekte in der Grundgesamtheit erwartet. Wie oben beschrieben, besteht bei der späteren Stichprobe 2021/2022 eine größere Unsicherheit. Dementsprechend ist das Konfidenzintervall in dem das Ergebnis mit 95-prozentiger Wahrscheinlichkeit liegen wird, für diese Stichprobe noch breiter. Hier kann die tatsächliche Erfolgsquote derzeit auf 88 Prozent bis 80 Prozent geschätzt werden. Das Konfidenzintervall wird mit den Ergebnissen weiterer ZPE kleiner werden, so dass die Aussagekraft der Erfolgsquote entsprechend zunimmt.

Abbildung 3: Erfolgsquote mit Konfidenzintervall

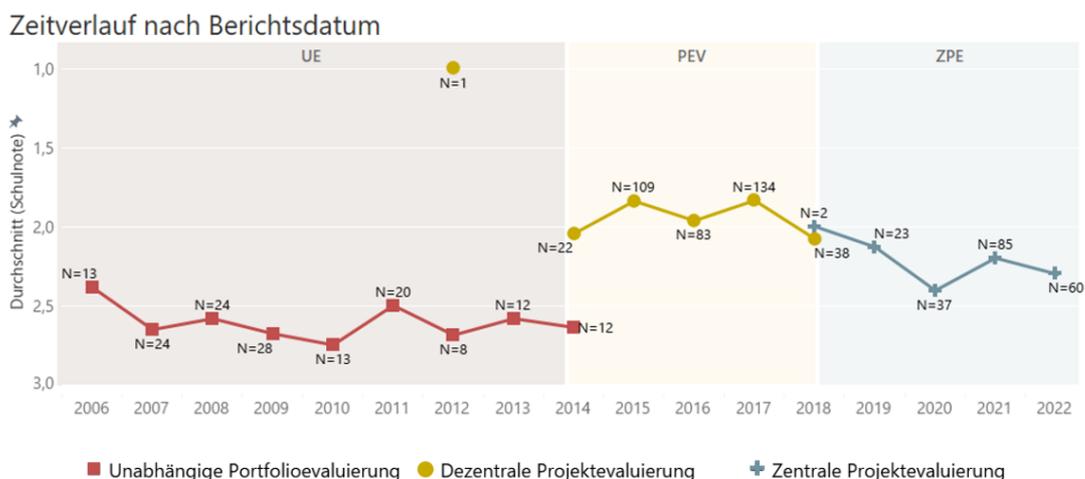


### 3.2 Auswertung der ZPE und früherer Instrumente im Zeitverlauf

Ein Blick auf den **zeitlichen Verlauf der GIZ-Evaluierungen und der zugehörigen Bewertungen** erlaubt eine Einordnung der beschriebenen Ergebnisse zur Erfolgsquote. Dabei lohnt sich auch ein Vergleich zwischen verschiedenen Evaluierungsinstrumenten, die die GIZ in ihrer langjährigen Evaluierungstradition etabliert hat. Die ZPE wurden Ende 2017 eingeführt und sind seitdem ein wesentliches Evaluierungsinstrument der GIZ. Neben den ZPE existieren bzw. existierten weitere Evaluierungsinstrumente, die durch ihre Zielsetzungen zum Anspruch der GIZ-Evaluierungen „Wissen, was wirkt“ beitragen bzw. beigetragen haben.<sup>10</sup> Zwei mit den ZPE vergleichbare Evaluierungsinstrumente waren die (1) unabhängigen Portfolioevaluierungen (UE) zwischen 2006 und 2014 und die (2) dezentralen Projektevaluierungen (PEV) zwischen 2012 und 2018. Auch wenn sich diese Formate in ihren jeweiligen Durchführungsprozessen von den ZPE unterschieden, wurden die GIZ-Projekte mit diesen Instrumenten ebenfalls nach einem 6-stufigen Bewertungssystem bewertet, so dass die entsprechenden Ergebnisse für einen (indirekten) Vergleich mit den Ergebnissen der ZPE herangezogen werden können.

**Ein Blick auf die in Abbildung 4 dargestellte Zeitachse zeigt, dass es deutliche Unterschiede im Vergleich der Durchschnittsbewertungen bei den Evaluierungsinstrumenten UE, PEV und ZPE gibt.** Zunächst fällt auf, dass die niedrigsten Durchschnittsbewertungen bei den unabhängigen Portfolioevaluierungen vergeben wurden. Bei insgesamt 162 Evaluierungen (Ø 18 pro Jahr) wurde jährlich eine Durchschnittsbewertung zwischen 2,39 (2006) und 2,75 (2010) vergeben. Im Gegensatz dazu wurden bei den dezentralen Projektevaluierungen deutlich höhere Durchschnittsbewertungen zwischen 1,84 (2017) und 2,08 (2018) vergeben. Gleichzeitig wurden mit durchschnittlich 97 PEV pro Jahr deutlich mehr Evaluierungen durchgeführt.<sup>11</sup> Letztlich liegen die Bewertungen der 207 ZPE im Vergleich zwischen den Bewertungen der beiden früheren Instrumenten: Hier wurde im Jahr 2018 die höchste Durchschnittsbewertung vergeben (2,00), während die Durchschnittsbewertung im Jahr 2020 mit 2,41 am niedrigsten war. Bisher wurden dabei im Schnitt 41 ZPE jährlich durchgeführt, wobei die Anzahl der ZPE stufenweise erhöht wurde: Während 2019 23 ZPE abgeschlossen wurden, waren es im Jahr 2021 85 ZPE. Die Unterschiede in den Bewertungen lassen jedoch keine direkt vergleichbaren Aussagen über den Erfolg oder Misserfolg der evaluierten Projekte zu, da die Bewertungsunterschiede auf unterschiedliche Prozesse der Evaluierungsinstrumente zurückzuführen sein könnten. So wurden beispielsweise die unabhängigen Portfolioevaluierungen (UE) meist als Ex-post-Evaluierungen zwei bis fünf Jahre nach Projektabschluss durchgeführt, während PEV und ZPE vorwiegend zu einem früheren Zeitpunkt angelegt wurden bzw. werden.

Abbildung 4: Bewertungen im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum)<sup>12</sup>



<sup>10</sup> GIZ, 2022: [https://www.giz.de/de/ueber\\_die\\_giz/516.html](https://www.giz.de/de/ueber_die_giz/516.html)

<sup>11</sup> Bis auf eine Evaluierung fallen alle PEV in den Zeitraum 2014 bis 2018. Eine PEV, die im Jahr 2012 fertiggestellt wurde, ist als „Ausreißer“ in Abbildung 4 zu erkennen.

<sup>12</sup> Zur besseren Darstellung zeigt die Abbildung nur einen Teil der Skala. Die Abbildung mit der vollständigen Skala befindet sich im Anhang.

**Ein genauerer Blick auf die Bewertungen der ZPE im Zeitverlauf zeigt, dass die durchschnittlichen Bewertungen der ZPE bis 2020 gesunken sind, ab 2021 aber anstiegen.** Im Startjahr 2018 wurden nur zwei ZPE abgeschlossen, in den Folgejahren jeweils deutlich mehr. Im Jahr 2021 sind mit 85 ZPE bisher die meisten Bewertungen abgeschlossen worden. Bei den Durchschnittswerten ist zu erkennen, dass die Werte von 2018 (2,00) bis 2020 (2,41) stetig sanken. Im Jahr 2021 verbesserte sich die Durchschnittsbewertung auf 2,20. Im Jahr 2022 liegt die durchschnittliche Bewertung bei 2,30. Auf diese Beobachtungen – und den möglichen Zusammenhang mit der COVID-19-Pandemie – wird in Kapitel 3.7 näher eingegangen.

Bei den bisher vorgestellten Ergebnissen zur Erfolgsquote sowie den durchschnittlichen Bewertungen im Zeitverlauf wurden alle abgeschlossenen Evaluierungen in Ihrer Gesamtheit betrachtet. Gleichzeitig gibt es bei einer näheren Analyse der ZPE-Bewertungen Unterschiede, die weitere Erkenntnisse zu den Evaluierungen und evaluierten Vorhaben liefern. Auf diese Unterschiede wird in den folgenden Abschnitten eingegangen.

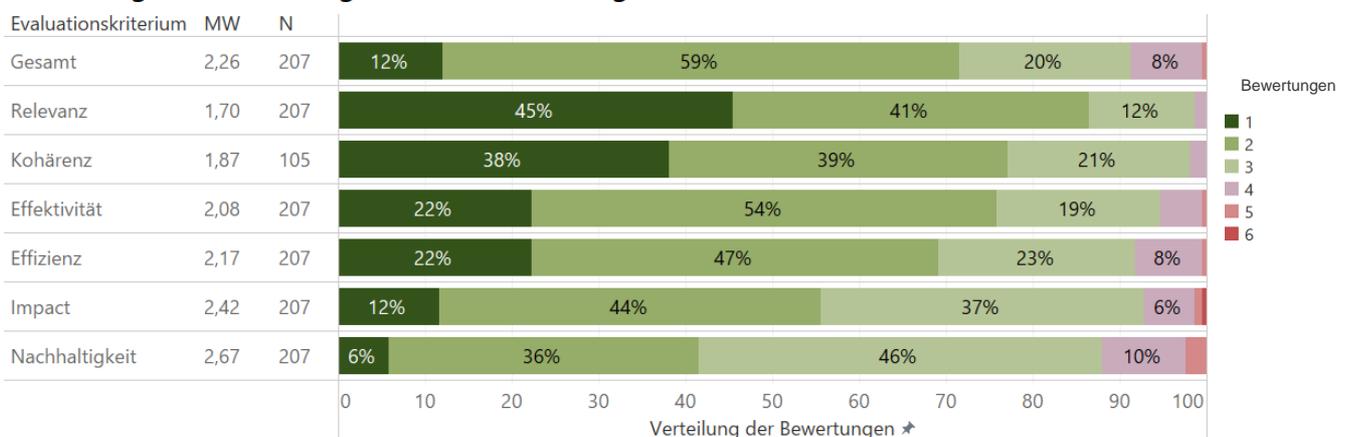
### 3.3 Auswertung der ZPE nach Evaluierungskriterien

Neben der bisher betrachteten Gesamtbewertung werden in jeder ZPE Punkte und Bewertungen für einzelne OECD DAC-Kriterien nach internationalen Evaluierungsstandards vergeben.<sup>13</sup> Dabei wird jedes Vorhaben anhand der sechs Evaluierungskriterien Relevanz, Kohärenz (seit Beginn 2021, vgl. Kapitel 3.7), Effektivität, Effizienz, Impact und Nachhaltigkeit bewertet. Die Bewertung anhand dieser Kriterien bildet dabei einen normativen Rahmen zur Bestimmung des Nutzens eines Projekts. Zusammen beschreiben sie dabei die gewünschten Eigenschaften von Projekten: Sie sollten für den Kontext relevant sein, kohärent mit anderen Interventionen sein, ihre Ziele erreichen, auf effiziente Weise Ergebnisse erzielen und positive Auswirkungen haben, die von Dauer sind.<sup>14</sup>

Ein Blick auf die **Verteilung der Beurteilungen in der Gesamtbewertung sowie nach Evaluierungskriterium** (Abbildung 5) lässt deutlich werden, wo Stärken und Schwächen in den von der GIZ durchgeführten Projekte gesehen wurden. Zunächst zeigt die genauere Betrachtung, dass die durchschnittliche Gesamtbewertung aller 207 ZPE (N) bei 2,26 (Mittelwert) liegt. Der weitaus größte Anteil aller ZPE (59 Prozent) wurde mit der Note 2 („erfolgreich“) bewertet. Bei den einzelnen Kriterien ergibt sich jedoch ein wesentlich differenzierteres Bild.

Abbildung 5: Bewertungen nach Evaluierungskriterien

#### Verteilung der Bewertungen nach Evaluierungskriterien



<sup>13</sup> Die **Gesamtbewertung** setzt sich letztlich aus den Bewertungen für die einzelnen Kriterien zusammen. Das Bewertungssystem sieht eine Vergabe von bis zu 100 Punkten pro Kriterium vor. Aus den vergebenen Punkten wird nach einer festgelegten Skala die Bewertung errechnet (Note 1: 100-92 Punkte; Note 2: 91-81 Punkte; Note 3: 80-67 Punkte; Note 4: 66-50 Punkte; Note 5: 49-30 Punkte; Note 6: 29-0 Punkte). Der Durchschnittswert dieser Bewertungen ergibt die Gesamtbewertung des Projekts. Ausnahmen gibt es durch die sogenannten KO-Kriterien (vgl. Kapitel 3.7).

<sup>14</sup> OECD, 2019

**Besonders positiv wurden die Projekte in Bezug auf die Evaluierungskriterien Relevanz und Kohärenz bewertet.** Die **Relevanz** der Projekte wurde im Durchschnitt mit Abstand am besten bewertet (Mittelwert 1,70). Bei knapp der Hälfte der ZPE wird mit der Bewertung 1 („sehr erfolgreich“) ausdrücklich bestätigt, dass die Projekte auf die Bedürfnisse, Strategien und Prioritäten der Partner oder Begünstigten abgestimmt sind. Ein großer Teil der Projekte ist laut den Evaluator\*innen so konzipiert, dass sie in ihrem jeweiligen Kontext relevant sind und wertvolle Ziele verfolgen. Zusätzlich wurde die Mehrheit der Projekte als sehr **kohärent** und damit kompatibel mit anderen Interventionen angesehen (Mittelwert 1,87). Hier erreichten 77 Prozent der Projekte die Bewertung bzw. Note 1 oder 2. Es ist zu beachten, dass dieses Kriterium erst seit 2021 hinzugezogen wurde und für 105 ZPE eine Bewertung vorliegt. Darüber hinaus ist es möglich, dass die Einführung des Kohärenzkriteriums einen Einfluss auf die Gesamtbewertung der ZPE hatte, da sich die Berechnung des Durchschnittswerts durch die Anpassung verändert hat (vgl. Kapitel 3.7). Dennoch wird deutlich, dass die Projekte nach Ansicht der Evaluierungsteams andere Interventionen eher unterstützen und Synergien sowohl intern mit anderen GIZ-Projekten als auch mit Interventionen anderer Akteure erzeugen. Die Komplementarität, Harmonisierung und Koordinierung mit anderen Akteuren wurde entsprechend positiv bewertet.

**Dagegen fallen die Bewertungen bei den Kriterien Impact und Nachhaltigkeit im Vergleich ab.** Der **Impact** der Projekte wurde im Mittel mit 2,42 und damit deutlich schlechter bewertet als der Gesamtdurchschnitt. Die Gutachter\*innen sehen demnach noch Verbesserungspotenzial bei den längerfristigen (transformativen) Wirkungen der Projekte und den Veränderungen von Systemen oder Normen sowie den möglichen Auswirkungen auf Entwicklungsziele. Das Kriterium **Nachhaltigkeit** wurde jedoch im Durchschnitt am schlechtesten bewertet (Mittelwert 2,67). In vielen Fällen sind die Evaluator\*innen skeptisch, ob die Auswirkungen der Projekte von Dauer sind. Dennoch ist zu sagen, dass sich die Bewertungen in einem erfolgreichen Bereich bewegen und nur ein Bruchteil der Projekte auch bei diesen Evaluierungskriterien als „nicht erfolgreich“ (Bewertung 4 oder schlechter) eingestuft wurde. Die im Vergleich schlechteren Bewertungen beim Kriterium Nachhaltigkeit decken sich zugleich mit den Ergebnissen einer Studie des Deutschen Evaluierungsinstituts der Entwicklungszusammenarbeit (DEval), wonach die Bewertung der Nachhaltigkeit in der deutschen Entwicklungszusammenarbeit sehr uneinheitlich erfolgt und daher als besonders herausfordernd angesehen werden muss.<sup>15</sup>

**Die Kriterien Effektivität und Effizienz liegen im Gesamtmittelwert.** Die **Effektivität** – das Ausmaß, in dem die Projekte ihre Ziele erreicht haben bzw. voraussichtlich erreichen werden – ist nach Ansicht der Gutachter\*innen im Durchschnitt „erfolgreich“ (Mittelwert 2,08). Die Wirtschaftlichkeit bzw. **Effizienz** der Projekte wird im Durchschnitt etwas schlechter bewertet (Mittelwert 2,17). Dementsprechend wurden in den ZPE in Bezug auf diese Kriterien sowohl erfolgreichere als auch weniger erfolgreiche Projekte evaluiert.

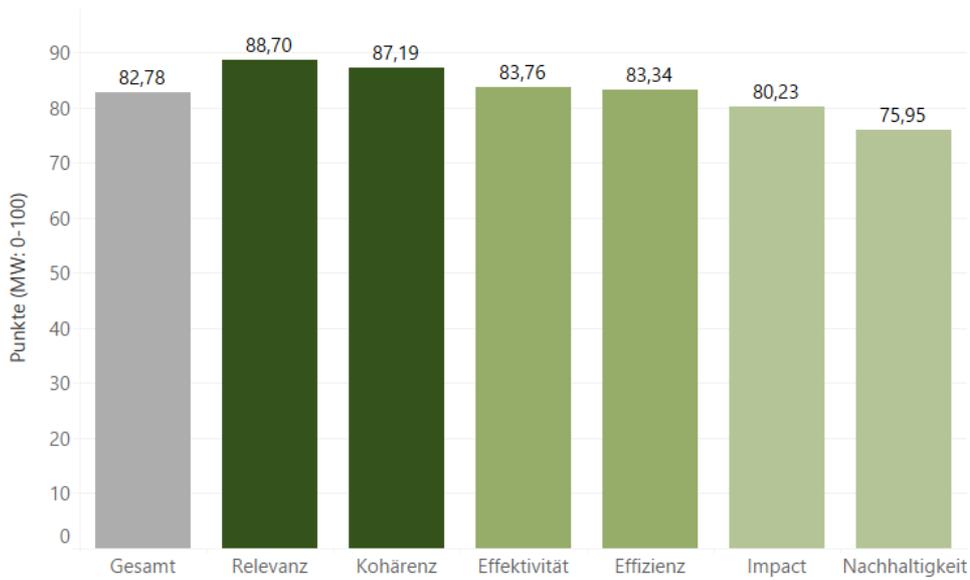
In einer alternativen Darstellung zeigt Abbildung 6 die **Durchschnittspunkte der Gesamtbewertung und der einzelnen OECD DAC-Kriterien auf der in der GIZ zusätzlich verwendeten 100-Punkte-Skala.** Die oben beschriebenen Unterschiede werden auch in dieser Darstellungsform deutlich.

---

<sup>15</sup> Deval, 2018: <https://www.deval.org/de/evaluierungen/laufende-und-abgeschlossene-evaluierungen/nachhaltigkeit-in-der-deutschen-entwicklungszusammenarbeit>

Abbildung 6: Durchschnittliche Punkteverteilung

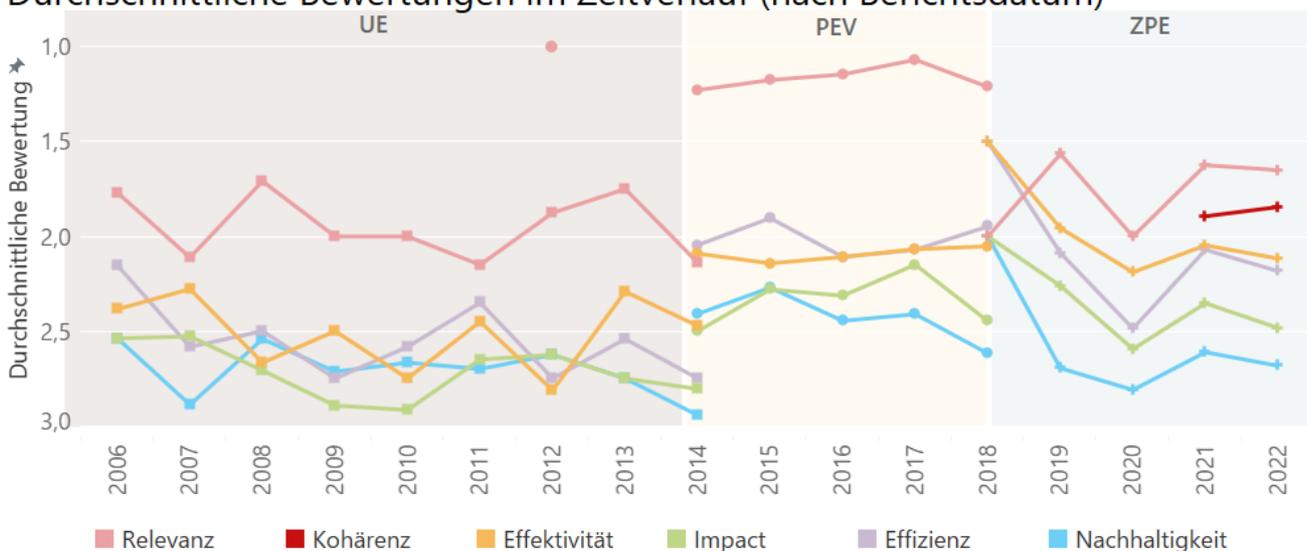
### Durchschnittliche Punkteverteilung



Diese in den ZPE für die einzelnen Evaluierungskriterien gemachten Beobachtungen sind insgesamt mit den Ergebnissen aus den bisherigen Evaluierungsinstrumenten vergleichbar. Um die Gemeinsamkeiten sichtbar zu machen, wurden in Abbildung 7 die Durchschnittsbewertungen anhand der einzelnen Kriterien im Zeitverlauf für ZPE, PEV und UE (vgl. Kapitel 3.2) visualisiert. Dabei zeigt sich, dass zum Beispiel das Kriterium Relevanz (in Rosa) auch bei den frühen Evaluierungsinstrumenten mit Abstand am positivsten bewertet wurde. Weitere vergleichbare Tendenzen lassen sich auch bei der relativ niedrigen Bewertung der Nachhaltigkeit (in Blau), insbesondere im Vergleich zwischen ZPE und PEV, erkennen. Zusätzlich wurde 2021 erstmals das Kriterium Kohärenz (in Rot) in die ZPE aufgenommen und ist dementsprechend nicht in früheren Zeiträumen und Evaluierungsinstrumenten zu finden.

Abbildung 7: Bewertungen nach Evaluierungskriterien im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum)

### Durchschnittliche Bewertungen im Zeitverlauf (nach Berichtsdatum)



tum)<sup>16</sup>

<sup>16</sup> Zur besseren Darstellung zeigt die Abbildung nur einen Teil der Skala. Die Abbildung mit der vollständigen Skala befindet sich im Anhang.

### 3.4 Auswertung der ZPE nach Projektmerkmalen

Bei BMZ-finanzierten Vorhaben lassen sich verschiedene Projekttypen unterscheiden. Darüber hinaus können die evaluierten Vorhaben nach weiteren Projektmerkmalen gegliedert werden. Betrachtet man die Ergebnisse nach diesen verschiedenen (ausgewählten) Projektmerkmalen, lassen sich weitere Unterschiede in den ZPE-Bewertungen feststellen.<sup>17</sup>

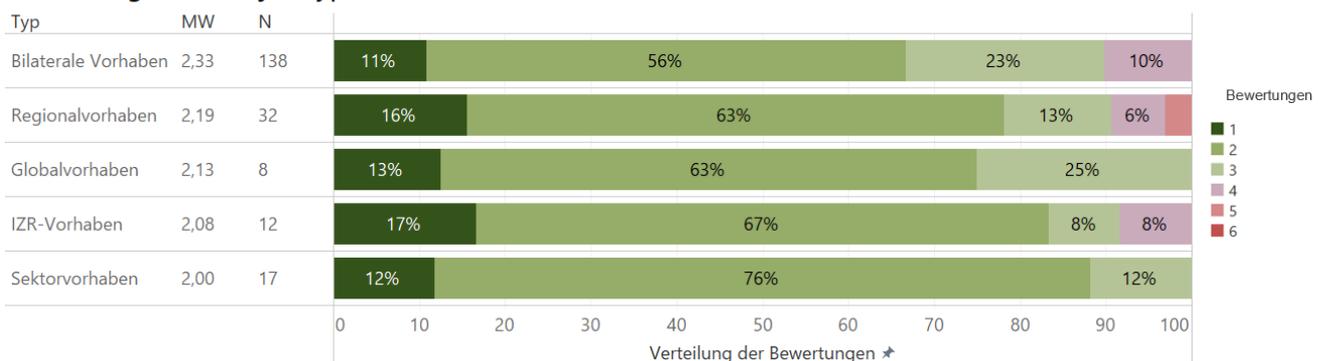
#### Projekttyp

Ein wesentliches Merkmal der evaluierten Projekte ist der **Projekttyp**, denn der Projekttyp ein zentraler Faktor für die gesamte Konzeption und Umsetzung eines Vorhabens und beeinflusst unter anderem die Zielsetzung und die Zielgruppen sowie die jeweiligen Kooperationspartner. Abbildung 8 zeigt die Verteilung der Bewertungen in Abhängigkeit von verschiedenen Projekttypen. Zunächst fällt auf, dass **bilaterale Vorhaben** – also Projekte, die bilateral in einem Zielland der deutschen Entwicklungszusammenarbeit (EZ) durchgeführt werden, um Veränderungsprozesse der Partner zu unterstützen – in der Mehrzahl der ZPE evaluiert wurden (N=138 von 207 oder 66 Prozent). Ein kleinerer Anteil der ZPE befasste sich mit **weiteren Projekttypen**, die sektorale oder globale Entwicklungsfragen in der länder- und regionenübergreifenden Zusammenarbeit behandeln. Dazu zählen Regionalvorhaben (N=32), Globalvorhaben (N=8), IZR-Vorhaben (N=12) und Sektorvorhaben (N=17).

**Im Vergleich der Projekttypen ist ersichtlich, dass die Durchschnittsbewertung für bilaterale Projekte mit 2,33 schlechter ausfällt als für alle anderen Projekttypen.** Außerdem sind vergleichsweise viele der „nicht erfolgreichen Projekte“ bilaterale Vorhaben. Diese Beobachtung wirft die Frage auf, welche Faktoren im Zusammenhang mit dem Projekttyp in den ZPE als besonders kritisch oder gut hervorgehoben wurden. In Trendbox 1 sind einige Faktoren aufgeführt, die die Unterschiede in den Durchschnittsbewertungen zwischen bilateralen Vorhaben und weiteren Projekttypen erklären können. Diese Faktoren wurden im Rahmen der qualitativen Analyse der ZPE-Berichte ermittelt (vgl. Kapitel 2.3). Gleichzeitig könnten die Unterschiede auch auf die Evaluierungen selbst zurückzuführen sein. So wurden beispielsweise bei Sektor- und Globalvorhaben hauptsächlich Datenerhebungen ohne Vor-Ort-Besuche durchgeführt. In diesem Kontext zeigt Kapitel 3.7, dass auch die Art der Datenerhebung einen Einfluss auf die durchschnittliche Bewertung hat.

Abbildung 8: Bewertungen nach Projekttyp (N=207)

#### Bewertung nach Projekttypen



<sup>17</sup> In der hier vorgestellten Analyse werden Bewertungsunterschiede bei ZPE anhand ausgewählter Projektmerkmale beschrieben. Die evaluierten Projekte unterscheiden sich anhand zahlreicher weiterer Merkmale, die sowohl Einfluss auf Konzeption und Durchführung als auch auf die ZPE-Bewertung haben könnten. Eine weitere Differenzierung der evaluierten Projekte, z.B. nach Haushaltstiteln sowie Kofinanzierung ist im Anhang dieses Berichts enthalten..

## Trendbox 1: Einfluss des Projekttyps auf ZPE-Bewertungen

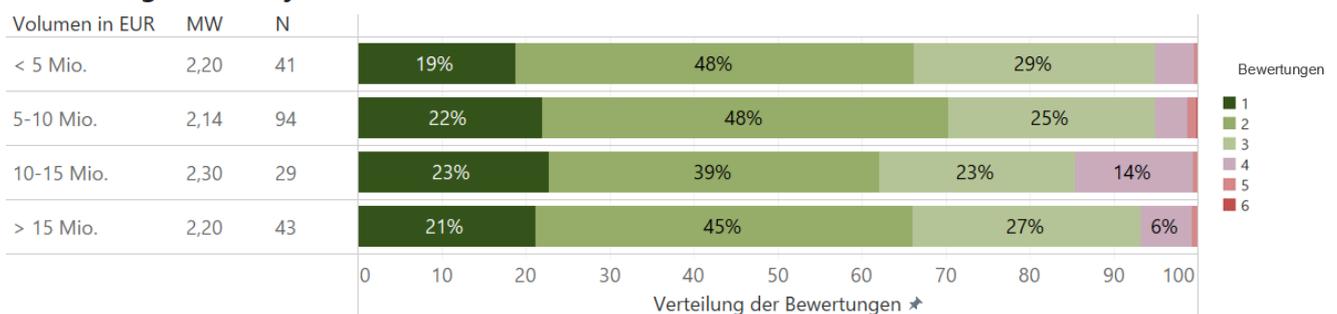
- Im Fall der **bilateralen Projekte** wird in den qualitativ analysierten ZPE-Berichten überdurchschnittlich oft auf den **Kontext** als Misserfolgswort verwendet. So werden in den ZPE-Berichten insbesondere Einschränkungen durch die COVID-19-Pandemie genannt, nämlich bei 51 % der qualitativ evaluierten Berichte bilateralen Vorhaben gegenüber 29 % der sonstigen qualitativ betrachteten Berichte zu nicht bilateralen Vorhaben. So konnten durch die pandemiebedingten Reiseeinschränkungen und Hygieneregeln Präsenzveranstaltungen nicht oder nur bedingt stattfinden. Dies betrifft nicht-bilaterale Projekte seltener, da zu deren Tätigkeiten häufiger Maßnahmen wie die Förderung von Netzwerken oder Beratungsleistungen gehören, die von den Einschränkungen weniger stark betroffen waren. Zudem wird als Misserfolgswort auch eine schlechte Sicherheitslage (17 % vs. 40 %) genannt, zum Beispiel wegen bewaffneter Konflikte, die dazu führen, dass bereits geplante Maßnahmen abgesagt werden mussten.
- Im Gegensatz dazu wird in **Globalvorhaben, Sektorvorhaben, IZR-Vorhaben und Regionalvorhaben** die Zusammenarbeit mit den beteiligten Stakeholdern als Erfolgsfaktor hervorgehoben, nämlich bei 82 % der qualitativ evaluierten ZPE-Berichte der Global-, Sektor-, IZR- und Regionalvorhaben gegenüber 56 % aller qualitativ betrachteten ZPE-Berichte zu bilateralen Vorhaben. Ein Beispiel für die erfolgreiche Zusammenarbeit ist die Anwendung partizipativer Ansätze. Hinderliche Kontextfaktoren, die diese **Kooperationssysteme** hätten negativ beeinflussen können, wurden hier nur selten genannt. Auch werden **interne Faktoren** überdurchschnittlich oft als Erfolgsfaktoren erwähnt: Zwar gibt es vermehrt Herausforderungen in der Steuerung (41 % vs. 33 %), zum Beispiel aufgrund von Steuerungsstrukturen, die die Auftraggeber einbeziehen und dadurch den Spielraum des Vorhabens ggf. einschränken. Allerdings wird die Fachkompetenz der Projektteams (68 % vs. 48 %) sowie das Engagement einzelner Projektmitarbeiter\*innen (14 % vs. 2 %) häufig hervorgehoben.

## Projektvolumen

Neben dem Projekttyp gibt es weitere Projektmerkmale, die einen wesentlichen Einfluss auf die Projektgestaltung haben. So ist das **Projektvolumen** für den Umfang und die Umsetzungsmöglichkeiten eines Projekts entscheidend. In Abbildung 9 sind die ZPE-Bewertungen nach der Projektgröße in vier Kategorien unterteilt. Hier zeigt sich zunächst, dass bislang sowohl vergleichsweise kleine Projekte (N=41 für Projekte mit Budget <5 Mio. Euro), mittelgroße Projekte (N=29 für Projekte mit Budget 10-15 Mio. EUR) sowie besonders große Projekte (N=43 für Projekte mit einem Budget >15 Mio. Euro) Teil der ZPE-Stichprobe waren und bewertet wurden. Die Mehrheit der bewerteten Projekte verfügte jedoch über ein Budget zwischen 5 und 10 Millionen Euro (N=94).

Abbildung 9: Bewertung Projektvolumen (Kategorien) (N=207)

### Bewertung nach Projektvolumen



**Aus der Auswertung anhand der Projektvolumina ergibt sich kein statistisch signifikanter Unterschied zwischen den Durchschnittsbewertungen in Abhängigkeit vom Projektvolumen.** Projekte mit einem Budget von bis zu 5 bis 10 Millionen Euro erreichten mit 2,14 die höchste Durchschnittsbewertung, während Projekte mit einem Budget von 10 bis 15 Millionen Euro oder mehr nur mit 2,30 bewertet wurden. Eine Korrelationsanalyse bestätigt jedoch keinen statistischen Zusammenhang zwischen Gesamtbewertung und Projektvolumen - eine statistische Aussage, dass die Gesamtbewertung einer ZPE umso niedriger ausfällt, je größer das Projekt ist, kann nicht getroffen werden. Diese Beobachtung steht im Kontrast zu einer früheren Auswertung zu den dezentralen Projektevaluierungen aus dem Jahr 2019, in der ein „*signifikanter Unterschied in der Beziehung zwischen dem Gesamtvolumen des Vorhabens und dem Erfolg*“ festgestellt wurde.<sup>18</sup> Dennoch wurden im Rahmen der qualitativen Analyse verschiedene, in den ZPE genannte Faktoren ermittelt, die mit der Projektgröße zusammenhängen. Diese sind in Trendbox 2 aufgeführt.

### **Trendbox 2: Einfluss des Projektvolumens auf ZPE-Bewertungen**

- **Vorhaben mit Projektvolumina bis 10 Millionen Euro** sind im Durchschnitt geringfügig erfolgreicher. In Bezug auf diese Vorhaben werden insbesondere die gute Zusammenarbeit mit den beteiligten Stakeholdern sowie die Fachkompetenz des Projektteams hervorgehoben. Außerdem wird auf die Steuerung des Vorhabens als Erfolgsfaktor verwiesen, zum Beispiel auf das genutzte Monitoring, auf die Flexibilität bei der Projektumsetzung sowie die schnelle Anpassung an die neuen Gegebenheiten wie die COVID-19-Pandemie.
- Im Gegensatz dazu sind **Vorhaben mit Projektvolumina ab 10 Millionen Euro** im Durchschnitt etwas weniger erfolgreich. In diesem Fall wird jedoch die Zusammenarbeit mit den beteiligten Stakeholdern, die in Projekten mit kleinerem Projektvolumen positiv hervorgehoben wird, als zentraler Misserfolgsfaktor genannt. Dabei wird in diesem Zusammenhang insbesondere auf bürokratische Hürden verwiesen.

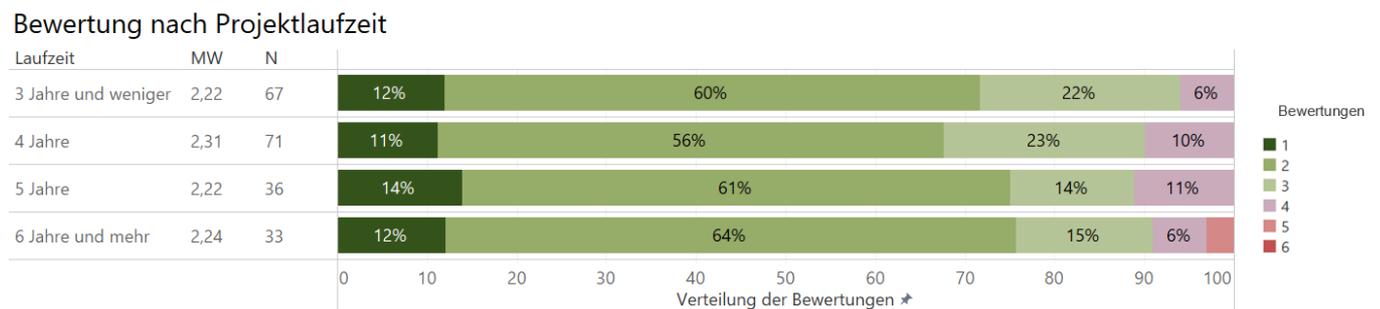
---

<sup>18</sup> GIZ, 2019

## Projektlaufzeit

Darüber hinaus ist die **Projektlaufzeit** ein wichtiges Projektmerkmal. Die Projektlaufzeit gibt den Zeitrahmen vor, innerhalb dessen ein Projekt seine Ziele erreichen kann. In Abhängigkeit von diesem Zeitrahmen können Ziele definiert und verfolgt werden. Dementsprechend hat auch dieses Merkmal einen wesentlichen Einfluss auf die Projektgestaltung und -durchführung. In Abbildung 10 sind die ZPE-Bewertungen für Projekte in zeitliche Kategorien unterteilt. 138 evaluierte Projekte hatten eine Laufzeit von bis zu vier Jahren. Nur 69 Projekte hatten eine Laufzeit von 5 Jahren oder länger. **Bei den Durchschnittsbewertungen in Abhängigkeit von der Projektlaufzeit lässt sich kein eindeutiger Trend erkennen.**<sup>19</sup> Diese Beobachtung deckt sich mit der bereits zitierten Auswertung dezentraler Projektevaluierungen aus dem Jahr 2019, in der ebenfalls kein Zusammenhang zwischen Projektlaufzeit und Projekterfolg festgestellt wurde.<sup>20</sup>

Abbildung 10: Bewertungen nach Projektlaufzeit (Kategorien) (N=207)



## 3.5 Auswertung der ZPE nach Ländern und Regionen

Die GIZ arbeitet weltweit mit einer Vielzahl von Partnerländern zusammen. Ebenso vielfältig sind die geographischen Rahmenbedingungen, unter denen die Projekte durchgeführt werden. Darüber hinaus lassen sich die Partnerländer weiteren Kategorien zuordnen: Zum einen werden sie nach dem Reformkonzept „BMZ 2030“ in verschiedene Länderkategorien eingeteilt. Zum anderen ist die Fragilität der Partnerländer ein herauszustellender Faktor für die Umsetzungsfähigkeit der Projekte. Eine geographische Analyse von ZPE-Bewertungen für verschiedene Kontinente, Weltregionen und Kontexte (BMZ-Länderkategorien und Fragilität der Partnerländer) kann daher Erkenntnisse über die Bedeutung der regionalen Rahmenbedingungen für den Projekterfolg liefern.<sup>21</sup>

### Kontinente

In Abbildung 11 sind die durchschnittlichen ZPE-Bewertungen für verschiedene **Kontinente** auf einer Weltkarte dargestellt. Zusätzlich ist die sektorale Verteilung der evaluierten Projekte (nach BMZ-Kernthemen) pro Kontinent in einem Kreisdiagramm dargestellt.<sup>22</sup> Überregionale Vorhaben, die keinem Kontinent eindeutig zugeordnet werden konnten, wurden separat dargestellt. Dies betrifft Globalvorhaben (GV), IZR-Vorhaben (IZR) sowie Sektorvorhaben (SV).

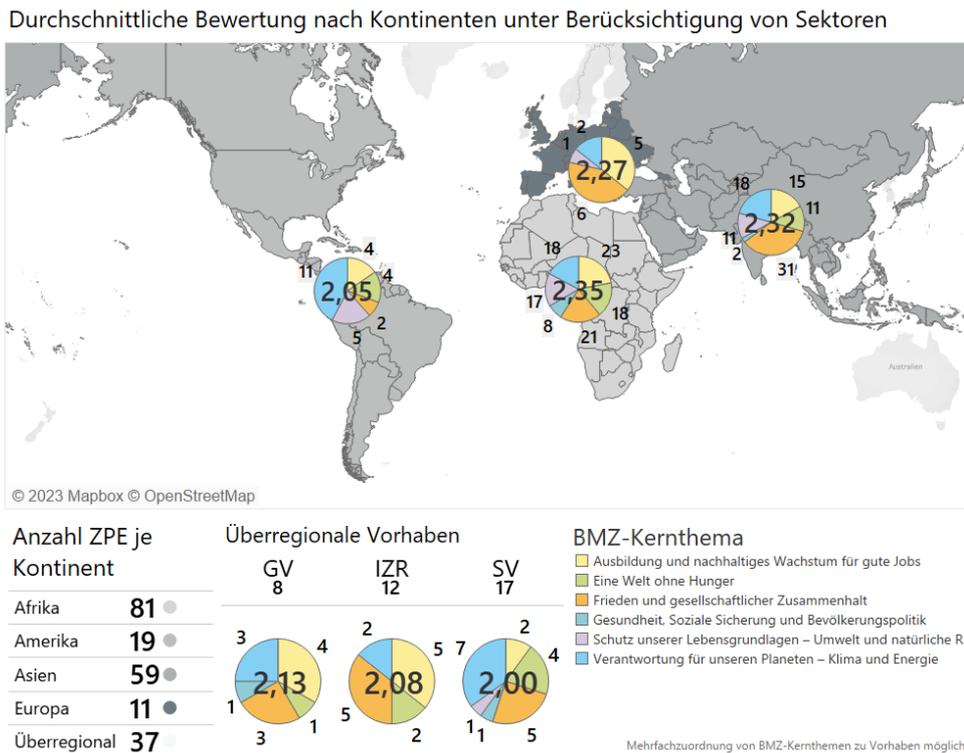
<sup>19</sup> Auch durch eine Korrelationsanalyse zwischen Projektlaufzeit und Gesamtbewertung der ZPE lässt sich kein statistisch signifikanter Zusammenhang finden ( $p=0,19$ ).

<sup>20</sup> GIZ, 2019

<sup>21</sup> Aufgrund der geringen Anzahl von ZPE auf Länderebene wurde keine Analyse pro Land durchgeführt. Gleichzeitig kann der Erfolg/Misserfolg und damit die Bewertung im Allgemeinen auch innerhalb einer Region variieren. Zudem sind die Ergebnisse pro Region auch von den in den anderen Kapiteln erörterten Kriterien abhängig (zum Beispiel den Projektmerkmalen).

<sup>22</sup> Es bestehen verschiedene Möglichkeiten, eine sektorale Verteilung der Projekte darzustellen (vgl. Kapitel 3.6). Zur übersichtlichen Darstellung wurde hier die Verteilung nach BMZ-Kernthemen gewählt.

Abbildung 11: Durchschnittliche Bewertungen nach Kontinent unter Berücksichtigung der Sektoren und BMZ-Kernthemen (N=207)



**Bei den meisten ZPE wurden Projekte in Afrika evaluiert** (insgesamt 81 Vorhaben). Ein großer Teil der Projekte diente der Förderung nachhaltiger wirtschaftlicher Perspektiven (N=23, gelb) sowie der Friedenssicherung (N=21, orange). Insgesamt war die inhaltliche Verteilung in diesem Kontext jedoch sehr vielfältig: In jedem BMZ-Kernthema wurden mehr als 10 Projekte evaluiert (mit Ausnahme des 6. Kernthemas „Gesundheit, Soziale Sicherung und Bevölkerungspolitik“, das allerdings erst seit 2021 als solches definiert ist (vgl. Kapitel 3.6)). In **Asien** wurden mit 51 ZPE die zweitmeisten ZPE durchgeführt. Dabei hatten die evaluierten Projekte einen klaren Schwerpunkt auf dem Bereich Stabilisierung und Friedensförderung (N=31, orange). Auf dem **amerikanischen Kontinent** (19 ZPE) und in **Europa** (11 ZPE) wurden deutlich weniger Projekte durchgeführt. Während die meisten Projekte auf dem amerikanischen Kontinent im Bereich Klima und Energie evaluiert wurden (N=11, blau), konzentrierten sich die Projekte in Europa auf die Bereiche Stabilisierung (N=6, orange) und nachhaltige wirtschaftliche Entwicklung (N=5, gelb). Darüber hinaus ist die sektorale Verteilung der Evaluierungen von überregional tätigen Vorhaben zu sehen. Bei den Global- und IZR-Projekten gehörte das Thema Nachhaltiges Wirtschaften zu den am häufigsten evaluierten Bereichen (N=4 für GV und N=5 für IZR, gelb). Sektorvorhaben wurden hingegen am häufigsten im Themenfeld Klima und Energie (N=5, blau) evaluiert. Insgesamt ist aber auch hier die sektorale Verteilung der ZPE vielfältig.

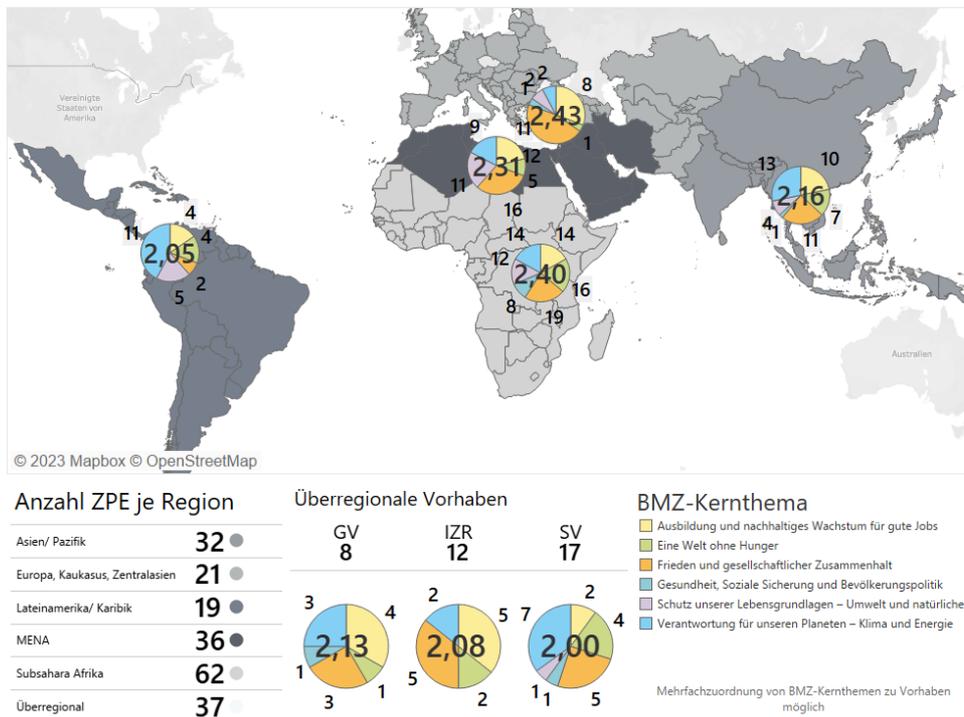
**Darüber hinaus zeigt die Karte, dass die Kontinente unterschiedliche Durchschnittsbewertungen aufweisen.** Die niedrigste Durchschnittsbewertung ist in Afrika zu finden, wo 81 Projekte mit einem Durchschnitt von 2,35 bewertet wurden. Der Kontinent mit der besten Durchschnittsbewertung ist Amerika (Mittelwert, kurz MW =2,05). Europa (MW=2,27) und Asien (MW=2,32) liegen in Bezug auf die durchschnittliche Bewertung dazwischen.

## Weltregionen

Die Betrachtung nach Kontinenten gibt einen guten Überblick über die geographische Verteilung und Bewertung der ZPE. Gleichzeitig fassen die Kontinente mehrere Regionen zusammen. Somit liefert eine differenziertere Analyse der **ZPE-Bewertungen für verschiedene Weltregionen**<sup>23</sup> eine etwas genauere Grundlage. Abbildung 12 zeigt eine alternative Kategorisierung nach den Weltregionen Asien/Pazifik (N=32), Europa, Kaukasus, Zentralasien (N=21), Lateinamerika/ Karibik (N=19), Mittelmeer und Naher Osten (MENA, N=36) sowie Subsahara-Afrika (N=62). Wie zuvor wird auch die sektorale Verteilung der evaluierten Projekte (nach BMZ-Kernthemen) dargestellt.

Abbildung 12: Durchschnittlichen Bewertungen nach Weltregionen unter Berücksichtigung von Sektoren und BMZ-Kernthemen (N=207)

### Durchschnittliche Bewertung nach Kontinenten unter Berücksichtigung von Sektoren



**Die Ergebnisse in dieser Karte zeigen Parallelen, aber auch neue Aspekte im Vergleich zur vorherigen Karte.** Die niedrigeren Durchschnittswerte für Afrika beziehen sich der Auswertung zufolge insbesondere auf Projekte in Subsahara-Afrika. Hier liegt der Durchschnittswert von 2,40 noch unterhalb des Durchschnittswerts für den Gesamtkontinent (vgl. Abbildung 11, MW=2,35). Projekte in der MENA-Region - die teilweise in Afrika liegt - werden mit 2,26 durchschnittlich besser bewertet. Die Bewertungen für Projekte in der Region Europa, Kaukasus und Zentralasien sind insgesamt am niedrigsten (2,43). Durch die Berücksichtigung von Projekten in Zentralasien verschlechtert sich der Durchschnittswert – dementsprechend hat die neue Kategorisierung auch einen Einfluss auf die durchschnittliche Bewertung der Projekte in der Region Asien/Pazifik, die hier deutlich besser abschneidet als das umfassendere „Asien“ auf der vorherigen Karte. Die Ergebnisse für Lateinamerika/Karibik sind die gleichen wie für den Kontinent „Amerika“ zuvor, da die deutsche EZ in Nordamerika nicht tätig ist.

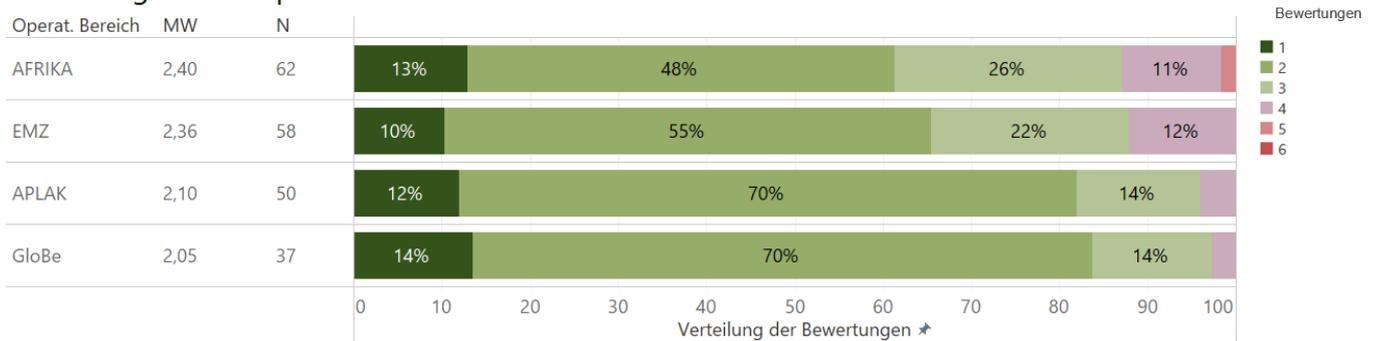
<sup>23</sup> Die Verteilung der Bewertungen nach Weltregionen ist als Balkendiagramm im Anhang des Berichts aufgeführt. Dort findet sich auch eine weitere Karte mit einer differenzierteren Ansicht der Weltregionen (Abbildung 29).

## Regionalbereiche

Eine dritte geographische Analyse der ZPE wurde für die GIZ-internen **Regionalbereiche** durchgeführt.<sup>24</sup> Dementsprechend ist diese Analyse für interne Auswertungen besonders nützlich. Die GIZ ist innerbetrieblich in vier operative Regionalbereiche unterteilt. Wie in Abbildung 13 zu sehen ist, wurden im Bereich Afrika 62 ZPE, in der Region Europa, Mittelmeerraum, Zentralasien (EMZ) 58 ZPE und in der Region Asien, Pazifik, Lateinamerika und Karibik (APLAK) 50 ZPE durchgeführt. Darüber hinaus wurden 37 Projekte im Bereich GLOBE bewertet, dem Sektor- und Globalvorhaben zugeordnet sind. Die in den vorangegangenen Karten dargestellten Trends (Afrika und Europa/Zentralasien mit niedrigeren durchschnittlichen Bewertungen, insbesondere im Vergleich zu Asien und Lateinamerika) werden dabei bestätigt.

Abbildung 13: Bewertungen nach operativen Bereichen (Regionalbereiche der GIZ) (N=207)

### Bewertungen nach operativen Bereichen



Unabhängig von der regionalen Kategorisierung nach Kontinenten, Weltregionen oder GIZ-Regionalbereichen wurden in der qualitativen Analyse der ZPE-Berichte (vgl. Kapitel 2.3) Faktoren analysiert, die im Zusammenhang mit den geographischen Rahmenbedingungen in den ZPE genannt wurden. In Trendbox 3 sind die Faktoren aufgeführt, die möglicherweise eine Erklärung für die Bewertungsunterschiede bieten.

<sup>24</sup> Dabei ist zu beachten, dass die Stichproben anteilig nach den operativen Bereichen der GIZ geschichtet sind, so dass eine repräsentative Auswahl gewährleistet wurde (vgl. Kapitel 2.2, GIZ, 2022: <https://www.giz.de/de/downloads/giz2022-de-instrumentenpapier-zentrale-projektevaluierungen-bmz.pdf>).

### Trendbox 3: Erklärungen für regionale Unterschiede in der ZPE-Bewertung

- Vorhaben in **Asien/Pazifik, Lateinamerika/Karibik und Überregional** sind im Durchschnitt erfolgreicher als andere. Dabei fällt auf, dass in den qualitativ betrachteten ZPE-Berichten die Zusammenarbeit mit den beteiligten Stakeholdern einerseits überdurchschnittlich oft als Erfolgsfaktor und andererseits überdurchschnittlich oft als Misserfolgswort aufgeführt wird. Dies zeigt, dass die Vorhaben zwar Schwierigkeiten haben, ein **permanentes und/oder ein temporäres Kooperationsystem** einzuführen. Sobald es jedoch erst einmal eingeführt wurde, erweist es sich als Erfolgsfaktor. So wird für Projekte in Lateinamerika/Karibik überdurchschnittlich häufig die Ownership der Partner als Erfolgsfaktor genannt: 33 % der qualitativ betrachteten ZPE-Berichte aus Lateinamerika/Karibik gegenüber 24 % aller anderen Berichte. Auch die Komponente **Projektteam** wird hervorgehoben: So werden die Fachkompetenz des Projektteams sowie das Engagement einzelner Projektmitarbeiter\*innen oft als interner Erfolgsfaktor genannt. Dies spiegelt sich zum Beispiel in der Planung der Vorhaben in der Region Asien/Pazifik wider, für die u. a. Nachhaltigkeitsstrategien überdurchschnittlich oft als positiver Faktor genannt werden (25 % vs. 8 %).
- Im Gegensatz dazu sind Projekte in den Regionen **Europa/Kaukasus/Zentralasien, Mittelmeer/Mittlerer Osten und Subsahara-Afrika** im Durchschnitt weniger erfolgreich. Bei diesen Vorhaben werden überdurchschnittlich oft Kontextfaktoren sowie Herausforderungen im Zusammenhang mit der Steuerung als Misserfolgswörter aufgeführt. Mit Blick auf den **Kontext** der Vorhaben, werden insbesondere Einschränkungen durch die COVID-19-Pandemie, die dazu führten, dass Maßnahmen verschoben oder in den digitalen Raum verlagert werden mussten, sowie die Sicherheitslage genannt. Mit Blick auf die **Steuerung** der Vorhaben, findet diese, wie in den qualitativ betrachteten ZPE-Berichten ausgeführt wird, zum Teil nicht systematisch statt, oder bindet relevante Stakeholder des Vorhabens nicht ausreichend ein.

### BMZ-Länderkategorien

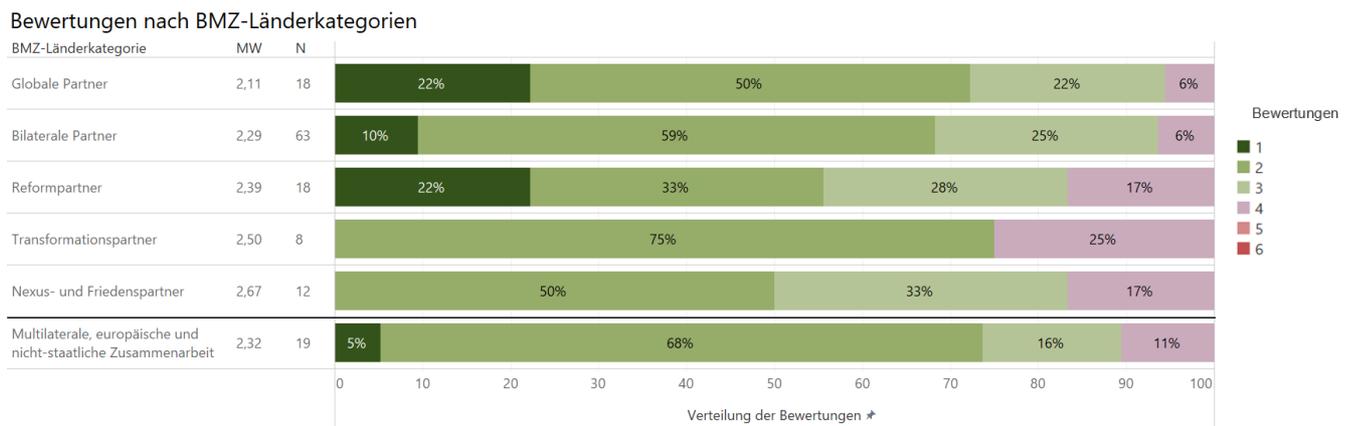
Die Rahmenbedingungen für die von der GIZ durchgeführten Projekte hängen auch von den politisch gesteuerten Formen der Zusammenarbeit durch die deutschen Entwicklungszusammenarbeit und dem jeweiligen Partnerland bzw. den Partnerländern ab. In diesem Kontext hat das BMZ mit dem Reformkonzept „BMZ 2030“ neue **Partnerschaftskategorien** eingeführt, die eine strategischere Zusammenarbeit mit den Partnerländern ermöglichen. Ziel ist es, die Art der Zusammenarbeit zu steuern und die Schwerpunkte und Themen in den Partnerländern möglichst wirksam umzusetzen.<sup>25</sup> Eine Analyse der ZPE-Bewertungen nach diesen Partnerschaftskategorien zeigt, wie die praktische Umsetzung der verschiedenen Arten der Zusammenarbeit von den Evaluierungsexpert\*innen im Durchschnitt bewertet wurde.<sup>26</sup>

**Insgesamt wurden Projekte mit globalen Partnern bei den bisher abgeschlossenen ZPE am besten bewertet. Projekte mit Nexus- und Friedenspartnern erhielten dagegen die niedrigsten Durchschnittsbewertungen** (Abbildung 14). Auf der einen Seite wird mit sogenannten **globalen Partnern** strategisch an der Lösung globaler Zukunftsfragen und dem Schutz globaler Güter (zum Beispiel Umwelt- und Klimaschutz) gearbeitet. Hier wurden 18 Projekte durchschnittlich mit 2,11 bewertet. Auf der anderen Seite werden **Nexus- und Friedenspartner** in Krisen- und Flüchtlingsregionen bei der Stabilisierung und der Bewältigung von Konfliktsachen unterstützt. Die 12 ZPE, die Projekte in diesen Partnerländern evaluiert haben, vergaben die im Vergleich niedrigste Durchschnittsbewertung von 2,67.

<sup>25</sup> BMZ, 2020: <https://www.bmz.de/de/entwicklungspolitik/reformkonzept-bmz-2030>

<sup>26</sup> Die Analyse umfasst nur ZPE, die bilateral einem Partnerland zugeordnet werden konnten. Dementsprechend wurden Regionale und Überregionale Vorhaben (N=69) in diesem Unterkapitel nicht berücksichtigt.

Abbildung 14: Bewertungen nach BMZ-Länderkategorien (N=138)



**Für die Projekte in Partnerländern anderer Kategorien liegt die durchschnittliche ZPE-Bewertung dazwischen.** Mit bilateralen Partnerländern werden langfristig gemeinsame Entwicklungsziele verfolgt. Hier werden neben den „klassischen“ **bilateralen Partnerschaften** (MW 2,29 aus 63 ZPE) zwei weitere Formen unterschieden: **Reformpartner** – Länder, die sich in den letzten Jahren als besonders reformorientiert erwiesen haben – werden besonders gefördert (MW 2,39 aus 18 ZPE). Darüber hinaus werden in Zusammenarbeit mit **Transformationspartnern** gezielt politische und wirtschaftliche Transformationsprozesse in der EU-Nachbarschaft unterstützt (MW 2,50 aus 8 ZPE). Mit Entwicklungsländern, die in keine der vorgenannten Kategorien fallen, gibt es keine direkte staatliche Zusammenarbeit. Allerdings werden auch hier Entwicklungsziele durch die Förderung der Zivilgesellschaft sowie durch die EU und multilaterale Institutionen unterstützt. Diese Art der Zusammenarbeit wurde in 19 ZPE mit durchschnittlich 2,32 bewertet. Erklärungsansätze für die Unterschiede in den Bewertungen – insbesondere für die vergleichsweise gut bewerteten Projekte mit globalen Partnern und

#### Trendbox 4: Einfluss der BMZ-Länderkategorien auf ZPE-Bewertungen

- Projekte in den Kategorien **Globale Partner, aber auch Bilaterale Partner, sowie Multilaterale, Europäische und Nicht-Staatliche Partner** sind im Durchschnitt erfolgreicher. Bei der qualitativen Betrachtung der ZPE-Berichte dieser Vorhaben fällt auf, dass Zusammenarbeit mit involvierten Stakeholdern einerseits überdurchschnittlich oft als Erfolgsfaktor, und gleichzeitig überdurchschnittlich oft als Misserfolgswort aufgeführt wird. Auch hier zeigt sich somit, dass ein **permanentes und temporäres Kooperationsystem**, einmal eingeführt, zum Erfolgsfaktor wird. Mit Blick auf die einzelnen Kategorien, stehen weitere unterstützende Faktoren hervor: So scheinen Vorhaben mit Globalen Partnern über eine geeignetere **Ressourcenausstattung** zu verfügen (38% der qualitativ betrachteten ZPE-Berichte der Vorhaben mit Globalen Partnern vs. 17%), zum Beispiel mit Blick auf die Einbindung erfahrener Langzeitfachkräfte.
- Projekte in den **Kategorien Reformpartner, Transformationspartner und Nexus- und Friedenspartner** sind im Durchschnitt weniger erfolgreich. Bei Vorhaben mit Reformpartnern fällt auf, dass in diesen das **Ownership der involvierten Stakeholder** (Partner ausgeschlossen) überdurchschnittlich oft kritisch aufgeführt wird, zum Beispiel als Konsequenz häufiger Personalwechsel. Bei Vorhaben mit Transformationspartnern werden hingegen überdurchschnittlich oft **Herausforderungen in der Steuerung** des Vorhabens (44% vs. 37%) genannt, die, laut Evaluator\*innen der ZPE, u.a. auf fehlendes systematisches Monitoring und Lernen zurückzuführen sind. Bei Vorhaben mit Nexus- und Friedenspartnern wird oft auf den **Kontext** als Hindernisfaktor verwiesen. Neben Einschränkungen durch die COVID-19-Pandemie (60% vs. 44%), betrifft das insbesondere die Sicherheitslage (80% vs. 14%). Konflikte und fragile Situationen vor Ort haben in diesen Vorhaben vergleichsweise oft zu Verspätungen und Mehraufwänden geführt, u.a. aufgrund Inflation und steigender Benzinpreise.

die schwächer bewerteten Projekte mit Nexus- und Friedenspartnern – wurden in der qualitativen Analyse erhoben (vgl. Kapitel 2.3). Die Ergebnisse sind in Trendbox 4 dargestellt.

## Fragilität

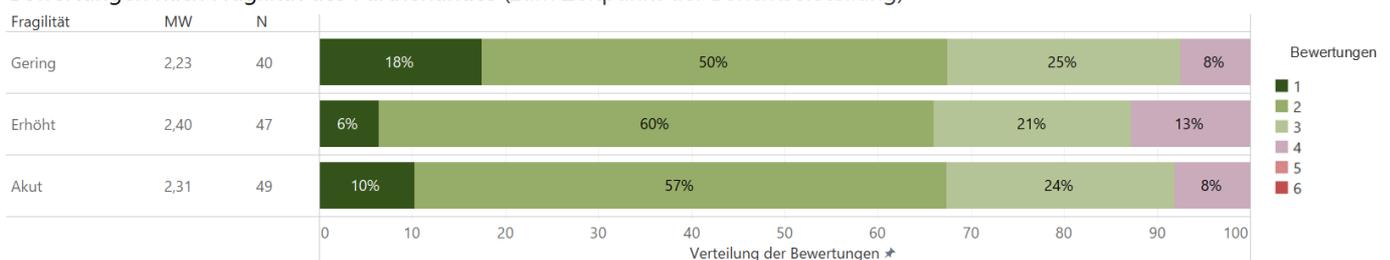
In den vorangegangenen Analysen wurden regionale Bewertungsunterschiede und die verschiedenen Arten der Zusammenarbeit aufgegriffen. Die politische und zivilgesellschaftliche Stabilität ist hier bereits ein wichtiger Faktor. Dennoch ist das Thema **Fragilität** noch einmal hervorzuheben – durch eine steigende Anzahl bewaffneter Konflikte leben mittlerweile rund ein Viertel der Weltbevölkerung in fragilen Staaten mit hohen Sicherheits- und Entwicklungsrisiken.<sup>27</sup> Fragilität hat demzufolge große Auswirkungen auf die Umsetzung der GIZ-Vorhaben, da unter anderem eine außerordentliche Anpassungsfähigkeit in Bezug auf die Planung und Steuerung der Projekte erforderlich ist, um effektive Unterstützungsleistungen zu erbringen.

**Abbildung 15 zeigt, wie sich die ZPE-Bewertungen von Projekten, die in einem mehr oder weniger fragilen Kontext durchgeführt wurden, von anderen Bewertungen unterscheiden.**<sup>28</sup> Die Kategorisierung erfolgte auf Grundlage interner und vertraulicher Daten anhand einer sicherheitsorientierten Fragilitätsanalyse der jeweiligen Partnerländer.<sup>29</sup> Dabei wird zwischen Ländern mit geringer Fragilität (gering), Partnerländern mit erhöhter Fragilität (erhöht) und Partnerländern mit akuter Fragilität unterschieden – hierzu zählen alle Länder, die sich in einem gewaltsamen Konflikt oder landesweiten Krisen befinden (akut).

**Hierbei kann festgestellt werden, dass kein eindeutiger Zusammenhang zwischen der Fragilität des Partnerlandes und den ZPE-Bewertungen besteht.** Die durchschnittliche Bewertung für Projekte in Kontexten mit geringer Fragilität liegt bei 2,23 (N=40) und ist damit etwas höher als die durchschnittliche Bewertung von Projekten in Kontexten mit erhöhter Fragilität (MW=2,40; N=47). Allerdings werden Projekte in Kontexten mit akuter Fragilität besser bewertet als Projekte in der Kategorie mittlerer Fragilität (MW=2,31; N=49). Auffällig ist, dass anteilig die meisten ZPE in Kontexten mit akuter Fragilität stattfanden, was durch den Fokus vieler Projekte auf Friedenssicherung und Stabilität erklärt werden kann (vgl. Kapitel 3.6). Insgesamt hat die Fragilität eines Partnerlandes nach diesen Daten keinen entscheidenden Einfluss auf die ZPE-Bewertung. Eine statistische Korrelationsanalyse zeigte ebenfalls keinen evidenzbasierten Zusammenhang.

Abbildung 15: Bewertungen nach Fragilität (N=136)

Bewertungen nach Fragilität des Partnerlandes (zum Zeitpunkt der Berichtserstellung)



Um den fehlenden Zusammenhang zwischen den ZPE-Bewertungen und der Fragilität weiter zu validieren, wurde ein **zweiter Fragilitätsindex** verwendet und mit den ZPE-Bewertungen korreliert. Mit dem **Fragile States Index (FSI)** wird die Fragilität von Ländern nach einem mehrdimensionalen Index bewertet, der nicht nur sicherheitsbezogene Aspekte berücksichtigt und jedem Land eine Gesamtpunktzahl zwischen 1 (nicht fragil) und 120 (sehr fragil) zuordnet.<sup>30</sup> Für die Zuordnung zu den ZPE wurde das Jahr der Evaluierungsmission anstelle des Berichtsjahrs herangezogen. Die Analyse lässt zwar einen schwachen Zusammenhang zwischen

<sup>27</sup> BMZ, 2020: <https://www.bmz.de/de/entwicklungspolitik/reformkonzept-bmz-2030>

<sup>28</sup> Die Analyse umfasst nur ZPE, die bilateral einem Partnerland zugeordnet werden konnten. Dementsprechend wurden in diesem Unterkapitel Regionale und Überregionale Vorhaben (N=69) nicht berücksichtigt. Zudem sind zwei bilaterale ZPE nicht abgebildet, da für das Land im Berichtsjahr der Evaluierung keine Risikobewertung vorgenommen wurde (Burundi).

<sup>29</sup> Eine Analyse nach dem Jahr der Evaluierungsmission (statt dem Berichtsjahr) hat keine wesentlichen Unterschiede ergeben.

<sup>30</sup> FSI 2022: Weitere Informationen zur Methodik des Indexes sind der Website zu entnehmen.

ZPE-Bewertungen und der Fragilität vermuten (je fragiler, desto schlechter die Bewertung), dieser ist jedoch statistisch nicht signifikant (vgl. Abbildung 29 im Anhang).

### 3.6 Auswertung der ZPE nach Themen und Sektoren

Neben den Projektcharakteristika und regionalen Kontextfaktoren lassen sich aus den ZPE auch Erkenntnisse im Zusammenhang mit der inhaltlichen Ausrichtung der evaluierten Projekte gewinnen. Die Frage, welche Themen und warum die Evaluierungsexpert\*innen GIZ-Projekte für besonders erfolgreich oder weniger erfolgreich halten, ist zentral für das organisationsbezogene Lernen der GIZ.

Dabei konzentrieren sich Vorhaben der GIZ (und anderer Durchführungsorganisationen der deutschen Entwicklungszusammenarbeit) seit dem Reformkonzept „BMZ 2030“ auf fünf **Kernthemen** (Friedenssicherung, Ernährungssicherung, Ausbildung und nachhaltiges Wachstum, Klima und Energie, Umwelt und natürliche Ressourcen).<sup>31</sup> Seit 2022 ist der Bereich Gesundheit und soziale Sicherung als sechstes Kernthema hinzugekommen.<sup>32</sup> Diese Kernthemen definieren die Schwerpunkte, die für den Beitrag Deutschlands zur Umsetzung der Agenda 2030 und der 17 globalen Nachhaltigkeitsziele (*Sustainable Development Goals* (SDGs)) gesetzt wurden. Für jedes dieser Kernthemen sind zusätzlich jeweils drei **Aktionsfelder** definiert, wodurch die inhaltliche Ausrichtung weiter differenziert wird. Die evaluierten Projekte wurden in Bezug auf die inhaltliche Ausrichtung bis zu drei Aktionsfeldern und Kernthemen zugeordnet.

#### BMZ-Kernthemen

**Ein Blick auf die Bewertungen nach BMZ-Kernthemen zeigt, dass die Projekte im Themenfeld Klima und Energie sowie Ausbildung und nachhaltiges Wachstum in den ZPE am positivsten bewertet wurden. Dagegen wurden Projekte im Bereich Friedensförderung und Ernährungssicherung am kritischsten gesehen.** Abbildung 16 zeigt, dass Projekte zu den Kernthemen „Verantwortung für unseren Planeten – Klima und Energie“ (N=61) und „Ausbildung und nachhaltiges Wachstum für gute Jobs“ (N=58) durchschnittlich mit 2,16 bewertet wurden. Im Gegensatz dazu erhielten die 73 evaluierten Projekte im Kernthema „Frieden und gesellschaftlicher Zusammenhalt“ eine im Vergleich niedrigere Bewertung von 2,37. Im Bereich der Friedensförderung und Stabilisierung wurde zudem die Mehrzahl der ZPE durchgeführt, was die Beobachtung aus dem vorherigen Kapitel über die große Anzahl von ZPE in fragilen Kontexten untermauert. Gleichzeitig wurde im vorherigen Kapitel kein eindeutiger Zusammenhang zwischen Fragilität und Evaluierung festgestellt, was die Frage aufwirft, warum Projekte in der Friedensförderung/Stabilisierung im Durchschnitt weniger erfolgreich sind als Projekte, die zu anderen Kernthemen arbeiten. Mögliche Erklärungen für dieses Phänomen werden im Textkasten aus der qualitativen Analyse am Ende dieses Kapitels aufgegriffen (vgl. Kapitel 2.3).

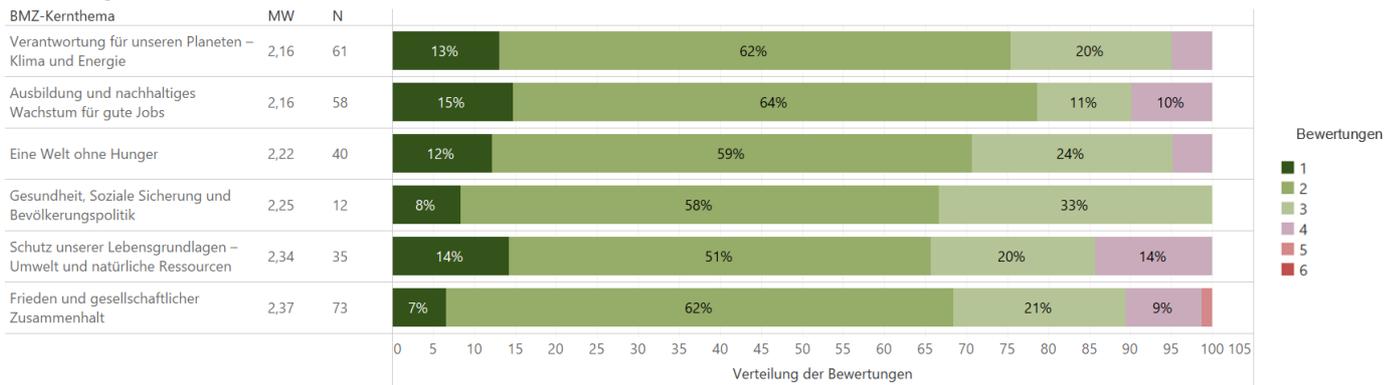
**Die anderen Kernthemen liegen in der Bewertung dazwischen.** Projekte im Bereich „Eine Welt Ohne Hunger“ liegen in der Durchschnittsbewertung an dritter Stelle (MW 2,22 aus 40 ZPE). Es folgen Projekte im neuen Kernthema „Gesundheit, soziale Sicherung und Bevölkerungspolitik“, in dem mit 12 ZPE bisher vergleichsweise wenige Projekte evaluiert wurden (MW=2,20). Da das Kernthema vorher nicht als solches definiert war, ist dies nachvollziehbar. In den Umweltprojekten „Schutz unserer Lebensgrundlagen – Umwelt und natürliche Ressourcen“ (MW 2,34 aus 35 ZPE) fällt zudem auf, dass sowohl viele Projekte als „sehr erfolgreich“ (14 Prozent mit der bestmöglichen Bewertung) als auch knapp 15 Prozent als „nicht erfolgreich“ (Bewertung 4) bewertet wurden.

<sup>31</sup> BMZ, 2020: <https://www.bmz.de/de/entwicklungspolitik/reformkonzept-bmz-2030>

<sup>32</sup> BMZ, Themenmodell 2022 (unveröffentlicht). In dem ab 03/2022 veränderten Themenmodell des BMZ bleiben die fünf Kernthemen und deren Aktionsfelder weitgehend unverändert. Ein Aktionsfeld im Kernthema „Nachhaltige Wirtschaftsentwicklung, Ausbildung und Beschäftigung“ wurde neu definiert und lautet statt „Handel und wirtschaftliche Infrastruktur“ jetzt „Sozial-ökologische Lieferketten, Handel und nachhaltige Infrastruktur“. Das neue sechste Kernthema und drei Aktionsfelder sind hinzugekommen. Das Aktionsfeld „Gesundheit, Pandemien und One Health“ war vorher als sogenanntes Initiativthema definiert.

Abbildung 16: Bewertungen nach BMZ-Kernthemen

**Bewertungen nach BMZ-Kernthemen**

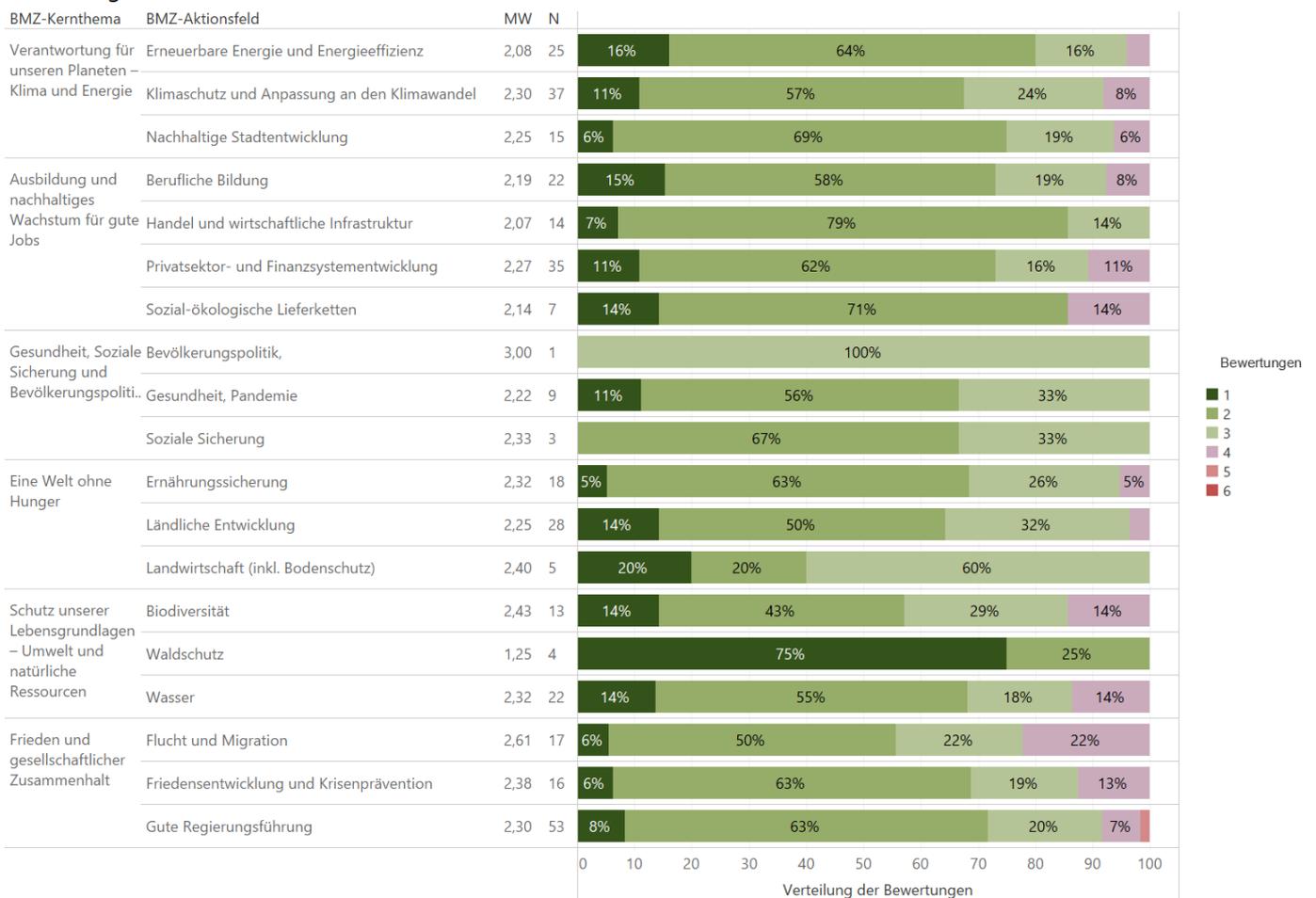


**BMZ-Aktionsfelder**

Eine vertiefte Betrachtung der Bewertungen in verschiedenen Themenfeldern ist über die **Aktionsfelder** möglich. Für jedes BMZ-Kernthema sind drei Aktionsfelder eindeutig definiert. Abbildung 17 zeigt die differenzierte Darstellung der Bewertung nach Aktionsfeldern.<sup>33</sup>

Abbildung 17: Bewertungen nach BMZ-Aktionsfeldern

**Bewertungen nach BMZ-Aktionsfeldern**



<sup>33</sup> Aktionsfeldern, denen kein evaluiertes Projekt zugeordnet wurde, werden in der Abbildung nicht berücksichtigt. Gleichzeitig können Projekte auch mehreren Aktionsfeldern (bzw. BMZ-Kernthemen) zugeordnet sein.

**Betrachtet man die Bewertungen nach der Zuordnung der Projekte zu Aktionsfeldern, wird deutlich, dass die zuvor diskutierte durchschnittliche Bewertung einzelner BMZ-Kernthemen durch Projekte in einigen Aktionsfelder zu erklären ist.** So lässt sich beispielsweise die gute Bewertung im Themenfeld Klima und Energie insbesondere durch Projekte im Aktionsfeld „Erneuerbare Energien und Energieeffizienz“ erklären, die zu 80 Prozent als „sehr Erfolgreich“ oder „erfolgreich“ bewertet wurden (N=25). Auf der anderen Seite des Bewertungsspektrums zeigt sich, dass im Bereich der Friedensförderung Projekte im Aktionsfeld „Flucht und Migration“ tendenziell als weniger erfolgreich bewertet werden. Die durchschnittliche Bewertung ist hier (mit Ausnahme von Aktionsfeldern mit <5 Bewertungen) die niedrigste im Vergleich aller Themen (MW=2,61). Es zeigt sich auch, dass die zuvor beschriebene Bandbreite der Bewertungen im Themenfeld Umwelt und natürliche Ressourcen auf die Themen „Biodiversität“ als auch „Wasser“ zurückzuführen ist. In beiden Aktionsfeldern werden jeweils knapp 15 Prozent der Projekt als „sehr erfolgreich“ (bestmögliche Bewertung) oder „erfolglos“ (Bewertung 4) bewertet. Wie zuvor erwähnt, werden in Textbox 5 mögliche Erklärungen aus der qualitativen Analyse in Bezug auf die Bewertungsunterschiede in Abhängigkeit vom Thema benannt (vgl. Kapitel 2.3).<sup>34</sup>

### **Trendbox 5: Erklärungen für unterschiedliche ZPE-Bewertungen in Abhängigkeit vom Themenfeld**

Die BMZ-Kernthemen „Frieden und gesellschaftlicher Zusammenhalt“, „Schutz unserer Lebensgrundlagen – Umwelt und natürliche Ressourcen“, sowie „Eine Welt ohne Hunger“ waren im Durchschnitt weniger erfolgreich:

- Bei der qualitativen Betrachtung der ZPE-Berichte über die Vorhaben zu den BMZ-Kernthemen „**Schutz unserer Lebensgrundlagen – Umwelt und natürliche Ressourcen**“ und „**Eine Welt ohne Hunger**“ wird deutlich, dass der Kontext die Vorhaben beeinträchtigt hat. So wird überdurchschnittlich oft auf die Einschränkungen durch die **COVID-19-Pandemie** sowie auf die **Sicherheitslage** vor Ort verwiesen. Im Fall der Vorhaben zu dem BMZ-Kernthema „Schutz unserer Lebensgrundlagen – Umwelt und natürliche Ressourcen“ hatte dies zudem Konsequenzen für die **Ownership der beteiligten Stakeholder** (Partner ausgenommen), deren Prioritäten sich im Rahmen der Pandemie-Bekämpfung teils stark gewandelt haben.
- Bei der qualitativen Betrachtung der ZPE-Berichte der Vorhaben des BMZ-Kernthemas „**Frieden und gesellschaftlicher Zusammenhalt**“ sind es nicht die Kontextfaktoren, die die unterschiedliche Bewertung erklären. Stattdessen wird in den ZPE-Berichten dieser Vorhaben vermehrt auf **Herausforderungen in der Steuerung** der Vorhaben verwiesen, zurückzuführen zum Beispiel auf unzureichendes Wissensmanagement oder Monitoringstrukturen.

Die inhaltliche Einordnung der Projekte ist nicht nur nach den vom BMZ definierten „Kernthemen“ und „Aktionsfeldern“ möglich. Die Projekte können zum Beispiel auch nach den von der GIZ intern definierten „**Positionierungsthemen**“ oder international etablierten **CRS-Codes**, die welche von Geberländern für die Berichterstattung über ihre Ausgaben in der Entwicklungszusammenarbeit verwendet werden, analysiert werden. Die ZPE-Bewertungen anhand dieser beiden Kategorien liegt im Anhang zu diesem Bericht bei.

<sup>34</sup> Neben den in diesem Bericht beschriebenen Erklärungen für die Bewertungsunterschiede erfolgte eine vertiefte inhaltliche Analyse nach 10 Themenfeldern in einem gesonderten Bericht (3. Zielsetzung von *QSA Insights*).

### 3.7 Abhängigkeit der Bewertung von Unterschieden bei der Evaluierung

In diesem letzten Ergebniskapitel wird untersucht, welchen Einfluss Veränderungen der Rahmenbedingungen bei der Durchführung der Zentralen Projektevaluierungen selbst auf die Bewertungen gehabt haben könnten. Demnach unterscheiden sich einige Zentrale Projektevaluierungen auch in ihren Umsetzungsprozessen und Bewertungskriterien. Seit der Einführung der ZPE sind dabei zwei Faktoren hervorzuheben: Zum einen wurden die Bewertungskriterien der ZPE im Jahr 2021 um das Kriterium Kohärenz (vgl. Kapitel 3.3) sowie die sogenannten **K.O.-Kriterien** ergänzt. Zum anderen hat die **COVID-19-Pandemie** nach wie vor einen erheblichen Einfluss auf den Verlauf der ZPE.

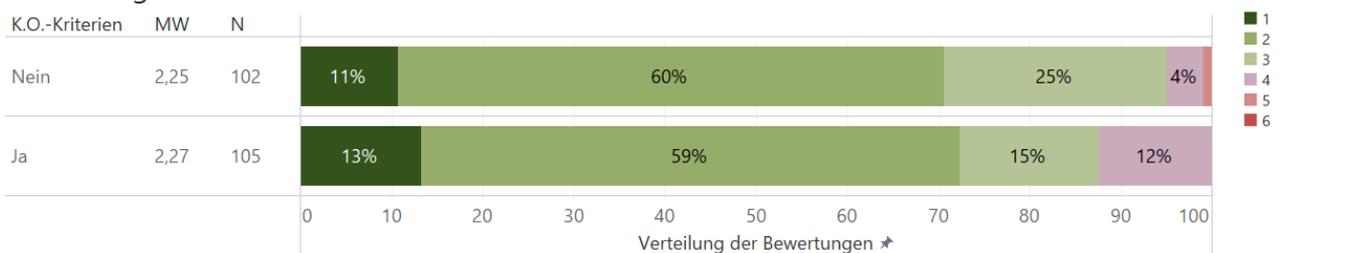
#### K.O.-Kriterien

Durch die Einführung der **K.O.-Kriterien** im Jahr 2021 durch das BMZ wurde sichergestellt, dass Projekte, die in ausgewählten (zentralen) Bewertungskriterien (vgl. Kapitel 3.3) als „nicht erfolgreich“ eingestuft wurden, auch in der Gesamtbewertung nicht als „erfolgreich“ eingestuft werden können. Genauer gesagt: Erhält ein Projekt in einem der drei Kriterien Effektivität, Impact oder Nachhaltigkeit die Bewertung 4 bewertet, so wird es insgesamt als „nicht erfolgreich“ eingestuft.<sup>35</sup>

**Welchen Einfluss die K.O.-Kriterien auf die Gesamtbewertung der ZPE haben würden, war vor der Einführung nicht klar:** Einerseits bestand die Möglichkeit, dass sich die durchschnittliche Gesamtbewertung mit der Einführung der K.O.-Kriterien verschlechtert, da Projekte, die bisher insgesamt als erfolgreich eingestuft wurden, nun aufgrund einer schlechteren Bewertung einzelner Kriterien insgesamt als „nicht erfolgreich“ gelten. Andererseits bestand die Möglichkeit, dass die Einführung von K.O.-Kriterien einen Einfluss auf die Bewertung durch die Gutachter\*innen hat (auch wenn dies ausdrücklich nicht beabsichtigt ist). So könnten die Expert\*innen ein Projekt in den Bewertungskriterien etwas besser und damit noch als „erfolgreich“ bewerten, weil sie das Projekt trotz der Schwächen bei bestimmten Kriterien nicht als „nicht erfolgreich“ einstufen wollen.

Abbildung 18: Bewertungen vor und nach Einführung der K.O.-Kriterien (N=207)

#### Bewertungen nach K.O.-Kriterien



**Ein Blick auf die Verteilung der Bewertungen zeigt, dass die durchschnittliche Gesamtbewertung mit der Einführung der K.O.-Kriterien bislang weder stark gestiegen noch stark gefallen ist (Abbildung 18).** Während vor der Einführung der K.O.-Kriterien 102 Projekte mit durchschnittlich 2,25 bewertet wurden, erhielten die 105 nach dem neuen Bewertungssystem evaluierten Projekte mit einer durchschnittlichen Bewertung von 2,27. Die Hypothese, dass die durchschnittliche Bewertung mit Einführung der K.O.-Kriterien wesentlich sinkt, kann demnach nicht bestätigt werden. Ein genauerer Blick auf die Bewertungen liefert jedoch Hinweise darauf, dass die zweite Hypothese einen Einfluss gehabt haben könnte. So hat sich zum einen der Anteil der in den K.O.-Kriterien Effektivität und Nachhaltigkeit mit 4 oder schlechter bewerteten Projekte verringert (im Kriterium

<sup>35</sup> GIZ, 2022: <https://www.giz.de/de/downloads/giz2022-de-instrumentenpapier-zentrale-projektevaluierungen-bmz.pdf>

Impact ist die Anzahl dagegen leicht gestiegen). Vor allem jedoch ist der Anteil der ZPE, die knapp mit 3 bewertet wurden für alle K.O.-Kriterien deutlich gestiegen.<sup>36</sup> Dies deutet darauf hin, dass die Evaluator\*innen häufiger die Bewertung 3 vergeben, damit kein K.O.-Kriterium greift. Insgesamt haben die K.O.-Kriterien die Gesamtbewertung bei bislang 9 ZPEs beeinflusst. Ein Vergleich der theoretischen ZPE-Bewertungen ohne die K.O.-Kriterien zeigt, dass die durchschnittliche Bewertung mit 2,17 deutlich besser ausgefallen wäre (siehe Abbildung 36 im Anhang). Da die Bewertung ohne das neu eingeführte Kriterium noch besser ausgefallen wäre, kann angenommen werden, dass die K.O.-Kriterien die durchschnittliche Gesamtbewertung gesenkt haben und dass andere Faktoren, die im Durchschnitt zu einer besseren Bewertung geführt haben, dadurch teilweise ausgeglichen wurden. Wichtig in diesem Zusammenhang ist, dass die Bewertung zusätzlich durch weitere Faktoren beeinflusst wird. Die zeitgleiche Einführung des sechsten Bewertungskriteriums Kohärenz könnte hier einen Einfluss haben, da dadurch der Anteil der in der Regel kritischer bewerteten Kriterien (v.a. Impact, Nachhaltigkeit) an der Bewertung verringert wurde. Gleichzeitig wurde das Kriterium Kohärenz bisher überdurchschnittlich gut bewertet, wodurch die durchschnittliche Bewertung angehoben wurde (vgl. Kapitel 3.3). Ein weiterer Faktor ist die zeitliche Überschneidung der Einführung der K.O.-Kriterien (Beginn 2021 bei den ZPE) mit der COVID-19-Pandemie (ab März 2020); darauf wird im nächsten Unterabschnitt eingegangen.

### Die COVID-19-Pandemie und ihr Einfluss auf die ZPE

Die weltweite **COVID-19-Pandemie** hat nach wie vor erhebliche Auswirkungen auf die internationale Entwicklungszusammenarbeit im Allgemeinen und die Arbeit der GIZ im Besonderen. Auch die ZPE blieben davon nicht verschont und mussten sich an Herausforderungen wie Reisebeschränkungen andere Maßnahmen zur Eindämmung der Pandemie anpassen.<sup>37</sup>

**Im Hinblick auf die während der Pandemie bei den Zentralen Projektevaluierungen vorgenommenen Anpassungen ist besonders ein Punkt hervorzuheben:** So wurde nach Ausbruch der Pandemie bei den Evaluierungsmissionen auf die ansonsten zur Datenerhebung üblichen Besuche vor Ort verzichtet und diese häufig durch *remote*-Missionen ersetzt. In Abhängigkeit von den landesspezifischen Regelungen und der jeweiligen Pandemiesituation wurde entweder eine *semi-remote*-Variante (die lokalen Evaluator\*innen bewegen sich vor Ort) oder eine reine *remote*-Variante (die Mission wird komplett virtuell durchgeführt) gewählt. Da diese Anpassungen in den Durchführungsprozessen der Evaluierungsmissionen Einfluss auf die zentrale Datenerhebung der Evaluierungen – und damit auch auf die Datenqualität – hatten, ist ein Einfluss auf die ZPE-Bewertungen nicht auszuschließen.

**Abbildung 19 zeigt die durchschnittliche ZPE-Bewertung nach dem Zeitpunkt der Evaluierungsmission und gibt einen Überblick über die Auswirkungen der Pandemie auf die ZPE-Bewertungen.** Die vor dem Ausbruch der Pandemie durchgeführten Evaluierungsmissionen sind links der Trennlinie dargestellt. Rechts von der Trennlinie befinden sich die ZPE, deren Evaluierungsmissionen nach 03/2020 durchgeführt wurden. Insgesamt ist zu erkennen, dass der größte Teil der abgeschlossenen ZPE in den Zeitraum nach der Pandemie fällt (N=145). Ein kleinerer Anteil der Missionen wurde vor der Pandemie durchgeführt (N=60). Bei dieser Betrachtung ist jedoch kein eindeutiger grafischer Trend zu erkennen, auch wenn nach der Pandemie etwas konstanter mit knapp 2 bewertet wurde. Betrachtet man die Mittelwerte im Vergleich der Gruppen, so fällt je-

<sup>36</sup> Als „knapp mit 3 bewertet“ gilt eine Bewertung, bei der für das jeweilige Kriterium zwischen 67 und 70 Punkte vergeben wurden, so dass die Bewertung knapp oberhalb von 4 lag (ab 66 Punkte oder weniger). Details zur Analyse:

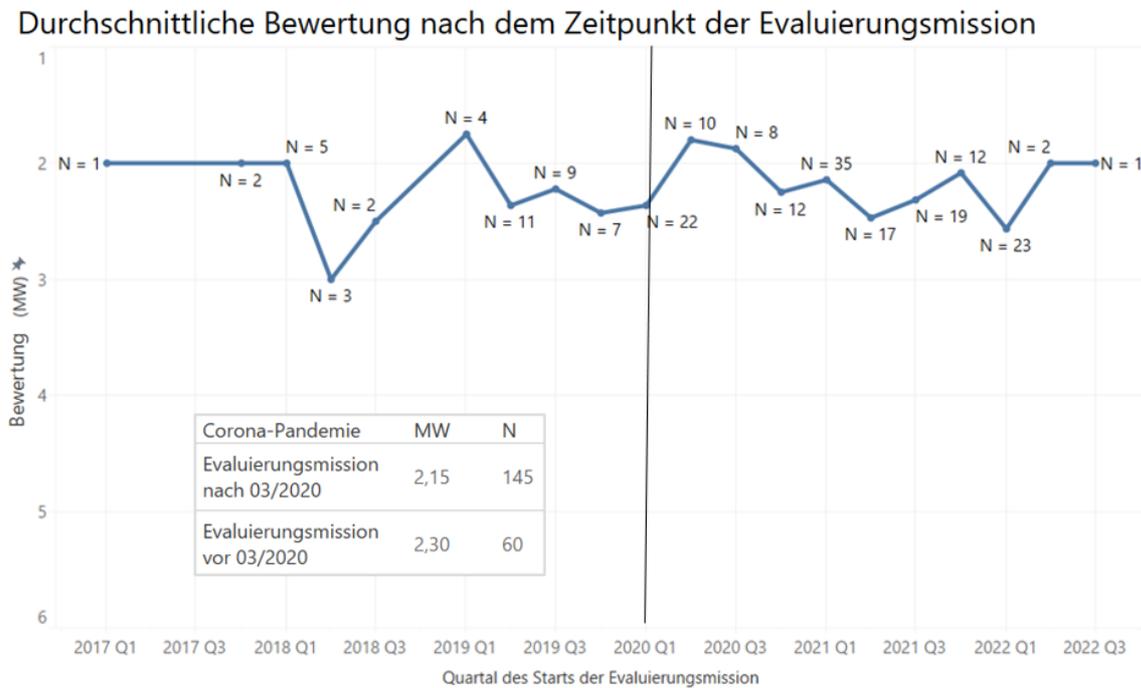
- Im Kriterium **Effektivität** wurde bei 4 % der ZPE seit Einführung der K.O.-Kriterien die Bewertung 4 oder schlechter vergeben (7 % vor der Einführung der K.O.-Kriterien). Zudem wurden 10 % der ZPE-bewerteten Projekte im Kriterium Effektivität knapp mit 3 bewertet (4 % vor Einführung der K.O.-Kriterien).
- Im Kriterium **Impact** wurde bei 8 % der ZPE seit Einführung der K.O.-Kriterien die Bewertung 4 oder schlechter vergeben (7 % vor Einführung der K.O.-Kriterien). Zudem wurden 15 % der ZPE-bewerteten Projekte im Kriterium Impact knapp mit 3 bewertet (9 % vor Einführung der K.O.-Kriterien).
- Im Kriterium **Nachhaltigkeit** wurde bei 11 % der ZPE seit Einführung der K.O.-Kriterien die Bewertung 4 oder schlechter vergeben (14 % vor Einführung der K.O.-Kriterien). Zudem wurden 21 % der ZPE-bewerteten Projekte im Kriterium Nachhaltigkeit knapp mit 3 bewertet (15 % vor Einführung der K.O.-Kriterien).

Die Daten sind in Anhang in den **Abbildungen 34 und 35** dargestellt.

<sup>37</sup> Der Einfluss der Corona-19 Pandemie auf die evaluierten Vorhaben sowie auf die Durchführung der ZPE wird durch eine vertiefte qualitative Analyse der ZPE-Berichte in einem separaten Produkt aufgearbeitet (3. Zielsetzung von QSA Insights)..

doch auf, dass die Durchschnittsbewertung von ZPE mit einer Evaluierungsmissionen nach der Pandemie etwas besser ausfällt (MW=2,15) als die Durchschnittsbewertung von ZPE mit einer Evaluierungsmission vor der Pandemie (MW=2,30).

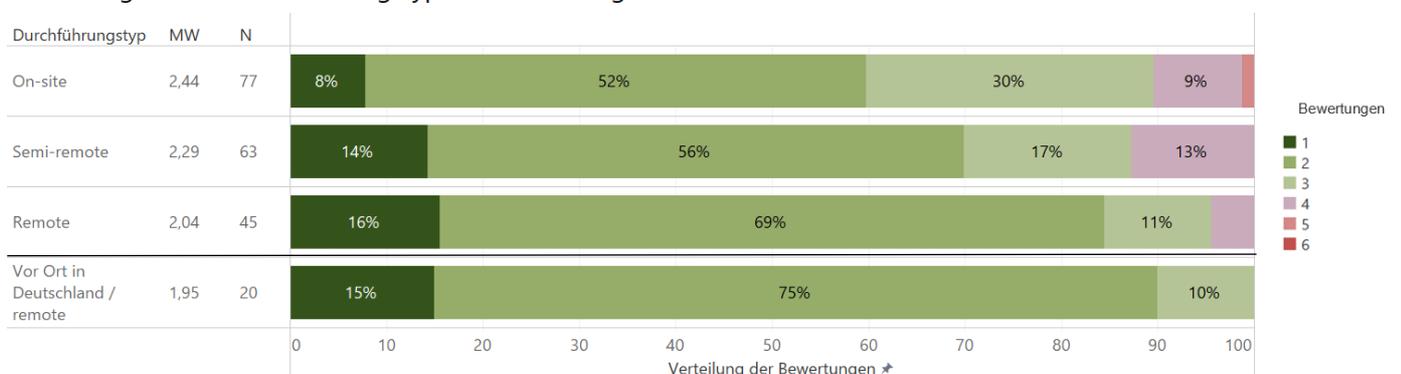
Abbildung 19: Durchschnittliche Bewertungen nach dem Zeitpunkt der Evaluierungsmission (Einfluss der COVID-19-Pandemie) (N=205)<sup>38</sup>



Ein Blick auf die verschiedenen Bewertungen in Abhängigkeit von den **Durchführungsmodalitäten der Evaluierungsmission** ist für eine genauere Analyse hilfreich. Insbesondere wurde in diesem Kontext die Hypothese aufgestellt, dass die Qualität der Daten bei reinen *remote*-Missionen nicht mit der Qualität der im Rahmen von Vor-Ort-Besuchen gewonnenen Daten vergleichbar ist. Denn bei der *remote*-Variante haben die Evaluator\*innen nicht die Möglichkeit, die Aktivitäten und Ergebnisse der evaluierten Projekte vor Ort gründlich zu untersuchen. Vielmehr müssen sie sich vor allem auf die in Telefonaten und Videokonferenzen getroffenen Aussagen der (projektinternen und -externen) Gesprächspartner\*innen „verlassen“. Bei *semi-remote*-Missionen können dagegen Vor-Ort-Besuche mit lokalen Expert\*innen durchgeführt werden – die Datenqualität hängt hier stark von der Interaktion zwischen internationalen Fachkräften (in Deutschland) und den lokalen Fachkräften vor Ort ab.

Abbildung 20: Bewertungen nach Durchführungstyp der Evaluierungsmission (N=205)

### Bewertungen nach Durchführungstyp der Evaluierungsmission



<sup>38</sup> In den Abbildungen 19 und 20 sind zwei ZPE nicht dargestellt, da sie durch Evaluierungen des Kofinanzierungspartners ersetzt wurden und keine vergleichbaren Daten zu den Evaluierungsmissionen vorlagen. Insgesamt wurden die Evaluierungen nach demselben Bewertungsmaßstab durchgeführt.

**Die Daten belegen einen klaren Zusammenhang zwischen dem Durchführungstyp der Evaluierungsmission und der durchschnittlichen Bewertung.** Aus Abbildung 20 wird ersichtlich, dass die Bewertungen von ZPE mit Vor-Ort-Missionen deutlich kritischer ausfallen (MW 2,44; N=77) als die Bewertungen von ZPE mit *remote*-Missionen (MW=2,04; N=45). ZPE mit *semi-remote*-Missionen liegen in der durchschnittlichen Bewertung im Mittelfeld (MW=2,29; N=63). Auffällig ist bei dieser Betrachtung ferner, dass fast alle Projekte mit *remote*-Mission als „erfolgreich“ (Bewertungen 1-3) eingestuft wurden, während jeweils mindestens 10 Prozent der im Rahmen von *semi-remote*-Missionen und Vor-Ort-Missionen bewerteten Projekte als „nicht erfolgreich“ (Bewertung 4 oder 5) eingestuft wurden. Demnach scheint die Qualität der Daten bei *semi-remote*- und Vor-Ort-Missionen einen tieferen Einblick in die Projekte zu bieten, so dass auch kritische Punkte stärker zu Tage treten. Eine statistische Analyse bestätigt diese Beobachtungen: Die ZPE-Bewertungen korrelieren signifikant mit dem Durchführungstyp der Evaluierungsmission (je höher der *remote*-Anteil, desto weniger kritisch die Bewertungen). Gleichzeitig besteht jedoch kein nachweisbarer statistischer Zusammenhang, wenn nur ZPE mit Vor-Ort- und *semi-remote*-Evaluierungsmissionen miteinander verglichen werden. Insgesamt bestätigen diese Beobachtungen die eingangs angeführte Hypothese über den Zusammenhang zwischen Evaluierungsmission und Datenqualität.

Letztlich ist ein weiterer Sonderfall von Evaluierungsmissionen die Durchführungsmodalität *Vor Ort in Deutschland / remote*. Hier werden sektorale und/oder globale Projekte aufgeführt, bei denen mit einem Vor-Ort-Besuch nicht der gleiche Mehrwert wie bei bilateralen Projekten erzielt werden kann, da die Projekte sektoral und geographisch weit verstreut arbeiten und oft keine direkten Zielgruppen in den Partnerländern haben.<sup>39</sup> Die hohe durchschnittliche Bewertung dieser Projekte lässt sich auch dadurch erklären, dass Projekte dieser Art insgesamt besser bewertet wurden, insbesondere im Vergleich zu bilateralen Projekten (siehe Kapitel 3.4).

**Die vorliegenden Daten plausibilisieren die Annahme, dass der zuvor beschriebene Anstieg der Bewertungen im Jahr 2021 (vgl. Kapitel 3.3) durch die hohe Anzahl an *remote*-Missionen in den Jahren 2020/2021 beeinflusst wurde.** Im Jahr 2020 wurden 17 Datenerhebungen erstmalig *remote* durchgeführt, im Jahr 2021 waren es bereits 26 ZPE mit diesem Durchführungstyp. Die Zahl der Vor-Ort-Missionen ist dagegen von 28 im Jahr 2019 auf 14 im Jahr 2020 und 9 im Jahr 2021 zurückgegangen. Dieser Trend hat sich im Jahr 2022 wieder umgekehrt, denn da weniger pandemiebedingte Einschränkungen bestehen, können die Evaluierungsmissionen wieder vermehrt vor Ort durchgeführt werden (vgl. Abbildung 35 im Anhang). Dies passt zu dem beschriebenen Bewertungstrend für die Jahre 2019 bis 2022 (vgl. Kapitel 3.3). Wie in den vorangegangenen Kapiteln beschrieben, lässt sich darüber hinaus jedoch feststellen, dass sich insbesondere die Einführung des Kohärenzkriteriums ab 2021 positiv auf die Gesamtpunktzahl ausgewirkt hat, während die Einführung des K.O.-Kriteriums zu einem Rückgang der durchschnittlichen Bewertung geführt hat. Somit gab es neben dem Durchführungstyp der Evaluierungsmission mehrere Faktoren, die die ZPE-Bewertungen seit Beginn der COVID-19-Pandemie beeinflusst haben.

---

<sup>39</sup> Nach Absprache mit der Stabsstelle Evaluierung wurden alle Sektorvorhaben zu dieser Kategorie gezählt. Ferner wurden Globalvorhaben entweder vor Ort, *semi-remote* oder vor Ort in Deutschland / *remote* kategorisiert.

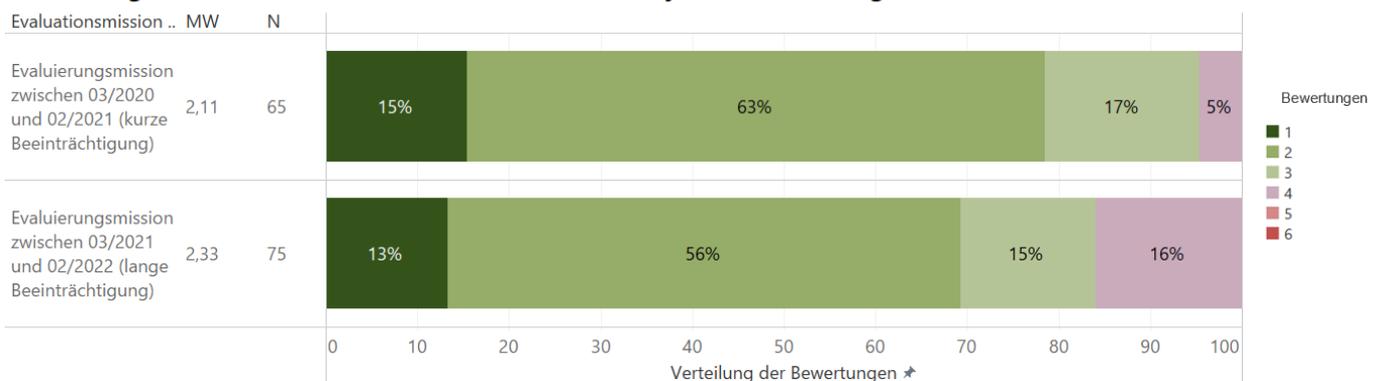
## Die COVID-19-Pandemie und ihr Einfluss auf die Durchführung der Vorhaben

Die COVID-19-Pandemie hatte auch erheblichen Einfluss auf die **Durchführung der Vorhaben**. Vor diesem Hintergrund ist davon auszugehen, dass die Pandemie auch die Bewertung beeinflusst hat, da die Umsetzung der Projektmaßnahmen durch die Pandemie beeinträchtigt wurde. Die Hypothese lautet daher: Je länger ein Projekt unter Pandemiebedingungen durchgeführt wurde, desto länger war die Durchführung eingeschränkt, wodurch wiederum die Bewertung negativ beeinflusst wurde.<sup>40</sup>

**Ein Vergleich der Bewertungen zwischen Projekten mit längerer und weniger langer Beeinträchtigung durch die Pandemie zeigt einen Zusammenhang zwischen Bewertung und Pandemie.** Im Rahmen des Vergleichs wurden die ZPE-Bewertungen von Vorhaben, die durch die Pandemie weniger lange eingeschränkt waren (Evaluierungsmissionen zu Beginn der Pandemie zwischen März 2020 und Februar 2021) und solchen, die deutlich länger eingeschränkt waren (Evaluierungsmission zwischen März 2021 und Februar 2022) einander gegenübergestellt. Der Vergleich bestätigt, dass die Dauer der Pandemie auch Auswirkungen auf die ZPE-Durchschnittsbewertungen hatte: Projekte, die über einen kürzeren Zeitraum von der Pandemie betroffen waren, werden im Durchschnitt eindeutig besser bewertet (MW 2,11) als solche, die über einen längeren Zeitraum betroffen waren (MW 2,33) (Abbildung 21). Dieses Ergebnis steht im Einklang mit anderen (qualitativen) Auswertungen, bei denen beispielsweise beobachtet wurde, dass die Erfüllung der Indikatoren durch die Pandemie negativ beeinflusst wurde und bei den ZPE letztlich zu negativeren qualitativen Analysen in Bezug auf die Kriterien Effektivität, Wirkung, Effizienz und Nachhaltigkeit geführt hat.<sup>41</sup>

Abbildung 21: Bewertungen nach Durchführungstyp der Evaluierungsmission (N=205)

### Bewertungen nach Einfluss der Pandemie auf die Projektdurchführung



<sup>40</sup> Wobei zu beachten ist, dass die Bewertungen von den vorher beschriebenen Einflüssen abhängig sind.

<sup>41</sup> Diese Aussage wurde in den weiteren Komponenten der Querschnittsauswertung vertieft diskutiert.

# Anhang

(1) Zusätzliche Abbildungen – unten aufgeführt

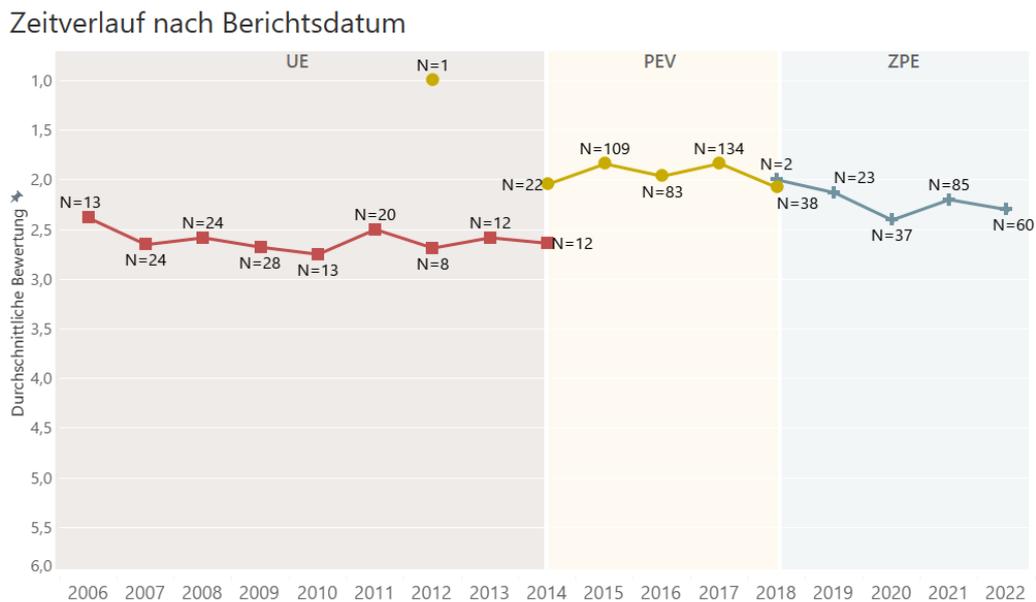
(2) Tabelle zu Erfolgs- und Misserfolgsk Faktoren der qualitativ untersuchten Trends – unten aufgeführt

## 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.

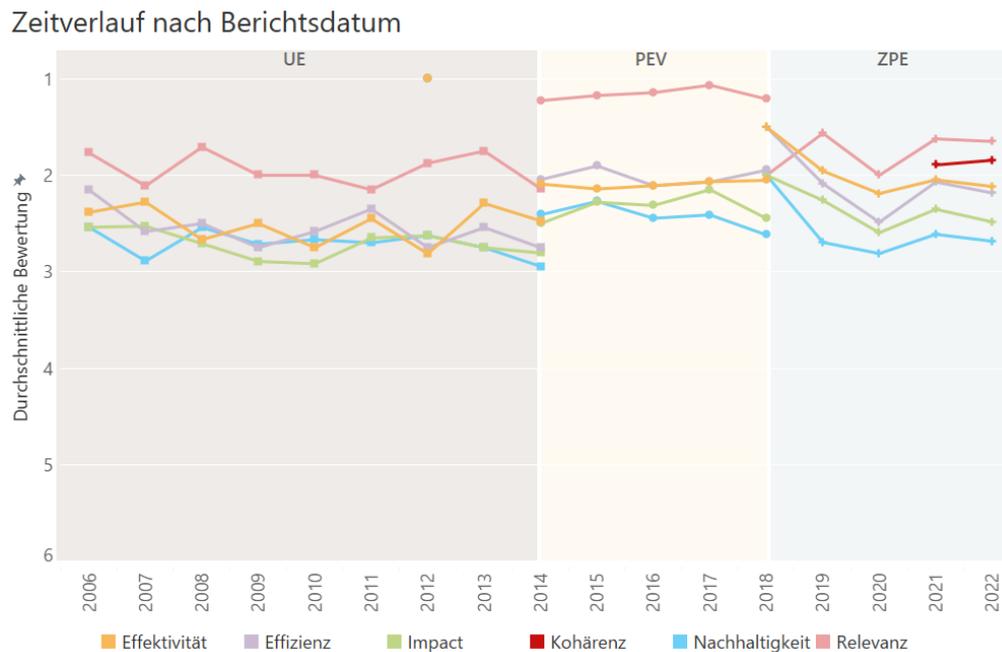
Abbildung 22: Bewertungen im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum) (N=207)



### 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.

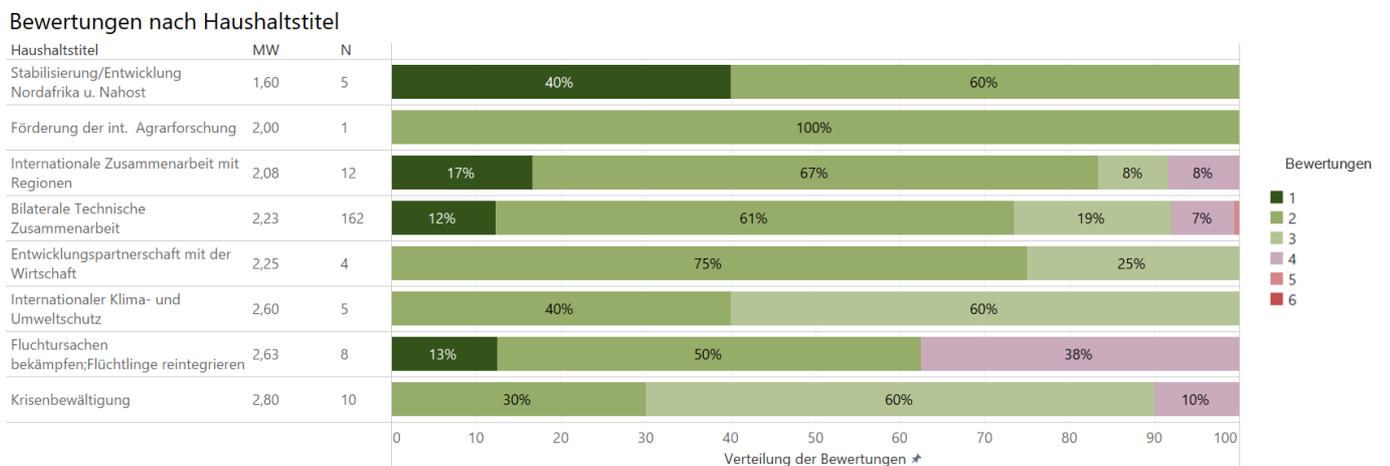
Abbildung 23: Bewertungen nach Evaluierungskriterien im Zeitverlauf einschließlich früherer Evaluierungen (nach Berichtsdatum) (N=207)



### 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.

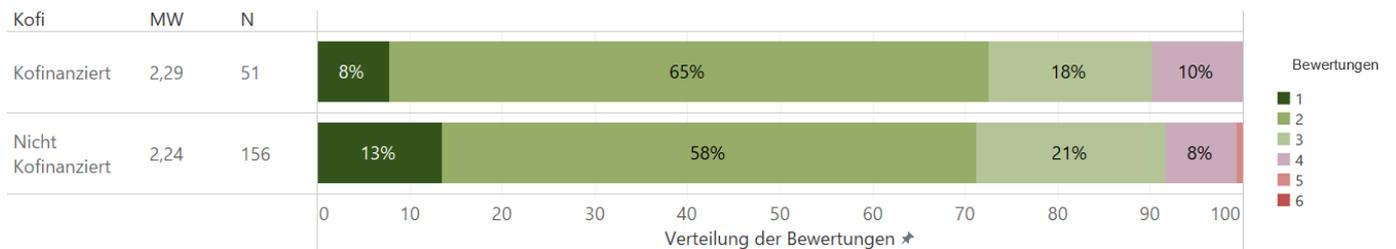
Abbildung 24: Bewertungen nach Haushaltstitel (N=207)



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)

### Bewertungen nach Kofinanzierung

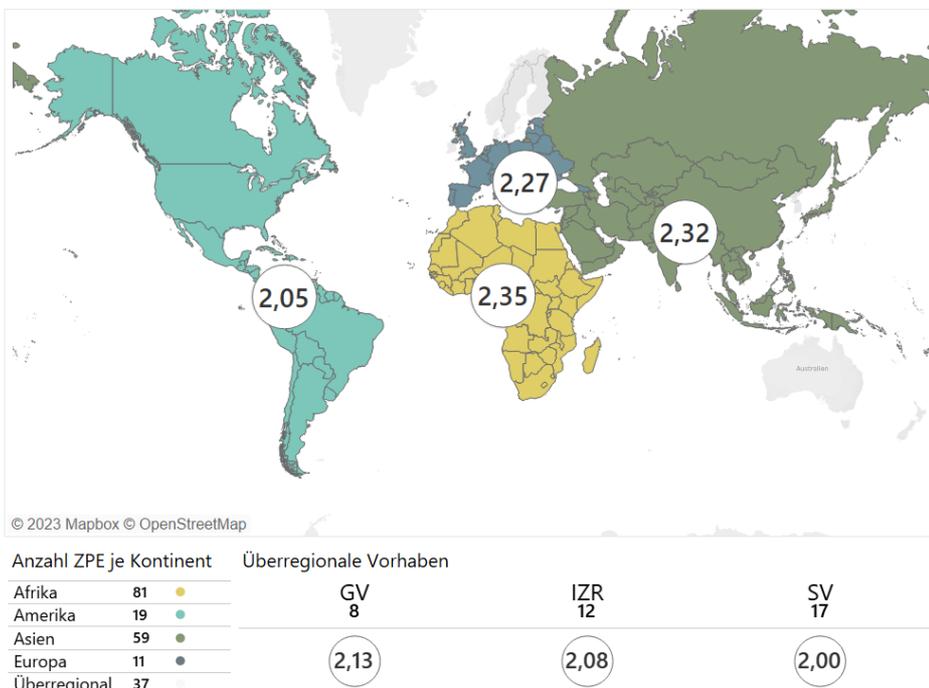


### Weitere Abbildungen zu Kapitel 3.5: Länder und Regionen

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**.

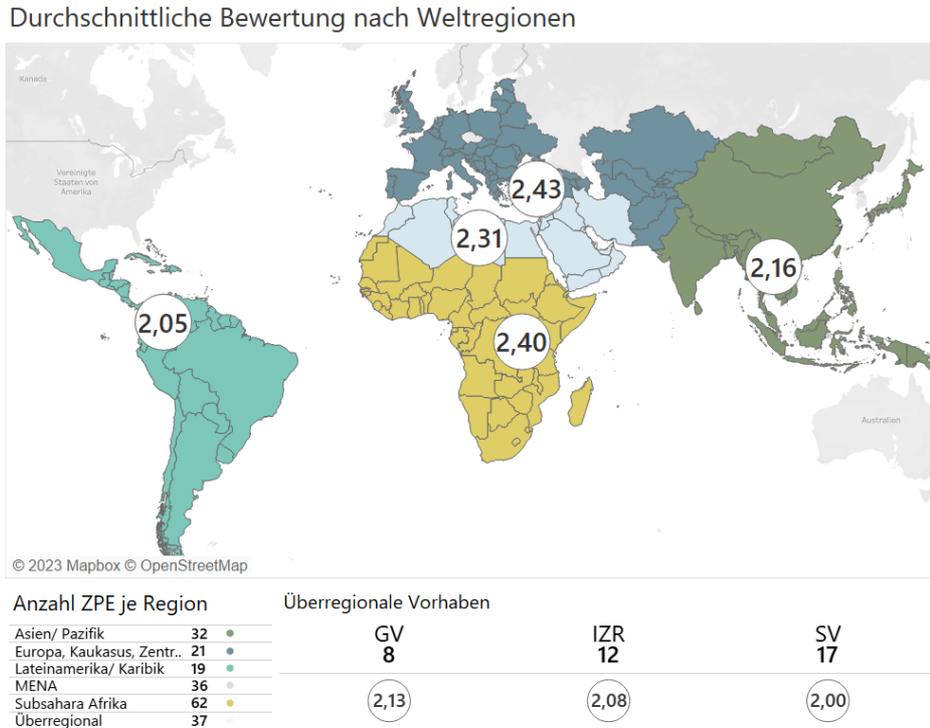
Abbildung 26: Durchschnittliche Bewertungen nach Kontinenten ohne Berücksichtigung von Sektoren und BMZ-Kernthemen (N=207)

### Durchschnittliche Bewertungen nach Kontinenten



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**.

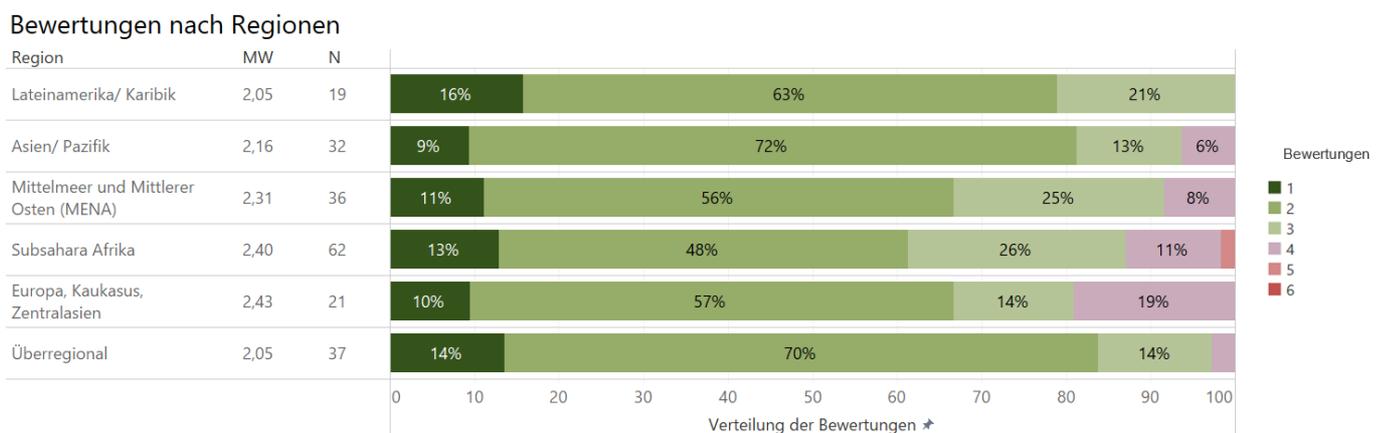
Abbildung 27: Durchschnittliche Bewertungen nach Weltregionen ohne Berücksichtigung von Sektoren und BMZ-Kernthe-



men (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.

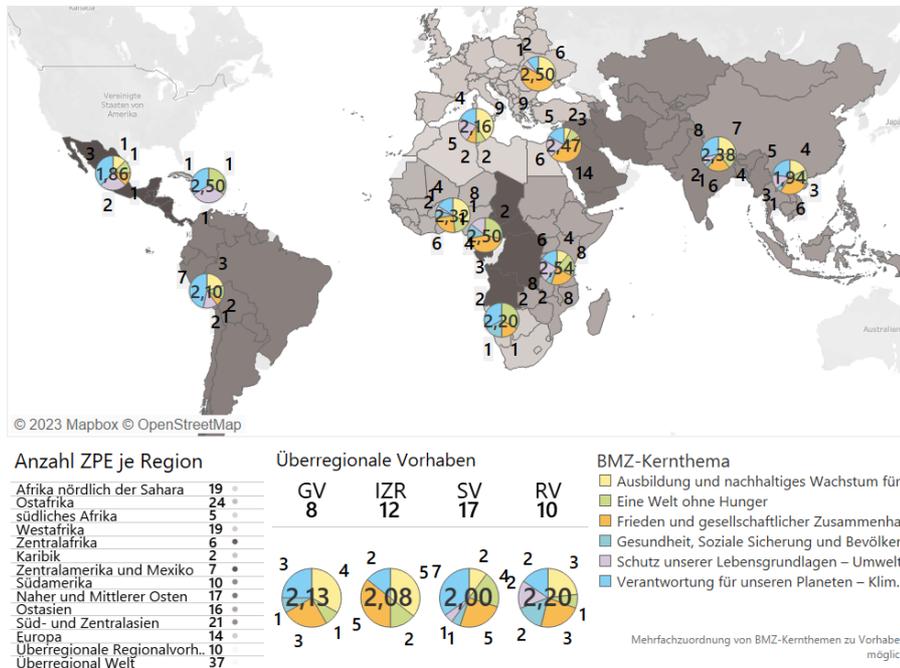
Abbildung 28: Bewertungen nach Weltregionen (N=207)



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)

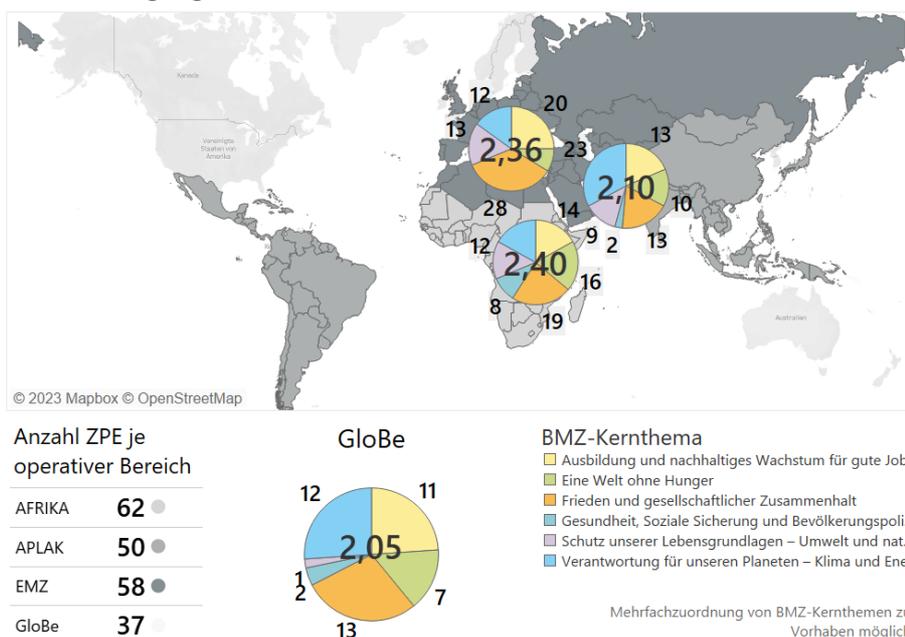
Durchschnittliche Bewertung (unter Berücksichtigung von Sektoren)



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.

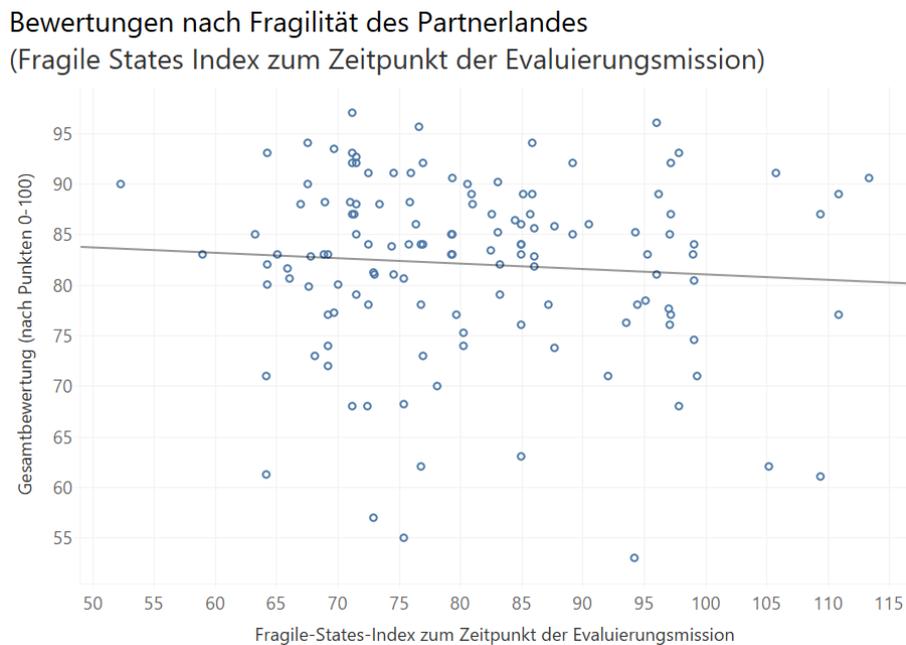
Abbildung 30: Durchschnittliche Bewertungen nach operativen Bereichen unter Berücksichtigung von Sektoren und BMZ-Kernthemen (N=207)

Durchschnittliche Bewertung nach operativen Bereichen (unter Berücksichtigung von Sektoren)



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)

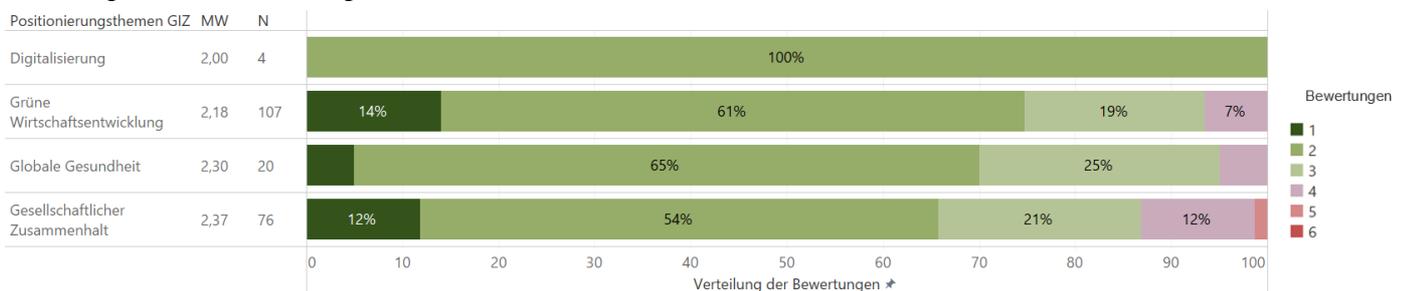


**Weitere Abbildungen zu Kapitel 3.6: Themen und Sektoren**

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 321: Bewertungen nach GIZ-Positionierung (N=207)

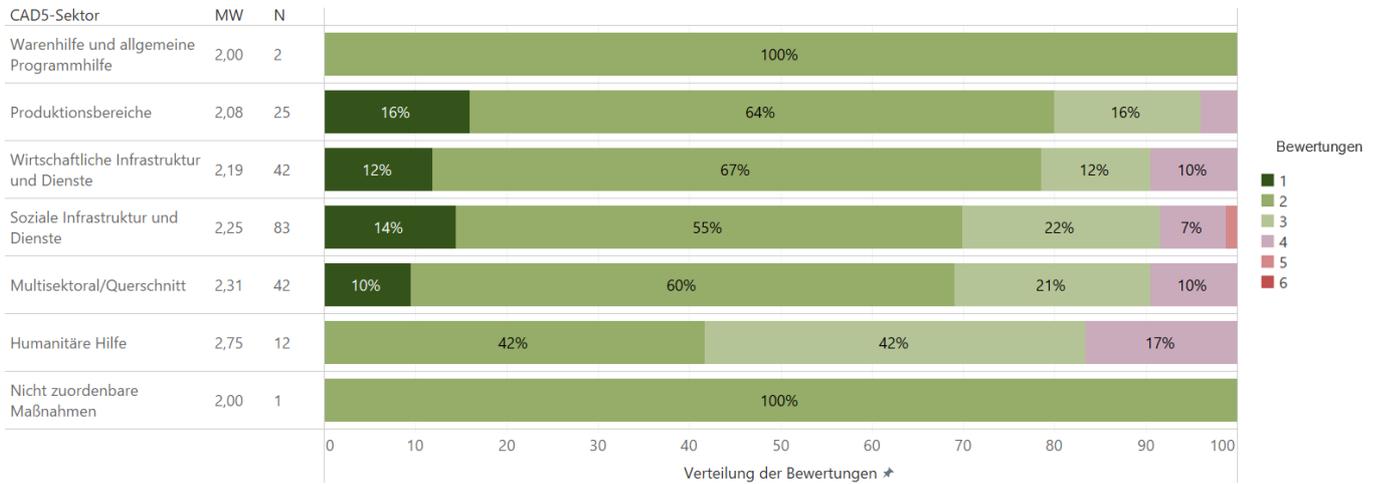
Bewertungen nach Positionierungsthemen der GIZ



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)

### Bewertungen nach Sektoren

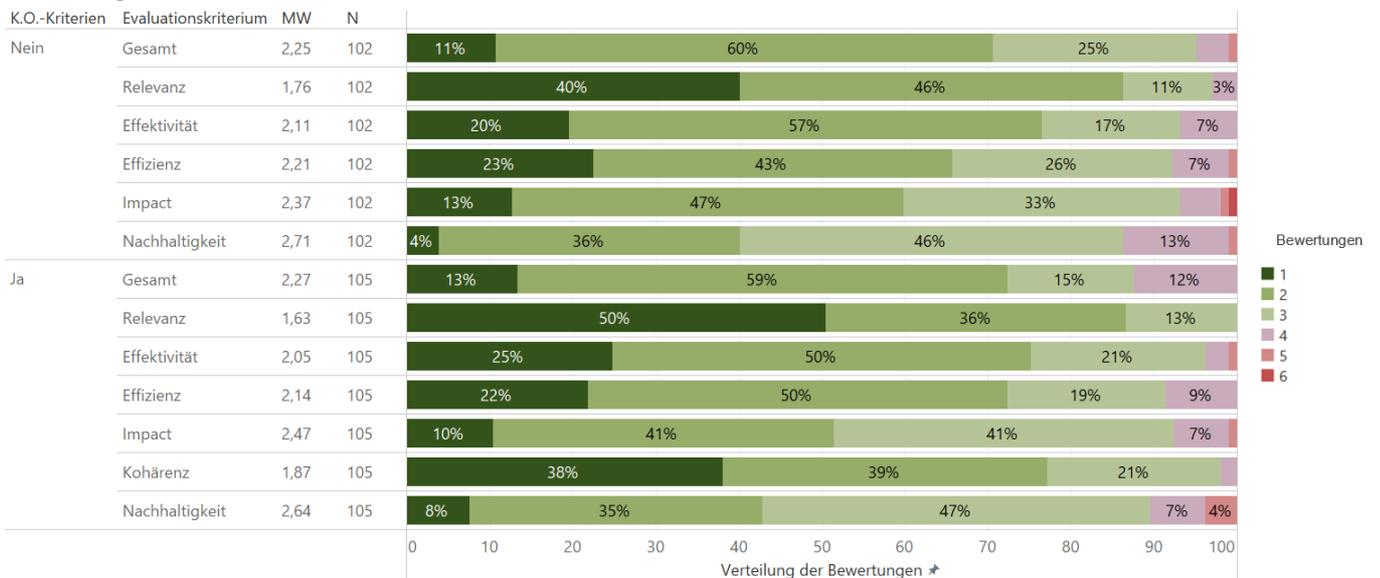


### Weitere Abbildungen zu Kapitel 3.7: Unterschiede in Evaluierungen

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)

### Bewertungen nach K.O.-Kriterien



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.

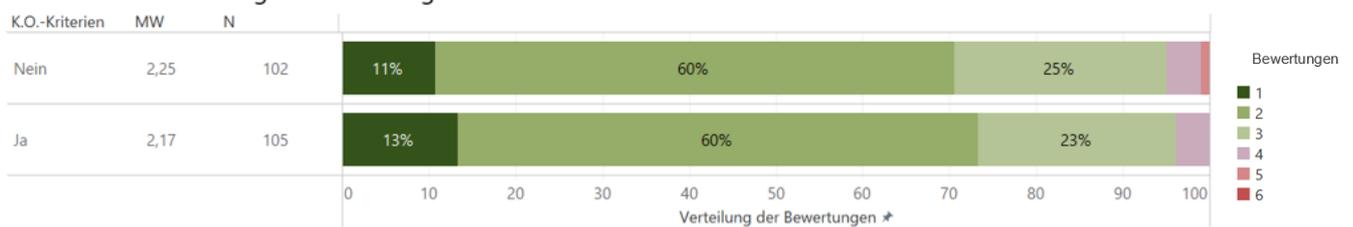
Abbildung 353: Anteil an „schlechten“ Bewertungen der Note 3 für die Evaluierungskriterien Effektivität, Impact und Nachhaltigkeit vor und nach Einführung der K.O.-Kriterien (N=207)

	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.

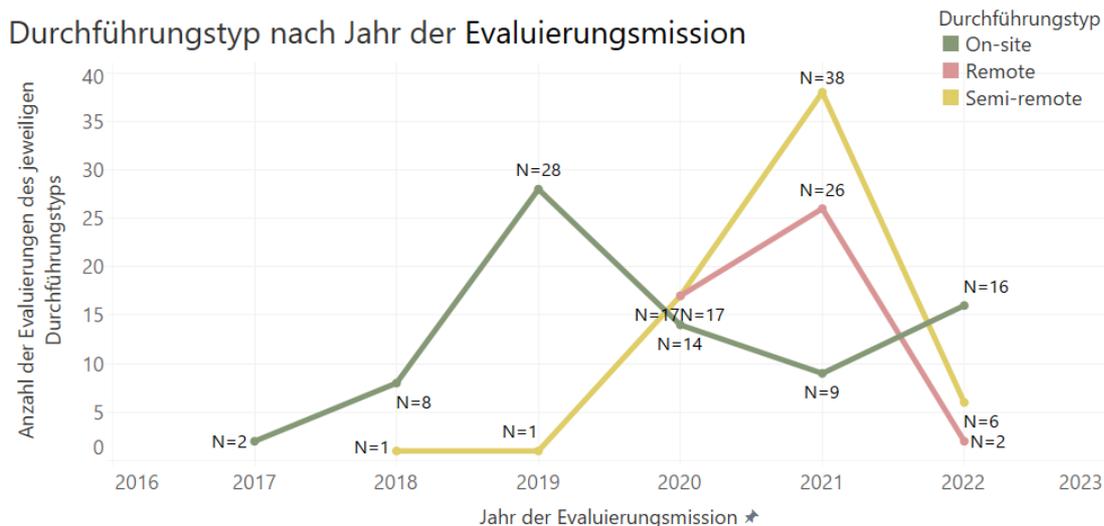
Abbildung 364: Bewertungen ohne K.O.-Kriterien (N=207)

### Theoretische Verteilung der Bewertungen K.O.-Kriterien



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 Misserfolgskfaktoren für die qualitativ untersuchten Trends

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

Trend		Erfolgsfaktoren	Misserfolgskfaktoren
<b>Projekttyp</b>	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 - Fachkompetenz des Projektteams - Engagement einzelner Projektmitarbeiter*innen - Synergien mit anderen Akteuren der deutschen EZ bzw. Akteuren der IZ	- Fehlende Ownership anderer Stakeholder (nicht Partner) - Einschränkungen durch den Projekttyp
<b>Projektvolumen</b>	Vorhaben mit kleinerem Projektvolumen (bis zu 5 Mio. EUR)	- Zusammenarbeit mit beteiligten Stakeholdern - Fachkompetenz des Projektteams - 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
<b>Operativer GIZ-Bereich</b>	Asien/Pazifik, Lateinamerika/Karibik und Überregional	- Zusammenarbeit mit beteiligten Stakeholdern - Ownership der Partner - Fachkompetenz des Projektteams - 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 - Herausforderungen bei der Steuerung des Vorhabens
<b>BMZ-Länderkategorie</b>	Globale Partner und bilaterale Partner sowie multilaterale, europäische und nicht-staatliche Partner	- Zusammenarbeit mit den beteiligten Stakeholdern - Ressourcenausstattung	- Zusammenarbeit mit beteiligten Stakeholdern
	Reformpartner, Transformationspartner sowie Nexus- und Friedenspartner	- Chancen im Partnerland	- Ownership der beteiligten Stakeholder - Herausforderungen bei der Steuerung des Vorhabens - Kontext, insbesondere Einschränkungen durch die COVID-19-Pandemie sowie die Sicherheitslage



## Bildnachweise und Quellen

### **Haftungsausschluss:**

Diese Publikation enthält Links zu externen Websites. Die Verantwortung für den Inhalt der aufgeführten externen Seiten liegt immer bei den jeweiligen Herausgebern. Bei der erstmaligen Veröffentlichung der Links zu diesen Seiten hat die GIZ den Inhalt der Drittanbieter daraufhin geprüft, ob er zu einer zivilrechtlichen oder strafrechtlichen Haftung führen könnte. Allerdings ist eine ständige Kontrolle der externen Links ohne konkrete Hinweise auf Rechtsverstöße nicht zumutbar. Kommt der GIZ zur Kenntnis oder wird sie von einem Dritten darauf hingewiesen, dass eine externe Seite, auf die sie verlinkt hat, zu einer zivilrechtlichen oder strafrechtlichen Haftung führen könnte, entfernt sie den Link zu dieser Seite unverzüglich. Die GIZ distanziert sich ausdrücklich von diesen Inhalten.

### **Kartenmaterial:**

Kartografische Darstellungen dienen nur zu Informationszwecken und stellen in keiner Weise eine völkerrechtliche Anerkennung von Grenzen und Gebieten dar. Die GIZ übernimmt keinerlei Gewähr für die Aktualität, Richtigkeit oder Vollständigkeit des bereitgestellten Kartenmaterials. Jegliche Haftung für Schäden, die direkt oder indirekt aus der Benutzung entstehen, wird ausgeschlossen.



Deutsche Gesellschaft für  
Internationale Zusammenarbeit (GIZ) GmbH

Sitz der Gesellschaft  
Bonn und Eschborn

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

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

E [info@giz.de](mailto:info@giz.de)  
I [www.giz.de](http://www.giz.de)