WIRKUNGSORIENTIERUNG UND EVALUIERBARKEIT VON EZ-PROGRAMMEN

Erfahrungen aus den Programmen zur nachhaltigen Wirtschaftsentwicklung in Ägypten, Mosambik und Myanmar

2021

DEval
DEUTSCHES EVALUIERUNGSINSTITUT DER ENTWICKLUNGS-ZUSAMMENARBEIT

Durch das Reformkonzept „BMZ 2030“ und die Gemeinsame Verfahrensreform wurden sowohl Inhalte als auch Verfahren der deutschen Entwicklungszusammenarbeit (EZ) überarbeitet. Die Steuerung der bilateralen staatlichen EZ erfolgt über den Dreiklang bestehend aus der Länderstrategie, EZ-Programmen und Modulen. EZ-Programme sollen Maßnahmen der Durchführungsorganisationen bündeln und stellen die maßgebliche Steuerungsebene für die Zusammenarbeit innerhalb eines Schwerpunkts dar.

Der vorliegende Bericht befasst sich mit der Wirkungsorientierung in der Programmkonzeption und -implementierung, der Evaluierbarkeit von EZ-Programmen sowie dem Erkenntnisinteresse und den möglichen Zeitpunkten zur Durchführung von EZ-Programmevaluierungen (EZ-PEs). Der Bericht soll einen Beitrag zur Weiterentwicklung der Formate und Verfahren sowie zur Entwicklung eines Evaluierungsformats auf Programmebene leisten.

Die Analyse zeigt, dass Verbesserungsbedarfe in der Programmkonzeption, beispielsweise der Festlegung von einheitlicheren Qualitätsstandards für die Formulierung von Programmzielen und -indikatoren, als auch in der Sicherstellung einer ausreichenden Datenbasis bestehen. Die identifizierten Defizite führen zu Einschränkungen im Monitoring und in der Berichterstattung und mindern die Evaluierbarkeit von EZ-Programmen. Die Einführung von EZ-PEs findet insgesamt viel Zuspruch und das Erkenntnisinteresse ist sehr vielfältig: Zentrale Fragestellungen betreffen die Analyse des Zielsystems und des Zusammenwirkens von Modulen.

WIRKUNGSORIENTIERUNG UND EVALUIERBARKEIT VON EZ-PROGRAMMEN

Erfahrungen aus den Programmen zur nachhaltigen Wirtschaftsentwicklung in Ägypten, Mosambik und Myanmar

2021

IMPRESSUM

Verfasst von

Miriam Amine
Judith Ihl
Michèle Kiefer
Dr. Nico Herforth

Verantwortlich

Amélie Gräfin zu Eulenburg

Gestaltung Umschlag und Grafiken

MedienMélange:Kommunikation!, Hamburg
www.medienmelange.de

Lektorat

Marcus Klein, PhD

Bildnachweis

Titelseite: VectorMine, Shutterstock

Bibliografische Angabe

Amine, M. et al. (2021), *Wirkungsorientierung und Evaluierbarkeit von EZ-Programmen. Erfahrungen aus den Programmen zur nachhaltigen Wirtschaftsentwicklung in Ägypten, Mosambik und Myanmar*, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

Druck

Bonifatius, Paderborn

© Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), 2021

ISBN 978-3-96126-148-2 (gebundene Ausgabe)
ISBN 978-3-96126-149-9 (PDF)

Herausgeber

Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval)
Fritz-Schäffer-Straße 26
53113 Bonn, Germany

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

Das Deutsche Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) ist vom Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) mandatiert, Maßnahmen der deutschen Entwicklungszusammenarbeit unabhängig und nachvollziehbar zu analysieren und zu bewerten.

Mit seinen Evaluierungen trägt das Institut dazu bei, die Entscheidungsgrundlage für eine wirksame Gestaltung des Politikfeldes zu verbessern und die Transparenz zu den Ergebnissen zu erhöhen.

Der vorliegende Bericht ist auch auf der DEval-Website als PDF-Download verfügbar unter: *https://www.deval.org/de/publikationen*.

Anfragen nach einer gebundenen Ausgabe richten Sie bitte an: *info@DEval.org*.

Eine Stellungnahme des BMZ findet sich unter: *https://www.bmz.de/de/ministerium/evaluierung/bmz-stellungnahmen-19404.*

DANKSAGUNG

Für den vorliegenden Bericht zur Wirkungsorientierung und Evaluierbarkeit von EZ-Programmen wurde das Team durch viele Personen unterstützt, bei denen wir uns herzlich bedanken möchten. Zunächst gilt unser Dank allen Mitgliedern der Referenzgruppe, deren Vertreter*innen aus dem Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ), der Deutschen Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, der KfW Entwicklungsbank, der Physikalisch-Technischen Bundesanstalt (PTB) und der Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) stammen. Die enge Zusammenarbeit und den fachlichen sowie konstruktiv-kritischen Austausch hat das Team während des gesamten Prozesses als gewinnbringend erachtet und daher sehr geschätzt.

Zudem danken wir allen Interviewpartner*innen für ihre interessanten und inhaltlich wertvollen Beiträge.

Ein besonderer Dank gilt auch Dr. Stefan Silvestrini, der den gesamten Prozess begleitet und das Team insbesondere methodisch unterstützt hat. Herzlich bedanken möchten wir uns gleichfalls bei Dr. Hans-Bernd Zöllner, der als externer Peer-Reviewer wertvolle Impulse gegeben und zur Qualitätssicherung des Berichts beigetragen hat. Dies gilt ebenso für unsere DEval-interne Peer-Reviewerin Kirsten Vorwerk.

Abschließend möchten wir uns bei Severin Flaig für seine kompetente Unterstützung über den gesamten Prozess hinweg bedanken.

ZUSAMMENFASSUNG

Hintergrund und Ziele

Mit der 2017 in Kraft getretenen **Gemeinsamen Verfahrensreform (GVR)** regelte das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) die Verfahren zur Planung und Steuerung mit den staatlichen Durchführungsorganisationen (DOs) neu. Die GVR enthält konkrete Verfahrensschritte für **die drei Ebenen (Länderstrategie [LS], EZ-Programme und Module) des Dreiklangs**, worüber die bilaterale deutsche Entwicklungszusammenarbeit (EZ) umgesetzt und gesteuert wird. Ziel ist es, die Wirksamkeit, Effizienz und Effektivität der bilateralen EZ zu verbessern und die Steuerungsfähigkeit des BMZ zu stärken (Hartmann et al., 2019; OECD DAC, 2021).

EZ-Programme bündeln die konkreten Maßnahmen (Module), stellen die Schnittstelle zwischen EZ-Programm- und Modulebene dar und operationalisieren somit die strategischen Vorgaben der LS (BRH, 2021; Hartmann et al., 2019). Das Ziel der EZ-Programme liegt darin, einen Beitrag zu übergeordneten Entwicklungsstrategien zu leisten und dadurch eine Veränderung in den Partnerländern anzustoßen (OECD DAC, 2021). Durch die GVR wird der EZ-Programmebene mehr Bedeutung beigemessen, insbesondere als Steuerungsebene. Dazu wird an der Verbesserung der EZ-Programmkonzeption und -implementierung angesetzt**.** Evaluierungen auf EZ-Programmebene werden bisher nicht durchgeführt, sind aber in der GVR vorgesehen.

Gegenstand des vorliegenden Berichts sind die **Wirkungsorientierung in der EZ-Programmkonzeption und -implementierung** sowie die **Evaluierbarkeit von EZ-Programmen**. Weitere Fragen beziehen sich auf das Erkenntnisinteresse (Ziele und Zwecke) und auf mögliche Zeitpunkte zur Durchführung einer EZ-Programmevaluierung (EZ-PE). Mit dem Bericht sollen zum einen Erkenntnisse zu Verbesserungsmöglichkeiten in der Planung und Steuerung von EZ-Programmen gewonnen und zum anderen inhaltliche und methodische Hinweise für ihre zukünftige Evaluierungen gegeben werden.

Wirkungsorientierung

Für eine verbesserte Wirkungsorientierung von EZ-Programmen ist es notwendig, dass **intendierte Wirkungen und notwendige Ressourcen bereits in der Planung** thematisiert werden. Im Sinne der Wirkungsbeobachtung beziehungsweise -messung müssen während der Projektimplementierung Fortschritte überprüft und Optimierungsbedarfe identifiziert und eingeleitet werden. Dies wird auch als **ergebnisorientiertes Managementsystem** (*results-based management*, RBM) bezeichnet (Bergmüller und Quiring, 2019; Caspari, 2012; Harten und Petrova, 2018; High Level Forum, 2005; Holzapfel, 2014; Klingebiel, 2011; OECD, 2019a; UNDG, 2011). In der bilateralen deutschen EZ ist bisher noch kein umfassendes RBM-System für die Wirkungsorientierung und -messung eingerichtet. Dies schließt EZ-Programme mit ein.

Wichtige Bestandteile einer optimalen Wirkungsorientierung und -messung sind:

1. die Definition von Zielen und Indikatoren gemäß den SMART-Kriterien (**S**[pezifisch], **M**[essbar], **A**[kzeptiert], **R**[elevant], **T**[erminiert]; Lehmann et al., 2020)
2. die Erhebung von Basiswerten und die Festlegung von Zielwerten im Zuge der EZ-Programmkonzeption sowie die Erhebung von Ist-Werten während der EZ-Programmimplementierung
3. die Überprüfung des EZ-Programmfortschritts beziehungsweise dessen Zielerreichung auf Basis einer ausreichenden Datengrundlage
4. die Partnereinbindung, zum Beispiel durch den Austausch von Daten und die gemeinsame Festlegung von Zielen (Holzapfel und Römling, 2020)

Kausale Wirkungsbeziehungen innerhalb des Zielsystems können mithilfe einer **Programmtheorie** eruiert und somit für die EZ-Programmsteuerung nutzbar gemacht werden (Caspari, 2012; Holzapfel, 2014). Ein zentrales Instrument der Wirkungsmessung ist das **Monitoring**, bei dem die Indikatoren aus der Konzeption mit aktuellen Daten bestückt werden und so der EZ-Programmfortschritt nachvollzogen werden kann (Holzapfel und Römling, 2020).

Wie sollen EZ-Programme ihre intendierten Wirkungen erreichen?

In der LS wird das übergeordnete entwicklungspolitische Ziel (langfristige Impact-Ebene) festgelegt, das durch die EZ-Programme erreicht werden soll. Die Umsetzung der EZ-Programme erfolgt jedoch ausschließlich über die Module. In der Regel werden **keine modulübergreifenden Maßnahmen** durchgeführt. Die zugrundeliegende Annahme ist somit, dass EZ-Programmziele (ausschließlich) durch das **Zusammenwirken der Module** erreicht werden. Eine wichtige Grundvoraussetzung zur Zielerreichung ist daher ein **konsistentes und realistisches Zielsystem** auf EZ-Programmebene (BMZ, 2020) unter Berücksichtigung der einzelnen Module.

Es wird zudem angenommen, dass alle Module gezielte und wesentliche Beiträge zur Erreichung des EZ-Programmziels leisten. **Module können innerhalb eines Handlungsfelds**, häufig bestehend aus einem Modul für Finanzielle Zusammenarbeit (FZ) und einem für Technische Zusammenarbeit (TZ), eine gemeinsame Wirkung entfalten. Module können auch aus mehreren **verschiedenen Handlungsfeldern** gemeinsam auf die Erreichung eines EZ-Programmziels hinwirken oder **losgelöst** auf übergeordneter Ebene (mittelfristige Impacts) gemeinsame Effekte mit anderen Handlungsfeldern entfalten.

Der **Grad des Zusammenwirkens der Module** variiert erheblich. Er kann sowohl eine grundsätzliche Koordinierung und Abstimmung in der Implementierung von Modulen zwischen den beteiligten Akteuren als auch ein enges operatives Ineinandergreifen der Inhalte und Aktivitäten sowie der Realisierung von Synergien bedeuten.

Evaluierung und Evaluierbarkeit von EZ-Programmen

Weder EZ-Programme noch LS werden gegenwärtig evaluiert. Die derzeit zur Verfügung stehenden Modulevaluierungen sowie die vorgegebenen Fragestellungen weisen kaum Bezüge zum jeweiligen EZ-Programm auf und tragen somit keine aussagekräftigen Erkenntnisse zur Relevanz und Wirksamkeit auf dieser Ebene bei. Dadurch entsteht eine erhebliche **Evidenzlücke auf EZ-Programmebene** und im Wirkungsgefüge zwischen Modul und EZ-Programm (BRH, 2021). Es fehlen wichtige Informationen zur strategischen Relevanz und Wirksamkeit der bilateralen EZ. Der Bundesrechnungshof (BRH) sowie der Entwicklungsausschuss (Development Assistance Committee, DAC) der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (Organisation for Economic Co-Operation and Development, OECD) empfehlen die Durchführung von Evaluierungen auf EZ-Programmebene mit dem **Ziel einer verbesserten Steuerung und Rechenschaftslegung** (BRH, 2021; OECD DAC, 2021).

Für die Konzeption und Anwendung eines Evaluierungsformats auf EZ-Programmebene ist eine **Beurteilung der Evaluierbarkeit** notwendig. Dabei wird vorab geprüft, inwiefern EZ-Programme die wünschenswerten Voraussetzungen für eine Evaluierung mitbringen. Dazu zählen beispielsweise die Existenz eines Monitoringsystems und eine angemessene Qualität von Indikatoren und Daten. Diese Aspekte wurden im vorliegenden Bericht mit Blick auf EZ-Programmen untersucht und somit ein Beitrag zur Gestaltung eines zukünftigen Evaluierungsformats geleistet.

Methodisches Vorgehen

Neben Vorgaben und Handreichungen zu EZ-Programmen wurden Programmdokumente der ausgewählten Fallstudien ausgewertet: drei EZ-Programme in Ägypten, Mosambik und Myanmar aus dem größten Förderbereich der bilateralen EZ „Nachhaltige Wirtschaftsentwicklung“ (NaWi). Zudem erfolgte eine Onlinebefragung von Länderreferent*innen im BMZ, und es wurden Interviews mit programmbeteiligten Akteuren sowie Fachexpert*innen und Verantwortlichen aus Grundsatzabteilungen geführt. Zur Einordnung und Validierung der Erkenntnisse diente eine Literaturauswertung.

Die Datenanalyse orientierte sich am Grounded-Theory-Forschungsansatz, sodass sich der Auswertungsprozess durch ein konstantes Wechselspiel aus Deduktion und Induktion kennzeichnete. Neue Erkenntnisse konnten während der Datenerhebung oder -analyse systematisch identifiziert und auf ihre Generalisierbarkeit hin überprüft werden.

Wesentliche Ergebnisse

In der Praxis erfüllen EZ-Programme ihren Zweck nur bedingt. Zwar werden Vorgaben der LS übernommen und Module formal unter dem Dach des EZ-Programms gebündelt, jedoch findet die strategische Planung und Steuerung in der Praxis bisher kaum über EZ-Programme statt, sondern auf Modulebene.

Erkenntnisinteresse auf EZ-Programmebene

Es besteht ein **umfangreiches und breitgefächertes Erkenntnisinteresse** an zukünftigen Evaluierungen von EZ-Programmen, sodass sich die Beteiligten in DOs und BMZ insgesamt einen großen Mehrwert davon versprechen. Als besonders wichtig wurde von den Befragten die **Analyse des Zielsystems** und **das Zusammenwirken der Module** hervorgehoben. EZ-PEs können grundsätzlich ex-ante oder ex-post beziehungsweise als Zwischen-, oder Schlussevaluierung durchgeführt werden, gegebenenfalls in Form von schlanken Review-Formaten sowie als übergeordnete strategische Evaluierung. Eine Mehrheit der Befragten hält eine **EZ-PE während der EZ-Programmdurchführung** (als Zwischenevaluierung) für sinnvoll, damit die Ergebnisse für das noch laufende EZ-Programm berücksichtigt werden können.

Für die bilaterale deutsche EZ lassen sich zwei grundsätzliche Möglichkeiten ableiten, wie EZ-Programme evaluiert werden könnten:

- als **übergeordnete strategische Evaluierung** mehrerer EZ-Programme, wobei deren Auswahl anhand geografischer oder thematischer Kriterien erfolgen kann
- als **Evaluierung einzelner EZ-Programme**, die konkrete und detaillierte Ergebnisse zu einem EZ-Programm liefern und in den Programmablauf integriert werden können

Wirkungsorientierung und Evaluierbarkeit von EZ-Programmen

Die EZ-Programmziele sollen die übergeordneten entwicklungspolitischen Ziele aus dem Strategiedokument (Impact-Ebene) reflektieren. In der Praxis werden die **EZ-Programmziele unterschiedlich konkret** formuliert. Dies liegt an ungenauen Angaben in den Vorgaben und unterschiedlichen Ansprüchen der Verantwortlichen. Die Qualitätssicherung wird von allen Beteiligten als herausfordernd beschrieben, unter anderem aufgrund des erforderlichen fachlichen und methodischen Know-hows. Zudem besteht ein sehr **heterogenes Anspruchsniveau** an das EZ-Programmziel, zum Beispiel, ob dieses grundsätzlich so formuliert sein sollte, dass es im festgelegten Zeitraum durch das Zusammenwirken der Module erreichbar ist oder als langfristiger Beitrag zu entwicklungspolitischen Veränderungen im Partnerland verstanden wird. Dies hat **Auswirkungen auf die Steuerung**, die Implementierung und das Monitoring von EZ-Programmen. Dementsprechend ist die gängige Praxis nicht ausreichend, um klare, spezifische und plausible intendierte Wirkungen zu formulieren, die zur Überprüfung im Rahmen von EZ-PE erforderlich sind.

Die **Festlegung von Indikatoren** wird in der Umsetzung aufgrund schwer zu operationalisierender Zielformulierungen und des Datenmangels als sehr herausfordernd erachtet. Die Partnerdaten, die für das Monitoring genutzt werden sollen, stehen häufig nicht rechtzeitig, in ausreichender Qualität oder im erforderlichen Aggregationsniveau zur Verfügung. Außerdem erfüllen die zum Einsatz kommenden EZ-Programmzielindikatoren nicht immer die SMART-Kriterien. In den Fallstudien sind sie zudem nicht lückenlos bestückt. Das Vorhandensein und die Qualität von Indikatoren und Daten für die Bewertung des Erreichens der intendierten Wirkungen sind **Voraussetzungen für die Evaluierbarkeit**, die derzeit nur eingeschränkt gegeben ist.

In der EZ-Programmkonzeption soll entsprechend den Vorgaben eine frühzeitige Auseinandersetzung mit den intendierten Wirkungen eines EZ-Programms (Programmtheorie) erfolgen. Doch **Wirkungsannahmen und insbesondere das Zusammenwirken zwischen den Modulen** werden derzeit nicht ausreichend und nachvollziehbar verschriftlicht. Demnach ist nicht immer evident, wie genau sich die **Wirkungen auf EZ-Programmebene** entfalten sollen**.** Insgesamt ist in der Praxis selten ein Zusammenwirken im Sinne des engen operativen Ineinandergreifens der Module vorzufinden.

In den untersuchten Fallstudien zeigt sich, dass zur Darstellung der inhaltlichen Schwerpunkte eines EZ-Programms die Strategiedokumente der Partner herangezogen werden, um die entsprechenden Bedarfe

abzuleiten und den Fokus zu identifizieren. Zudem werden die Entwicklungsherausforderungen des Partnerlandes und deren Konsequenzen für bestimmte Zielgruppen beschrieben. Jedoch ist nicht immer nachvollziehbar, ob die beabsichtigten Veränderungen für die Zielgruppen innerhalb der Laufzeit des EZ-Programms tatsächlich erreicht werden können.

Annahmen und Risiken werden vorwiegend im Hinblick auf externe, nicht durch das EZ-Programm verursachte Faktoren beschrieben. Es mangelt an einer Reflexion möglicher unbeabsichtigter sowohl positiver wie negativer Effekte (nicht intendierter Wirkungen), die das EZ-Programm für bestimmte Zielgruppen, Räume oder auch Sektoren hat. Eine angemessene **Strategie zur Risikoerkennung und -minderung als** Voraussetzung für die Evaluierbarkeit wird somit bislang eingeschränkt berücksichtigt.

Die wesentlichen **Verantwortlichkeiten** in den Hauptschritten der EZ-Programmkonzeption und -implementierung sind festgelegt. Für die Umsetzung einzelner Prozessschritte ist zum Teil **weitere Klärung erforderlich,** da Unklarheiten zu Unstimmigkeiten und teilweise auch Qualitätsverlusten führen. Dies betrifft sowohl die Rollenklärung zwischen BMZ und DO als auch zwischen den DOs und kommt insbesondere in der operativen EZ-Programmsteuerung und Qualitätssicherung zum Tragen.

Ein enger **Austausch mit Partnern** findet über die Module statt. Langfristige Zielsetzungen und Zusagen werden zudem in Regierungsverhandlungen und -konsultationen mit der Partnerregierung besprochen. EZ-Programme stellen jedoch lediglich eine **interne Planungs- und Steuerungsebene** dar und werden in der Regel nicht mit den Akteuren der Partnerländer diskutiert.

Wie auch in der EZ-Programmkonzeption findet während der EZ-Programmimplementierung und in den dazugehörigen Berichterstattungen **keine angemessene Auseinandersetzung** mit der **Überprüfung von Wirkungsannahmen** (insbesondere zur gemeinsamen Wirkungsentfaltung der Module) statt.

In der Berichterstattung fehlt es häufig sowohl an **Primär- wie an Sekundärdaten.** Eigene Erhebungen durch das EZ-Programm finden kaum statt, da keine expliziten Ressourcen dafür vorgesehen sind. In den drei Fallstudien wird in unterschiedlichem Maße auf Sekundärdaten (Studien und Berichte) zurückgegriffen. Diese stammen von Partnerinstitutionen, deren Daten teilweise nicht planmäßig oder nur in unzureichender Qualität überlassen werden, was zu **Datenlücken** führt. Das Monitoring auf EZ-Programmebene reicht in den betrachteten Fallstudien nicht aus, um aktuelle Informationen zu den EZ-Programmzielindikatoren bereitzustellen. Die **unzureichende Datenverfügbarkeit** stellt eine enorme **Herausforderung für die Messung** des EZ-Programmfortschritts dar. Die Datenlücken als auch die bereits beschriebenen Qualitätsunterschiede in den Ziel- und Indikatorenformulierungen führen dazu, dass die Voraussetzung des **Monitorings** für die Evaluierbarkeit derzeit nur selten gegeben ist.

In der Praxis enthalten die EZ-Programmdokumente der Implementierungsphase Informationen zu den aktuellen **Einflussfaktoren** der EZ-Programmzielerreichung. Zudem konnten in allen Fällen auch **interne Herausforderungen** identifiziert werden, zum Beispiel durch (zeitliche) Verzögerungen in einzelnen Modulen oder unterschiedlichen Prozessen in der FZ und der TZ. Diese internen Faktoren können die Zielerreichung zum Teil erheblich beeinflussen, werden aber in aktuellen Formaten zum Austausch und zur Berichtslegung zu wenig berücksichtigt.

Die identifizierten Defizite in der EZ-Programmkonzeption setzen sich in der Implementierung fort. Damit werden sowohl die Monitoringaktivitäten als auch die Bewertung des EZ-Programmfortschritts im Programmverlauf erschwert, und die Steuerungsfähigkeit des BMZ wird beeinträchtigt.

Empfehlungen

Ziele und Mehrwert von EZ-Programmen

Empfehlung 1: Das BMZ sollte Ziele und Zwecke der EZ-Programme und die damit verbundenen Steuerungsaufgaben eindeutiger definieren, um EZ-Programme insgesamt für alle Beteiligten sichtbarer zu machen und die Voraussetzungen für die strategische, ergebnisorientierte Steuerung über Schwerpunkte (zukünftig Kernthemen und Aktionsfelder) zu verbessern.

Umsetzungshinweis 1.1: Das BMZ sollte in Absprache mit den DOs prüfen, ob Ziele und Zwecke sowie Kohärenzansprüche von EZ-Programmen in bestimmten fragilen Kontexten oder Aktionsfeldern unterschiedlich definiert sein sollten.

Umsetzungshinweis 1.2: Es sollte nachvollziehbarer dargestellt werden, inwiefern Partnerinteressen berücksichtigt werden – insbesondere, wenn die DOs mit verschiedenen Partnerinstitutionen zusammenarbeiten, die möglicherweise unterschiedliche Interessen und Prioritäten verfolgen. Die Beiträge der Partner sollten klarer beschrieben werden, da sie für die erfolgreiche Umsetzung von EZ-Programmen wichtige Akteure sind.

Wirkungsorientierung und Evaluierbarkeit von EZ-Programmen

Empfehlung 2: Die Wirkungsorientierung und die Evaluierbarkeit von EZ-Programmen sollten durch das BMZ und die DOs verbessert werden, um dem BMZ steuerungsrelevante Informationen liefern zu können und eine aussagekräftige Überprüfung und Evaluierung von EZ-Programmen zu ermöglichen.

Die Verbesserungsbedarfe bestehen insbesondere in der EZ-Programmkonzeption und betreffen:

- **das Zielsystem von EZ-Programmen auf den verschiedenen Wirkungsebenen**
- **eine angemessene und nachvollziehbare Darstellung des angestrebten Zusammenwirkens und der Wechselwirkungen zwischen Modulen**
- **die Definition angemessener EZ-Programmzielindikatoren oder gegebenenfalls -zielgrößen**
- **die Verbesserung der Datenbasis**

Umsetzungshinweis 2.1: Um Aspekte zur Wirkungsorientierung und Evaluierbarkeit in der Planung und Steuerung zukünftig systematisch zu gewährleisten, sollten in der Erarbeitung entsprechender Prozesse und Verfahren die Evaluierungseinheiten des BMZ und der DOs miteinbezogen werden.

Auf die identifizierten Schwachstellen und Verbesserungsmöglichkeiten der einzelnen aufgeführten Punkte wird in den Empfehlungen 3 bis 5 näher eingegangen.

Heterogene Programmzielformulierungen

Empfehlung 3: Das BMZ sollte die Vorgaben und Qualitätsstandards für die Formulierung von EZ-Programmzielen weiter verbessern und die Umsetzung der Vorgaben sicherstellen.

Umsetzungshinweis 3.1: Die Ziele sollten als langfristiger Beitrag zu entwicklungspolitischen Veränderungen (langfristige Impact-Ebene) oder konkret und messbar für einen klar definierten Zeitraum (eher Outcome- oder mittelfristige Impact-Ebene) formuliert werden. Es sollte möglichst ein konkretes und messbares Ziel festgelegt werden. Falls dies aufgrund der Rahmenbedingungen nicht umsetzbar ist, sollte in der Programmkonzeption eine Begründung erfolgen.

Umsetzungshinweis 3.2: Wenn Module inhaltlich nicht ausreichend in ein Programm integriert werden können und keine messbaren Beiträge zu Programmziel- und Programmzielindikatoren leisten, sollten diese auch nicht als Teil des Programms dargestellt werden.

Umsetzungshinweis 3.3: Die Handreichungen und Verfahrensinformationen sollten hier klarere Vorgaben machen. Derzeit werden Fragen zur Messbarkeit und Konkretheit der Ziele erst in der EZ-Programmkonzeption thematisiert und je nach Ermessen der Beteiligten diskutiert.

Empfehlung 4: Vor Inkrafttreten der EZ-Programme sollte das BMZ die Qualitätssicherung der formulierten Ziele und Indikatoren durch ausreichend methodisch qualifizierte Ansprechpartner*innen gewährleisten. Die Qualitätssicherung der Ziele sollte mit der Prüfung der Indikatoren einhergehen und entweder durch das BMZ oder die DOs erfolgen.

Umsetzungshinweis 4.1: Für die Qualitätssicherung sollten konkrete personelle Zuständigkeiten sowie finanzielle und zeitliche Ressourcen festgelegt und bereitgestellt werden.

Umsetzungshinweis 4.2: Damit die Qualitätssicherung der Ziele und Indikatoren über die DOs gewährleistet werden kann, sollten diese auch stärker in die Festlegung der EZ-Programmziele miteinbezogen werden. Zurzeit erfolgt die Zielformulierung häufig auf Ebene der LS mit wenig Einbindung der DO.

Verbesserung der Datenbasis

Empfehlung 5: Es sollte ein angemessenes Monitoring auf Programmebene etabliert werden. Die DOs sollten zum Zeitpunkt der Programmkonzeption verdeutlichen, inwiefern erforderliche Daten für das Monitoring durch Partner, Sekundärliteratur und die Module zur Verfügung stehen werden und welche zusätzlichen Erhebungen auf Programmebene erforderlich sind. Das BMZ sollte entsprechende Ressourcen bereitstellen.

Umsetzungshinweis 5.1: Während der EZ-Programmkonzeption und der Finalisierung von Zielformulierungen sollten folgende Maßnahmen umgesetzt:

- BMZ und DOs sollten sich darauf einigen, welche Daten zum Monitoring des EZ-Programmfortschritts erforderlich sind. Auf dieser Basis sollte durch die DOs ein angemessenes Monitoringsystem erarbeitet werden.
- Es sollte, sofern möglich, auf Partnerdaten zurückgegriffen werden.
- Schließlich sollten, falls notwendig, ergänzende und gezielte Erhebungen zur Ermittlung von Baseline- und Monitoringdaten erfolgen. Benötigte Ressourcen sollten durch das BMZ zur Verfügung gestellt werden. Hierfür ist es notwendig, dass zum Zeitpunkt der EZ-Programmkonzeption durch die DOs die erforderlichen Mittel eingeplant werden.

Umsetzungshinweis 5.2: In den DOs sollte für jedes EZ-Programm qualifiziertes und mit ausreichenden Mitteln ausgestattetes Personal mit dem Monitoring betraut werden.

Erkenntnisinteresse und Nutzen von EZ-Programmevaluierungen

Empfehlung 6: Um einen größtmöglichen Mehrwert und Nutzen von EZ-PEs gewährleisten zu können, müssen in der Konzeptionsphase des Evaluierungsformats folgende Fragen durch das BMZ geklärt werden:

- **Wer soll die Ergebnisse von EZ-PEs nutzen?**
- **Wofür sollen die Ergebnisse verwendet werden (Rechenschaftslegung, [Um-]Steuerung von laufenden EZ-Programmen, programmübergreifendes Lernen)?**
- **Wann und in welcher Frequenz sollen EZ-PEs durchgeführt werden, um den vorgesehenen Zweck zu erfüllen (für jedes EZ-Programm, zufällige oder kriterienbasierte Stichprobe)?**
- **Wie können Partner in die Evaluierungen einbezogen und in welcher Form Ergebnisse mit ihnen geteilt werden?**
- **Welche der genannten Themen und Fragestellungen sollen durch EZ-PEs beantwortet werden und welche nicht? Gibt es andere Formate, um die nicht berücksichtigten Fragen beantworten zu können**?

Umsetzungshinweis 6.1: Je nach Erkenntnisinteresse müssten EZ-PEs zu unterschiedlichen Zeitpunkten und für unterschiedliche Adressatenkreise durchgeführt werden. Damit Ergebnisse von EZ-PEs von Programmverantwortlichen aus BMZ und DOs in der Planung und Steuerung berücksichtigt werden können, müssten die Evaluierungen möglichst während der Laufzeit des EZ-Programms entweder für die vorgesehene Reflexion nach drei Jahren Programmlaufzeit oder für die Planung eines Folgeprogramms durchgeführt und abgeschlossen werden (abhängig von Programmdauer und -umfang als Zwischenevaluierung oder -review). Alternativ oder ergänzend wären strategische EZ-PEs möglich, in denen mehrere EZ-Programme anhand übergeordneter Fragestellungen untersucht werden. Dieses Evaluierungsformat würde das programmübergreifende Lernen in den Vordergrund stellen.

Umsetzungshinweis 6.2: Mit Blick auf die Sequenzierung von EZ-PEs sollte im Fall von Zwischenevaluierungen geklärt werden, ob Evaluierungen grundsätzlich als Informationsquelle zur Steuerung und Reflexion von Programmen eingesetzt werden sollten und somit jedes EZ-Programm innerhalb eines Programmzyklus evaluiert werden sollte. Zum Zweck der Rechenschaftslegung sowie zur Beantwortung von übergeordneten und strategischen Fragestellungen wäre eine gezielte Stichprobe möglich. Eine andere Option wären geografische und thematische Cluster.

Umsetzungshinweis 6.3: Das Evaluierungsformat sollte regelmäßig überprüft und weiterentwickelt werden. Hier sollte insbesondere reflektiert werden, ob es zukünftig weitere Fragen zu beantworten gilt oder zusätzliche Datenerhebungen innerhalb von EZ-Programmevaluierungen reduziert werden können, falls erforderliche Daten dann in ausreichendem Maße durch das Monitoring und/oder die Berichterstattungen bereitgestellt werden.

EXECUTIVE SUMMARY

Background and objectives

The **Joint Procedural Reform (Gemeinsame Verfahrensreform, GVR)** enacted by the Federal Ministry for Economic Cooperation and Development (BMZ) in 2017 redefined the procedures to be followed by the Ministry and its governmental implementing organisations[1] in their German development cooperation (GDC) activities. The GVR provides specific procedural steps for use by GDC actors at the **three triad levels (country strategy, programmes and modules)** for planning, implementation and steering purposes. The aim is to enhance the impact, efficiency and effectiveness of bilateral development cooperation while strengthening the BMZ's steering capacity (Hartmann et al., 2019; OECD DAC, 2021)[2].

GDC programmes are supposed to pool the specific measures (modules) managed by the implementing organisations. By forming an interface between the programme and module levels, they operationalise the BMZ strategic objectives set out in the country strategy (BRH, 2021; Hartmann et al., 2019). The objective of GDC programmes is to contribute to high-level development strategies to initiate change in partner countries (OECD DAC, 2021). The GVR increases the importance attached to the programme level, especially for steering purposes, and is intended to improve GDC programme design and implementation. Although the GVR stipulated programme-level evaluations, they have not yet been conducted.

This report examines whether and to what degree **GDC programmes are oriented towards results. Special attention is given to GDC programme design, reporting and evaluability**. Further matters of interest include the specific demand for and the purpose of future GDC programme evaluations and the most suitable timing for such evaluations. With the insights generated, this report aims to contribute to improved GDC programme planning and steering while providing content-based and methodological guidance for the GDC programme evaluations to come.

Results orientation

To ensure that GDC programmes are adequately oriented towards results, the **intended results and necessary resources** need to be assessed from the planning and design phase onwards. To address those aspects, proper procedures for verifying implementation progress, identifying where improvement is necessary and initiating corresponding optimisation measures throughout the implementation process are necessary. Such an approach can also be referred to as **results-based management** (RBM) (Bergmüller and Quiring, 2019; Caspari, 2012; Harten and Petrova, 2018; High Level Forum, 2005; Holzapfel, 2014; Klingebiel, 2011; OECD, 2019; UNDG, 2011). No comprehensive RBM system has yet been established in bilateral GDC, including within GDC programmes.

The following aspects are key components for an effective orientation towards results:

1. **Definition of objectives and indicators** in accordance with the SMART criteria (**S**[pecific], **M**[easurable], **A**[chievable], **R**[elevant] and **T**[ime-bound]; Lehmann et al., 2020)
2. **Identification of baseline and target values** as part of programme design, with regular **monitoring** of current values during programme implementation
3. Sufficient data available for verification of programme progress and achievement of objectives
4. Close cooperation with development partners by, for example, exchanging data and working together to define objectives (Holzapfel and Römling, 2020)

[1] The governmental implementing organisations are the Federal Institute for Geosciences and Natural Resources (BGR), Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), KfW Development Bank (KfW), the Physikalisch-Technische Bundesanstalt (PTB) and Engagement Global. For further information on these implementing organisations, see the evaluation guidelines published by the BMZ: *https://www.bmz.de/resource/blob/92884/08507d1204d093141b5f00bf5cbb8db7/bmz-leitlinien-evaluierung-2021.pdf* .

Cause–effect relationships can be illustrated via a **programme theory**. As programme theory allows for the systematic definition and presentation of objectives at all impact levels, it also serves as a programme management tool (Caspari, 2012; Holzapfel, 2014). Subsequent **monitoring activities** are essential for measuring progress and should provide current data (such as current values) for each programme indicator defined during the design phase.

How should the intended results of GDC programmes be achieved?

The country strategy defines the high-level development policy objective (long-term impact level) that GDC programmes are intended to achieve. However, GDC programme implementation is conducted entirely through individual modules without any additional or **overarching measures at the programme level**. Hence, the underlying assumption is that programme objectives are achieved (exclusively) through outcomes of and **interactions between its modules**. A key prerequisite for achieving objectives is therefore a **consistent and realistic definition of those objectives** at the programme level (BMZ, 2020) that takes into account the individual modules.

It is assumed that every individual module makes a specific and essential contribution to the programme objectives. **Modules may have a joint effect within a certain field of action** – which often includes a module for financial cooperation and one for technical cooperation – or collectively contribute to the achievement of a programme objective common to **several different fields of action**. Some modules may initially also have effects (and results) **in isolation from other modules** before developing joint effects with other fields of action at a higher level (medium-term impacts).

The **degree of interaction can vary greatly between modules**, which are usually implemented and steered by different implementing organisations, from logistical coordination, through joint planning to in-depth intertwining of activities to maximise synergy effects.

Evaluation and evaluability of GDC programmes

Neither GDC programmes nor country strategies are currently subject to systematic and regular evaluation. The module evaluations currently available contain only a few specific questions and little reference to the respective programme. In other words, module evaluations do not yield any significant findings regarding GDC programme relevance and effectiveness. A considerable **evidence gap** results, both **at the programme level** and **between the module and programme levels** (BRH, 2021), and important information is missing regarding the strategic relevance and effectiveness of bilateral development cooperation. Therefore, and with a view to **improving the steering and accountability** of GDC programmes, the Federal Court of Auditors (Bundesrechnungshof, BRH) and the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD) both recommend the introduction of evaluations at the programme level (BRH, 2021; OECD DAC, 2021).

To design and apply an evaluation format at the programme level, it is necessary to first **assess its evaluability**, which means examining whether GDC programmes meet the requirements for an evaluation. The relevant aspects should be consistent with the key components for effective results orientation, as described above, and include performance indicators, a monitoring system and the availability of data that is both adequate and reliable. This report examines the evaluability of programmes and is aimed at providing relevant information for anyone piloting and introducing a standardised evaluation format for GDC programmes.

Methodological procedure

The DEval evaluation team assessed the existing BMZ guidelines and specific programme documents for the three case studies selected: Egypt, Mozambique and Myanmar. Those GDC programmes were chosen for their relevance to the main focal area of German bilateral development cooperation – sustainable economic development. An online survey of BMZ country officers was also conducted and interviews were held with BMZ and implementation organisation staff involved in or in charge of the case study programmes,

along with experts and officers from other relevant departments, especially evaluation units. A literature review helped in the classification and validation of the findings.

The data analysis followed the grounded theory research approach, hence the assessment process was characterised by constant interplay between deduction and induction. New findings were systematically reviewed during data collection and analysis to determine the generalisability of the results.

Essential results

GDC programmes only partially fulfil their purpose as set out in the GVR. GDC programmes serve as an envelope encompassing the high-level objectives of country strategies and the operational activities of programme modules. However, strategic planning and steering is currently limited at the programme level, with most such activities being conducted at the module level.

Demand for evidence at the GDC-programme level

The demand for programme evaluation in general and the resulting evidence specifically regarding programme progress is high and covers a wide range of topics.

Actors at the BMZ and the implementation organisations expect substantial benefits from conducting such evaluations. It is considered particularly important to **learn more about the definition of objectives at the programme level and the interconnections between modules**. GDC programme evaluations can be performed at various points in time, for instance: ex ante, mid-programme or at the end of a programme (final or ex-post evaluations). They could either be performed as reviews, following a streamlined format, or as fully fledged strategic evaluations. The majority of survey respondents and interviewees prefer **evaluations during programme implementation** (as mid-programme evaluations) to allow the evaluation results to feed into the implementation of the rest of the programme.

Two main options exist for the evaluation of GDC programmes:

- **Higher-level strategic evaluations**: dealing with overarching strategic thematic or regional questions based on multiple programmes to generate evidence for strategic decision-making, institutional learning and accountability
- **Evaluations of individual programmes**: focusing on one programme and aimed at providing specific results at the operational level, thus informing programme management and steering while improving accountability

Results orientation and evaluability of GDC programmes

The intended results at the programme level should reflect the development policy objectives stipulated in the country strategy document (impact level). In practice, **some variation occurs in the way and the degree of detail to which programme objectives are defined**. That is due to imprecise specifications and differing standards expected by those responsible. Quality assurance is described as challenging due to the necessary technical and methodical expertise. Moreover, **programme objective expectations are very heterogeneous**. For instance, disagreements arise over whether objectives should generally be formulated in such a way that they can be achieved through interaction between the modules within a specified period or whether they should be understood and defined as a long-term contribution to development policy changes in the partner country. However, an **aligned understanding of GDC programme objectives** is crucial for **monitoring and steering activities**. The current practice of formulating intended results is thus found to be insufficient for adequately evaluating GDC programme objectives.

The **definition of indicators** is regarded as a very challenging process, partly due to the previously described weaknesses in defining objectives and partly because of a lack of data. Data for measuring indicators is generally to be provided by the development partners in each country. However, such data is often not available when needed for reporting, at least not at the required aggregation or quality levels. Furthermore, the programme objective indicators used do not always meet the SMART criteria. The case studies

also show that the necessary data is not provided consistently, thus confirming the finding that the **evaluability of the GDC programmes** reviewed varies and is only partially ensured.

According to the guidelines, programme theory specifies the intended results of a GDC programme and should be developed early in the design phase. However, there is currently insufficient documentation regarding the **impact assumptions and, especially, the foreseen intertwining of modules (interconnections)**. Consequently, it is not always apparent how exactly the modules should take **effect at the programme level**. Overall, barely any interaction in the sense of close operational intertwining of modules is found in practice.

Moreover, the three case studies reviewed show that the partner strategy documents are used to define the thematic priorities and focus of GDC programmes. The development challenges in the partner country and the related consequences for certain target groups are also described. However, it is not always clear whether the intended changes for the target groups are realistic and achievable within the foreseen programme duration.

Although external factor **assumptions and risks** are generally highlighted and discussed, less information is available on the potential unintended (positive or negative) effects of GDC programmes on, for example, the target groups. Consequently, it is hard to find information on anticipated **strategies and measures for detecting and reducing risks** during programme implementation. As a result, it is also hard to assess and evaluate those aspects in retrospect.

Regarding organisational structures, the essential steps and **responsibilities within GDC programme design and implementation** are set out in various guidelines. However, as any unclear aspects can lead to inconsistency and sometimes even a loss of quality, **further clarification and specification is needed** in some places for the implementation of individual procedural steps. That applies to the clarification of roles not only between the BMZ and its implementing organisations but also among the implementing organisations themselves. That is particularly important for improving operational programme management and quality assurance.

Throughout the whole GDC programme planning and implementation process, close **exchange takes place with various implementing partners** at the module level. Long-term GDC objectives and commitments are discussed with the government in a respective partner country as part of regular government negotiations and consultations. However, GDC programmes merely serve as **internal planning and steering tools**; they are not usually subject to discussions with actors in the partner countries.

Much in the same way as in the programme design phase, programme implementation and **corresponding reporting** grapples with the **verification of impact assumptions** (especially in terms of detecting the joint impact achieved by modules).

Programme reporting activities also frequently suffer from a lack of both **primary and secondary data**. A GDC programme collects little data itself, as no resources are specifically assigned to that task. The three case studies draw on secondary data (studies and reports), to varying degrees. Issues appear with the provision of data by partner institutions and consequently cause **data gaps**, which make the monitoring of indicators and the overall reporting of programme progress difficult. Generally, only limited exchange occurs with partner organisations in the given countries regarding the necessary data, its quality and when it is needed. In the case studies reviewed, programme-level monitoring is insufficient for providing up-to-date information on programme objective indicators. In other words, the **lack of data** presents a huge **challenge for measuring** the progress of GDC programmes. Given the previously mentioned weakness regarding the defining of objectives and indicators, the prerequisite for effective programme **monitoring** is rarely met under the current procedural approaches.

Although GDC programme reports in the implementation phase contain information on **factors influencing** the achievement of programme objectives, **internal risks or challenges** were identified in all cases. One practical example of that is provided by the implementation of individual modules relating to either financial or technical cooperation measures. Despite being planned to run simultaneously, those modules are

ruled by different procedures and timelines, thus they generate the risk of implementation gaps or delays. Such internal risk factors are not systematically or sufficiently displayed in current reporting formats.

The deficiencies identified in GDC programme design continue to pose implementation issues, especially in terms of monitoring and reporting. Assessing the progress of GDC programmes over the course of their implementation is thus a huge challenge, which can have detrimental effects on the BMZ's steering capability.

Recommendations

GDC programme objectives and added value

Recommendation 1: The BMZ should more clearly define the objectives and purposes of programmes and their associated steering tasks to enhance transparency and clarity for everyone involved and to increase programme benefits for strategic, results-oriented steering regarding thematic priorities.

Implementation guidance 1.1: The BMZ should work in coordination with the implementing organisations to examine whether the requirements and standards for defining objectives, purposes and coherence should vary depending on the context or fields of action.

Implementation guidance 1.2: More information should be provided on how partner interests are accounted for, especially when the implementing organisations work with various partner institutions potentially pursuing different interests and priorities. Partners' contributions should be described more clearly to better reflect their status as key actors for the successful implementation of GDC programmes.

Programme results orientation and evaluability

Recommendation 2: The BMZ and implementing organisations should improve the results orientation and evaluability of programmes to provide the BMZ with steering-relevant information and allow for sound verification and evaluation of GDC programmes.

Improvements are needed in the following areas, especially in the design phase:

- **Coherent definition of GDC programme objectives at the various impact levels**
- **Appropriate and transparent presentation of the envisaged intertwining and interconnections between modules**
- **Definition of suitable programme objective indicators or, where appropriate, target values**
- **Data availability**

Implementation guidance 2.1: To ensure the systematic embedding of results orientation and evaluability in future planning and steering, BMZ and implementing organisation evaluation units should be involved in developing the corresponding processes and procedures.

Recommendations 3 to 5 expand on the identified weaknesses and possible improvements regarding the individual points listed above.

Heterogeneous formulation of GDC programme objectives

Recommendation 3: The BMZ should further improve the guidelines and quality standards for formulating GDC programme objectives and ensure that those guidelines are implemented.

Implementation guidance 3.1: The objectives should be formulated as a long-term contribution to development policy changes (long-term impact level) or specifically and measurably for a clearly defined timeframe (rather than at the outcome or medium-term impact level). When feasible, specific and measurable objectives should be formulated. If any hindering framework conditions require adjustments, the rationale should be discussed and specified in the planning documents of the given GDC programme.

Implementation guidance 3.2: If the content of modules cannot be sufficiently integrated into a GDC programme and they do not make any measurable contributions to the programme objective or programme objective indicators, those modules should not be presented as part of that GDC programme.

Implementation guidance 3.3: The guidelines and procedural information should provide clearer guidelines prior to programme planning. Questions regarding whether the objectives are measurable and specific are currently only addressed during programme design and at the discretion of those involved.

Recommendation 4: Before GDC programmes are launched, the BMZ should assign contact people with sufficient methodical skills to ensure the quality of the formulated objectives and indicators. Quality assurance for the objectives should also include the revision of the indicators and be performed by either the BMZ or the implementing organisations.

Implementation guidance 4.1: Specific responsibilities, financial resources and time requirements should be defined and made available for quality assurance purposes.

Implementation guidance 4.2: If implementing organisations are to perform the quality assurance of objectives and indicators, they should also be more involved in defining the GDC programme objectives. Objectives are currently often formulated at the country-strategy level, with little involvement of the implementing organisations.

Improved data availability

Recommendation 5: Suitable monitoring should be established at the programme level. The implementing organisations should clearly indicate during GDC programme design to what extent the data required for monitoring is available from the partners, secondary literature and modules, and what additional data needs to be collected at the programme level. The BMZ should provide corresponding resources.

Implementation guidance 5.1: During programme design and finalisation of objectives, the following measures should be implemented:

- The BMZ and implementing organisations should agree on what data is required for monitoring GDC programme progress. On that basis, the implementing organisations should develop a suitable monitoring system
- Partner data should be used wherever possible
- Supplementary targeted data collection measures should be adopted to determine baseline and monitoring data, where necessary. The BMZ should provide the required resources. It is important that the implementing organisations provide for the necessary resources during GDC programme design

Implementation guidance 5.2: The implementing organisations should assign qualified staff with sufficient resources for the monitoring of each GDC programme.

Demand for evidence and benefits of GDC programme evaluations

Recommendation 6: To ensure maximum added value and the most benefit from GDC programme evaluations, the BMZ needs to clarify the following issues when planning and designing the evaluation format:

- **Who will use the GDC programme evaluation results?**
- **What should the results be used for (accountability, steering/redirecting ongoing programmes, learning across different programmes)?**

- **When and how frequently should GDC programme evaluations be conducted to fulfil the intended purpose (for each programme; a random or a criteria-based sample)?**
- **How can partners be involved in the evaluations and in what form can results be shared with them?**
- **Which of the specified issues and questions should GDC programme evaluations address and which should they omit? Can other formats be used to answer those omitted questions?**

Implementation guidance 6.1: Depending on the demand for evidence, GDC programme evaluations need to be conducted at various times and for different target groups. If the results are to be used by programme managers from the BMZ and the implementing organisations for planning and steering, the evaluations need to be conducted and completed during programme delivery, wherever possible. A suitable time would be during the regular reflection phase after three years of programme implementation or a little later if the results are to be used to inform the planning process for a follow-on programme. In that case, GDC programme evaluations would be conducted as mid-programme evaluations or reviews, depending on the programme duration and scope. Strategic programme evaluations investigating multiple programmes in relation to overarching issues could be performed as an alternative or in addition. Such evaluations would place the emphasis on learning across different programmes.

Implementation guidance 6.2: With a view to the sequencing of programme evaluations, it is necessary not only to clarify whether evaluations should generally be used as a source of information for steering and reflecting on programmes but also whether a mid-programme evaluation should be conducted as part of every GDC programme. A targeted sample could be evaluated for accountability purposes and to answer higher-level and strategic questions. Using geographical and thematic clusters would be another option.

Implementation guidance 6.3: The GDC programme evaluation format should be subject to regular review and enhancement. Particular reflection is needed on whether further questions can be answered in future or whether additional data collection within programme evaluations can be reduced if the necessary data is made sufficiently available through improved monitoring and/or reporting.

INHALT

Impressum iv
Danksagung v
Zusammenfassung vi
Executive Summary xiii
Inhalt xx
Abkürzungen und Akronyme xxii

1. Einleitung 1
2. EZ-Programme 3
3. Wirkungsorientierung 7
4. Evaluierung von EZ-Programmen 11
5. Methodik 14
6. Erkenntnisinteresse und Evaluierungsansätze 17
   6.1 Erkenntnisinteresse auf EZ-Programmebene in der deutschen EZ 18
   6.2 Evaluierungspraktiken anderer Geber auf Länderebene 20
7. Erkenntnisse zur Wirkungsorientierung und Evaluierbarkeit von EZ-Programmen 24
   7.1 Wirkungsorientierung 25
      7.1.1 EZ-Programmkonzeption 25
      7.1.2 EZ-Programmimplementierung 29
   7.2 Beurteilung der Evaluierbarkeit 30
8. Schlussfolgerungen und Empfehlungen 32
9. Literatur 38
10. Anhang 42
   10.1 Überblick zu den untersuchten Programmen 43
   10.2 Evaluierungspraktiken anderer Geber auf Länderebene 43
   10.3 Bestückung der EZ-Programmzielindikatoren 45
   10.4 Voraussetzungen für die EZ-Programmbeurteilung 46
   10.5 Team und Mitwirkende 47
   10.6 Zeitplan 47

Abbildungen

Abbildung 1 Bestandteile und Wirkungsebenen eines EZ-Programms 5
Abbildung 2 Wirkungsannahmen von EZ-Programmen 10
Abbildung 3 Datenerhebungs- und Datenanalysemethoden 15
Abbildung 4 Mögliche Zeitpunkte einer EZ-Programmevaluierung 19
Abbildung 5 Erkenntnisinteresse von EZ-Programmevaluierungen nach Evaluierungstypen 20
Abbildung 6 Evaluierungsmöglichkeiten von EZ-Programmen 22
Abbildung 7 Grade des Zusammenwirkens der Module in der Praxis 27
Abbildung 8 Grundlagen zur Wirkungsmessung nach EZ-Programmphase 34
Abbildung 9 Voraussetzungen für die EZ-Programmbeurteilung 46

Tabellen

Tabelle 1 Überblick zu den untersuchten Programmen 43
Tabelle 2 Evaluierungspraktiken anderer Geber auf Länderebene 43
Tabelle 3 Bestückung der EZ-Programmzielindikatoren 45

Kästen

Kasten 1 Gemeinsame Verfahrensreform (GVR) 4
Kasten 2 Erkenntnisse zum Zusammenwirken der Module im EZ-Programm 27

ABKÜRZUNGEN UND AKRONYME

ADB	Asian Development Bank (Asiatische Entwicklungsbank)
BGR	Bundesanstalt für Geowissenschaften und Rohstoffe
BMZ	Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung
BRH	Bundesrechnungshof
CAPE	Country Assistance Programme Evaluation
CPS	Country Partnership Strategy
DAC	Development Assistance Committee (Entwicklungsausschuss)
DEval	Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit
DO	Durchführungsorganisation
EU	Europäische Union
EZ	Entwicklungszusammenarbeit
EZ-PE	EZ-Programmevaluierung
FZ	Finanzielle Zusammenarbeit
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH
GG	Grundgesamtheit
GVR	Gemeinsame Verfahrensreform
IEG	Independent Evaluation Group (Unabhängige Evaluationsgruppe) der Weltbank
KfW	KfW Entwicklungsbank
LS	Länderstrategie
NaWi	Nachhaltige Wirtschaftsentwicklung
OECD	Organisation for Economic Co-Operation and Development (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung)
PTB	Physikalisch-Technische Bundesanstalt
RBM	Results-based Management
SMART	Spezifisch, Messbar, Akzeptiert, Relevant, Terminiert
TZ	Technische Zusammenarbeit
UNDP	United Nations Development Programme (Entwicklungsprogramm der Vereinten Nationen)
UNIDO	United Nations Industrial Development Organization (Organisation der Vereinten Nationen für industrielle Entwicklung)
WiLo	Wirkungslogik

1. EINLEITUNG

In den letzten Jahren hat das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) wesentliche Reformen der bilateralen deutschen Entwicklungszusammenarbeit (EZ) initiiert. Mit der 2017 in Kraft getretenen Gemeinsamen Verfahrensreform (GVR) wurden Verfahren zur Planung und Steuerung sowie zur Zusammenarbeit mit den staatlichen Durchführungsorganisationen (DOs) neu geregelt. Mit „BMZ 2030“ wurde 2020 zudem ein umfassendes Reformkonzept mit dem Ziel vorgelegt, die Maßnahmen und Mittel der Entwicklungspolitik zukünftig noch strategischer, wirksamer und effizienter einzusetzen. Neben neuen inhaltlichen Schwerpunkten, Partnerschaftskategorien und einem integrierten Planungsverfahren wird der Wirksamkeit und damit einhergehend Wirkungsanalysen ein hoher Stellenwert beigemessen (BMZ, 2020).

Die Planung und Steuerung der bilateralen deutschen EZ erfolgt über den sogenannten Dreiklang bestehend aus einer Länderstrategie (LS), EZ-Programmen und Modulen. EZ-Programme stellen darin die Schnittstelle zwischen den strategischen Vorgaben des BMZ aus der LS und der operativen Implementierung der Module durch die staatlichen DOs dar und operationalisieren somit die strategischen Vorgaben aus der Länderstrategie. Regelmäßige Evaluierungen von EZ-Programmen sind zwar in der GVR vorgesehen, werden bisher aber nicht durchgeführt.

Im vorliegenden Bericht wird die Wirkungsorientierung und Evaluierbarkeit von EZ-Programmen betrachtet. Damit soll ein Beitrag zur Entwicklung und Einführung eines Evaluierungsformats geleistet werden. Um eine aussagekräftige Untersuchung des Steuerungsinstruments „EZ-Programme“ gewährleisten zu können, wurden zunächst intensiv bestehende Vorgaben und Strukturen und deren Umsetzung im Rahmen von drei Fallstudien untersucht. Eine Sekundärdatenanalyse, Interviews und eine Onlinebefragung dienten der Validierung und Ergänzung übergeordneter Erkenntnisse.

Mit dem Bericht sollen zum einen Erkenntnisse zu Verbesserungsmöglichkeiten in der Planung und Steuerung von EZ-Programmen gewonnen und zum anderen inhaltliche und methodische Hinweise für die zukünftigen Evaluierungen und Wirkungsanalysen auf Programmebene gegeben werden. Dazu wurde die Frage behandelt, welche Ziele und Zwecke Evaluierungen von EZ-Programmen (EZ-PEs) erfüllen sollten und wann Ergebnisse vorliegen müssten, um bestmöglich genutzt zu werden. Außerdem wurden bestehende Daten- und Berichtsformate untersucht, um zu eruieren, inwieweit die zur Verfügung stehenden Informationen eine hinreichende Grundlage zur Durchführung von EZ-PEs liefern.

Der vorliegende Bericht soll darüber hinaus der interessierten Fachöffentlichkeit dabei helfen, die Bedeutung von EZ-Programmen innerhalb der bilateralen deutschen EZ nachzuvollziehen und den potenziellen Mehrwert und Nutzen von Programmevaluierungen verstehen zu können. Zudem soll ein Beitrag zum verbesserten Verständnis für notwendige Voraussetzungen einer Evaluierung und Wirkungsmessung von komplexen EZ-Programmen geleistet werden.

Im Folgenden wird zunächst der Evaluierungsgegenstand „EZ-Programme“ genauer definiert (Kapitel 2). Darauffolgend wird das Konzept der Wirkungsorientierung (Kapitel 3) erläutert, bevor auf die gegenwärtige Evaluierungspraxis in der deutschen EZ eingegangen wird (Kapitel 4). In diesem Zusammenhang werden auch der Mehrwert und das Erkenntnisinteresse von Evaluierungen sowie von verschiedenen Ansätzen zur Evaluierung auf Programmebene unter Berücksichtigung der Erfahrungen anderer Geber diskutiert. Anschließend wird das methodische Vorgehen (Kapitel 5) dargestellt. Danach werden das Erkenntnisinteresse und die Evaluierungsansätze (Kapitel 6) sowie die Ergebnisse aus den Erhebungen (Kapitel 7) präsentiert. Abschließend werden Schlussfolgerungen und Empfehlungen (Kapitel 8) zur Stärkung von EZ-Programmen und für die Entwicklung eines Evaluierungsformats ausgesprochen.

2. EZ-PROGRAMME

Was ist ein EZ-Programm?

Die bilaterale deutsche EZ wird durch das Zusammenwirken drei verschiedener Ebenen (bezeichnet als Dreiklang aus der LS, EZ-Programmen und Modulen) gesteuert und umgesetzt. Das Steuerungsmodell Dreiklangs wurde durch die Einführung der GVR im Jahre 2017 weiterentwickelt (siehe Kasten 1). Die strategische Planung und Steuerung der Maßnahmen im Partnerland soll dementsprechend über die LS-Ebene erfolgen. Die LS enthält maßgebliche strategische Vorgaben (unter anderem zur Zusammenarbeit mit dem Partnerland) und bestimmt dadurch die übergeordnete Zielsetzung für die untergeordneten EZ-Programme (BRH, 2021; Hartmann et al., 2019), über die die eigentlichen Maßnahmen umgesetzt und gesteuert werden.

Kasten 1 Gemeinsame Verfahrensreform (GVR)

Mit der GVR verfolgt das BMZ zwei eng miteinander verbundene Ziele. Erstens soll die Wirksamkeit, Effizienz und Effektivität des Engagements der bilateralen deutschen Zusammenarbeit verbessert und zweitens die eigene Steuerungsfähigkeit gestärkt werden, um den deutschen Beitrag zur Erreichung der Agenda 2030 in den Partnerländern zu erhöhen (Hartmann et al., 2019; OECD DAC, 2021). Im Zuge der GVR wurden die gesamten Prozesse der Planung und Implementierung von Programmen und Modulen angepasst, was für die programmbeteiligten Akteure zu Änderungen in ihren Verfahren führte. Bestehende Vorgaben wurden weiterentwickelt sowie zum Teil neue Anforderungen definiert und ergänzt. Dies umfasst beispielsweise die stärkere Einbindung der Programme in die entwicklungspolitischen Strategien des BMZ und der Partnerregierung (Syspons, 2018).

Für jede dieser Ebenen des Dreiklangs enthält die GVR konkrete Verfahrensschritte, die den jeweils Verantwortlichen und auch den beteiligten Akteuren eine Orientierung zur operativen Umsetzung bieten. Solche Angaben beziehen sich unter anderem auf die Übertragung von Zuständigkeiten, die Nutzung von Formaten sowie die Frage, wann spezifische Akteure eingebunden werden sollen (Bundesregierung, 2019).

Die GVR zielt darauf ab, die Bedeutung der Programmebene als Steuerungsebene insgesamt zu stärken. Hierzu setzt sie bei der Verbesserung der Programmkonzeption und -implementierung an. Eine wesentliche Neuerung stellt die sogenannte Wirkungslogik (WiLo) dar, die sowohl in der Programmkonzeption als auch in der Implementierung die Wirkungszusammenhänge auf Outcome- sowie zwischen Outcome- und Impact-Ebene visualisieren soll (Syspons, 2018).

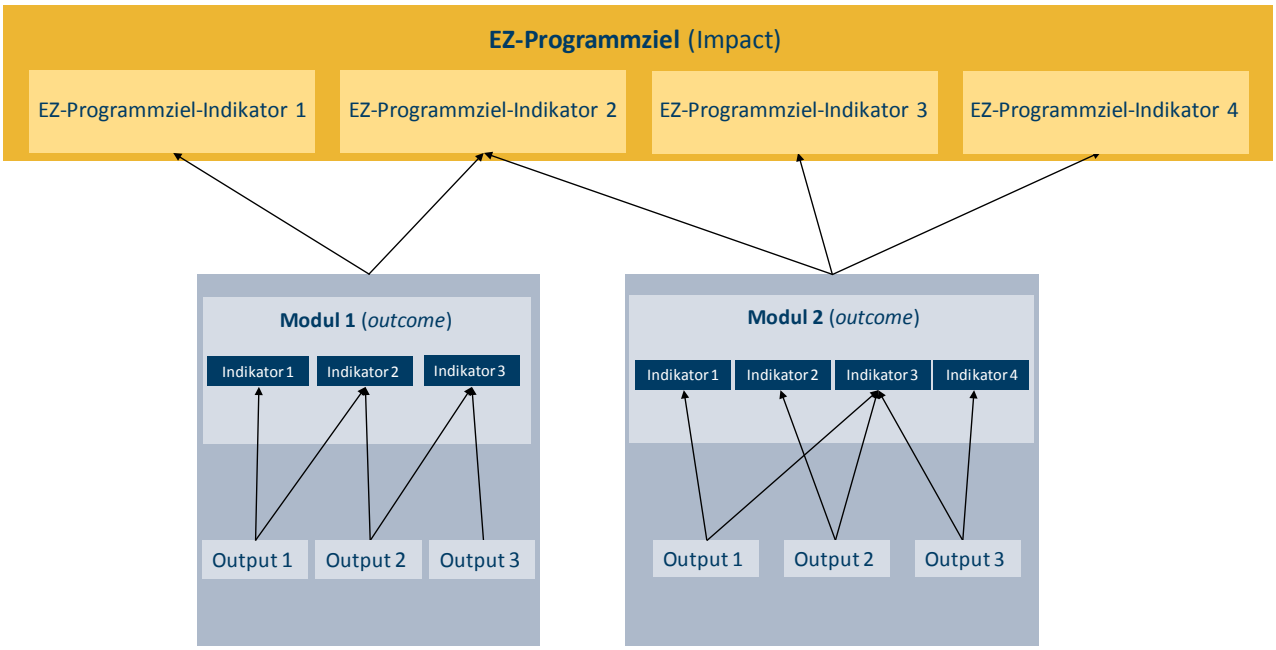
In ihrer Gesamtheit bündeln EZ-Programme die Vorhaben der DOs (sogenannte Module), die jeweils in Form von Finanzieller Zusammenarbeit (FZ) oder Technischer Zusammenarbeit (TZ) durchgeführt werden. Sie tragen somit zur übergeordneten entwicklungspolitischen Zielsetzung bei (Dok.[1] 1, 2; BRH, 2021; Hartmann et al., 2019; BMZ, 2021). Das Ziel der EZ-Programme liegt darin, einen Beitrag zu übergeordneten Entwicklungsstrategien des BMZ zu leisten und dadurch eine Veränderung in den Partnerländern anzustoßen (OECD DAC, 2021). Die Agenda 2030 für nachhaltige Entwicklung der Vereinten Nationen, die jeweiligen nationalen Entwicklungsbedarfe des Partnerlandes und die übergeordneten Ziele des BMZ bilden den groben Orientierungsrahmen für die Inhalte und Ziele der LS und der EZ-Programme (International Atomic Energy Agency, 2021; Dok. 1).

Abbildung 1 stellt zusammenfassend die wesentlichen Bestandteile von EZ-Programmen dar. Die Operationalisierung und Messung des EZ-Programmziels erfolgt über Programmzielindikatoren. Die Abbildung macht deutlich, dass die Erreichung des Programmziels beziehungsweise der Indikatoren über die Beiträge der Module erfolgt. Die Modulziele, die auf der untergeordneten Zielebene (*Outcome*) liegen, werden ana-

[1] Um die Vertraulichkeit der an das DEval weitergegebenen nicht veröffentlichten Dokumente zu gewährleisten, werden sie in der Form „Dok." plus einer fortlaufenden Zahl im Text zitiert und erscheinen nicht im Literaturverzeichnis.

log zur Programmebene durch eigene Modulzielindikatoren gemessen und resultieren aus den erbrachten Outputs (BMZ, 2021).

Abbildung 1 Bestandteile und Wirkungsebenen eines EZ-Programms



Quelle: eigene Darstellung

Die TZ-Module werden grundsätzlich innerhalb von drei Jahren implementiert, wobei eine Verlängerung um ein weiteres Jahr normalerweise möglich ist. Für FZ-Module gibt es hingegen keine konkreten Implementierungsfristen. Ähnlich wie für die FZ-Module enthalten die Vorgaben für EZ-Programme prinzipiell, das heißt unabhängig vom jeweiligen Sektor, keine Laufzeiten. Diese sind demnach zeitlich weniger stark gebunden[2] (Dok. 1, 2).

Bis zur Aussteuerung, also der Beendigung von EZ-Programmen in einem Partnerland, durchlaufen EZ-Programme im Normalfall einen iterativen Prozess, bei dem die während der Modulimplementierung gewonnenen Erkenntnisse idealerweise in die Konzeption und Gestaltung eines Folgeprogramms einfließen und somit einen neuen Programmzyklus anstoßen (Dok. 2).

Gesteuert und verantwortet werden EZ-Programme durch das BMZ, das die Programme inhaltlich definiert und entsprechende finanzielle und rechtliche Voraussetzungen für die Implementierung[3] schafft (BMZ, 2021). Die vier DOs – die Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, die KfW Entwicklungsbank (KfW), die Physikalisch-Technische Bundesanstalt (PTB) und die Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) – sind dagegen für die Vorbereitung und eigenverantwortliche Umsetzung der Module innerhalb des Programms zuständig (BMZ, 2021).

Formate zum Informationsaustausch und zur Rechenschaftslegung bestehen für alle Ebenen des Dreiklangs. Vorbereitend für die Erstellung von LS können Länderportfolioreviews erstellt werden.[4] Zusätzlich gibt es

2 Eine reguläre Laufzeit für EZ-Programme ist nicht festgelegt. Durch die Erhebungen im Rahmen dieser Evaluierung konnte jedoch festgestellt werden, dass EZ-Programme meistens vier bis sechs Jahre laufen.

3 Hierzu zählen unter anderem die Bereitstellung der Mittel und die Koordination und Aushandlung völkerrechtlicher Vereinbarungen mit dem Partner.

4 Die Verfahrensleitlinie des DEval zur Durchführung von Länderportfolioreviews ist abrufbar unter: *https://www.deval.org/fileadmin/Redaktion/PDF/05-Publikationen/Berichte/2019_Laenderportfolioreviews/DEval_Verfahrenleitlinie_LPR_2019_DE.pdf*.

vereinzelt Evaluierungen auf Länderebenen.[5] Auf Programm- wie auch auf Modulebene ist die Berichtslegung der DOs an das BMZ verbindlich festgelegt. Die Berichte liefern während der Implementierung Informationen über Programmfortschritte und zentrale Anpassungs-/Umsteuerungsbedarfe, Rahmenbedingungen im Partnerland sowie die Zielerreichung des Programms. Sie dienen somit der Rechenschaftslegung gegenüber dem politischen Raum (BMZ, 2021; BRH, 2021). Die DOs führen außerdem Modulevaluierungen von einer jährlichen Stichprobe an Modulen durch. Auf Programmebene finden solche Evaluierungen derzeit noch nicht statt, sind jedoch aktuell in Planung (siehe Kapitel 3).

Während der Dreiklang insgesamt zur besseren strategischen Steuerung und Wirksamkeit der bilateralen staatlichen EZ beitragen soll, werden auf Programmebene entwicklungspolitische Ziele beziehungsweise Wirkungen festgelegt. Im nachfolgenden Kapitel wird dargestellt, was unter Wirkungsorientierung allgemein und der konkreten Konzeption und Durchführung von Programmen verstanden wird.

[5] Derzeit führt das DEval eine ressortgemeinsame strategische Evaluierung des Engagements von Auswärtigem Amt und BMZ in Irak durch.

3. WIRKUNGSORIENTIERUNG

In der EZ hat die Debatte um die Wirkungsorientierung ihren Anfang durch die Aid-Effectiveness-Agenda im Rahmen der Erklärung von Rom (2003), der Erklärung von Paris (Paris Declaration) aus dem Jahre 2005, dem Aktionsplan von Accra (2008) und der Entwicklungspartnerschaft von Busan (2011) erfahren (Bergmüller und Quiring, 2019; Holzapfel, 2014; OECD, 2019b). In diesem Kontext wurde das gemeinsame Ziel festgelegt, die Wirksamkeit der EZ zu steigern. Die Verantwortung hierfür soll sowohl beim Geber als auch beim Partnerland liegen (Holzapfel, 2014). Maßnahmen gelten einer stärkeren Harmonisierung der Leistungen, einer verbesserten Partnerorientierung und der Einführung beziehungsweise der Stärkung ergebnisorientierter Ansätze, zu dem das ergebnisorientierte Managementsystem (*results-based management*, RBM) zählt (High Level Forum, 2005; Klingebiel, 2011).

Das RBM umfasst zum einen eine sorgfältige Planung von EZ-Maßnahmen und eine frühzeitige Auseinandersetzung mit geplanten Ergebnissen und notwendigen Ressourcen (Wirkungsorientierung)[6]. Zum anderen zählt dazu auch die fortwährende Überprüfung von Fortschritten und weiteren Optimierungsbedarfen, also die Wirkungsbeobachtung und -messung (Bergmüller und Quiring, 2019; Caspari, 2012; Harten und Petrova, 2018; Holzapfel, 2014; OECD, 2019a; UNDG, 2011). Die Wirkungsorientierung zielt demnach darauf ab, die transformative Kraft der EZ-Maßnahmen und insbesondere deren Ergebnisse zu erfassen (Harten und Petrova, 2018). Durch die Nutzung von RBM können die vereinbarten Ziele der internationalen Gemeinschaft erreicht werden, vor allem die Verbesserung der Effizienz und Wirksamkeit als auch die Rechenschaftspflicht der EZ (Harten und Petrova, 2018; Holzapfel, 2014).

Die folgende theoretische Einordung über effektive Instrumente der Wirkungsorientierung und -messung orientiert sich an den durch Holzapfel und Römling (2020) beschriebenen Schritten des RBM sowie zum Teil auch am Logical Framework Approach. Ein wichtiger Bestandteil ist die **Definition von Zielen und Indikatoren**. Indikatoren unterliegen klaren Qualitätskriterien, die als spezifisch, messbar, akzeptiert, relevant und terminiert bezeichnet werden. Entsprechend diesen **SMART-Kriterien** müssen Programmzielindikatoren auf der richtigen Ergebnisebene formuliert und zudem so spezifisch definiert sein, dass sie das Programmziel beziehungsweise entsprechende Veränderungen tatsächlich messen können (Holzapfel und Römling, 2020). Dies bietet die Grundlage, um eine Programmtheorie zu erarbeiten, die intendierte und auch nicht intendierte Wirkungsannahmen abbildet.

Baselinedaten (Basis-/Zielwerte) sind in der Programmkonzeption zu definieren und sollten demnach erhoben beziehungsweise festgelegt werden, bevor das Programm implementiert wird. Sie sollen sowohl im Umsetzungsverlauf als auch danach als Vergleichswerte zur **Überprüfung des Programmfortschritts und seiner Zielerreichung** dienen (Holzapfel und Römling, 2020). Hierzu zählt die Festlegung der Datengrundlage zur Wirkungsorientierung von Programmen. Dabei können sowohl Primär- wie Sekundärdaten genutzt werden. Im Sinne der internationalen Vereinbarungen (zum Beispiel der Erklärung von Paris über die Wirksamkeit der EZ) sollten vorrangig Daten aus nationalen Systemen der Partner genutzt werden, um deren Eigenverantwortung zu stärken.

Datenquellen müssen auf passender Ebene aggregiert und in angemessener Qualität bereitgestellt werden, um Wirkungen im Interventionsbereich des Programms attribuieren zu können (Holzapfel und Römling, 2020). Zudem erfordert es eine angemessene Programmkonzeption und ein adäquates Evaluierungsdesign. Ein weiteres Ziel der Wirkungsorientierung der gemeinsam mit einem Partnerland durchgeführten Programme besteht darin, eine bestmögliche **Partnereinbindung** zu gewährleisten, um die jeweiligen Bedarfe angemessen zu berücksichtigen und die Kontextbedingungen zu adressieren. Eine Möglichkeit hierzu bietet die gemeinsame Definition des Ziels (Holzapfel und Römling, 2020).

Herausforderungen entstehen hier insbesondere hinsichtlich der Datenbeschaffung und -bereitstellung. So wird die Qualität der Daten aus den Partnersystemen häufig als unzureichend beschrieben. Eine weitere Problematik besteht in der Attribution, also der Zurechenbarkeit der Resultate zum jeweiligen Pro-

[6] Anstelle des Begriffs „Wirkung“ wird hier bewusst als Übersetzung des englischen Begriffs „results“ Ergebnisse verwendet. Results umfasst gemäß offizieller OECD-DAC-Definition Outputs, Outcomes und Impacts.

gramm/Modul (Holzapfel, 2014). Ein angemessenes RBM während der Programmimplementierung ist Voraussetzung für die Wirkungsbeobachtung und -messung. Hierfür bieten das Zielsystem und die Programmtheorie eine wichtige Grundlage. Kausale Wirkungsbeziehungen können in der Durchführung eruiert werden und informieren die Programmsteuerung (Caspari, 2012; Holzapfel, 2014).

Ein zentrales Instrument der Wirkungsmessung und -beurteilung ist laut Holzapfel und Römling (2020) das Monitoring. Im Rahmen des Monitorings werden die Indikatoren aus der Programmkonzeption und -implementierung mit aktuellen Daten bestückt (zum Beispiel Ist-Werte), die aus den vorab festgelegten Datenquellen herangezogen werden (Holzapfel und Römling, 2020). Darüber hinaus spielt der vorausschauende Umgang mit Risiken, das heißt die Auseinandersetzung mit möglichen Risiken der Zielerreichung, eine wichtige Rolle. In der Implementierung sind diese entsprechend zu reflektieren und mit Risikomanagementmaßnahmen zu hinterlegen (Holzapfel und Römling, 2020).

Laut Holzapfel und Römling (2020) hat das BMZ für die bilaterale deutsche EZ bislang noch kein umfassendes RBM-System eingerichtet[7]. Die Autorinnen der Studie verweisen darauf, dass zwar grundsätzliche Instrumente (unter anderem zur Berichterstattung) und Systeme (beispielsweise Evaluierungssysteme) eingeführt worden sind und genutzt werden, die Wirkungsbeobachtung und -messung jedoch auf dieser Grundlage noch nicht vollständig ermöglicht wird. In diesem Zusammenhang wird auch auf den Prüfbericht des Entwicklungsausschusses (Development Assistance Committee, DAC) der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (Organisation for Economic Co-Operation and Development, OECD) aus dem Jahr 2015 verwiesen (OECD, 2015). Dies gilt ebenso für EZ-Programme.

Wie sollen EZ-Programme ihre intendierten Wirkungen erreichen?

Das übergeordnete entwicklungspolitische Ziel wird in der LS festgelegt (langfristige Impact-Ebene) und soll durch die EZ-Programme erreicht werden. Auf EZ-Programmebene werden jedoch keine gezielten modulübergreifenden Maßnahmen durchgeführt, um dies zu gewährleisten. Die **Umsetzung der EZ-Programme erfolgt ausschließlich über die Module der DOs**. Die zugrunde liegende Wirkungsannahme ist somit, dass EZ-Programmziele ausschließlich durch das (Zusammen-)Wirken der Module erreicht werden. Eine wichtige Grundvoraussetzung zur Zielerreichung ist daher, ein **konsistentes und realistisches Zielsystem auf EZ-Programmebene** zu erarbeiten (BMZ, 2020).

Die Wirkungsentfaltung und die Erreichung der EZ-Programmziele können über **verschiedene Wirkpfade** (Möglichkeiten, wie Wirkungen von der Modul- bis zur LS-Ebene entfaltet werden können) erfolgen, die in Abbildung 2 abgebildet sind. Module können zunächst gemeinsam mit anderen Modulen innerhalb eines Handlungsfelds wirken, um dann gemeinsam mit den Effekten der anderen Handlungsfelder auf die Erreichung des EZ-Programmziels hinzuarbeiten. Darüber hinaus besteht die Möglichkeit, dass Module zunächst losgelöst von weiteren Modulen agieren, um auf übergeordneter Ebene (mittelfristige Impacts) gemeinsame Wirkungen mit anderen Handlungsfeldern zu entfalten. Insbesondere die Verzahnung von Modulen in gleichen Handlungsfeldern (üblicherweise aus mindestens einem FZ- und einem TZ-Modul) soll komplementär wirken. In den meisten EZ-Programmen sind unterschiedliche Wirkpfade zu identifizieren. Unabhängig von den Wirkpfaden wird jedoch immer davon ausgegangen, dass alle Module gezielte und wesentliche Beiträge zur Erreichung des EZ-Programmziels leisten.

Die Vorgaben für EZ-Programme legen drei verschiedene Pfade beziehungsweise Möglichkeiten nahe, wie Module gemeinsame Wirkungen im Programm entfalten können und somit auf die Erreichung des Programmziels hinwirken. Diese umfassen:

A) ein enges operatives Zusammenwirken im Sinne eines intensiven Ineinandergreifens der Module (Verzahnung). Demzufolge sollen Module sich sowohl zeitlich und räumlich wie thematisch und hinsichtlich der involvierten Akteure gegenseitig bestmöglich ergänzen. So wird zum Beispiel nachvollziehbar be-

[7] Bei der Etablierung eines RBM-Systems in der deutschen EZ sind die bereits in der Wissenschaft identifizierten Kritikpunkte und alternativen Ansätze zu berücksichtigen (OECD, 2019a).

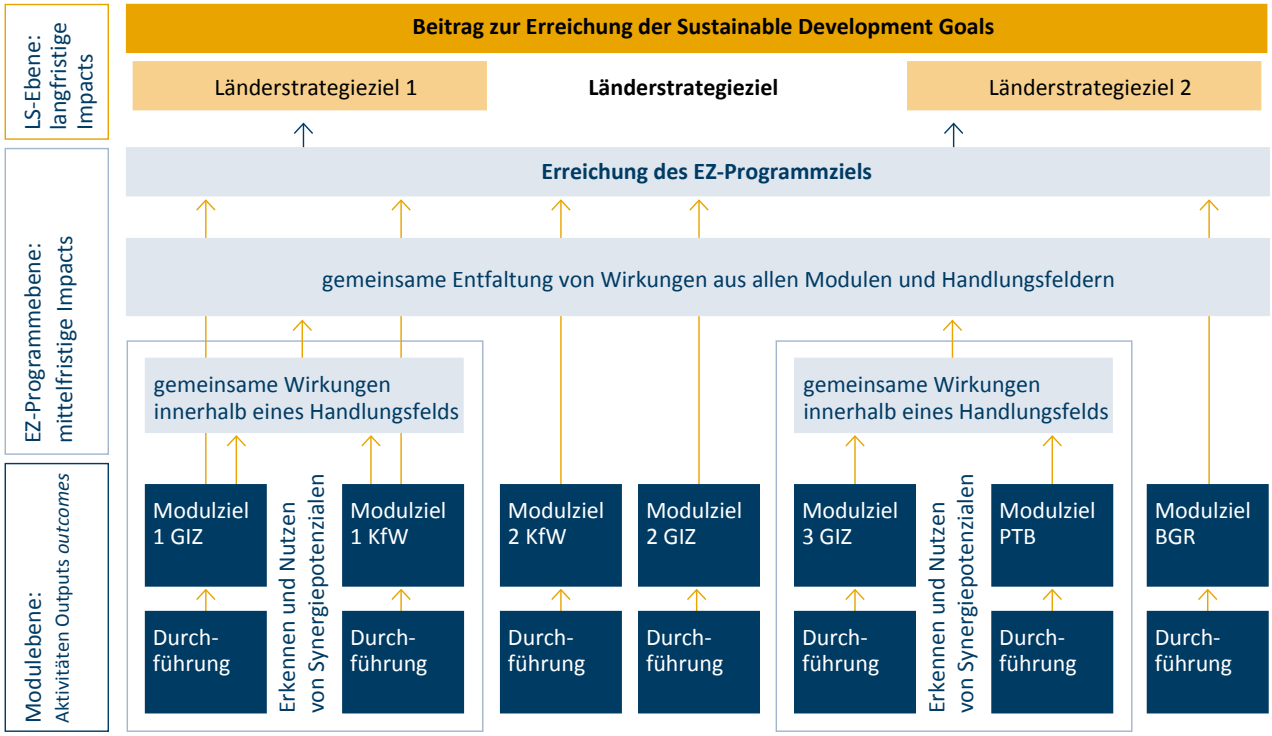
schrieben, wie eine Aktivität der TZ in einer Schule (beispielsweise Kapazitätstraining von Lehrkräften) durch die Aktivitäten der FZ (Bereitstellung von Ausstattung und Materialien für den Unterricht) in derselben Schule ergänzt wird und wie diese Module gemeinsam auf das Programmziel hinwirken.

B) eine bestmögliche Abstimmung der Inhalte einzelner Module durch Austausch-/Dialogformate zwischen den DOs. Diese zielen ebenfalls auf die Gewährleistung einer optimalen Verstärkung der Module ab.

C) einen Einzelbeitrag eines Moduls. Das Zusammenwirken mit anderen Modulen erfolgt hierbei, wie in Abbildung 2 ersichtlich, frühestens auf Ebene der Handlungsfelder.

Gemäß den Vorgaben wird besonders Option A (enges operatives Zusammenwirken) hervorgehoben, um die gemeinsame Entfaltung von Wirkungen bestmöglich gewährleisten zu können.

Abbildung 2 Wirkungsannahmen von EZ-Programmen



Quelle: eigene Darstellung

Zur Messung des Programmziels sollen **Monitoringdaten** herangezogen werden. Diese stellen eine unabdingbare Grundlage dar, um Fortschritte und die Zielerreichung messen zu können. In Ergänzung können Evaluierungen verwendet und mit den Monitoringdaten trianguliert werden, um ein vollständiges Bild über den Programmfortschritt und die Zielerreichung zu erhalten (Holzapfel und Römling, 2020).

4. EVALUIERUNG VON EZ-PROGRAMMEN

Unter Evaluierung von Programmen versteht man allgemein die systematische Sammlung von Informationen über Aktivitäten, Merkmale und Ergebnisse. Ziel von Programmevaluierungen ist es, das Programm zu beurteilen, um die **Wirksamkeit zu verbessern**, die **Inhalte weiterzuentwickeln**, Entscheidungen über **zukünftige Programmplanungen** zu treffen oder auch die **Rechenschaftslegung**.

Programme sind meist durch eine hohe Komplexität gekennzeichnet, da sie durch eine Vielzahl an Aktivitäten auf verschiedene *Outcomes* und Impacts abzielen, die sich nicht direkt beobachten lassen (von Werthern, 2019). Das gilt auch für deutsche EZ-Programme und rührt daher, dass sie durch viele unterschiedliche Akteure und Maßnahmen gestaltet werden, die zum Teil in verschiedenen Fachbereichen (des BMZ und der DOs) verortet sind. Die Prozesse der Akteure werden demnach nicht selten durch unterschiedliche Verfahren geregelt.

In Evaluierungen muss die Programmtheorie zunächst nachvollzogen und häufig ebenso rekonstruiert werden (theoriebasierte Evaluierung). Dementsprechend schlägt sich die geschilderte Komplexität von Programmen – und hier insbesondere von deutschen EZ-Programmen – gleichfalls auf ihre Evaluierung nieder.

Unter Länderprogrammevaluierungen versteht man in der internationalen EZ in der Regel die Evaluierung des Portfolios eines Gebers in einem Partnerland. Einige Geber (zum Beispiel die Europäische Union [EU]) führen jedoch pro Land nicht nur ein Länderprogramm durch, sondern haben verschiedene Strategien und mehrere Schwerpunkte, sodass es mehrere Programme pro Land gibt, die entsprechend durch verschiedene Evaluierungsformate (wie strategische oder geografische Evaluierungen) untersucht werden (siehe Abschnitt 6.2).

In der bilateralen deutschen EZ hat das BMZ erstmals 2006 eine Pilotphase zur Evaluierung von Länderprogrammen, in diesem Fall definiert als das gesamte BMZ-finanzierte bilaterale Portfolio in einem Partnerland, gestartet. Auf Basis von Untersuchungen der damaligen Länderprogramme zu Bangladesch, Honduras und Uganda wurden unter anderem erste Erkenntnisse und Empfehlungen zur Ausarbeitung einer Methodik von Länderprogrammevaluierungen formuliert (Hartmann et al., 2019; Dok. 10).

Wie bereits erwähnt, werden im Zuge des Reformkonzepts „BMZ 2030“ vermehrt Anstrengungen unternommen, um die Wirkungsorientierung und -messung von EZ-Maßnahmen weiter zu stärken. Die regelmäßige Durchführung von Evaluierungen ist im Rahmen des Dreiklangs bisher lediglich auf Modulebene durch die DOs vorgesehen (siehe Kapitel 2). Die Modulevaluierungen berücksichtigen jedoch nur teilweise systematisch die Programmebene. Vor dem Hintergrund der bisher nicht stattfindenden Länder- und EZ-Programmevaluierungen vergrößert dies die **Evidenzlücke** zwischen der Modul- und der EZ-Programmebene (BRH, 2021).

Auch der Bundesrechnungshof (BRH) stellt in seiner Prüfung der Evaluierung von Maßnahmen in der EZ (Teil I – Steuerung im BMZ) fest, dass aufgrund dieser Evidenzlücke Informationen zur strategischen Relevanz und Wirksamkeit für die bilaterale staatliche EZ nur bedingt vorliegen. Daher wird empfohlen, Evaluierungen auf Programm- wie auf Länderebene in Verbindung mit einer systematischen Berichterstattung mit dem **Ziel einer verbesserten Steuerung und Rechenschaftslegung** durchzuführen (BRH, 2021). Der BRH weist zudem darauf hin, dass für die Modul- und die Programmebene der Verfahrensschritt „Evaluierung“ im Regelwerk lediglich grafisch dargestellt, jedoch nicht inhaltlich beschrieben ist (BRH, 2021). Der BRH (2021) macht dazu konkrete Empfehlungen:

- **Evaluierungen auf Länder- und EZ-Programmebene** inhaltlich und ablauforganisatorisch definieren, einschließlich der Festlegung, wer die unabhängigen Evaluierungen in einem für die regelmäßige Steuerung der EZ notwendigen Umfang ausführen soll
- interne Kompetenzen und Abläufe für die **übergeordnete Auswertung der Berichterstattung und der Evaluierungen** regeln
- **Schnittstellen des Systems** zu den Projekt- und Querschnittsauswertungen der DOs, zu Länderreviews und Evaluierungen des DEval (Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit) definieren

Die Evidenzlücke auf EZ-Programmebene wird auch im OECD-DAC-Prüfbericht 2021 mit entsprechenden Empfehlungen, diese zu schließen, konstatiert (OECD DAC, 2021). Zudem stellt der Bericht fest, dass „die übergeordneten Ziele […] jedoch nicht so formuliert [sind], dass ihre Umsetzung gemessen und evaluiert werden könnte." Dazu bedarf es eines **verbesserten ergebnisorientierten Managements (RBM)** und einer **verankerten Ergebniskultur** in der deutschen EZ, zum Beispiel in Form eines integrierten Datenmanagementsystems und einer Erweiterung der Anzahl von Indikatoren. Bisher werden Evaluierungsergebnisse für die **Rechenschaftslegung und Kommunikation** genutzt, sie sollten aber auch für die **strategische Ausrichtung und das Management** Berücksichtigung finden (OECD DAC, 2021).

Für die Konzeption und Anwendung eines Evaluierungsformats auf EZ-Programmebene ist eine **Beurteilung der Evaluierbarkeit** notwendig. Ziel ist es, vorab zu überprüfen, ob der vorgesehene Evaluierungsgegenstand – in diesem Fall EZ-Programme – evaluierbar ist, das heißt, ob er bestimmte Voraussetzungen zur Evaluierung mitbringt (*quality at entry*). Solche Vorbedingungen knüpfen an die Aspekte der Wirkungsorientierung und -messung an und umfassen laut Leonard und Eulenburg (2012) folgende Aspekte:

- Die **Indikatoren und auch die Daten**, die zur Bewertung der Zielerreichung herangezogen werden, sind von **hoher Qualität** (SMARTe Ziele, Baseline liegt vor).
- Die Programmdokumente enthalten ausreichend Informationen über die **Ausgangssituation des Programms** und auch über dessen Rahmenbedingungen.
- Das Programm enthält eine angemessene Strategie zur Reflexion und zum Management von **Risiken**.
- Für das Programm liegt ein durchdachtes **Monitoringsystem** vor, das eine klare Regelung der Zuständigkeiten, die Einplanung entsprechender Ressourcen und valide Datenquellen umfasst.

Ergänzend zu diesen zentralen Punkten von Leonard und Eulenburg werden zudem **verbesserte beziehungsweise angemessenere Zielformulierungen** als wichtig erachtet und im Rahmen dieser Arbeit als weiterer Punkt aufgenommen und untersucht.

Mit der Beurteilung der Evaluierbarkeit soll erfasst werden, ob eine Evaluierung zu validen Erkenntnissen führen kann. Die Ergebnisse dieser Bewertung können demnach einen gewinnbringenden **Beitrag für die zukünftige Gestaltung einer Evaluierung, des Monitorings oder des Projekts, des Programms oder der Strategie leisten** (Davies, 2013). Sie enthalten Einschätzungen über die „technische" Möglichkeit als auch den praktischen Mehrwert einer Evaluierung und zu erwartende Limitationen. Im vorliegenden Bericht wird die Evaluierbarkeit von EZ-Programmen untersucht und somit eine wichtige Vorarbeit für die Gestaltung eines Evaluierungsformats auf EZ-Programmebene geleistet.

5. METHODIK

Die methodische Herangehensweise wurde im Rahmen der Inception-Phase erarbeitet und im entsprechenden Bericht festgelegt. Parallel zur Daten- und Dokumentenbeschaffung wurden **Kriterien zur Fallstudienauswahl** festgelegt und insgesamt drei EZ-Programme als Fallstudien ausgewählt. Aufgrund von eingeschränkten Ressourcen kam eine Prüfung aller laufenden EZ-Programme (circa 64) nicht infrage. Aus diesem Grund wurde ebenfalls entschieden, Fallstudienprogramme aus dem Schwerpunkt „Nachhaltige Wirtschaftsentwicklung“ (NaWi) auszuwählen, dem größten Förderbereich der bilateralen deutschen EZ. Die Fallauswahl erfolgte in einem mehrstufigen Verfahren.

i. Identifizierung der EZ-Programme, die

- gemäß der nach 2017 durch die GVR in Kraft getretenen Vorgaben geplant wurden
- alle nach GVR verpflichtend zu erstellenden Dokumente aus der Planung und Berichterstattung vorweisen konnten
- mindestens eine Berichterstattung zum Programmfortschritt erstellt hatten

Daraus ergab sich eine Grundgesamtheit von 14 EZ-Programmen.

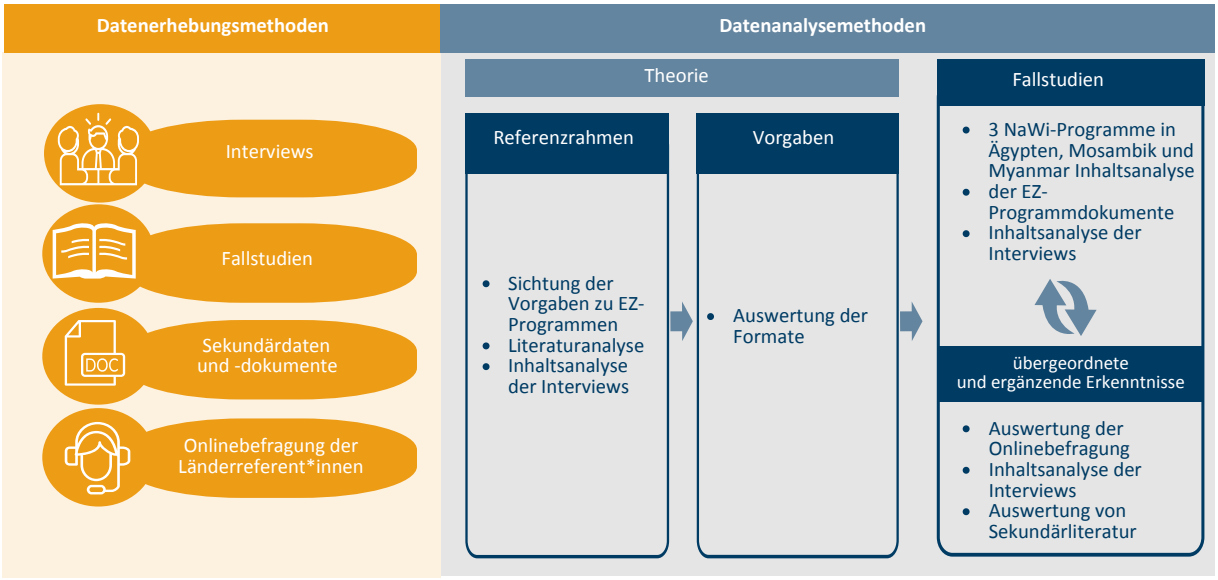
ii. Unter Berücksichtigung der Regeln der *distribution-based case selection* bestand das Ziel darin, möglichst diverse Fälle, das heißt je ein eher kleines, durchschnittliches und großes EZ-Programm, zu identifizieren. Entsprechend wurden die folgenden Kriterien herangezogen:
   - Anzahl der Module
   - Anzahl der beteiligten DOs
   - Fördervolumen

Demzufolge wurden drei EZ-Programme als Fallstudien ausgewählt (siehe Anhang 10.1): „Berufsbildung und Entwicklung des Arbeitsmarktes in Ägypten“, „Nachhaltige Wirtschaftsentwicklung in Mosambik“ und „Nachhaltige Wirtschaftsentwicklung in Myanmar“.

Datenerhebung

Die Datenerhebung wurde von Februar bis April 2021 durchgeführt. Diese umfasste die **Sammlung von Vorgaben und Handreichungen** zu EZ-Programmen und den **Programmdokumenten** zur Planung und Berichterstattung im Rahmen der drei ausgewählten Fallstudien. Zudem erfolgte eine **Onlinebefragung von Länderreferent*innen des BMZ**, **Interviews** mit programmbeteiligten Akteuren und mit Personen, die übergreifende Aufgaben innehaben. Ergänzend diente eine **Literaturauswertung** zur Einordnung und Validierung der gewonnenen Erkenntnisse.

Abbildung 3 Datenerhebungs- und Datenanalysemethoden



Quelle: eigene Darstellung

Zusätzlich wurden insgesamt 28 Interviews mit Vertreter*innen des BMZ, der DOs, der Botschaften (Referent*innen für wirtschaftliche Zusammenarbeit [WZ-Referent*innen]) und einzelnen Fachexpert*innen/ Wissenschaftler*innen geführt. Sieben davon fanden als Gruppeninterviews (mindestens zwei Interviewpartner*innen) statt. Die Onlinebefragung erzielte eine Rücklaufquote von 34 Prozent. Dies entspricht 30 (von 88 kontaktierten Länderreferent*innen) beantworteten Fragebogen. Knapp ein Drittel (acht Personen) gab an, derzeit nicht für EZ-Programme zuständig zu sein oder dass die jeweiligen EZ-Programme die vorgegebenen Kriterien (EZ-Programm nach GVR geplant und mindestens eine Berichterstattung) nicht erfüllen würden. Somit ist davon auszugehen, dass die infrage kommende Grundgesamtheit niedriger ist als angenommen, da nicht alle Länderreferent*innen für EZ-Programme verantwortlich sind, die die Kriterien erfüllen. 19 der befragten Länderreferent*innen sind für EZ-Programme zuständig, die alle Kriterien erfüllten und den Fragebogen vollständig beantwortet haben.

Datenanalyse

Die übergeordnete methodische Herangehensweise ist an den Grounded-Theory-Forschungsansatz angelehnt, der häufig für die Durchführung explorativer Mixed-Methods-Studien angewendet wird (Strübing, 2014; Walsh et al., 2015). Ein konstantes Wechselspiel aus Deduktion (vorher festgelegte Kategorien durch theoretische und/oder empirische Ableitung)) und Induktion (theorieentwickelndes Verfahren) kennzeichnet den Prozess und äußert sich in einer fortwährenden Abfolge aus Datenerhebung und -analyse. Dies bedeutet, dass neue Erkenntnisse während der Datenerhebung oder -analyse systematisch identifiziert und auf ihre Generalisierbarkeit hin überprüft werden. Die Theoriebildung erfolgt durch die Ableitung des Referenzrahmens, wohingegen die Theorieprüfung auf den unterschiedlichen Analyseebenen geschieht, durch die der Referenzrahmen und die Bewertungskriterien verbessert werden können.

Hinweise zur Generalisierbarkeit der Ergebnisse

Die Ergebnisse des Berichts sind in **erster Linie für idealtypische EZ-Programme generalisierbar**. Zudem muss darauf hingewiesen werden, dass noch nicht für alle Länder eine LS erarbeitet und verabschiedet wurde und die Erkenntnisse auf diese Länder nicht gänzlich übertragbar sind.

Es wurden ausschließlich EZ-Programme aus dem Sektor „NaWi“ berücksichtigt. Vor allem aus den Interviews ging jedoch hervor, dass einige Schwerpunkte und Vorhabentypen besondere Rahmenbedingungen und Herausforderungen mit sich bringen. Demnach ist unklar, inwiefern die Erkenntnisse gleichermaßen auch für andere Schwerpunkte gelten. Es können insbesondere keine Aussagen zu Fragestellungen der Verzahnung oder zeitlichen Dimension in der Planung berücksichtigt werden, wenn kurzfristige und sehr langfristige Maßnahmen (zum Beispiel kurzfristige Beratung und Kapazitätsentwicklung versus langfristige Finanzierungs-, Bau- und Infrastrukturvorhaben) innerhalb eines EZ-Programms vorgesehen sind.

6. ERKENNTNISINTERESSE UND EVALUIERUNGSANSÄTZE

6.1 Erkenntnisinteresse auf EZ-Programmebene in der deutschen EZ

Mithilfe von Interviews mit programmbeteiligten Akteuren und einer Onlinebefragung der BMZ-Länderreferent*innen wurde zum einen das Erkenntnisinteresse (Ziele und Zwecke) einer EZ-PE eruiert. Zum anderen wurde erfragt[8], zu welchem Zeitpunkt deren Ergebnisse am ehesten benötigt und bestmöglich genutzt werden könnten.

Die Untersuchung ergab, dass insgesamt ein großes Erkenntnisinteresse an EZ-Programmen besteht und der EZ-PE **viele unterschiedliche Aufgaben beziehungsweise Funktionen** beigemessen werden. Ein Großteil der Befragten erhofft sich eine **Analyse des Zusammenwirkens von Modulen** (insbesondere von FZ- und TZ-Modulen und zum Teil auch zwischen TZ-Modulen). In den Interviews wird zudem deutlich, dass derzeit Informationen zum Zusammenwirken für Umsteuerungsempfehlungen fehlen. Eine **Analyse des Zielsystems** auf EZ-Programmebene hinsichtlich entwicklungspolitischer Wirksamkeit, Effektivität, Relevanz (siehe auch BRH, 2021)[9], einschließlich Empfehlungen zur Umsteuerung, wird ebenfalls von einer Mehrheit als zentrales Erkenntnisinteresse für eine EZ-PE angeführt. Ein Drittel der Interviewten ergänzt dazu, dass im Rahmen einer EZ-PE die Angemessenheit der Indikatoren, die Qualität der Daten und die Annahmen in der Wirkungsmatrix überprüft werden sollten.

Darüber hinaus wird die **Überprüfung des Dreiklangs** in der Praxis vorgeschlagen, das heißt inwiefern die verschiedenen Ebenen und deren Zielsetzungen plausibel und sinnvoll ineinandergreifen und strategische Interessen nachvollziehbar im EZ-Programm berücksichtigt wurden. Ein besonderes Erkenntnisinteresse liegt hier in der Verzahnung der Länder-, EZ-Programm- und Modulebenen als auch in der Reflexion der Zeitpunkte, zu denen EZ-Programme und LS erstellt werden. Erkenntnisse zur Überprüfung des Dreiklangs könnten in Anpassungen der LS mit einfließen und Informationen zur Zielerreichung beisteuern. Neben dem inhaltlichen Schwerpunkt **Nachhaltigkeit** geben zwei Drittel der Interviewten an, dass die **Balance zwischen Anforderungen und Ressourcen/Effizienz** von EZ-Programmen sowie der **Dialog** zwischen BMZ, DOs und den Partnern während der EZ-Programmdurchführung untersucht werden sollten.

In drei der 28 geführten Interviews wird geäußert, dass in EZ-PEs der **politische Kontext und entwicklungspolitische Trends** (zum Beispiel im Partnerland; BMZ 2030) analysiert werden sollten. Demzufolge scheinen derzeit zur Verfügung stehende Darstellungen nicht immer als ausreichend erachtet zu werden. Die Ausarbeitung von **Empfehlungen** für weitere Phasen eines EZ-Programms wird in der Onlinebefragung von einem Großteil befürwortet, während sich in den Interviews ebenfalls drei Personen für eine Zusammenfassung von **gewonnenen Erkenntnissen** (*lessons learnt*) für den Transfer der Erfahrungen auf andere EZ-Programme aussprechen Diese Einschätzung wird vom BRH geteilt, der eine modulübergreifende Analyse der Lernfolge und die Standardisierung bewährter Abläufe empfiehlt (BRH, 2021).

Vereinzelt wird in den Interviews vorgeschlagen, den **EZ-Programmumfang** hinsichtlich der thematischen Bandbreite und Kohärenz der Module im EZ-Programm zu analysieren und zu messen, welchen Beitrag Deutschland in Abgrenzung zu anderen Gebern für die EZ-Programmzielerreichung leistet. Ein weiterer Vorschlag umfasst die Analyse einer möglichen **Verknüpfung des bilateralen und des regionalen Portfolios in EZ-Programmen sowie die Einbeziehung von globalen Vorhaben**.

Je nach Erkenntnisinteresse der Befragten werden unterschiedliche Durchführungszeitpunkte für EZ-PEs als nützlich identifiziert. Die Mehrheit erhofft sich jedoch Evaluierungsergebnisse für das jeweils noch laufende EZ-Programm, insbesondere für die vorgesehene EZ-Programmreflexion nach spätestens drei Jahren, oder vorbereitend für die Planung eines Folgeprogramms. Abhängig von der Programmdauer und vom Programmumfang wäre dies als **Zwischenevaluierung oder -review** durchzuführen. Die Untersuchung in Form

8 In diesem Kapitel wird zunächst das Erkenntnisinteresse hinsichtlich EZ-PEs vorgestellt. Dies beinhaltet jedoch keine detaillierte Einschätzung, mit welchem Aufwand diese beantwortet werden könnten (Voraussetzungen, Herausforderungen).

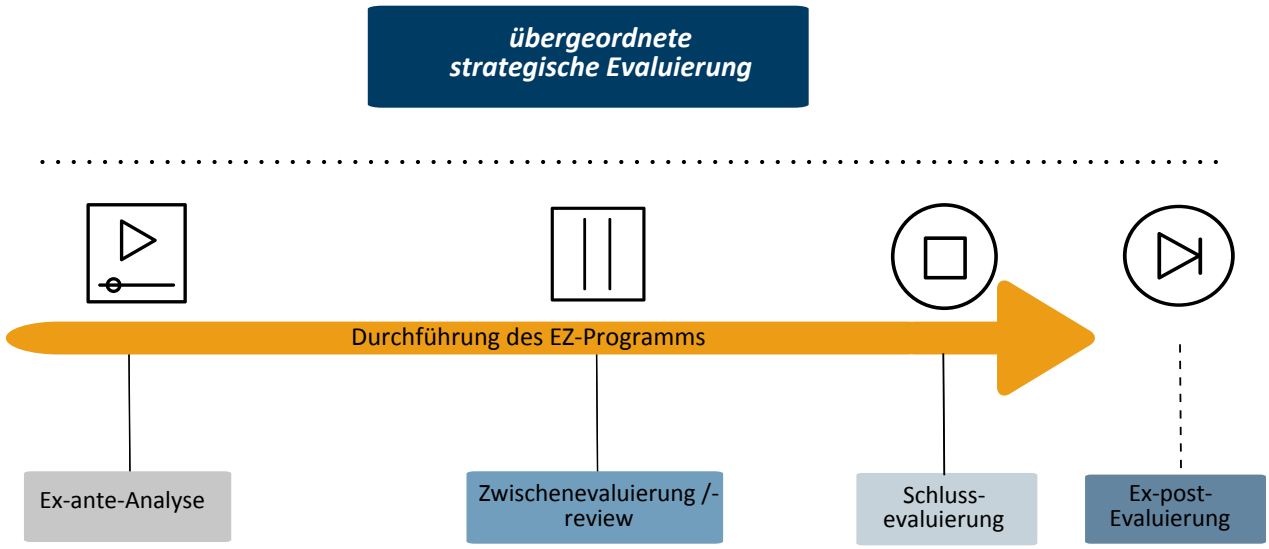
9 Im Rahmen der Onlinebefragung wird direkt nach dem Erkenntnisinteresse zu den sechs OECD-DAC-Evaluierungskriterien (Relevanz, Effektivität, Effizienz, Wirkung, Nachhaltigkeit und Kohärenz) gefragt. In den Interviews wird der Mehrwert und das Erkenntnisinteresse für EZ-PEs offen abgefragt.

einer **Ex-ante-Analyse** wird von mehr als einem Drittel befürwortet. In den Interviews geben die Befragten an, dass dabei die Evaluierungsergebnisse zum Zeitpunkt der EZ-Programmplanung vorliegen sollten.

Eine **Ex-post-Evaluierung** wird lediglich von einer Minderheit (11 Prozent) als erforderlich erachtet. Eine Schluss- oder Ex-post-Evaluierung kann die Zielerreichung eines EZ-Programms untersuchen und die Wirkungen eines EZ-Programms messen und würde somit für die Programmverantwortlichen die Rechenschaftsfunktion von EZ-PEs in den Vordergrund stellen. Übergeordnete Erkenntnisse zu den in EZ-Programmen eingesetzten Wirkungen und Funktionsweisen von Instrumenten sowie den inhaltlichen Fragestellungen könnten darüber hinaus zum programmübergreifenden Lernen beitragen.

Zudem wird von den Befragten die Option einer **übergeordneten strategischen EZ-Programmevaluierung** hervorgehoben, bei der Aspekte wie der BMZ-2030-Prozess, der Dreiklang sowie gewonnene Erkenntnisse für andere EZ-Programme untersucht werden könnten. Unter übergeordneter strategischer EZ-Programmevaluierung wird die Auseinandersetzung mit strategischen Fragen verstanden, die weniger für die Steuerung des einzelnen EZ-Programms, sondern für EZ-Programme insgesamt von Interesse sind. Auf die Evaluierungsmöglichkeiten und Unterschiede wird in Abschnitt 6.2 noch näher eingegangen.

Abbildung 4 Mögliche Zeitpunkte einer EZ-Programmevaluierung



Anmerkung: Die Zwischenevaluierung erhielt die größte Zustimmung (80 Prozent) durch die Befragten.
Quelle: eigene Darstellung

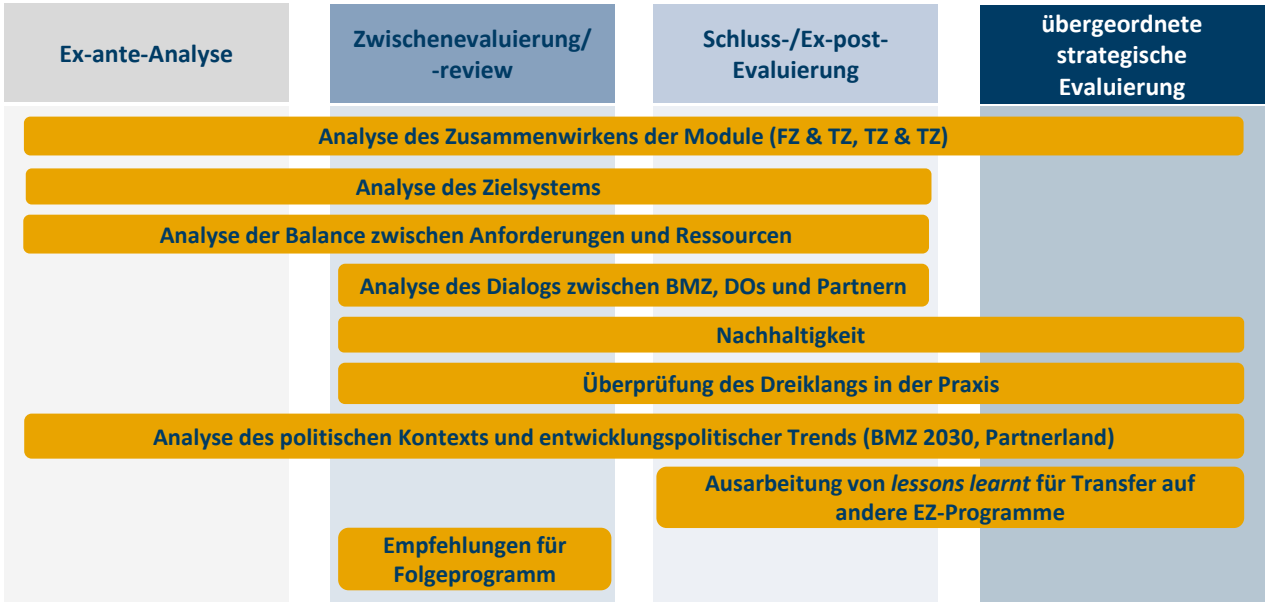
Die benannten Ziele und Zwecke lassen sich teilweise zu unterschiedlichen Zeitpunkten im EZ-Programmablauf untersuchen und können somit **verschiedenen Evaluierungstypen** zugeordnet werden (siehe Abbildung 5). Ein Großteil des Erkenntnisinteresses lässt sich jedoch durch verschiedene Evaluierungstypen zu mehreren Zeitpunkten der Implementierung oder in einer übergeordneten strategischen Evaluierung unabhängig vom Durchführungszeitpunkt einzelner EZ-Programme befriedigen.

Die Analyse des Zusammenwirkens von Modulen als zentraler Bestandteil von EZ-PEs kann zu allen Zeitpunkten einer EZ-Programmevaluierung sowie im Rahmen einer übergeordneten strategischen Untersuchung durchgeführt werden. Besonders gewinnbringend ist sie jedoch bei einer **Ex-ante-Analyse**, da hier Erkenntnisse zum bestmöglichen Zusammenwirken bereits in die EZ-Programmkonzeption einfließen können. Gleiches gilt für das Zielsystem, das während der EZ-Programmplanung und -durchführung analysiert werden kann, um zum einen Empfehlungen in der EZ-Programmsteuerung zu berücksichtigen und zum anderen – rückblickend – übergeordnete Fragen zum Zielsystem und zu bestehenden Herausforderungen zu beantworten. Die Erkenntnisse können zur Verbesserung von Monitoring- und Evaluierungssystemen inklusive der Formulierung von Indikatoren beitragen. Dies würde die strategischen Steuerungsmöglichkeiten erhöhen.

Die Auseinandersetzung mit dem Dialog zwischen BMZ, DOs und Partnern kann sowohl in einer **Zwischenevaluierung** wie in einer **Schluss- oder Ex-post-Evaluierung** stattfinden. Aspekte wie Nachhaltigkeit und die Überprüfung des Dreiklangs bieten sich für eine übergeordnete strategische Evaluierung als auch für eine Zwischen- und Schluss- oder Ex-post-Evaluierung an. Eine Analyse des politischen Kontexts und entwicklungspolitischer Trends (im Partnerland; BMZ 2030) könnte in verschiedenen Evaluierungstypen verortet werden. Um Empfehlungen für ein Folgeprogramm nutzen zu können, sollten diese bereits zur Planung dieses EZ-Programms vorliegen. Dies kann insbesondere durch Zwischenevaluierungen gewährleistet werden.

Das Zielsystem könnte auch nach Beendigung des EZ-Programms evaluiert werden, um die Zielerreichung messen zu können. Im Falle einer Schluss- oder Ex-post-Evaluierung stünden die Ergebnisse jedoch nicht rechtzeitig für die Planung eines sich anschließenden EZ-Programms zur Verfügung. Hier gilt zu eruieren, ob übergeordnete Erkenntnisse für EZ-Programme in einem ähnlichen Kontext oder Sektor generiert werden können. Ansonsten kann eine Beschäftigung mit gewonnenen Erkenntnissen für den Transfer auf andere EZ-Programme eher in einer übergeordneten strategischen EZ-PE erfolgen, da Erkenntnisse und Empfehlungen nicht unbedingt von den jeweiligen Programmverantwortlichen aufgegriffen werden können.

Abbildung 5 Erkenntnisinteresse von EZ-Programmevaluierungen nach Evaluierungstypen



Quelle: eigene Darstellung

6.2 Evaluierungspraktiken anderer Geber auf Länderebene

In den 1980er-Jahren begannen Geber, ihre EZ-Aktivitäten in Form von Länderprogrammen oder LS zu formulieren, um die nationalen Entwicklungsstrategien der Partnerländer besser zu berücksichtigen und zu adressieren (Dahlgren, 2007). **Evaluierungen von Länderprogrammen** werden aktuell von bilateralen und vor allem multilateralen Geberorganisationen durchgeführt.

Länderprogrammevaluierungen oder **Länderportfolioevaluierungen** anderer Geber bilden entweder das **gesamte Portfolio eines Gebers** ab und messen somit ihren Beitrag im Partnerland, oder es werden Evaluierungen auf Länderprogrammebene mit **thematischen Schwerpunkten** durchgeführt. Die EU hat beispielsweise standardisierte Evaluierungsformate entwickelt, bei denen mehrere Programme mit strategisch-thematischen oder strategisch-geografischen Schwerpunkten untersucht werden. Die Direktion für Entwicklung und Zusammenarbeit der Schweiz führt neben länderspezifischen Evaluierungen auch Untersuchungen regionaler Strategien durch. Das finnische Außenministerium untersucht indes mit seinen Policy and Strategic Evaluations Politikmaßnahmen (*policies*), LS, Finanzierungsinstrumente und thematisch breit angelegte Programme. Das britische Foreign, Commonwealth & Development Office führt wiederum the-

matische Syntheseevaluierungen auf Länderprogrammebene in fragilen Staaten durch. Eine Auflistung relevanter Geberorganisation und ihres Ansatzes zur Durchführung von Länderprogrammevaluierungen findet sich in Anhang 10.2.

Andere Geberorganisationen verfolgen **unterschiedliche Zielsetzungen** und bestimmen **verschiedene Adressat*innen** für ihre Länderprogrammevaluierungen. Häufig werden jedoch Evaluierungsansätze und **standardisierte methodische Herangehensweisen** für die Durchführung dieser Evaluierungen festgelegt. Im Folgenden werden beispielhaft die beiden unterschiedlichen Ansätze der Asian Development Bank (Asiatische Entwicklungsbank, ADB) sowie der Independent Evaluation Group (Unabhängige Evaluationsgruppe, IEG) der Weltbank und der United Nations Industrial Development Organization (Organisation der Vereinten Nationen für industrielle Entwicklung, UNIDO) vorgestellt.

- **Die Country Assistance Program Evaluation (CAPE)** der ADB deckt eine Zeitperiode von sieben bis zwölf Jahren und sämtliche Aktivitäten im jeweiligen Land ab. Ziel der Evaluierung ist eine Relevanzanalyse der Country Partnership Strategy (CPS) und die Überprüfung, inwieweit die Strategie durch ADB-Beiträge verfolgt wurde. Sofern die Beiträge in der CAPE als nicht ausreichend deklariert werden, erfolgt eine Relevanzprüfung des Programms. Eine CAPE wird vor der Erstellung jeder neuen CPS durchgeführt, um Evaluierungsergebnisse inklusive gewonnenen Erkenntnissen und Empfehlungen mit einfließen zu lassen. Es besteht ein standardisierter Prozess zur Entwicklung von CAPEs:
    - Identifikation der Ziele, des Umfangs sowie des Zeitraums des Programms
    - Entwicklung eines Evaluierungsgdesigns inklusive zentraler Evaluierungsfragen, der Anwendung von Evaluierungskriterien und der Darstellung der Bewertungsskala (in der Regel 20 Prozent je OECD-DAC-Kriterium: Relevanz, Effektivität, Effizienz, Nachhaltigkeit, Impact)
    - Planung finanzieller und personeller Ressourcen, Ausarbeitung von Evaluierungsinstrumenten wie Dokumentenstudien, Datenanalysen, Interviews und Konsultationen, Feldbeobachtung, Fragebogen und Fallstudien
    - keine standardisierte Methodik, häufig jedoch theoriebasierte Evaluierungen mit einer Wi-Lo (ADB, 2015)

- Von der IEG der Weltbank und der UNIDO werden **Cluster (Country) Program Evaluations** durchgeführt. Die IEG nutzt dieses Evaluierungsformat seit drei Jahren, wobei meist vier Länder in einem Cluster zusammengefasst werden. Für die Identifizierung relevanter Querschnittsergebnisse erfolgt die Länderauswahl in einem ex ante identifizierten Analyserahmen (Independent Evaluation Group, 2021). Bisher wurden beispielsweise kleine Staaten (Dschibuti, Kap Verde, Mauritius, Seychellen) oder Länder als Cluster ausgewählt, die reich an nicht erneuerbaren natürlichen Ressourcen und gleichzeitig abhängig von deren Ausbeutung sind (Bolivien, Kasachstan, Mongolei, Sambia). Ziel ist es, ein besseres länderübergreifendes Lernen zwischen Ländern und Regionen mit ähnlichen Herausforderungen zu etablieren sowie die Abdeckung und Effizienz von Evaluierungen zu verbessern (IEG, 2015, 2016).

  Die UNIDO führt Projekt- als auch Programmevaluierungen auf Länderebene durch. Die Evaluierungen umfassen einen Projektzeitraum zwischen sechs und zehn Jahren. Statt Projektevaluierungen zusammenzufassen, hat die UNIDO Clusterevaluierungen eingeführt, um Projekte thematisch zu gruppieren. Der Clusteransatz wurde aufgrund der hohen Anzahl von Projekten (durchschnittlich zehn Projekte) eines Portfolios ausgewählt. Die Evaluierungen sind theoriebasiert und untersuchen auf Grundlage expliziter Wirkungsannahmen das Engagement der Geber im Partnerland. Zunächst werden im Inception-Report Evaluierungsfragen, Informationsquellen sowie die Methodik zur Datenerhebung und -analyse festgehalten. Im Anschluss werden Evaluierungsergebnisse so weit wie möglich aggregiert, um mithilfe von Clustern Muster zu identifizieren. Zudem sollen in der Evaluierung insbesondere nicht intendierte Wirkungen aufgezeigt werden (UNIDO, 2016).

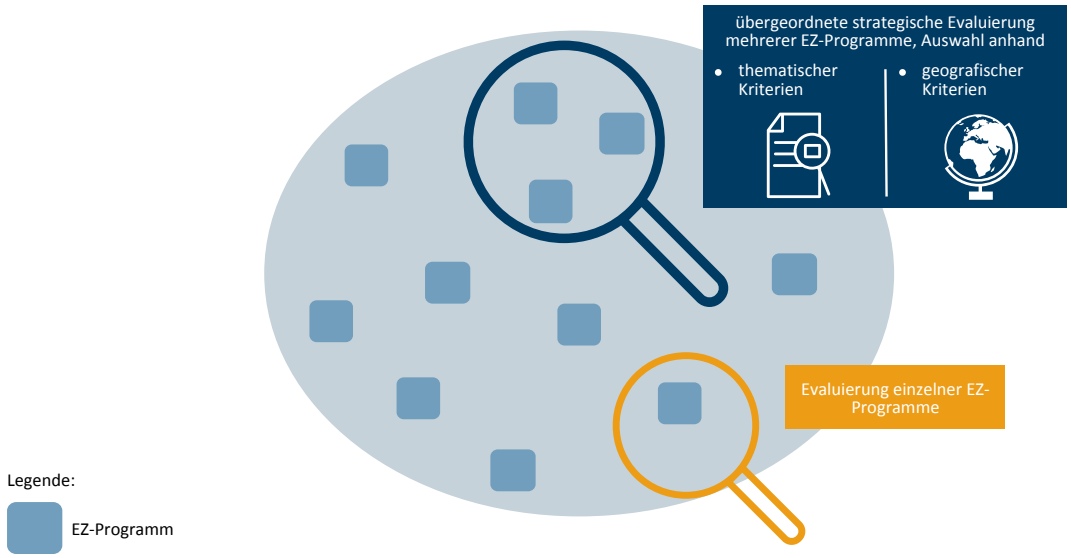
Die Länderprogrammevaluierungen anderer Geberorganisationen weisen einen **hohen Grad an Gemeinsamkeiten** auf. Sie operieren mit sich meist an den OECD-DAC-Kriterien orientierenden Evaluierungsfragen, nutzen Programmtheorien/Theorien des Wandels (ToC) und sind, zumindest teils, durch Vorgaben standardisiert. So gibt zum Beispiel das Independent Evaluation Office des United Nations Development Programme (Entwicklungsprogramm der Vereinten Nationen, UNDP) feste Evaluierungsfragen vor (IEO, o. J.). Län-

derprogrammevaluierungen werden in der Regel als Schlussevaluierungen durchgeführt, mitunter auch als Zwischenevaluierungen. Zentrale Funktionen umfassen die Herausstellung von gewonnenen Erkenntnissen und deren Nutzung für Folgeprogramme, allgemein für die Programmstruktur oder zur Überprüfung beziehungsweise (Weiter-)Entwicklung einer LS. Zudem werden sie ebenfalls zum Zweck der Rechenschaftslegung realisiert.

Möglichkeiten für die Evaluierung von Programmen in der deutschen EZ

Im Gegensatz zum vorherigen Abschnitt, in dem die Ansätze und Erfahrungen anderer Geber im Rahmen von Länderprogrammevaluierungen, das heißt von Evaluierungen, die das gesamte Portfolio eines Gebers betrachten, vorgestellt wurden, sollen beim geplanten Evaluierungsformat für die deutsche EZ zunächst EZ-Programme beleuchtet werden. Aus den dargestellten Vorgehensweisen lassen sich aber auch zwei Möglichkeiten für Evaluierungen auf dieser Ebene ableiten: übergeordnete strategische EZ-Programmevaluierungen sowie Evaluierungen einzelner EZ-Programme.

Abbildung 6 Evaluierungsmöglichkeiten von EZ-Programmen



Quelle: eigene Darstellung

Die vorgestellten Beispiele deuten auf **zwei grundsätzliche Möglichkeiten**, wie bilaterale EZ-Programme in der deutschen EZ evaluiert werden könnten:

i) **übergeordnete strategische Programmevaluierungen**, in denen strategische Fragestellungen untersucht werden, die von übergeordnetem Interesse sind, und möglichst im Kontext mehrere EZ-Programme berücksichtigen, um Erkenntnisse für die zuständigen Grundsatz- und Sektorreferate und allgemein für politisch-strategische Entscheidungen zu liefern. Zudem können Ergebnisse für die Verbesserung der strategischen Steuerung auf Programmebene insgesamt von Nutzen sein. Zur Durchführung dieser Evaluierungen erscheint der vorgestellte Clusteransatz sinnvoll.

ii) **Evaluierungen einzelner EZ-Programme**, die konkrete und detaillierte Ergebnisse zu einem EZ-Programm liefern und in den EZ-Programmzyklus integriert werden können, um von BMZ-Regionalreferaten und DO-Verantwortlichen für EZ-Programme für die Steuerung oder Planung von aktuell laufenden und zu planenden EZ-Programmen genutzt zu werden (siehe Abbildung 6).

Je nach Möglichkeit ergeben sich somit **Unterschiede im Nutzen und mit auf Blick auf die primären Adressat*innen** der Evaluierung. Es sollte geklärt werden, inwiefern Evaluierungsergebnisse mit Partnern geteilt werden. Bei der Entwicklung eines Evaluierungsformats von EZ-PEs sollte zunächst zwischen den beiden Möglichkeiten abgewogen werden.

Darüber hinaus ergeben sich **Konsequenzen für die Auswahl der zu betrachtenden EZ-Programme**. Bei untergeordneten strategischen Evaluierungen würde die Auswahl der EZ-Programme anhand geografischer

oder thematischer Kriterien erfolgen. So könnten entsprechende Cluster identifiziert werden mit dem Ziel des programm- beziehungsweise länderübergreifenden Lernens. Hier wäre es auch möglich, EZ-Programme im gleichen Kernthema oder Aktionsfeld in einem Cluster zu betrachten. Die Evaluierungen könnten die Übertragbarkeit der Ergebnisse erhöhen, und für Strategiefragen auf EZ-Programmebene würde somit eine bessere Evidenzbasis geschaffen werden. Bei Evaluierungen einzelner EZ-Programme müsste entweder ein bestimmter Zeitpunkt im Programmzyklus festgelegt oder nach Bedarf und Interesse der zuständigen Regionalreferate ausgewählt werden. Denkbar für die Evaluierung einzelner Programme ist eine Vollerhebung, eine stichprobenbasierte, eine kriterienbasierte oder eine nachfrageorientierte Auswahl.

7. ERKENNTNISSE ZUR WIRKUNGSORIENTIERUNG UND EVALUIERBARKEIT VON EZ-PROGRAMMEN

In den verschiedenen Phasen eines EZ-Programms werden jeweils unterschiedliche Voraussetzungen zur Wirkungsorientierung und Evaluierbarkeit geschaffen. In diesem Kapitel wird zunächst die Berücksichtigung der Wirkungsorientierung in der Programmkonzeption und -implementierung von EZ-Programmen untersucht. Die Analyse orientiert sich dabei an den in Kapitel 3 dargelegten Elementen von Holzapfel und Römling (2020) zum RBM-Ansatz. Anschließend werden die Erkenntnisse synthetisiert und ihre Bedeutung für die Evaluierbarkeit von EZ-Programmen übertragen.

7.1 Wirkungsorientierung

7.1.1 EZ-Programmkonzeption

In der Konzeptionsphase eines EZ-Programms soll eine Auseinandersetzung mit den Kontextbedingungen sowie den strategischen Zielvorgaben des BMZ und des Partnerlandes (unter anderem Ziele der Agenda 2030) erfolgen. Außerdem sollen das Zielsystem (Programmziel, Indikatoren) und inhaltliche Schwerpunkte auf EZ-Programmebene festgelegt werden. Dabei geht es insbesondere darum, die intendierte gemeinsame Wirkungsentfaltung über das Zusammenwirken der Module explizit und nachvollziehbar darzustellen.

Zieldefinition

Die Zielformulierung erfolgt in der Programmkonzeption gemeinsam zwischen BMZ und DOs und unter Berücksichtigung der Absprachen mit der Partnerregierung, jedoch ohne deren ausdrückliche Beteiligung. Programmziele sollen laut den Vorgaben die übergeordneten entwicklungspolitischen Ziele aus dem Strategiedokument (Impact-Ebene) reflektieren und somit das längerfristige Ergebnis aller Module darstellen. In der Praxis stellt sich heraus, dass es schwierig ist, die Balance zwischen der langfristig angestrebten (aber nur bedingt beeinflussbaren) Wirkung einerseits und der Formulierung operativer und messbarer Ziele auf der mittelfristigen EZ-Programmebene andererseits zu finden.

In den untersuchten Fallstudien werden in der EZ-Programmkonzeption die EZ-Programmziele ausgearbeitet. Die Analyse der EZ-Programmbeispiele macht jedoch deutlich, dass diese **unterschiedlich konkret formuliert sind**. Hierfür scheinen **ungenaue Formulierungen in den Vorgaben,** aber auch **individuell unterschiedliche Ansprüche der Verantwortlichen** ursächlich zu sein. Es besteht vor allem Uneinigkeit darüber, ob EZ-Programme in einem bestimmten Zeitraum auf ein konkretes, messbares Ziel hinarbeiten sollten oder es sich um einen bestmöglich abgestimmten Beitrag der bilateralen deutschen EZ handeln sollte, der sich an langfristigen entwicklungspolitischen Zielen des Partnerlandes orientiert.

Auch die Befragung der Länderreferent*innen im BMZ hat die Auswertung der Fallstudien bestätigt und deutet auf ein sehr **heterogenes Anspruchsniveau an das EZ-Programmziel** hin:

- Etwa ein Drittel der Befragten gibt an, dass ein EZ-Programmziel grundsätzlich so formuliert sein sollte, dass es im festgelegten Zeitraum durch das Zusammenwirken der Module erreichbar ist (33 Prozent „trifft voll und ganz zu“, 17 Prozent „trifft größtenteils zu“, 6 Prozent „weder noch“, 44 Prozent „trifft weniger zu“).
- Knapp die Hälfte der Befragten gibt an, dass ein EZ-Programmziel grundsätzlich so formuliert sein sollte, dass es zu entwicklungspolitischen Zielen des Partnerlandes beiträgt und daher nur im Zusammenwirken mit anderen Gebern und Akteuren erreichbar ist (28 Prozent „trifft voll und ganz zu“, 22 Prozent „trifft größtenteils zu“, 5 Prozent „weder noch“, 28 Prozent „trifft weniger zu“, 17 Prozent „weiß nicht“).

Solche unterschiedlichen Anspruchsniveaus an EZ-Programmziele und ihre Erreichbarkeit unter den Programmplanenden und -verantwortlichen haben in der Folge auch Auswirkungen auf die Steuerung, die Implementierung und das Monitoring von EZ-Programmen. Es deutet zudem darauf hin, dass zum Planungszeitpunkt kein gemeinsames Verständnis darüber besteht, wann EZ-Programme konkret als Erfolg angesehen und beurteilt werden und wann nicht.

Festlegung von Programmzielindikatoren und Datenquellen

Die Vorgaben sehen vor, dass alle EZ-Programmziele durch mehrere (in der Regel drei bis fünf) EZ-Programmzielindikatoren zu operationalisieren sind. Um die **Qualität der Indikatoren** sicherzustellen, sollen die SMART-Kriterien berücksichtigt werden. Darüber hinaus sollen möglichst nationale Indikatoren der Partnerstrategien genutzt werden. Erhebungen von gezielten Daten durch die Akteure der deutschen EZ sind nur in Ausnahmefällen vorgesehen.

Die Analyse der Fallstudien und die Gespräche mit Verantwortlichen aus Grundsatzabteilungen machen deutlich, dass die Festlegung von Indikatoren in der Praxis als sehr herausfordernd erachtet wird. Dies liegt nicht nur an den schwer zu operationalisierenden Zielformulierungen, sondern auch am **Mangel an erforderlichen Daten** zur Messung der Indikatoren.

Zum Monitoring der Indikatoren sollen möglichst Partnerdaten herangezogen werden. Doch die Qualität und die Beschaffung von entsprechenden Daten stellen eine große Herausforderung dar. Diese stehen häufig nicht rechtzeitig, in ausreichender Qualität oder im erforderlichen Aggregationsniveau zur Verfügung, sodass die Daten nur selten eine umfassende Bewertung des EZ-Programms und die Festlegung von konkreteren Indikatoren erlauben. Die Schwierigkeiten in der Festlegung passender Indikatoren erschweren zudem die Messung von EZ-Programmfortschritten und der Zielerreichung anhand ebendieser Indikatoren.

In der Praxis erfüllen die zum Einsatz kommenden EZ-Programmzielindikatoren nicht immer die SMART-Kriterien. Überdies gelingt es nicht immer, die EZ-Programmziele durch Indikatoren plausibel zu operationalisieren. Die EZ-Programmzielindikatoren in den Fallstudien sind zu 80 Prozent wertbestückt. Das ist zwar die Mehrheit, bedeutet aber auch, dass die Basiswerte in den anderen 20 Prozent der EZ-Programme fehlen. Außerdem sind anvisierte Fortschritte und Zielwerte, die in der EZ-Programmimplementierung zu überprüfen sind, zum Teil nicht mit Indikatoren unterlegt, wodurch ein gängiges Mittel zur Erfolgsüberprüfung fehlt.

Erarbeitung einer Programmtheorie

Die zugrunde liegende Wirkungsannahme von EZ-Programmen ist, dass diese ausschließlich durch ihre Module und deren Wechselbeziehungen übergeordnete Wirkungen entfalten (siehe Abbildung 2) oder gezielte Einzelbeiträge leisten. In der EZ-Programmkonzeption soll entsprechend den Vorgaben eine frühzeitige Auseinandersetzung mit den intendierten Wirkungen eines EZ-Programms erfolgen: Die wesentlichen Inhalte und die Programmtheorie sollen sowohl im Narrativ als auch grafisch dargestellt werden.

Die Analyse der Vorgaben und der Umsetzung im Rahmen von drei Fallbeispielen macht jedoch deutlich, dass Wirkungsannahmen und Wechselbeziehungen zwischen den Modulen derzeit nicht ausreichend und nachvollziehbar dargestellt werden. Insbesondere das zur Wirkungsentfaltung **erforderliche Zusammenwirken wird zu wenig detailliert** erläutert. Es wird zwar darauf hingewiesen, dass Synergien und ein Zusammenwirken angestrebt werden, es ist aber meist nicht vollständig und plausibel nachvollziehbar, wie die Module miteinander verzahnt werden und sich inhaltlich in ein EZ-Programm einfügen sollen, um gemeinsam übergeordnete Veränderungen anzustoßen. Dies begründet sich unter anderem durch die unterschiedlichen Beauftragungsprozesse und Laufzeiten der Module (siehe Kasten 2), durch den jeweiligen Sektor, in dem das EZ-Programm verankert ist, sowie auch durch Rahmenbedingungen im Partnerland. Vor allem in volatilen Kontexten sind mehr Flexibilität und kurzfristige Anpassungen erforderlich. Dies stellt sowohl für die Steuerung über den Dreiklang wie für die Umsetzung von Synergien wie für die Gewährleistung des Zusammenwirkens der Module eine große Herausforderung dar.

Daraus resultierend geht die Wirkungsbeziehung der Modul- und der EZ-Programmebene häufig nicht klar hervor, sodass nicht nachvollziehbar ist, **wie genau sich die Wirkungen auf EZ-Programmebene entfalten sollen**. Die Darstellung der Programmtheorie ist derzeit wenig aussagekräftig, da in den drei Fallstudien fast alle Modulziele mit fast allen EZ-Programmzielen verbunden sind. Es findet keine inhaltliche Auseinandersetzung oder ergänzende Erläuterung der dahinterliegenden **Wirkungsannahmen oder -hypothesen** statt, aus denen hervorgeht, welche kausalen Zusammenhänge zwischen der Erreichung des Modulziels und der Erreichung des EZ-Programmziels bestehen.

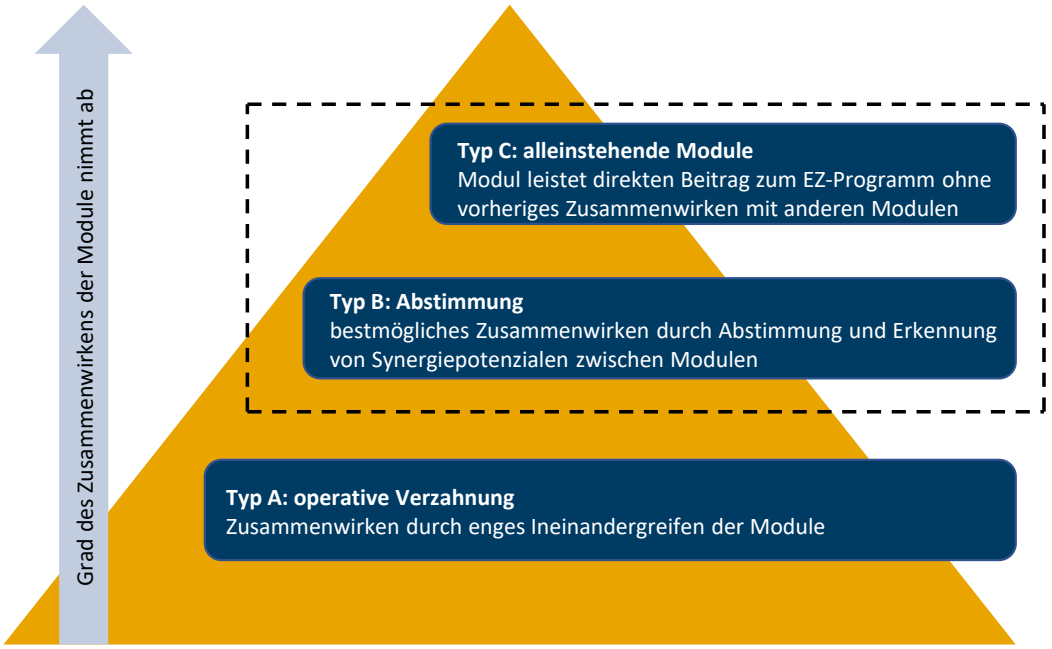
Kasten 2 Erkenntnisse zum Zusammenwirken der Module im EZ-Programm

Die Analyse der drei Fallstudien zeigt, dass die in Kapitel 3 dargelegten Möglichkeiten zur Erreichung des Zusammenwirkens der Module in der Praxis nicht entsprechend berücksichtigt werden. Die untersuchten Programmdokumente deuten darauf hin, dass kein einheitliches Verständnis zum Zusammenwirken der Module besteht und sehr unterschiedliche Begrifflichkeiten verwendet werden, um dieses zu beschreiben. So wird in den Programmdokumenten vom „Zusammenwirken", „der Verzahnung", der „Bildung von Synergien" und von „Kooperationen zwischen den DO[s] beziehungsweise de[n] Module[n]" gesprochen.

In der Praxis begreifen die DOs das Zusammenwirken der Module – sowohl innerhalb der TZ als auch zwischen TZ und FZ – primär als bestmögliche Abstimmung zwischen den DOs und einem entsprechenden Austausch darüber (Typ B, siehe Abbildung 27). Zudem finden sich in den untersuchten EZ-Programmen jeweils einzelne Module, die gänzlich unabhängig von weiteren Modulen im Programm wirken (Typ C). Im Fall von Modulen in verschiedenen Handlungsfeldern, die gemeinsam auf ein Programmziel hinwirken sollen, werden die konkreten Aktivitäten zwar oftmals sehr detailliert aufgeschlüsselt, die Module werden aber eher als parallel denn komplementär dargestellt, und es fehlen konkrete Schnittstellen zwischen ihnen. Dementsprechend wird das als idealtypisch beschriebene Ineinandergreifen der Module (FZ & TZ, TZ & TZ), wie es Möglichkeit A in der Umsetzung vorsieht, meist als unrealistisch und nicht praktikabel bezeichnet.

Die Interviews bestätigen die Erkenntnisse aus der Analyse der Programmdokumente. Das Zusammenwirken wird in der Praxis als nicht ausreichend, teilweise sogar als nicht realisierbar empfunden. In dieser Hinsicht wird angeführt, dass synergetische Wirkungen in der Theorie zwar möglich sind, der tatsächliche Grad des Zusammenwirkens aber stark kontextabhängig ist und dieser nur selten eine enge operative Verzahnung ermöglicht. Eine Herausforderung hierfür stellt neben den grundlegenden Kontextbedingungen im Partnerland unter anderem der unterschiedliche Beauftragungsprozess der FZ und der TZ mit unterschiedlichen Zeithorizonten dar. In der Folge ist die Implementierung der Module nicht immer bestmöglich aufeinander abgestimmt und die systematische Identifikation und Realisierung von Synergien wird als herausfordernd verstanden.

Abbildung 7 Grade des Zusammenwirkens der Module in der Praxis



Anmerkung: Typ A ist in den Handreichungen das präferierte Verzahnungsmodell. In der Praxis sind jedoch vorwiegend die Typen B und C vorzufinden.
Quelle: eigene Darstellung

Darstellung der Inhalte eines EZ-Programms

In der EZ-Programmkonzeption sollen entsprechend den Vorgaben die **inhaltlichen Schwerpunkte** des EZ-Programms unter Berücksichtigung bestehender strategischer Richtlinien gesetzt werden. Vorgesehen ist zudem, wie auf **Entwicklungsbedarfe und -strategien der Partner** eingegangen wird, sowie die Festlegung von **Zielgruppen** und möglichen **räumlichen Schwerpunkten**. Des Weiteren sollen laut den Vorgaben mögliche **Einflussfaktoren** und Risiken beschrieben werden, die dazu führen können, dass die gesteckten EZ-Programmziele nicht erreicht werden, zum Beispiel Schwierigkeiten in der Zusammenarbeit mit Partnern oder Regierungswechsel.

In den untersuchten Fallstudien zeigt sich, dass die Strategiedokumente der Partner herangezogen werden, um die entsprechenden Bedarfe abzuleiten und den Schwerpunkt zu identifizieren. Zudem werden die Entwicklungsherausforderungen des Partnerlandes und deren Konsequenzen für bestimmte Zielgruppen beschrieben. Die Analyse macht dabei deutlich, dass nicht immer nachvollziehbar ist, ob die beabsichtigten Veränderungen für die Zielgruppen innerhalb der Laufzeit des EZ-Programms tatsächlich erreicht werden können. Die Umsetzung aller Module eines EZ-Programms in vorab vereinbarten Regionen erweist sich zum Teil als schwierig, weil nicht immer die erforderlichen Voraussetzungen gegeben sind, wie das Vorfinden von qualifizierten Partnerbanken, Verbänden oder Trainingsinstitutionen zur Entwicklung und Einführung von bestimmten Angeboten und Dienstleistungen.

Annahmen und Risiken

Annahmen und Risiken werden vorwiegend im Hinblick auf externe, nicht durch das EZ-Programm verursachte Faktoren beschrieben. Es mangelt an einer Reflexion möglicher unbeabsichtigter sowohl positiver wie negativer Effekte (nicht intendierter Wirkungen), die das EZ-Programm für bestimmte Zielgruppen, Räume oder auch Sektoren hat.

Rollentrennung der Akteure auf Geberseite

In den Vorgaben ist festgeschrieben, dass das BMZ die Steuerung und Verantwortung von EZ-Programmen übernimmt. Die DOs sind dagegen für die Vorbereitung und eigenverantwortliche Umsetzung der Module innerhalb des EZ-Programms zuständig. Konkrete Verfahrensschritte werden vorgegeben, um den jeweils Verantwortlichen und auch den beteiligten Akteuren eine **Orientierung und Hilfestellung zur Implementierung des jeweiligen EZ-Programms** zu bieten. Solche Angaben beziehen sich unter anderem auf die Übertragung von Zuständigkeiten, die Nutzung von Formaten oder die Zeitpunkte zur Einbindung spezifischer Akteure (Bundesregierung, 2019)**.**

In der Praxis zeigt sich, dass grundlegende Zuständigkeiten und Rollen in den Hauptphasen der EZ-Programmplanung und -umsetzung (das heißt zur Erarbeitung der Formate für Konzeption oder Berichterstattung) eindeutig festgelegt sind. Für die Umsetzung einzelner Prozessschritte sind dennoch zum Teil **weitere Klärungen erforderlich**, da Unklarheiten zu Unstimmigkeiten und teilweise auch Qualitätsverlusten führen. Dies betrifft die Rollenklärung sowohl zwischen BMZ und DOs als auch zwischen den DOs und kommt insbesondere in der operativen EZ-Programmsteuerung, -koordination und Qualitätssicherung zum Tragen, beispielsweise hinsichtlich der finalen Beurteilung und Abnahme von EZ-Programmzielindikatoren gemäß den bestehenden Qualitätskriterien.

Partnereinbindung

Es ist vorgesehen, mit EZ-Programmen zu übergeordneten und langfristigen entwicklungspolitischen Veränderungen im Partnerland beizutragen. Nationale Strategien des Partnerlandes werden dabei in der EZ-Programmkonzeption berücksichtigt. In den Fallstudien gibt es auf Modulebene häufig eine Reihe von bereits laufenden Maßnahmen und Zusagen sowie enge Arbeitsbeziehungen mit Durchführungspartnern, die es für die EZ-Programmkonzeption zu berücksichtigen gilt. Somit findet ein enger **Austausch mit Partnern über die Module** statt. Langfristige Zielsetzungen und Zusagen werden zudem in Regierungsverhandlungen und -konsultationen mit der Partnerregierung besprochen. EZ-Programme selbst stellen jedoch ein **internes Planungs- und Steuerungsinstrument** dar und werden in der Regel nicht mit den Akteuren der Partnerlän-

der diskutiert. Dies erweist sich insbesondere dann als problematisch, wenn Partnerinteressen und -wünsche zum Beispiel hinsichtlich des Einsatzes von FZ und TZ sowie regionaler Schwerpunkte nicht in Einklang mit den zugrunde liegenden Annahmen für EZ-Programme (Programmziele sind nur durch das Zusammenwirken der Module erreichbar) sind.

7.1.2 EZ-Programmimplementierung

Während der EZ-Programmimplementierung erfolgen regelmäßige und umfangreiche Berichterstattungen durch die DOs an das BMZ. Zweck der Berichte ist es, das BMZ über aktuelle und sich ändernde Rahmenbedingungen des EZ-Programms, Inhalte, EZ-Programmfortschritte und Herausforderungen zu informieren. Falls nötig, soll auf dieser Grundlage steuernd eingegriffen werden, um die Erreichung der Ziele weiterhin sicherzustellen. Folgende Aspekte zur Wirkungsorientierung in der EZ-Programmimplementierung ergänzen die bereits aufgeführten Punkte aus der EZ-Programmkonzeption.

Analyse der Wirkungstheorie und des Zielsystems hinsichtlich EZ-Programmfortschritten

Im Rahmen der Berichte sind die DOs entsprechend den Vorgaben dazu aufgefordert, eine vollständige Beschreibung des **EZ-Programmfortschritts**, eine nachvollziehbare Darstellung des **Handlungsbedarfs** zur EZ-Programmsteuerung und zum **Zusammenwirken der Module** sowie Hinweise auf die **Kohärenz** des EZ-Programmziels mit Blick auf strategische Zielvorgaben an das BMZ zu übermitteln.

Bereits in der EZ-Programmkonzeption deutet sich an, dass in der Praxis die Beschäftigung mit den Synergien und Wirkungen auf EZ-Programmebene nicht intensiv genug ist. Auch während der Implementierung und in dazugehörigen Berichterstattungen findet im Rahmen der Fallstudien **keine angemessene Auseinandersetzung** mit und Überprüfung von Wirkungsannahmen statt. Vielmehr ist vornehmlich eine Befassung mit den Inhalten und Beiträgen einzelner Module festzustellen. Der wichtige Aspekt der gemeinsamen Wirkungsentfaltung und der Veränderungen auf EZ-Programmebene wird hingegen zu wenig ausdifferenziert und erläutert. Dementsprechend kann nur selten nachvollzogen werden, ob Synergien zwischen Modulen ausreichend genutzt werden und sie den erhofften Mehrwert bringen, um gemeinsam und effizient auf das EZ-Programmziel hinzuwirken.

Bestückung von Indikatoren (Ist-Werte) und Verfügbarkeit von Monitoringdaten

Entsprechend den Vorgaben ist eine tabellarische Auflistung der Indikatorwerte (Ist-Werte, angestrebte Zielwerte, Einschätzung der Erreichbarkeit) erforderlich. Zudem sollen die vorgesehenen Datenquellen zur Messung der Indikatoren dargelegt werden.

In den untersuchten Fallstudien fehlt es häufig sowohl an **Primär- als auch an Sekundärdaten** in den Berichterstattungen. Eigene Erhebungen durch das EZ-Programm finden kaum statt, da keine expliziten Ressourcen dafür vorgesehen sind. Demnach besteht die Datenbasis für die Indikatoren nur in seltenen Fällen aus durch die deutsche EZ erhobenen Primärdaten. In den drei Fallstudien wird daher in unterschiedlichem Maße auf Sekundärdaten (Studien und Berichte) zurückgegriffen.

Die Beschreibungen basieren teilweise auf externen wissenschaftlichen Erkenntnissen, Ergebnisse von Evaluierungen werden hingegen nicht angemessen berücksichtigt. Die in einigen Fällen nicht planmäßig oder nur in unzureichender Qualität überlassenen Sekundärdaten stammen von Partnerinstitutionen, was zu **Datenlücken** führt. In allen Fallstudien wird in unterschiedlichem Umfang auf fehlende Statistiken und mangelnde Daten von Partnern verwiesen. Gezielte Absprachen zum Datenaustausch finden nicht oder zu selten statt. Das Monitoring auf EZ-Programmebene reicht in den betrachteten Fallstudien folglich nicht aus, um aktuelle Informationen zu den EZ-Programmzielindikatoren bereitzustellen. EZ-Programmfortschritte lassen sich aufgrund teils fehlender Basiswerte aus EZ-Programmkonzeption und/oder fehlender Ist-Werte aus der EZ-Programmimplementierung nicht vollständig darstellen.

Die unzureichende Datenverfügbarkeit stellt somit eine enorme **Herausforderung für die Messung** des EZ-Programmfortschritts dar. Neben der erforderlichen Verbesserung des Datenaustauschs mit den Partnern ist auch ein Monitoring auf EZ-Programmebene nötig. In den analysierten Fällen wurde in der Konzeption

und frühen Implementierung nicht ausreichend thematisiert, wie mit Datenlücken umgegangen werden soll. Zudem stehen keine eigenen Ressourcen für Erhebungen durch das EZ-Programm zur Verfügung. Dementsprechend ist auch nicht davon auszugehen, dass erforderliche Daten im weiteren Programmverlauf für eine vollständige Messung der EZ-Programmzielindikatoren generiert werden.

Risikoanalyse und -management

Die Vorgaben sehen konkret vor, Risiken der EZ-Programmzielerreichung zu antizipieren und einzuschätzen und in den Berichten fortlaufend Stellung hierzu zu beziehen. In der Praxis enthalten die EZ-Programmdokumente zur Implementierung Informationen zu den aktuellen **Einflussfaktoren** der EZ-Programmzielerreichung. Dazu zählen **kontextbezogene** (beispielsweise Covid-19) wie auch wirtschaftliche, politische sowie naturräumliche **Risiken**. Außerdem konnten in allen Fällen ebenso **interne Herausforderungen** identifiziert werden, zum Beispiel durch Verzögerungen in einzelnen Modulen oder durch unterschiedliche Prozesse in FZ und TZ. Diese internen Faktoren können die Zielerreichung zum Teil erheblich beeinflussen, werden aber in aktuellen Formaten zum Austausch und zur Berichtslegung zu wenig berücksichtigt.

Zu einem ähnlichen Ergebnis ist auch die Evaluierung des Abwasserprogramms in Vietnam gekommen. Dort wird darüber hinaus darauf verwiesen, dass besonders die internen Herausforderungen oft bereits zum Zeitpunkt der Planung absehbar gewesen wären (Meyer, 2019).

7.2 Beurteilung der Evaluierbarkeit

Die Beurteilung der Evaluierbarkeit von EZ-Programmen bezieht sich auf die von Leonard und Eulenburg (2012) identifizierten Bestandteile zur Messung/Bewertung der Evaluierbarkeit von Entwicklungsmaßnahmen (siehe Kapitel 4).

Klarheit, Spezifität und Plausibilität der angegebenen intendierten Wirkungen: Diese Voraussetzung für die Evaluierbarkeit ist bei EZ-Programmen aktuell nicht gegeben. Das Anspruchsniveau an das EZ-Programmziel ist derzeit sehr heterogen. Zunächst müsste ein einheitliches Verständnis erarbeitet werden, ob ein EZ-Programmziel im festgelegten Zeitraum erreicht werden muss und es im Zusammenwirken mit anderen Gebern und Akteuren auch zu entwicklungspolitischen Zielen des Partnerlandes beitragen soll. Ungenaue Vorgaben zur Formulierung des EZ-Programmziels führen dazu, dass Ziele unterschiedlich formuliert werden. Dies schränkt die Klarheit, die Spezifität und die Plausibilität des EZ-Programmziels erheblich ein. Intendierte Wirkungen werden in der Programmtheorie definiert, diese weist jedoch selbst noch Defizite auf (nicht ausreichend explizite Darstellung des Zusammenwirkens von Modulen sowie der Verknüpfung der Modul- und der EZ-Programmebene). Dies verschlechtert somit die Qualität der Beschreibung der intendierten Wirkungen. Nicht intendierte Wirkungen werden auf Ebene des EZ-Programms nicht angemessen berücksichtigt.

Vorhandensein und Qualität von Indikatoren und Daten für die Bewertung des Erreichens der intendierten Wirkungen: Diese Voraussetzung für die Evaluierbarkeit ist derzeit eingeschränkt gegeben. Es werden umfangreiche Informationen über die Ausgangssituation vor Beginn des EZ-Programms bereitgestellt, die Basiswerte derzeit allerdings nicht vollständig bestückt. Diese Datenlücken stellen auch für das Monitoring eine Herausforderung dar. Zudem sind die Verantwortlichkeiten und Prozesse nicht ausreichend geklärt. Dies ist jedoch für die Evaluierbarkeit von zentraler Bedeutung. Die drei untersuchten Fallstudien bestätigen diese Erkenntnisse. Darüber hinaus ist es erforderlich, dass Indikatoren vollständig den SMART-Kriterien entsprechen, um später Wirkungen messen zu können.

Angemessenheit der Strategie zur Risikoerkennung und -minderung: Dieser Aspekt wird bislang eingeschränkt berücksichtigt. Risiken beziehungsweise Einflussfaktoren, die der EZ-Programmzielerreichung entgegenstehen, werden insbesondere im Hinblick auf externe Faktoren (beispielsweise politische Umbrüche) berücksichtigt. Die Auseinandersetzung mit eher programminternen Faktoren wird dagegen nicht gleichermaßen vollzogen. Risikomanagementmaßnahmen werden zwar dargelegt, jedoch nicht fortwährend im Hinblick auf deren Erfolg und Nutzen reflektiert.

Angemessenheit der Bestimmungen für das Monitoring: Diese Voraussetzung ist aktuell häufig nicht gegeben, jedoch von zentraler Bedeutung, da die Evaluierbarkeit im hohen Maße von den Monitoringdaten abhängt. Zum einen gibt es Qualitätsunterschiede in den Ziel- und Indikatorenformulierungen, die als Ausgangspunkt für die Entwicklung von Monitoringsystemen eine Herausforderung darstellen. Zum anderen soll überwiegend auf Sekundärdaten von Partnern zurückgegriffen werden. Wie damit umzugehen ist, wenn diese nicht zum erforderlichen Zeitpunkt, in ausreichender Qualität oder in hinlänglicher Aggregation zur Verfügung stehen, ist unklar. Zusätzliche Ressourcen zur Erhebung von Daten für das EZ-Programmmonitoring gibt es meist nicht.

8. SCHLUSSFOLGERUNGEN UND EMPFEHLUNGEN

Ziele und Mehrwert von EZ-Programmen

In der Theorie des Dreiklangs kommt EZ-Programmen eine klare Funktion zu: Sie bilden die Schnittstelle zwischen den übergeordneten Vorgaben der LS und den einzelnen Modulen der DOs. EZ-Programme bün-deln die gemeinsam auf ein übergeordnetes Ziel hinwirkenden Module inhaltlich und sollen somit dem BMZ die Steuerung auf der Schwerpunkt- beziehungsweise der Mesoebene ermöglichen.

In der Praxis geschieht das aber nur bedingt. Zwar werden Vorgaben der LS übernommen und Module unter dem Dach des EZ-Programms gebündelt, jedoch findet bisher kaum eine strategische Planung und Steuerung über EZ-Programme statt. Die Implementierung der Maßnahmen und auch die Steuerung erfolgen auf Modulebene. Es ist meist nicht klar, wie gemeinsame Veränderungen durch das Zusammenwirken der Module eintreten sollen oder welche Aufgaben mit der Steuerung über EZ-Programme durch das BMZ einhergehen. Ebenso ist die Rolle der Partner unpräzise beschrieben. Sofern EZ-Programme dem politischen Dialog und kohärenteren Auftreten der deutschen EZ in einem Themenbereich dienen sollen, müsste der Partner über das Bestehen der EZ-Programme informiert werden.

Empfehlung 1: Das BMZ sollte Ziele und Zwecke der EZ-Programme und die damit verbundenen Steuerungsaufgaben eindeutiger definieren, um EZ-Programme insgesamt für alle Beteiligten sichtbarer zu machen und die Voraussetzungen für die strategische, ergebnisorientierte Steuerung über Schwerpunkte (zukünftig Kernthemen und Aktionsfelder) zu verbessern.

Umsetzungshinweis 1.1: Das BMZ sollte in Absprache mit den DOs prüfen, ob Ziele und Zwecke sowie Kohärenzansprüche von EZ-Programmen in bestimmten (fragilen) Kontexten oder Aktionsfeldern unterschiedlich definiert sein sollten.

Umsetzungshinweis 1.2: Die Interessen und Beiträge der Partner sollten stärker berücksichtigt werden, da sie für die erfolgreiche Umsetzung von EZ-Programmen wichtige Akteure sind.

Wirkungsorientierung und Evaluierbarkeit von EZ-Programmen

Voraussetzungen für eine stringente Ergebnis- oder Wirkungsorientierung sind konkrete und messbare Ziele, das Zusammenwirken der Module, nachvollziehbare Wirkungsannahmen, eine mit belastbaren Daten dargelegte Ausgangssituation des EZ-Programms sowie ein Monitoring. Diese werden durch die bestehende Praxis in der EZ-Programmkonzeption und-implementierung jedoch nur teilweise gewährleistet.

Die identifizierten Defizite in der EZ-Programmkonzeption, zum Beispiel ein unzureichend ausformuliertes Zielsystem, setzen sich in der Implementierung fort und erschweren sowohl die Monitoringaktivitäten als auch die Bewertung des EZ-Programmfortschritts im Programmverlauf. Insgesamt beeinträchtigen die festgestellten Mängel nicht nur die Steuerungsfähigkeit des BMZ, sondern auch die Evaluierbarkeit von EZ-Programmen. Sie sollten angegangen werden, um effektive Beiträge zur Umsetzung zum einen der UN-Agenda 2030 und zum anderen des Reformkonzepts „BMZ 2030“ zu gewährleisten. Die zu schaffenden Voraussetzungen pro Phase werden in Abbildung 8 zusammenfassend dargestellt.

Abbildung 8 Grundlagen zur Wirkungsmessung nach EZ-Programmphase

Quelle: eigene Darstellung

Empfehlung 2: Die Wirkungsorientierung und die Evaluierbarkeit von EZ-Programmen sollten durch das BMZ und die DOs verbessert werden, um dem BMZ steuerungsrelevante Informationen liefern zu können und eine aussagekräftige Überprüfung und Evaluierung von EZ-Programmen zu ermöglichen.

Die Verbesserungsbedarfe bestehen insbesondere in der EZ-Programmkonzeption und betreffen:

- **das Zielsystem von EZ-Programmen auf den verschiedenen Wirkungsebenen**
- **eine angemessene und nachvollziehbare Darstellung des angestrebten Zusammenwirkens und der Wechselwirkungen zwischen Modulen**
- **die Definition angemessener EZ-Programmzielindikatoren oder gegebenenfalls -zielgrößen**
- **die Verbesserung der Datenbasis**

Umsetzungshinweis 2.1: Um Aspekte zur Wirkungsorientierung und Evaluierbarkeit in der Planung und Steuerung zukünftig systematisch zu gewährleisten, sollten in der Erarbeitung entsprechender Prozesse und Verfahren die Evaluierungseinheiten des BMZ und der DOs miteinbezogen werden.

Auf die identifizierten Schwachstellen und Verbesserungsmöglichkeiten der einzelnen aufgeführten Punkte wird in den Empfehlungen 3 bis 5 näher eingegangen.

Heterogene Programmzielformulierungen

Die Untersuchung der Vorgaben und der Fallstudien deutet auf ein sehr heterogenes Verständnis und uneinheitliches Anspruchsniveau bezüglich der Formulierung von EZ-Programmzielen hin. Zum einen sollen EZ-Programmziele möglichst konkret und messbar sein und durch die Maßnahmen des EZ-Programms erreicht werden, das heißt primär durch die gemeinsame Entfaltung von Wirkungen der jeweiligen Module. Zum anderen werden EZ-Programme und EZ-Programmziele als abgestimmter deutscher Beitrag zu entwicklungspolitischen Veränderungen im Partnerland verstanden, die in Zusammenarbeit mit diversen Entwicklungs- und Durchführungspartnern sowie der Partnerregierung erreicht werden sollen. Dies führt in der Praxis zu sehr unterschiedlichen Zielformulierungen. Die Ziele selbst sind wiederum mittels Indikatoren unterschiedlich gut operationalisiert und damit messbar.

Es existieren zwar Qualitätsstandards für die Festlegung von EZ-Programmzielen und EZ-Programmzielindikatoren, diese werden aber momentan nicht einheitlich verstanden und umgesetzt. Zudem erfordert die Qualitätssicherung ein fachliches und methodisches Know-how. Gegenwärtig sind Zuständigkeiten nicht immer klar und die Qualitätssicherung wird von allen Beteiligten als herausfordernd beschrieben. Die Verbindlichkeit und Sicherstellung einheitlicher und gleichzeitig umsetzbarer Qualitätsstandards über alle EZ-Programme hinweg sind eine wichtige Voraussetzung für die Einführung eines standardisierten Evaluierungsformats. Andernfalls muss die methodische Vorgehensweise in EZ-PEs individuell festgelegt werden, wodurch es nur in unterschiedlichem Maße möglich wäre, die Erreichung von Zielen beziehungsweise Beiträgen beurteilen zu können.

Empfehlung 3: Das BMZ sollte verbindliche Vorgaben und Qualitätsstandards für die Formulierung von EZ-Programmzielen festlegen.

Umsetzungshinweis 3.1: Die Ziele sollten als langfristiger Beitrag zu entwicklungspolitischen Veränderungen (langfristige Impact-Ebene) oder konkret und messbar für einen klar definierten Zeitraum (eher Outcome- oder mittelfristige Impact-Ebene) formuliert werden. Es sollte möglichst ein konkretes und messbares Ziel festgelegt werden. Falls dies aufgrund der Rahmenbedingungen nicht umsetzbar ist, sollte in der Programmkonzeption eine Begründung erfolgen.

Umsetzungshinweis 3.2: Wenn Module inhaltlich nicht ausreichend in ein Programm integriert werden können und keine messbaren Beiträge zu Programmzielen- und Programmzielindikatoren leisten, sollten diese auch nicht als Teil des Programms dargestellt werden.

Umsetzungshinweis 3.3: Die Handreichungen und Verfahrensinformationen sollten hier klarere Vorgaben machen. Derzeit werden Fragen zur Messbarkeit und Konkretheit der Ziele erst in der EZ-Programmkonzeption thematisiert und je nach Ermessen der Beteiligten diskutiert.

Empfehlung 4: Vor Inkrafttreten der EZ-Programme sollte das BMZ die Qualitätssicherung der formulierten Ziele und Indikatoren durch ausreichend methodisch qualifizierte Ansprechpartner*innen gewährleisten. Die Qualitätssicherung der Ziele sollte mit der Prüfung der Indikatoren einhergehen und entweder durch das BMZ oder die DOs erfolgen.

Umsetzungshinweis 4.1: Für die Qualitätssicherung sollten konkrete personelle Zuständigkeiten sowie finanzielle und zeitliche Ressourcen festgelegt und bereitgestellt werden.

Umsetzungshinweis 4.2: Sofern die Qualitätssicherung der Ziele und Indikatoren über die DOs gewährleistet werden soll, sollten diese auch stärker in die Festlegung der EZ-Programmziele mit einbezogen werden. Zurzeit erfolgt die Zielformulierung häufig auf Ebene der LS mit wenig Einbindung der DO.

Verbesserung der Datenbasis

Je höher die Ansprüche an belastbare und zuordenbare Aussagen zu EZ-Programmfortschritten und zur Zielerreichung sind, desto wichtiger ist eine ausreichende Datenbasis. Derzeit sind für die koordinierende DO kaum personelle und keine finanziellen Ressourcen für ein Monitoringsystem auf EZ-Programmebene vorgesehen, wohingegen die grundsätzliche Aufgabe, ein Monitoringsystem zu entwickeln und Daten bereitzustellen, bei den DOs liegt. Dementsprechend liegen wichtige Primärdaten zur Beurteilung des EZ-Programmfortschritts und zur Messung der Zielerreichung nicht in ausreichender Form vor. Es sollen fast ausschließlich Daten von Partnerinstitutionen zum Monitoring der EZ-Programmzielindikatoren herangezogen werden. Diese stehen jedoch häufig nicht in der jeweils erforderlichen aggregierten Form und zum notwendigen Zeitpunkt zur Verfügung. Das schränkt die Möglichkeit zur Formulierung von angemessenen EZ-Programmindikatoren von vornerein deutlich ein.

Empfehlung 5: Es sollte ein angemessenes Monitoring auf Programmebene etabliert werden. Die DOs sollten zum Zeitpunkt der Programmkonzeption verdeutlichen, inwiefern erforderliche Daten für das Monitoring durch Partner, Sekundärliteratur und die Module zur Verfügung stehen werden und welche zusätzlichen Erhebungen auf Programmebene erforderlich sind. Das BMZ sollte entsprechende Ressourcen bereitstellen.

Umsetzungshinweis 5.1: Während der EZ-Programmkonzeption und der Finalisierung von Zielformulierungen sollten folgende Maßnahmen ergriffen werden:

- BMZ und DOs sollten sich darauf einigen, welche Daten zum Monitoring des EZ-Programmfortschritts erforderlich sind. Auf dieser Basis sollte ein angemessenes Monitoringsystem erarbeitet werden.
- Es sollte, sofern möglich, auf Partnerdaten zurückgegriffen werden.
- Schließlich sollten, falls notwendig, ergänzende und gezielte Erhebungen zur Ermittlung von Baseline- und Monitoringdaten erfolgen. Benötigte Ressourcen sollten durch das BMZ zur Verfügung gestellt werden. Hierfür ist es notwendig, dass zum Zeitpunkt der EZ-Programmkonzeption durch die DOs die erforderlichen Mittel eingeplant werden.

Umsetzungshinweis 5.2: In den DOs sollte für jedes EZ-Programm qualifiziertes und mit ausreichenden Mitteln ausgestattetes Personal mit dem Monitoring betraut werden.

Erkenntnisinteresse und Nutzen von EZ-Programmevaluierungen

Es besteht ein umfangreiches und vielfältiges Erkenntnisinteresse an EZ-Programmevaluierungen, sodass sich die Beteiligten in DOs und BMZ insgesamt einen großen Mehrwert davon versprechen. Als besonders relevant wird die Auseinandersetzung mit dem Zielsystem (von der Messung von erreichten Zielen bis hin zur Überprüfung des Dreiklangs aus LS, EZ-Programmen und Modulen) sowie die Bewertung des Zusammenwirkens von Modulen erachtet. Darüber hinaus wird der Wunsch nach Dokumentation von gewonnenen Erkenntnissen und Empfehlungen für die Planung neuer EZ-Programme oder Fragen zur Zusammenarbeit und Kommunikation zwischen BMZ, DOs und Partnern hervorgehoben.

Empfehlung 6: Um einen größtmöglichen Mehrwert und Nutzen von EZ-PEs gewährleisten zu können, müssen in der Konzeptionsphase des Evaluierungsformats folgende Fragen durch das BMZ geklärt werden:

- **Wer soll die Ergebnisse von EZ-PEs nutzen?**
- **Wofür sollen die Ergebnisse verwendet werden (Rechenschaftslegung, [Um-]Steuerung von laufenden EZ-Programmen, programmübergreifendes Lernen)?**
- **Wann und in welcher Frequenz sollen EZ-PEs durchgeführt werden, um den vorgesehenen Zweck zu erfüllen (für jedes EZ-Programm, zufällige oder kriterienbasierte Stichprobe)?**
- **Wie können Partner in die Evaluierungen einbezogen und in welcher Form Ergebnisse mit ihnen geteilt werden?**
- **Welche der genannten Themen und Fragestellungen sollen durch EZ-PEs beantwortet werden und welche nicht? Gibt es andere Formate, um die nicht berücksichtigten Fragen beantworten zu können?**

Umsetzungshinweis 6.1: Je nach Erkenntnisinteresse müssten EZ-PEs zu unterschiedlichen Zeitpunkten und für unterschiedliche Adressatenkreise durchgeführt werden. Damit Ergebnisse von EZ-PEs von Programmverantwortlichen aus BMZ und DOs in der Planung und Steuerung berücksichtigt werden können, müssten die Evaluierungen möglichst während der Laufzeit des EZ-Programms entweder für die vorgesehene Reflexion nach drei Jahren Programmlaufzeit oder für die Planung eines Folgeprogramms durchgeführt und abgeschlossen werden (abhängig von Programmdauer und -umfang als Zwischenevaluierung oder -review). Alternativ oder ergänzend wären strategische EZ-PEs möglich, in denen mehrere EZ-Programme anhand übergeordneter Fragestellungen untersucht werden. Dieses Evaluierungsformat würde das programmübergreifende Lernen in den Vordergrund stellen.

Umsetzungshinweis 6.2: Mit Blick auf die Sequenzierung von EZ-PEs sollte im Fall von Zwischenevaluierungen geklärt werden, ob Evaluierungen grundsätzlich als Informationsquelle zur Steuerung und Reflexion von Programmen eingesetzt werden sollten und somit jedes EZ-Programm innerhalb eines Programmzyklus evaluiert werden sollte. Zum Zweck der Rechenschaftslegung sowie zur Beantwortung von übergeordneten und strategischen Fragestellungen wäre eine gezielte Stichprobe möglich. Eine andere Option wären geografische und thematische Cluster.

Umsetzungshinweis 6.3: Das Evaluierungsformat sollte regelmäßig überprüft und weiterentwickelt werden. Hier sollte insbesondere reflektiert werden, ob es zukünftig weitere Fragen zu beantworten gilt oder zusätzliche Datenerhebungen innerhalb von EZ-Programmevaluierungen reduziert werden können, falls erforderliche Daten dann in ausreichendem Maße durch das Monitoring und/oder die Berichterstattungen bereitgestellt werden.

9. LITERATUR

ADB (2015), „2015 Guidelines for the Preparation of Country Assistance Program Evaluations and Country Partnership Strategy Final Review Validations“, Asian Development Bank, o.O.

Bergmüller, C. und E. Quiring (2019), „Einleitung“, in Bergmüller, C., B. Causemann, S. Höck, J.-M. Krier und E. Quiring (Hrsg.), *Wirkungsorientierung in der entwicklungspolitischen Inlandsarbeit*, Waxmann, Münster/New York, 12. Aufl., S. 13–22.

Betts, J. et al. (2020), „Evaluation. Adapting for Change: Country Strategy. Approach in Fragile Contexts“, Report 2020/2a, 2b, Ministry for Foreign Affairs of Finland, o.O.

BMZ (2020), „Reformkonzept „‚BMZ 2030‘. Umdenken – Umsteuern“, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn/Berlin.

BMZ (2021), „Leitlinien für die bilaterale finanzielle und technische Zusammenarbeit mit Kooperationspartnern der deutschen Entwicklungszusammenarbeit“, *BMZ-Konzept*, Nr. 165, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn/Berlin.

BRH (2021), „Abschließende Mitteilung an das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung über die Prüfung Evaluierung von Maßnahmen der Entwicklungszusammenarbeit“, Teil I – Steuerung im BMZ, Bundesrechnungshof, Bonn.

Bundesregierung (2019), „Antwort der Bundesregierung auf die Kleine Anfrage der FDP. Strukturprobleme der deutschen Entwicklungszusammenarbeit und ihrer Steuerung am Beispiel des Programms, Abwassermanagement in Provinzstädten in Vietnam‘“, Berlin.

Caspari, A. (2012), „Chancen der Wirkungsorientierung für die entwicklungspolitische Bildungsarbeit*“, Zeitschrift für internationale Bildungsforschung und Entwicklungspädagogik*, Nr. 35, S. 11–17.

Chapman, N. und C. Vaillant (2010), „Synthesis of Country Programme Evaluations Conducted in Fragile States“, *Evaluation Report EV709*, Department for International Development, London/Glasgow.

Dahlgren, S. (2007), „Evaluations of Country Strategies. An Overview of Experiences and a Proposal for Shaping Future Country Programme Evaluations“, *Sida Studies in Evaluation 2007:04*, Swedish International Development Cooperation Agency, Stockholm.

Davies, R. (2013), „Planning Evaluability Assessments. A Synthesis of the Literature with Recommendations“, *Working Paper*, Nr. 40, Department for International Development, Cambridge.

DEZA (2016), „Country and Regional Strategy Evaluation“, Direktion für Entwicklung und Zusammenarbeit, Bern.

DEZA (2018), „Evaluationspolitik“, Direktion für Entwicklung und Zusammenarbeit, Bern.

DFID (2013), „International Development Evaluation Policy“, Department for International Development, London.

Directorate-General for International Cooperation and Development (o.J.), „Work Programme for Strategic Evaluations“, European Commission, o.O.

EU (2017), „Geographic, thematic and other complex evaluations“, European Union, Capacity4dev, https://europa.eu/capacity4dev/evaluation_guidelines/wiki/en-geographic-thematic-and-other-complex-evaluations-0 (zugegriffen 13.09.2021).

Harten, S. und M. Petrova (2018), „Die deutsche EZ braucht mehr Wirkungsorientierung*“, Zeitschrift für Evaluation*, Vol. 17, Nr. 1, S. 145–452.

Hartmann, C. et al. (2019), „Länderportfolioreviews: Ein Analyseinstrument für die deutsche Entwicklungszusammenarbeit“, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit, Bonn.

High Level Forum (2005), „Paris Declaration on Aid Effectiveness. Ownership, Harmonisation, Alignment, Results and Mutual Accountability“, Paris.

Holzapfel, S. (2014), „The Role of Indicators in Development Cooperation. An overview study with a special focus on the use of key and standard indicators“, *Studies*, Nr. 81, Deutsches Institut für Entwicklungspolitik, Bonn.

Holzapfel, S. und C. Römling (2020), „Monitoring in German Bilateral Development Cooperation. A Case Study of Agricultural, Rural Development and Food Security Projects“, *Discussion Paper*, Nr. 18/2020, Deutsches Institut für Entwicklungspolitik, Bonn.

IADB (2009), „Protocol for Country Program Evaluation (CPE) 2008“, Inter-American Development Bank, Washington, D.C.

IAEA (2021), „Country Programme Frameworks“, International Atomic Energy Agency, https://www.iaea.org/services/technical-cooperation-programme/country-programme-frameworks (zugegriffen 08.09.2021).

IEG (2014), „Tunisia Country Program Evaluation, FY2005-13, Volume 1: Main Report“, Evaluation of the World Bank Group Program, Independent Evaluation Group, World Bank, Washington, D. C.

IEG (2015), „World Bank Group Engagement in Resource-Rich Developing Countries: The Cases of the Plurinational State of Bolivia, Kazakhstan, Mongolia, and Zambia“, Clustered Country Program Evaluation Synthesis Report, Independent Evaluation Group, World Bank, Washington, D. C.

IEG (2016), „World Bank Group Engagement in Small States“, Executive Summary, Independent Evaluation Group, World Bank, Washington, D. C.

IEO (o.J.), „Independent Country Programme Evaluations“, Independent Evaluation Office, UNDP, New York.

Independent Evaluation Group (2021), „IEG Methodology“, https://ieg.worldbankgroup.org/methodology (zugegriffen 15.09.2021).

Klingebiel, S. (2011), „Der Gipfel von Busan: Neue Ansätze der Entwicklungszusammenarbeit?“, *Die aktuelle Kolumne*, 04.10., Deutsches Institut für Entwicklungspolitik, Bonn.

Lehmann, M. et al. (2020), „Gute Indikatoren: Ein Crashkurs“, pme-campus, Mainz.

Leonard, K. und A. Eulenburg (2012), „Evaluability – is it Relevant for EBRD?“, Evaluation Brief, Evaluation Department, European Bank for Reconstruction and Development, o.O.

Meyer, L. R. (2019), „Abwassermanagement in Provinzstädten in Vietnam“, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit, Bonn.

OECD (2015), „DAC-Prüfbericht über die Entwicklungszusammenarbeit: Deutschland 2015“, Organisation for Economic Co-Operation and Development, Paris.

OECD (2019a), „Learning from Results-Based Management evaluations and reviews*“, Discussion paper for the OECD/ DAC Results Community workshop on 29-30 October 2018*, Paris, Organisation for Economic Co-Operation and Development, Paris.

OECD (2019b), „The High Level Fora on Aid Effectiveness: A history“, https://www.oecd.org/dac/effectiveness/thehighlevelforaonaideffectivenessahistory.htm#Rome (zugegriffen 15.10.2021).

OECD DAC (2021), „DAC-Prüfbericht über die Entwicklungszusammenarbeit: Deutschland 2021 (Kurzfassung): Wichtigste Ergebnisse und Empfehlungen des Entwicklungsausschusses (DAC)“, Organisation for Economic Co-Operation and Development, Development Assistance Committee, Paris.

Strübing, J. (2014), Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung eines pragmatistischen Forschungsstils, Springer VS, Wiesbaden, 3. Aufl.

Syspons (2018), „Externe Qualitätskontrolle der GIZ – Ergebnisbericht 2017“, Syspons GmbH, Berlin.

UNDG (2011), „Results-based Management. Harmonizing RBM concepts and approaches for improved development results at country level“, United Nations Development Group, o. O.

UNFPA (2019), „Evaluation Handbook. How to Design and Conduct a Country Programme Evaluation at UNFPA“, United Nations Population Fund, New York.

UNIDO (2016), „Independent terminal evaluation of the UNIDO project: GEF-UNIDO Cleantech Programme for SMEs in Malaysia“, United Nations Industrial Development Organization, o. O.

UNIDO (2018), „Evaluation Manual“, Office of Evaluation and Internal Oversight Independent Evaluation Division UNIDO, Wien.

Von Werthern, A. (2019), Theoriebasierte Evaluation. Entwicklung und Anwendung eines Verfahrensmodells zur Programmtheoriekonstruktion, Springer VS, Wiesbaden.

Walsh, I. et al. (2015), „What Grounded Theory Is... A Critically Reflective Conversation Among Scholars“, *Organizational Research Methods*, Vol. 18, Nr. 4, S. 581–599.

10. ANHANG

10.1 Überblick zu den untersuchten Programmen

Tabelle 1 Überblick zu den untersuchten Programmen

Programm	Programmziel	beteiligte DOs
Berufsbildung und Entwicklung des Arbeitsmarktes in Ägypten	Die Beschäftigungssituation junger Erwachsener in Ägypten ist quantitativ und qualitativ verbessert.	GIZ, KfW
Nachhaltige Wirtschaftsentwicklung in Mosambik	Die verbesserte Wettbewerbsfähigkeit einzelner Unternehmen führt zum breitenwirksamen Wachstum kleinster, kleiner und mittlerer Unternehmen.	GIZ, KfW, PTB
Nachhaltige Wirtschaftsentwicklung in Myanmar	Die personellen, institutionellen, politischen sowie infrastrukturellen Voraussetzungen für ein nachhaltiges und sozial ausgewogenes Wirtschaftswachstum in Myanmar sind verbessert. (NaWi I) Die personellen, institutionellen, politischen sowie infrastrukturellen Voraussetzungen für ein nachhaltiges und sozial ausgewogenes Wirtschaftswachstum in Myanmar, insbesondere im Shan Staat, sind verbessert. (NaWi II)	GIZ, KfW, BGR, PTB

10.2 Evaluierungspraktiken anderer Geber auf Länderebene

Tabelle 2 Evaluierungspraktiken anderer Geber auf Länderebene

Geberinstitution	Titel des Instruments	Zielsetzung
Länderprogrammevaluierungen eines (gesamten) Portfolios		
Asian Development Bank (ADB)	Country Assistance Program Evaluation (CAPE) Country Partnership Strategies Final Review (CPSFR)	Analyse der Relevanz der CPS-Ziele (Country Partnership Strategy) Analyse des Beitrags der ADB zur Erfüllung von Strategiezielen im Partnerland; wenn nicht erfüllt, Relevanzprüfung eines Programms
Direktion für Entwicklung und Zusammenarbeit (DEZA), Schweiz	Programm-/Projektevaluation	Lernen Entscheidungsfindung und Steuerung von Programmen und Projekten Rechenschaftslegung
Europäische Union (EU)	Project and programme evaluation	Analyse der Projekt-/Programmbeiträge
Finnland	Project and Programme Evaluation (Decentralized evaluation)	Lernen Rechenschaftslegung

Geberinstitution	Titel des Instruments	Zielsetzung
Independent Evaluation Group (IEG), World Bank	Country-Focused Evaluation	*lessons learnt* (auch auf übergeordneter Ebene für Weltbank) Rechenschaftslegung
Inter-American Development Bank (IADB)	Country Program Evaluation (CPE)	*lessons learnt* und Empfehlungen für Effektivitätsverbesserung der übergeordneten Strategie und der Länderprogramme Rechenschaftslegung und Lernprozesse
United Nations Development Programme (UNDP)/Independent Evaluation Office (IEO)	Country Program Evaluation (CPE)	Evidenz für Entwicklung neuer Länderprogramme Rechenschaftslegung gegenüber UNDP und nationalen Stakeholdern
United Nations Population Fund (UNFPA)	Country Program Evaluation (CPE)	Rechenschaftslegung Nutzung der Evaluierungsergebnisse für nächsten Programmzyklus
United Nations Industrial Development Organization (UNIDO)	Country Programme Evaluation, Country Evaluation	*lessons learnt* Nutzung der Evaluierungsergebnisse für nächsten Programmzyklus
thematische/geografische/strategische Ansätze der Länderprogrammevaluierung		
Direktion für Entwicklung und Zusammenarbeit (DEZA), Schweiz	Country and Regional Strategy Evaluation	Lernen Entscheidungsfindung und Steuerung von Programmen und Projekten Rechenschaftslegung
Europäische Union (EU)	Geographic evaluation (strategic evaluation)	Analyse der Politikmaßnahmen (*policies*) und der Projekt-/Programmbeiträge in einem Land oder Region
	Thematic evaluation (strategic evaluation)	Analyse der Politikmaßnahmen (*policies*), Modalitäten und Projekt-/Programmbeiträge anhand eines Schwerpunkts oder Sektors
finnisches Außenministerium	Policy and Strategic Evaluations (Centralized evaluation; evaluations of policies, country strategies, financing instrument, and wide thematic programmes)	Lernen und Rechenschaftslegung
Foreign, Commonwealth & Development Office (FCDO), Vereinigtes Königreich	Country Programme Evaluation (CPE) Synthesis of Country Programme Evaluations conducted in Fragile States	Überprüfung, inwieweit die *corporate policy in fragile states* von Programmen umgesetzt wurde (CPE-Anwendung vor allem in Staaten mit fragilem Kontext)

Geberinstitution	Titel des Instruments	Zielsetzung
Independent Evaluation Group (IEG), World Bank	Cluster Country Program Evaluations	Cross-Country-Lernen erweiterte Abdeckung der Evaluierungen und Effizienz von Programmevaluierungen
United Nations Industrial Development Organization (UNIDO)	Country Evaluation (Project and programme evaluations) als Clusterevaluierungen	Analyse der Beiträge auf Länderebene Erkenntnisse zur Gestaltung der zukünftigen Partnerschaft mit Partnerland Nachhaltigkeitsbewertung früherer Aktivitäten

Anmerkung: Die Liste erhebt keinen Anspruch auf Vollständigkeit. Die Evaluierungsansätze anderer Geber wurden unter anderem hinsichtlich der Informationsverfügbarkeit und der eventuellen Übertragbarkeit beziehungsweise als Hinweise für die deutsche EZ ausgewählt.

Quelle: eigene Darstellung, basierend auf ADB, 2015; Betts et al., 2020; Chapman und Vaillant, 2010; DEZA, 2016, 2018; DFID, 2013; Directorate-General for International Cooperation and Development, o.J.; EU, 2017; IADB, 2009; IEG, 2014, 2016; IEO, o.J.; UNFPA, 2019; UNIDO, 2018

10.3 Bestückung der EZ-Programmzielindikatoren

Tabelle 3 Bestückung der EZ-Programmzielindikatoren

		Ägypten		Mosambik		Myanmar		Gesamt	
		absolute Anzahl/GG	%	absolute Anzahl/GG	%	absolute Anzahl/GG	%	absolute Anzahl/GG	%
Konzeption	Basiswerte EZ-Programmindikatoren	2 von 3	66,66	3 von 3	100	3 von 4	75	8 von 10	80
	Zielwerte EZ-Programmindikatoren	2 von 3	66,66	3 von 3	100	4 von 4	100	9 von 10	90
Implementierung	Basiswerte EZ-Programmindikatoren	2 von 3	66,66	3 von 3	100	3 von 4	75	8 von 10	80
	Zielwerte EZ-Programmindikatoren	2 von 3	66,66	3 von 3	100	4 von 4	100	9 von 10	90
	Ist-Werte der EZ-Programmindikatoren	0 von 3	0	1 von 3	33,33	2 von 4	50	3 von 10	30

GG = Grundgesamtheit

Quelle: eigene Darstellung

10.4 Voraussetzungen für die EZ-Programmbeurteilung

Im Bericht wird herausgestellt, dass die EZ-Programmkonzeption als auch die EZ-Programmberichterstattung Evidenzlücken aufweisen, die die Möglichkeit zur Beurteilung des EZ-Programmfortschritts und zur Erreichung von EZ-Programmzielen einschränken.

Abbildung 9 Voraussetzungen für die EZ-Programmbeurteilung

Aspekt	EZ-Programmkonzeption	EZ-Programmberichterstattung	Schlussfolgerungen für EZ-Programmbeurteilung
Plausibilität des EZ-Programmziels	zum Teil erfüllt	zum Teil erfüllt	eingeschränkt möglich
SMARTe EZ-Programmzielindikatoren	zum Teil erfüllt	zum Teil erfüllt	eingeschränkt möglich
Festlegung der Basis-/Zielwerte	zum Teil erfüllt	zum Teil erfüllt	eingeschränkt möglich
Bestückung der Ist-Werte	nicht relevant	zum Teil erfüllt	eingeschränkt möglich
Nachvollziehbarkeit/ Attribuierbarkeit der Datenquellen	zum Teil erfüllt	zum Teil erfüllt	eingeschränkt möglich
Zusammenwirken der Module/Darstellung von Wirkungsannahmen (WiLo)	nicht enthalten	nicht enthalten	nicht möglich
Kausalität der Wirkungszusammenhänge (Narrativ)	zum Teil erfüllt	zum Teil erfüllt	eingeschränkt möglich
Stand der Umsetzung	nicht relevant	zum Teil erfüllt	eingeschränkt möglich
EZ-Programmzielerreichung	nicht relevant	zum Teil erfüllt	eingeschränkt möglich

Quelle: eigene Darstellung

10.5 Team und Mitwirkende

Kernteam	Funktion
Miriam Amine	Senior-Evaluatorin und Teamleiterin
Michèle Kiefer	Evaluatorin
Dr. Nico Herforth (bis Juni 2021)	Evaluator
Judith Ihl (ab Juni 2021)	Evaluatorin
Rebecca Maicher	Projektadministratorin
Severin Flaig	studierender Beschäftigter

Verantwortlich	
Amélie Gräfin zu Eulenburg	Abteilungsleiterin

Mitwirkende	Funktion
Dr. Stefan Silvestrini	externer Gutachter
Kirsten Vorwerk	DEval-interne Peer-Reviewerin (Evaluatorin und Teamleiterin)
Dr. Hans-Bernd Zöllner	externer Peer-Reviewer

10.6 Zeitplan

Zeitrahmen	Aufgaben
06–09/2020	Erstellung der Concept Note
09/2020	1. Referenzgruppensitzung zur Concept Note
10/2020–01/2021	Erstellung des Inception-Reports
02/2021	2. Referenzgruppensitzung zum Inception-Report
02–04/2021	Datenerhebung
05–07/2021	Auswertung und Synthese der Ergebnisse
06/2021	3. Referenzgruppensitzung in Form von Work-shops zu vorläufigen Ergebnissen
07–10/2021	Berichtslegung
11/2021	4. Referenzgruppensitzung zum Bericht
12/2021	Abschluss der Evaluierung nach Layout und Druck