

Dörrbecker, Nicola M.

Working Paper

Lernen aus Ex-post-Evaluierungen der KfW? Wie Zielkonflikte den Nutzen einschränken können

IDOS Discussion Paper, No. 12/2023

Provided in Cooperation with:

German Institute of Development and Sustainability (IDOS), Bonn

Suggested Citation: Dörrbecker, Nicola M. (2023) : Lernen aus Ex-post-Evaluierungen der KfW? Wie Zielkonflikte den Nutzen einschränken können, IDOS Discussion Paper, No. 12/2023, ISBN 978-3-96021-213-3, German Institute of Development and Sustainability (IDOS), Bonn, <https://doi.org/10.23661/idp12.2023>

This Version is available at:

<https://hdl.handle.net/10419/275726>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Lernen aus Ex-post-Evaluierungen der KfW?

Wie Zielkonflikte den Nutzen einschränken können

Nicola M. Dörrbecker



Lernen aus Ex-post-Evaluierungen der KfW?

Wie Zielkonflikte den Nutzen einschränken können

Nicola M. Dörrbecker

Nicola M. Dörrbecker ist Studentin des Studiengangs Kultur und Gesellschaft Afrikas M.A. an der Universität Bayreuth mit den Schwerpunkten Entwicklungssoziologie/Entwicklungspolitik und Ethnologie.

E-Mail: nicoladoerrbecker@gmail.com

Mit finanzieller Unterstützung des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ).

Zitationsvorschlag:

Dörrbecker, N.M. (2023). *Lernen aus Ex-post-Evaluierungen der KfW? Wie Zielkonflikte den Nutzen einschränken können* (IDOS Discussion Paper 12/2023). Bonn: German Institute of Development and Sustainability (IDOS). <https://doi.org/10.23661/idp12.2023>

Disclaimer:

Die Publikation gibt die Meinung der Autor*innen wieder und stellt nicht unbedingt jene des German Institute of Development and Sustainability (IDOS) dar.



Sofern nicht anders angegeben, ist der Inhalt dieser Publikation unter einer Creative Commons Attribution 4.0 International License lizenziert. Es ist erlaubt, dieses Werk zu vervielfältigen, zu verbreiten und zu bearbeiten, solange das German Institute of Development and Sustainability (IDOS) gGmbH und der/die Autor*in(nen) genannt werden.

IDOS Discussion Paper / German Institute of Development and Sustainability (IDOS) gGmbH

ISSN 2751-4439 (Print)

ISSN 2751-4447 (Online)

ISBN 978-3-96021-213-3 (Print)

DOI: <https://doi.org/10.23661/idp12.2023>

© German Institute of Development and Sustainability (IDOS) gGmbH

Tulpenfeld 6, 53113 Bonn

E-Mail: publications@idos-research.de

<https://www.idos-research.de>



Gedruckt auf umweltfreundlichem, zertifiziertem Papier

Vorwort

Das German Institute of Development and Sustainability (IDOS), vormals Deutsches Institut für Entwicklungspolitik (DIE), unterstützt durch interdisziplinäre Forschung und wirkungsorientierte Beratung eine globale Gemeinwohlpolitik, die am Leitbild nachhaltiger Entwicklung orientiert ist. Das von mir geleitete Forschungsprojekt zu „Wirksamkeit, Wissensmanagement und Lernen“ untersucht verschiedene Entwicklungsorganisationen hinsichtlich der Frage, wie Entwicklungsbürokratien Wissen verwalten und daraus lernen.

Im Kontext dieses Forschungsprojektes bin ich auf die Masterarbeit von Nicola Dörrbecker im Studiengang Kultur und Gesellschaft Afrikas an der Universität Bayreuth aufmerksam geworden. Ich halte Frau Dörrbeckers Untersuchung zum Thema „Ex-post-Evaluierungen in der KfW Entwicklungsbank“ aufgrund der Qualität ihrer Analyse und der Praxisrelevanz des Themas für wissenschaftlich wertvoll und habe Frau Dörrbecker auf dieser Basis mit der Erstellung einer gekürzten Version der Arbeit als IDOS Discussion Paper beauftragt.

Das Discussion Paper hat einen internen IDOS Peer-Review durchlaufen, und Frau Dörrbecker hat zahlreiche Kommentare in einer Überarbeitung der Studie adressiert. Das nun vorliegende Discussion Paper stellt eine grundlegend überarbeitete Fassung der ursprünglichen Masterarbeit dar. Alle hier zitierten Interviewten haben ihre Einverständniserklärung zur Verwendung ihrer Daten erteilt.

Ich bin sicher, dass Frau Dörrbeckers Arbeit einen wertvollen Beitrag zur theoretischen Auseinandersetzung mit und praktischen Umsetzung von Evaluierungen in der deutschen Entwicklungszusammenarbeit leistet, und wünsche viel Vergnügen bei der Lektüre.

Bonn, August 2023

Heiner Janus

Danksagung

Ich möchte mich bei allen Interviewten für ihre Zeit und Unterstützung bei der Beantwortung meiner zahlreichen Fragen bedanken.

Außerdem gilt mein Dank Heiner Janus, Daniel Esser, Miriam Amine, Max-Otto Baumann, Stephan Klingebiel und Anna-Katharina Hornidge für die zahlreichen hilfreichen und wertvollen Hinweise sowie konstruktiven Kommentare. Mein Dank gilt ebenfalls Katharina Schramm und Alexander Stroh-Steckelberg an der Universität Bayreuth, unter deren Betreuung die Masterarbeit entstand, aus der dieses Discussion Paper hervorgegangen ist.

Bayreuth, 11.08.2023

Nicola M. Dörrbecker

Zusammenfassung

Im Politikfeld der Entwicklungszusammenarbeit (EZ) wird die Frage nach deren Wirksamkeit umfangreich diskutiert. Doch trotz zahlreicher Kontroll- und Evaluierungsformate, die zu Lernprozessen und damit zu einer Steigerung der Wirksamkeit führen sollen, ist diese Steigerung häufig nicht abbildbar. Dieses Paper hat vor diesem Hintergrund das Ziel, den Nutzen von Ex-post-Evaluierungen (EPE) der Entwicklungsbank der Kreditanstalt für Wiederaufbau (KfW) zu analysieren – sowohl innerhalb der KfW Entwicklungsbank als auch bei ihrem Auftraggeber, dem Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ).

Aus der Forschung lässt sich schließen, dass EPE mit hoher Sorgfalt betrieben werden. Zudem können EPE zur Legitimität der (finanziellen) EZ beitragen, da Projektergebnisse strukturiert reflektiert und präsentiert werden. Dennoch schätzen die für diese Studie Interviewten die EPE als eine (sehr) subjektive Bewertung ein und halten diese Evaluierungen unter Umständen auch nicht für untereinander vergleichbar. Die Vergleichbarkeit von EPE ist jedoch erforderlich, da aus ihren Gesamtnoten eine Erfolgsquote errechnet wird, die zurzeit bei ca. 81 % liegt. Dies wiederum hat Einfluss auf die Erfolgsberichterstattung der KfW gegenüber dem BMZ und der Öffentlichkeit. Die Daten aus den Interviews zeigen, dass durch Zielkonflikte, die bei der Erstellung und Anwendung der EPE entstehen, der Nutzen dieses Formats eingeschränkt zu sein scheint. EPE sollen sowohl der Rechenschaftslegung gegenüber der Öffentlichkeit bzw. dem BMZ als auch dem Lernen innerhalb der KfW dienen. Allerdings konfliktieren diese beiden Ziele miteinander, da für jedes eine andere Herangehensweise erforderlich wäre.

EPE werden laut den Befragten von KfW und BMZ nur selten gelesen oder genutzt. Interviewte beschreiben, dass EPE selten Relevanz für die operativen Bereiche haben, da sie erst mehrere Jahre nach Abschluss des Projektes erscheinen und nur vereinzelt relevante Informationen für aktuelle Projekte enthalten. Diese Zeit muss jedoch abgewartet werden, da sonst die Nachhaltigkeit und die entwicklungspolitische Wirkung eines Projektes nicht abgebildet werden können. Zudem haben EPE laut den Interviews und der Evidenz aus anderen Untersuchungen kaum politische Steuerungsrelevanz im BMZ, auch nicht in aggregierter Form. EPE nicht mehr durchzuführen ist jedoch nach Ansicht der Autorin keine Option, da lediglich auf diese Weise kostengünstig und für eine repräsentative Projektanzahl entwicklungspolitische Wirkung und Nachhaltigkeit überprüft und somit die Basis für die Rechenschaftslegungsfunktion gebildet werden kann.

Den Zielkonflikt zwischen Lernen und Rechenschaftslegung zu lösen stellt eine Herausforderung dar. Für die Lernkomponente erscheint es sinnvoll, verstärkt auf Querschnittsauswertungen zu setzen sowie eine zentrale Unterstützungsstruktur für alle Durchführungsorganisationen und das BMZ zu etablieren, um alle Kerninformationen aus den Evaluierungen zu sammeln und – zugeschnitten auf die Bedürfnisse an BMZ, KfW, aber auch an die Partnerländer – weiterzugeben. Für die Rechenschaftslegungskomponente sollte außerdem die Transparenz erhöht werden, indem abgeschlossene Evaluierungsberichte zeitnah und in voller Länge der Öffentlichkeit zur Verfügung gestellt werden. Das Paper basiert neben einer Auswertung internationaler Forschungsliteratur insbesondere auf empirischen Interviewdaten. Insgesamt wurden 13 gezielt ausgewählte Expert*innen aus dem System der deutschen Entwicklungszusammenarbeit interviewt. Diese Interviewdaten stellen damit eine illustrative, aber keine repräsentative Stichprobe dar.

Inhaltsverzeichnis

Vorwort	III
Danksagung	IV
Zusammenfassung	V
Abkürzungen	VII
1 Ex-post-Evaluierungen der KfW im Zeichen globaler Wirksamkeitsbemühungen	1
2 Methodik und Datengenerierung	3
3 Theoretische Grundlagen und Anwendungen von Ex-post-Evaluierungen	4
4 Wie wird eine Ex-post-Evaluierung in der KfW erstellt? Ein Überblick	7
5 Herausforderungen in Ex-post-Evaluierungen: Was ergeben die Interviewdaten?	10
6 Ex-post-Evaluierungen und ihr Nutzen: Wie werden Ergebnisse verwendet?	14
6.1 Wer liest die Ex-post-Evaluierung?	14
6.2 Wie nutzt die KfW ihre Ex-post-Evaluierungen?	15
6.3 Die Frage nach der politischen Wirkung einer hohen Erfolgsquote	17
6.4 Politische Steuerungsrelevanz: Nutzt das BMZ die Ex-post-Evaluierung?	19
6.5 Wer lernt nun aus der Ex-post-Evaluierung?	20
7 Wie Ex-post-Evaluierungen in Zukunft genutzt werden können	23
Literaturverzeichnis	26
Anhang	31
Abbildungen	
Abbildung 1: Gesamtnotenhäufigkeit in EPE 2007–2022 in der KfW	18

Abkürzungen

BMZ	Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung
BRH	Bundesrechnungshof
DAC	Development Assistance Committee
DEval	Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit
EPE	Ex-post-Evaluierung
EZ	Entwicklungszusammenarbeit
FZ	Finanzielle (Entwicklungs-)Zusammenarbeit
FZ-E	Evaluierungsabteilung der KfW (Finanzielle Zusammenarbeit – Evaluierungseinheit)
GIZ	Gesellschaft für internationale Zusammenarbeit
IDOS	German Institute of Development and Sustainability (vormals Deutsches Institut für Entwicklungspolitik – DIE)
KfW	Kreditanstalt für Wiederaufbau
OECD	Organisation for Economic Co-operation and Development
PM	Portfoliomanager*in
TSV	Technische/r Sachverständige/r

1 Ex-post-Evaluierungen der KfW im Zeichen globaler Wirksamkeitsbemühungen

Die Frage der Wirksamkeit ist in der Entwicklungszusammenarbeit (EZ) allgegenwärtig. Im Politikfeld der EZ wird daher so umfangreich evaluiert wie nur in wenigen anderen Politikfeldern (Faust, 2020, S. 64). Ferguson, Mchombu und Cummings (2008, S. 38) beschreiben dieses Politikfeld als regelrechte „Wissensindustrie“. Die hohe Evaluierungsdichte lässt sich aus einer funktionalen Perspektive heraus begründen, wie auch aus jener der Legitimität des Politikfeldes (Stockmann, 2004, S. 3). Historisch betrachtet begann die gemeinsame Evaluierungskultur von den Mitgliedstaaten des Development Assistance Committee (DAC) der Organisation for Economic Cooperation and Development (OECD) Anfang der 1990er Jahre (Meyer et al., 2019, S. 165). Das OECD/DAC Network on Development Evaluation¹ definierte 1991 fünf Kriterien und mit ihnen verbundene Leitfragen als Leitfaden für Evaluierungen (OECD, 1991). Diese Leitfragen wurden 2019 überarbeitet, und ein sechstes Kriterium (Kohärenz) wurde hinzugefügt (OECD [Organisation for Economic Cooperation and Development], 2020, S. 1; OECD, 2021a, S. 13). Die nun sechs DAC-Kriterien, nach denen Evaluierungen durchgeführt werden sollen, lauten: Relevanz, Kohärenz, Effektivität, Effizienz, entwicklungspolitische Wirkungen und Nachhaltigkeit (OECD/DAC Network on Development Evaluation, 2019, S. 7–12).²

Ein Instrument, um Wirkungen festzustellen, ist die Ex-post-Evaluierung (EPE). Sie bezeichnet eine Evaluierung nach Abschluss eines Entwicklungsprojektes. Ihr Hauptzweck ist es, dessen entwicklungspolitische Wirkung und Nachhaltigkeit zu erfassen; dafür soll sie erst einige Zeit nach Abschluss des Projektes durchgeführt werden (BMZ [Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung], 2021b, S. 3; Cracknell, 2002, S. 74).

Solche EPE werden in der Entwicklungsbank der KfW durchgeführt. Sie ist ein zentraler Akteur in der deutschen EZ.³ Aufgrund des – zumindest politisch angenommenen – hohen Rechtfertigungsdrucks in der EZ wird in der KfW Entwicklungsbank⁴ seit 1966 eine Erfolgskontrolle aller geförderten Entwicklungsmaßnahmen durchgeführt (Terberger, 2011, S. 223). Neben anderen Formaten werden in der KfW heute ca. 50 % der Projekte einer EPE unterzogen (KfW [Kreditanstalt für Wiederaufbau], 2021c, S. 6–7).

Die KfW Entwicklungsbank ist Teil der KfW Bankengruppe und für die finanzielle Zusammenarbeit (FZ) in der Entwicklungspolitik verantwortlich. Die FZ orientiert sich an den Entwicklungsstrategien des Partnerlandes und an den Länderstrategien des BMZ. In den EZ-Programmen, die sich an den Länderstrategien orientieren, sind die Kernsektoren festgehalten, die im Empfängerland unterstützt werden. Ein einzelnes Projekt, auch Vorhaben genannt, zählt in der Regel in dieses Programm ein und wird im Detail von der KfW ausgearbeitet. In der FZ sind drei Akteure an der Umsetzung eines Projektes beteiligt. Das ist zum einen die KfW, die mit einem operativen Team aus Portfoliomanager*innen (PM) und ggf. technischen Sachverständigen (TSV) das Projekt planen, den Projektträger bei der Umsetzung beraten und an das BMZ berichten. Der zweite Akteur ist der Projektträger oder Projektpartner im Empfängerland. Dies ist

1 Das Network on Development Evaluation ist eine Arbeitsgemeinschaft von Vertreter*innen von Evaluierungseinheiten der Entwicklungsorganisationen der OECD/DAC-Mitglieder.

2 Ausführliche Beschreibungen der DAC-Kriterien bieten das BMZ (2020, S. 1) und die OECD (2020, S. 2; 2021a, S. 10). Kritisch diskutiert wird die Anwendung der DAC-Kriterien z. B. von Noltze, Euler und Verspohl (2018b) sowie von Schönhuth und Jerrentrup (2019, S. 52).

3 Die KfW war im Jahr 2020 mit rund 11 Mrd. Euro an Entwicklungsprojekten beteiligt, von denen ca. 4 Mrd. Euro aus dem Bundeshaushalt stammten (KfW, 2021b, S. 4); im Vergleich dazu beliefen sich die gesamten Official-Development-Assistance-Flüsse der Bundesrepublik Deutschland im Jahr 2020 auf knapp 25 Mrd. Euro (BMZ, 2022).

4 Im Folgenden nur noch als „KfW“ bezeichnet.

meist ein Ministerium oder eine staatliche Institution. Der dritte Akteur kann der Umsetzungsconsultant sein, der für die technische Umsetzung des Projektes verantwortlich ist (KfW, 2022b; Terberger, 2011, S. 228).⁵

EPE dienen als Wirkungsnachweis und tragen zur Legitimität der EZ bei, da mit ihnen Projektergebnisse strukturiert reflektiert und präsentiert werden. Sie erwecken jedoch auch den Eindruck eines Repräsentationsobjektes der FZ, denn nur wenige Dokumente über Entwicklungsprojekte der KfW (außer den EPE) sind öffentlich zugänglich. Auf der Webseite finden sich seit Ende 2021 in einer Projektdatenbank Projektinformationen von etwas mehr als 2.100 aktiven Projekten der KfW, jedoch lediglich in komprimierter Form.⁶ Die EPE bilden insgesamt etwa die Hälfte der FZ-Projekte ab. Sie bieten Informationen zu Zielsetzung, Vorgehensweise, eingesetzten Mitteln, Partnern vor Ort und zur kritischen Auseinandersetzung mit diesen Informationen. Durch dieses Evaluierungssystem soll – nach Angaben der KfW – gelernt werden, die zukünftige FZ zu verbessern (KfW, 2022c).

Da die EPE für die Außendarstellung von großer Bedeutung sind und in der EZ so umfangreich evaluiert wird, stellt diese Studie die empirische Frage: Wie, wofür und in welchen Bereichen werden EPE im System der EZ genutzt? Dies wird durch eine Theorie-geleitete Untersuchung der aktuellen Nutzungspraxis der EPE in der KfW beantwortet. Um den potenziellen Nutzen zu ermitteln, werden das Ex-post-Evaluierungssystem und konkrete EPE der KfW analysiert. Dazu wurden 15 Expert*inneninterviews geführt und sechs Langfassungen von EPE der KfW ausgewertet. Zudem wurde ermittelt, ob die vom BMZ vorgeschlagenen Evaluierungszwecke den theoretischen Evaluierungsfunktionen entsprechen und wie die KfW mit den Evaluierungszwecken sowie dem Soll-Nutzen von EPE umgeht. Es wird herausgearbeitet, welche Verzerrungen bei der Erstellung entstehen können, wie die Ergebnisse in der KfW in der Realität verarbeitet werden und ob die EPE den eigenen Ansprüchen gerecht werden. Die Frage nach der Nutzung stellt sich ebenfalls für die politische Steuerung der Programme im BMZ, welches EZ-Vorhaben durch Wirkungsnachweise explizit verbessern will (Amine & Eulenburg, 2022, S. 1) sowie aggregierte Erkenntnisse aus Projektevaluierungen als politisches Steuerungsinstrument wie auch zum Wissensmanagement nutzen will (BMZ, 2021a, S. 40).

Die Ergebnisse dieser Arbeit zeigen, wie Zielkonflikte den Nutzen der EPE begrenzen können, denn häufig stehen verschiedene ihrer Funktionen im Konflikt zueinander – aber auch die Umstände, unter denen EPE entstehen. Die Ergebnisse sind relevant für Mitarbeitende der KfW sowie des BMZ – im Hinblick darauf, auf welche Details sie in Zukunft bei der Nutzung von EPE achten können, um die Wirksamkeit bzw. Lernfähigkeit der FZ zu verbessern. Außerdem trägt diese Forschung zu einem breiteren wissenschaftlichen Diskurs zum Thema „Lernen von EZ-Organisationen“ bei (vgl. z. B. Dexis Consulting Group, 2020; Hovland, 2003; Kogen, 2018; Krohwinkel-Karlsson, 2007; Yanguas, 2021) und ermittelt, wie und ob die KfW institutionell aus EPE lernt.

Im folgenden Abschnitt wird zunächst die Methodik der Forschung aufgezeigt, die dieser Analyse zugrunde liegt. Darauf folgen die offiziellen Erwartungen an die EPE des BMZ und der KfW, die Methodik, mit der EPE erstellt werden, und Herausforderungen, die mit der Methodik einhergehen können. Nachfolgend wird der unterschiedliche, hier ermittelte Nutzen der EPE dargestellt und aufgezeigt, welche Lernerfahrungen aus ihnen gezogen werden können. Schlussendlich wird diskutiert, welchen Nutzen die EPE in Zukunft haben können.

5 In manchen Projekten gibt es keinen Umsetzungsconsultant, in anderen existieren große Vor-Ort-Büros des Consultingunternehmens.

6 Die Projektdatenbank findet sich unter: kfw-entwicklungsbank.de/s/dezKK_4. Die Anzahl der enthaltenen Projekte wurde zuletzt am 31.10.2022 aktualisiert.

2 Methodik und Datengenerierung

Die Fragestellung, wie EPE in der Praxis genutzt werden, wird mit einer qualitativen Methodik beantwortet. Angelehnt an die „Following Objects“-Methode von Czarniawska (2007) aus den Science and Technology Studies wird eine Analyse des Evaluierungsprozesses durchgeführt. Die EPE stellt dabei das „Objekt“ dar, dem in Erstellung und Nutzung gefolgt wird. Das Ziel ist festzustellen, welche Funktionen und welchen möglichen Nutzen EPE innerhalb des Systems haben und durch welche Zielkonflikte diese Funktionen eingeschränkt sein können. Zur Triangulation der Daten über solche Zielkonflikte werden insbesondere die Interpretationen und Wahrnehmungen von EPE aus verschiedenen Perspektiven unterschiedlicher Akteure herausgearbeitet.

In erster Linie stützt sich die Analyse in diesem Paper auf semi-strukturierte Expert*inneninterviews mit KfW-Mitarbeitenden. Diese Interviewmethode der qualitativen empirischen Forschung (nach Meuser & Nagel, 1991) ist eine viel diskutierte qualitative Methode in der Politik- und Sozialforschung. Die Interviews basieren dabei auf offenen Fragen, deren Analyse Erkenntnisse und Einsichten zur Beantwortung der Forschungsfrage liefert. Des Weiteren wurde eine Analyse der sechs EPE-Langfassungen durchgeführt, die von der KfW zur Verfügung gestellt wurden. Informationen von der Webseite (KfW, 2022c) und die alle zwei Jahre erscheinenden Synthesenberichte aus der Evaluierungsabteilung der KfW (FZ-E) wurden dabei berücksichtigt. Durch die Covid-19-Pandemie und damit einhergehende Einschränkungen waren analoge persönliche Gespräche und teilnehmende Beobachtung nicht möglich. Daher liegt der Fokus auf der o. g. Dokumentenanalyse sowie auf den 15 semistrukturierten Online-Interviews.

Da die Interviewten ein gewisses Grundverständnis über EPE haben mussten, wurde die Auswahl gezielt und nicht randomisiert getroffen. Hierfür wurde das Schneeballsystem (nach Kalton & Anderson, 1986, bzw. Merton, 1949) angewandt, nach dem bereits Befragte um interne Referenzen gebeten wurden. Insgesamt wurden neun KfW- und zwei BMZ-Mitarbeitende, eine Person im Deutschen Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) und eine Person im German Institute for Development and Sustainability (IDOS) befragt. Die Interviews wurden inhaltlich jeweils auf die Interviewten zugeschnitten, je nachdem, aus welcher Abteilung bzw. aus welchem Bereich sie stammten und ob sie bereits EPE durchgeführt hatten.

Die Suche nach gesprächsbereiten Personen stellte eine Herausforderung dar. Fünf Interviewanfragen wurden von der KfW mit der Begründung des Zeitmangels abgelehnt und sieben weitere gar nicht beantwortet. Interne Dokumente, wie eine kommentierte Mustergliederung einer EPE, wurden von der KfW nicht zur Verfügung gestellt, da die Datenschutzabteilung nicht bereit war, sie freizugeben. Wie der Forschenden mitgeteilt wurde, hätte sie bereits vor der Datenerhebung eine entsprechende vertragliche Übereinkunft mit der KfW erzielen müssen. Erst nach Berufung auf das Informationsfreiheitsgesetz über das BMZ wurden sechs angefragte EPE-Langfassungen von der KfW zur Verfügung gestellt. Im Gegenzug wurde die Forschende gebeten, eine Datenschutzvereinbarung zu den EPE zu unterzeichnen. Bei der abschließenden Vorlage der Zitate äußerten Ansprechpersonen in der KfW die Erwartung, dass Inhalte der Forschungsarbeit anzupassen seien, sowohl bei den Zitaten von KfW-Mitarbeitenden als auch beim Inhalt der Arbeit insgesamt. KfW-Interviewte baten um mehrere Feedbackschleifen. Infolge einer erneuten Abfrage der Einverständniserklärung für Zitate in dieser Publikation zogen drei von ihnen ihre Erlaubnis wieder zurück. Sie begründeten diese Entscheidung primär mit der Sorge, keinen Einfluss auf die Interpretation ihrer (oder anderer) Aussagen zu haben. Für die Forschende war die Datengenerierung in der KfW somit insgesamt schwierig. Gleichzeitig ist zu betonen, dass einige Mitarbeitende sich sehr bemühten und ihrerseits geringes Verständnis für die ungewöhnlich komplexen internen Prozesse aufbrachten.

Das Forschungsmaterial gibt einen Einblick, welchen Nutzen die EPE in der KfW und im BMZ potenziell haben können. Zudem vermittelt es einen Eindruck davon, welche Herausforderungen mit der Erstellung und Verwendung von EPE einhergehen können. Aufgrund der Auswahl

der Befragten ist das Material jedoch nicht repräsentativ und hat keinen Anspruch auf Vollständigkeit. Potenziell ist es möglich, dass es noch weitere Nutzungsmöglichkeiten gibt, die hier nicht abgebildet sind. Außerdem kann die Studie nicht zeigen, ob die Herausforderungen, die das Evaluierungssystem mit sich bringt, systematisch gefördert bzw. ermöglicht werden, oder ob es sich dabei um Einzelfälle handelt. Es kann nur gezeigt werden, welche Defizite theoretisch vorhanden sind, aber nicht in welchem Ausmaß sie möglicherweise „ausgenutzt“ werden könnten. Da die Interviewten überwiegend der KfW angehörten, können über andere Akteure, die mit EPE in Berührung kommen, nur begrenzt Aussagen getätigt werden. Vor allem bei den Länder- und Sektorreferaten des BMZ hätten mehr Interviews geführt werden müssen, um deren Leseverhalten und Umgang mit EPE abbilden zu können. Die in den folgenden Abschnitten geschilderten Einschätzungen sind somit illustrativ. Dennoch können aus der Analyse Verbesserungspotenziale für die Prozessausgestaltung bei der Erstellung und Nutzung der EPE abgeleitet werden.

Mit MAXQDA, einer Software zur systematischen Analyse qualitativer Datenmengen, wurden alle Interviews, alle EPE-Langfassungen, das Feldtagebuch, E-Mail-Konversationen mit den KfW-Mitarbeitenden, der KfW-Synthesenbericht 2019–2020 (KfW, 2021a) und die „Evaluierungskriterien für die deutsche bilaterale Entwicklungszusammenarbeit“ des BMZ (2020) kodiert. Dabei wurde eine offene Kodierung⁷ angewandt, die sich an der Grounded Theory (Glaser & Strauss, 2010) orientiert. Die theoretische Sättigung konnte mit 15 Interviews jedoch nicht erreicht werden und wird auch nicht beansprucht.

Kodiert wurde nach Mayring (2010) mit einer induktiven Kategorienentwicklung und einer Kategorienbildung im Prozess. Entsprechend wurden die ersten zwei Interviews zuerst kodiert, um daraus präzierte Fragen für die nachfolgenden Interviews abzuleiten. Diese beiden Befragten wurden daher im späteren Verlauf noch einmal interviewt, um eine Vergleichbarkeit sicherzustellen. Nach Abschluss aller Interviews wurden aus allen Kategorien Oberkategorien gebildet und diese in einer „Mindmap“ sortiert, um dadurch Zusammenhänge zwischen den Oberkategorien aufzuzeigen. Die Mindmap findet sich in Anhang 1. Nach Abschluss der Forschung wurden die Ergebnisse allen Interviewten zur Verfügung gestellt, um eventuelle Missverständnisse bei der Deutung der Antworten auszuräumen.

3 Theoretische Grundlagen und Anwendungen von Ex-post-Evaluierungen

Gemäß der OECD (1991, S. 5) bezeichnet der Ausdruck Evaluierung eine systematische und objektive Bewertung eines laufenden oder bereits abgeschlossenen Projektes hinsichtlich seiner Konzeption, Durchführung und Ergebnisse. In der internationalen Literatur werden primär zwei Ziele von Evaluierungen unterschieden: Rechenschaftslegung und Lernen (vgl. Armytage, 2011; Cracknell, 2002; Kogen, 2018; OECD, 1991; Reinertsen, Bjørkdahl & McNeill, 2022).⁸ Im deutschen Raum unterscheiden Faust (2020, S. 64) und Stockmann (2007, S. 37) ebenfalls zwischen Erkenntnis-, Lern- und Kontroll- bzw. Rechenschaftslegungsfunktion. Je nachdem, welches Ziel primär verfolgt wird, muss in der Evaluierung unterschiedlich vorgegangen werden (Stockmann 2007, S. 37). Da Rechenschaftslegung und Lernen einen Zielkonflikt in der EPE bilden, muss – je nachdem, welches Ziel Vorrang hat – die Methodik angepasst werden (Armytage, 2011, S. 267; Stockmann, 2007, S. 38). Für die Funktion der Rechenschaftspflicht sollten

7 Die offene Kodierung ist eine induktive Herangehensweise, bei der – ohne Beachtung bestehender Theorien – das Interviewmaterial Satz für Satz ausgewertet und kategorisiert wird.

8 Jenseits einer eurozentrischen Sichtweise gibt es weitere Einordnungen zum Thema, welche Ziele Evaluierungen in der EZ verfolgen. Dazu äußern sich z.B. Shallwani und Dossa (2023).

die Evaluierenden extern und unabhängig sein, für die Lernfunktion möglichst intern, um die gelernten Inhalte besser in die Praxis umsetzen zu können (Cracknell, 2002, S. 56). Wenn die Rechenschaftspflicht im Vordergrund steht, sind neben der Unabhängigkeit auch Replizierbarkeit und Effizienz von Bedeutung. Steht dagegen das Lernen im Vordergrund, so muss der Fokus auf dem Prozess liegen. Auch Aktualität ist ein erheblicher Faktor, da die Erfahrungen akut weitergegeben werden müssen. Beide Ziele zu vereinen ist nahezu unmöglich, und auch Kompromissmethoden führen dazu, beide Bereiche nicht ausreichend abzudecken (Cracknell, 2002, S. 55–56; OECD, 2001, S. 65–67; siehe auch Hoey, 2015, S. 9; Kogen, 2018; Reinertsen et al., 2022, S. 363–364).

Das BMZ definiert Evaluierungen in der EZ als systematische und objektive Analysen und Bewertungen von laufenden oder abgeschlossenen Entwicklungsprojekten. Diese Untersuchungen erstrecken sich auf die Konzeption, die Durchführung und insbesondere die Ergebnisse der Entwicklungsprojekte. Sie sollen handlungsrelevante Erkenntnisse und ggf. Empfehlungen für Verbesserungen in der Gestaltung von Projekten enthalten (BMZ, 2021b, S. 3). Der Begriff „Evaluierung“ beinhaltet damit die Anwendung grundlegender Prinzipien sowie die methodisch fundierte und nachvollziehbare Analyse und Bewertung der empirischen Daten. Die grundlegenden Prinzipien sind die Nützlichkeit der Erkenntnisse (z. B. Schaffung von Transparenz und handlungsrelevanten Empfehlungen), die Glaubwürdigkeit der Befunde, die Unabhängigkeit der Bewertung, eine Partnerschaftlichkeit gegenüber den Projektpartnern sowie ethische Standards nach menschenrechtlichen Prinzipien (BMZ, 2021a, S. 7–8). EPE sind nur *eine* Form von Evaluierungen.

In der Theorie ziehen Durchführungsorganisationen wie die KfW das Ziel des Lernens aus Evaluierungen dem der Rechenschaftslegung vor, da sie primär ein Interesse daran haben, Effizienz und Wirksamkeit ihrer Organisation zu verbessern. Das Ziel der Rechenschaftslegung überwiegt bei den Akteuren, die die Mittel der Steuerzahlenden verwalten (z. B. Ministerien oder Parlamente), um diesen Rechenschaft darüber ablegen zu können, *ob* die Zusammenarbeit erfolgreich war oder nicht. Sie sind jedoch seltener daran interessiert, *warum* die Zusammenarbeit (nicht) erfolgreich war (Cracknell, 2002, S. 55–56). Das Lernen wird auch in der Praxis zunehmend in den Mittelpunkt der EZ gestellt. Die darauf beruhende evidenzbasierte Politikgestaltung zeigt sich auch in der Agenda 2030 für nachhaltige Entwicklung, die auf einer rigorosen ergebnisbasierten Managementlogik beruht (Yanguas, 2021, S. 1). Es ist jedoch nicht klar, ob Wissen (wozu Evaluierungen führen sollen) auch tatsächlich zum Lernen und zu Anpassungen in der Organisation führt (Dexis Consulting Group, 2020, S. 13, 33; Hovland, 2003, S. 7; Krohwinkel-Karlsson, 2007, S. 6–9). Trotzdem wurden Wissens- und Lernansätze in die EZ übernommen, obwohl es keine ausreichenden Beweise für ihre Wirksamkeit gibt (Yanguas, 2021, S. 19). Lernen in Organisationen hat oft einen informellen Charakter, weswegen eine wissenschaftliche Bewertung komplex ist (King & McGrath, 2003, S. 12). Die EZ nimmt dennoch seit Jahrzehnten an, dass ein größerer Wissensstand zu einer höheren Durchführungseffizienz von Entwicklungsprojekten führen muss, indem über das Lernen aus vergangenen Projekten die Praktiken und Ansätze angepasst und die Ressourcen besser eingesetzt werden, um somit die gewünschten Ergebnisse zu erzielen (Yanguas, 2021, S. 1).

Für die Rechenschaftslegungsfunktion ist Transparenz unausweichlich. Die Offenlegung der Ergebnisse von Evaluierungen sorgt für Transparenz über Erfolge und Defizite des Projektes. Daraus können Konsequenzen gezogen werden (Stockmann, 2007, S. 39). Bei der OECD herrscht ein breiter Konsens darüber, dass nicht nur Ergebnisse, sondern vollumfängliche Evaluierungsberichte offengelegt werden müssen (OECD, 2001, S. 26). Transparenz kann auch zu organisationsübergreifenden Lernerfahrungen führen und ermöglicht es der Öffentlichkeit und dem Auftraggeber (in diesem Fall das BMZ), die Wirksamkeit der EZ glaubwürdig zu überprüfen (Borrmann & Stockmann, 2009a, S. 135–136). Für das Legitimierungsziel sollen Daten belegen, mit welchem Input welcher Output und welche Wirkung durch ein Projekt erzeugt wurden. Die Mittelgeber und Durchführungsorganisationen können dadurch ihre Effizienz und ihren Wir-

kungsgrad nachweisen. EPE können zudem die Nachhaltigkeit einer Wirkung aufzeigen (Stockmann, 2007, S. 39–40). Evaluierungen werden auch wegen ihrer Symbolkraft genutzt. Dies ist jedoch nicht ihr eigentlicher Sinn (Stockmann, 2007, S. 40).

Evaluierungen sind in Deutschland gesetzlich verankert: In der Bundeshaushaltsordnung (BHO) ist die Vorgehensweise für Ausgaben und Einnahmen des Bundes geregelt. Für Evaluierungen relevant sind § 7 BHO – in dem es um Wirtschaftlichkeit und Sparsamkeit geht, die bei allen staatlichen Aufgaben und Maßnahmen bedacht werden müssen – sowie § 44 BHO und die dazugehörige Verwaltungsvorschrift 11a, die eine Erfolgskontrolle mit Nachweisprüfung fordert (BMZ, 2021a, S. 12–13). Dies ist in den Leitlinien „Evaluierung der Entwicklungszusammenarbeit“ vom BMZ (erstmalig 2007, in neuer Auflage 2021) festgehalten, die u. a. die Berichterstattung und Evaluierung regeln (BMZ, 2021a). Die Leitlinien sind verbindlich für das BMZ, die Durchführungsorganisationen und das DEval (BMZ, 2021a, S. 5). Evaluierungen sind dabei nur *ein* Instrument der Erfolgskontrolle. In den „Leitlinien für die bilaterale Finanzielle und Technische Zusammenarbeit mit Kooperationspartnern der deutschen Entwicklungszusammenarbeit“ (BMZ, 2021c) ist zudem angegeben, dass die Durchführungsorganisationen eigenverantwortlich in Anlehnung an die OECD/DAC-Kriterien Evaluierungen durchführen, um die Wirksamkeit der Entwicklungsmaßnahme beurteilen und Lehren für zukünftige Vorhaben ziehen zu können. Die Berichte von Schluss- und Ex-post-Evaluierungen⁹ werden an die Bundesregierung übersandt sowie veröffentlicht (BMZ, 2021c, S. 27).

Die EPE in der deutschen EZ hat eine Reihe von Zielen und Funktionen, die 2021 vom BMZ bestätigt wurden und sich an den OECD-Standards für Entwicklungsevaluierungen (OECD, 2010) orientieren. Für das BMZ sind Evaluierungen generell ein Mittel, um aus Erfahrungen zu lernen und Rechenschaft über die erzielten Ergebnisse abzulegen. Sie sollen durch möglichst objektive und empirisch fundierte Analysen und Bewertungen des Erfolgs von Entwicklungsmaßnahmen zur Verbesserung von deren Wirksamkeit und Legitimität beitragen sowie für die Hauptnutzenden Evidenz generieren (*Erkenntnisfunktion*). Evaluierungen sind eine Grundlage für Steuerungsentscheidungen sowohl für Einzelprojekte als auch für übergeordnete Fragestellungen in der EZ (*Lern- und Steuerungsfunktion*) (BMZ, 2021a, S. 7–8). Die aus der Evaluierung entstehende Evidenz soll zur Entscheidungsfindung sowie zum Wissens- und Datenmanagement beitragen (BMZ, 2021a, S. 37). Der Schwerpunkt der zentralen Evaluierungseinheiten der Durchführungsorganisationen (z. B. der FZ-E der KfW) soll auf Schluss- und Ex-post-Evaluierungen liegen, da entwicklungspolitische Wirkungen sowie Nachhaltigkeit erst nach Abschluss der Förderung fundiert beurteilt werden können (BMZ, 2021a, S. 27).

Aufgrund dieser konsequenten Wirkungsorientierung erhalten die Fach- und Grundsatzreferate des BMZ zentrale Aufgaben entlang eines Projektzyklus:¹⁰ Die BMZ-Referate sollen von Beginn an die Evaluierbarkeit der Projekte sichern. Dafür muss jedes Projekt mit Indikatoren hinterlegte Ziele auf der Wirkungsebene aufweisen. Das Evaluierungssystem des BMZ setzt nicht nur auf Evaluierungen der einzelnen Durchführungsorganisationen nach Abschluss der Projekte, sondern auch auf Monitoring und Evaluierung während der Projektdurchführung (BMZ, 2021a, S. 41). Das BMZ schlägt zur besseren Nutzbarkeit der Evaluierungserkenntnisse vor, Vorkehrungen zu treffen, um diese Erkenntnisse besser aufzubereiten und damit in die Praxis zu bringen, um wiederum Lernprozesse zu fördern (BMZ, 2021a, S. 40). Die aggregierten Erkenntnisse aus Evaluierungen sollen zudem von den Fachreferaten für die zukünftige strategische Gestaltung der EZ genutzt werden (BMZ, 2021a, S. 37–38). Der Bundesrechnungshof setzt

9 Schlussevaluierungen werden direkt nach Projektabschluss durchgeführt (wie es bei der GIZ [Gesellschaft für internationale Zusammenarbeit] meist der Fall ist), wohingegen EPE erst einige Jahre nach Projektabschluss abgewickelt werden.

10 Der Projektzyklus besteht aus Planung, Implementierung, Überwachung, Lernen und Anpassung von Folgeprojekten.

dem entgegen, dass es dem BMZ derzeit nicht gelingt, bestehende Evidenzen zu aggregieren und für Steuerungs- oder Lernzwecke zu nutzen (BRH [Bundesrechnungshof], 2021, S. 20).

Die KfW verdeutlichte 2021 ihre Ziele einer EPE: Sie arbeite seit 20 Jahren mit EPE und trage dadurch Wissen zusammen, um über Wirkung und Nachhaltigkeit der FZ Aussagen treffen zu können. Positive wie negative Wirkungen sollen untersucht und quantifiziert werden, um dadurch die Wirksamkeit der FZ zu steigern (KfW, 2021a, S. 10, 28). Auch gehe es um die Feststellung der konkreten Wirkung: Das Ziel sei Wissensgenerierung über die Erfolge und Wirkungen der FZ. Die Evaluierungen sollen analysieren, ob Projekte ihre gesetzten Ziele erreichen und die Hintergründe reflektieren. Damit solle die Qualität der Arbeit der KfW gesichert und aus Erfahrungen gelernt werden (KfW, 2022c). Es gehe um institutionelles sowie operatives Lernen, da die Lernerfahrungen systematisch in den operativen Bereich zurückgespielt werden sollen (KfW, 2021a, S. 2). Dass die EPE, wie vom BMZ betont, auch der Rechenschaftslegung dient, wird auf der KfW-Webseite wie auch in KfW 2021a nicht explizit dargelegt – genauso wenig wie der Umstand, dass die EPE politische Steuerungsrelevanz haben kann (KfW, 2022c).

Zusammenfassend formuliert, sind die vom BMZ definierten Ziele der EPE die Rechenschaftslegung, das Lernen für zukünftige Projekte, der Nachweis von Nachhaltigkeit und entwicklungspolitischer Wirkung, die institutionelle Qualitätssicherung und (auf die KfW bezogen) die Wirksamkeitsanalyse der FZ. Die Erkenntnisse aus (Meta-)Evaluierungen¹¹ dienen zudem beim BMZ der strategischen Gestaltung der EZ. Die Literatur sieht ähnliche Ziele und Funktionen wie BMZ und KfW, die Leitlinien sind dementsprechend theoretisch gut begründet. An manchen Stellen bezweifelt die Literatur, ob alle diese Ziele einer Evaluierung (gleichzeitig) erfüllt werden können, und zeigt Zielkonflikte auf. Der Bundesrechnungshof kritisiert die vom BMZ definierten Ziele als zu ungenau und fordert Nachbesserungen (BRH, 2021, S. 13). Die KfW konzentriert sich in ihren EPE (nach eigener Aussage) überwiegend auf den Lernaspekt.

4 Wie wird eine Ex-post-Evaluierung in der KfW erstellt? Ein Überblick

Nach Klärung des Zwecks einer EPE wird im Folgenden kurz der Ablauf einer EPE erläutert, um im darauffolgenden Kapitel die Herausforderungen der Methodik bewerten zu können: Für die Erstellung der EPE gibt es genaue Leitlinien vom BMZ (siehe BMZ, 2021a, S. 18–20), an denen sich die KfW orientiert. Die wichtigsten Schritte sollen im Folgenden aufgezeigt werden. Idealtypisch besteht ein Evaluierungsprozess aus vier Schritten: der Vorbereitung und Konzeption, der Datenerhebung und -analyse, der Berichtserstellung sowie der Umsetzung von Empfehlungen (BMZ, 2021a, S. 17).¹²

Vorbereitung der Ex-post-Evaluierung

Zu Beginn wird entschieden, welche Projekte ex-post evaluiert werden. Dafür wird bei der Evaluierungsabteilung FZ-E am Anfang jeden Jahres eine nach Sektoren geschichtete Zufallsstichprobe aus den abgeschlossenen sogenannten „evaluierungsreifen Vorhaben“ gezogen. In der KfW beträgt die Zufallsstichprobe 50 % der Grundgesamtheit (2019/20 waren es 149 gezogene

11 Meta-Evaluierungen sind übergeordnete Evaluierungen, die z. B. aus mehreren einzelnen Projektevaluierungen eines Sektors erstellt werden und diese bewerten bzw. die Ergebnisse zusammenfassen.

12 Wie Evaluierungen idealtypisch durchgeführt werden, kann in Stockmann (2007, S. 108) und angewandt auf die KfW in Borrmann und Stockmann (2009b) sowie bei Noltze, Euler und Verspohl (2018a, S. 12) nachgelesen werden.

Projekte).¹³ Evaluierungsreife Vorhaben sind solche, bei denen die Abschlusskontrolle durchgeführt wurde und seitdem ca. drei bis fünf Jahre vergangen sind. Für diese Projekte werden EPE durchgeführt (KfW, 2021a, S. 42). Die Zufallsstichprobe garantiert deren Unabhängigkeit und Repräsentativität. Damit kann eine aussagekräftige Durchschnittsnote gebildet werden, die repräsentativ für Sektor und Jahr ist (BRH, 2022, S. 14; Zintl, 2009, S. 247).

Drei verschiedene Personengruppen können eine EPE in der KfW durchführen: interne Projektmanager*innen der FZ-E; externe Evaluierungsconsultants, die auf Ausschreibungen der FZ-E reagieren; und sogenannte Abgeordnete. Letztere sind Mitarbeitende aus anderen Abteilungen der KfW, die sich bereiterklären, eine EPE durchzuführen. Die Verteilung der Gruppen beläuft sich auf ca. 25 % interne Projektmanager*innen, 25 % externe Consultants und 50 % Abgeordnete. Im Folgenden wird erläutert, wie Abgeordnete eine EPE durchführen.¹⁴ Die zu evaluierenden Projekte werden nach mehreren Kriterien besetzt: Die/Der Abgeordnete darf nicht an diesem Projekt beteiligt gewesen sein. Es darf keine direkte Verbindung zum Team oder zur Abteilungsleitung bestehen, um Neutralität bzw. Unabhängigkeit zu gewährleisten. Die/Der Abgeordnete sollte die Landessprache sprechen und über Sektorkenntnisse verfügen (KfW¹⁵; Borrmann & Stockmann, 2009b, S. 250).

Während der Durchführung der EPE werden die Abgeordneten von den Betreuenden der Evaluierungsabteilung in Fragen der Methodik, Datengenerierung, Benotung etc. unterstützt. Die Betreuenden garantieren die Qualitätssicherung und sind für den Bericht verantwortlich (BRH, 2022, S. 4). Verantwortlich bedeutet in diesem Fall, „Objektivität“ bzw. Vergleichbarkeit in der EPE zu generieren, da FZ-E-Mitarbeitende, laut KfW2 und KfW4, einen größeren Erfahrungsschatz haben und besser Vergleiche ziehen können. Sie wissen z. B., welche Kriterien ein Projekt nicht erfüllt haben muss, um die Note 5¹⁶ zu erhalten (KfW2, KfW4).

Methodik der Ex-post-Evaluierung

In der KfW werden EPE mit der Rapid-Appraisal-Methode durchgeführt. Eine Beschreibung des Rapid Appraisal findet sich bei Beebe (1995) und Chambers (1981). Rapid Appraisal ist ein Ansatz, der sich auf mehrere (häufig qualitative) Bewertungsmethoden und -techniken stützt, um schnell und systematisch Daten zu sammeln, wenn die Zeit vor Ort bzw. das Budget limitiert sind oder es eine begrenzte Verfügbarkeit von zuverlässigen Sekundärdaten gibt. Vindrola-Padros und Johnson (2020) führten einen Literatur-Review zu „rapiden“ Evaluierungsmethoden durch. Neben der Auflistung der Vorteile von z. B. Zeit- und Kostenersparnis fanden sie auch Einschränkungen, beispielsweise die Herausforderung, nicht dieselbe „Tiefe“ oder „Interpretationsebene“ wie bei herkömmlichen Datenanalysemethoden zu erreichen bzw. einem Detailverlust zu unterliegen, sowie eine ggf. höhere Voreingenommenheit der Evaluierenden (Vindrola-Padros & Johnson, 2020, S. 1600). Sie weisen darauf hin, dass diese „rapiden“ Techniken nicht

13 Die anderen 50 % werden nicht evaluiert, die Wirkung wird daher nicht überprüft. Wie jedes Projekt in der KfW erhalten sie eine Abschlusskontrolle, die vom operativen Team nach Projektabschluss durchgeführt wird.

14 Diese Schritte durchlaufen die internen Evaluierenden von FZ-E in ähnlicher Weise. Es ist davon auszugehen, dass externe Evaluierungsconsultants auf ähnliche Art evaluieren.

15 Anonymisierte direkte und indirekte Zitate wurden zu Beginn der Primärdatenerhebung von allen 13 Befragten erlaubt; diese Erlaubnis wurde im Vorfeld dieser Publikation aber von vier Interviewten wieder zurückgezogen. Zur Anonymisierung der verbliebenen Interviews wurden Abkürzungen der in der Studie berücksichtigten Institution genutzt und von KfW1 bis KfW6, als BMZ1, DEV1 und IDOS1 durchnummeriert. Die Stellenbeschreibungen wurden gekürzt und finden sich in Anhang 2.

16 Das Notenspektrum orientiert sich an Schulnoten und geht von 1 bis 6, wobei bereits die Note 4 als „nicht erfolgreich“ gilt.

für alle Forschungsumgebungen geeignet sind; diese Techniken erfordern methodische Anpassungen, für die es mehrere (erfahrene) Forschende brauche (Vindrola-Padros & Johnson, 2020, S. 1601).

Der Rapid-Appraisal-Ansatz wird bei der KfW mit quantitativen Methoden unterstützt bzw. hinterlegt (BRH, 2022, S. 12). Die Evaluierenden besuchen das Projektland und befragen die Zielgruppe, die Träger bzw. Partner, andere am Projekt beteiligte Personen sowie Wissenschaftler*innen oder Sektorexpert*innen und andere Entwicklungsorganisationen. Eine Sichtung interner und externer Informationsquellen,¹⁷ wissenschaftlicher Berichte und Studien sowie ggf. quantitativer Datenquellen wird vorgenommen. Schlussendlich werden die Ergebnisse und Erkenntnisse trianguliert (KfW, 2019, S. 53–57).

Das operative Team erstellt bei der Projektkonzipierung eine Wirkungsmatrix. Diese ist ein im Zusammenhang der Beauftragung festgelegter Rahmen, wie Output, Outcome und Impact gemessen werden (KfW5), und baut auf der Wirkungslogik auf (DEV1; GIZ, 2012, S. 28). Dabei bezeichnen die Inputs die Ressourcen, die von der KfW in ein Projekt hineinfließen, um Outputs zu erzielen. Outputs sind Leistungen, die durch den Einsatz dieser Ressourcen entstehen. Der Outcome zeigt direkte positive wie negative Wirkungen des Projektes auf, die aus der Nutzung des Outputs für die Zielgruppe entstanden sind (GIZ, 2012, S. 25). Der Impact gibt den entwicklungspolitischen Nutzen eines Projektes an (GIZ, 2012, S. 24). Er ist nicht auf der Ebene eines einzelnen Projektes definiert, sondern auf jener des ganzen EZ-Programms, also pro Land und Sektor. Das einzelne Projekt muss mit dem Projektziel und den dazugehörigen Indikatoren in das EZ-Programm „ein zahlen“ (KfW5; vgl. dazu auch Amine & Eulenburg, 2022, S. 2). In der EPE gilt die Wirkungslogik als Bewertungsgrundlage: Die/Der Evaluierende hinterfragt sie, überprüft anhand dessen, ob und inwieweit das Projekt einen Beitrag zu den Zielen geleistet hat, und untersucht, wie sich die Indikatoren aus der Wirkungsmatrix über die Zeit entwickelt haben (DEV1; BMZ, 2020).

Um zweifelsfrei festzustellen, ob eine Wirkung durch das Projekt erzielt wurde, müsste eine rigorose Impact-Evaluierung¹⁸ durchgeführt werden (BRH, 2022, S. 13), denn in der EPE wird häufig mit Annahmen darüber gearbeitet, welche Wirkung das Projekt entfalten *könnte*. Laut KfW lohnen sich rigorose Impact-Evaluierungen für „normale“ EPE aufgrund der hohen Kosten und des verhältnismäßig geringen Nutzens nicht. Eine Untersuchung vom Krämer et al. zeigt jedoch, dass der Nutzen von rigorosen Impact-Evaluierungen ihre Kosten häufig übersteigt (Krämer, Jechel, Kretschmer & Schneider, 2021, S. 36). Die Kontrollgruppenmethode wird in den EPE zwar nicht angewandt, allerdings können z. B. mittels Satellitendaten Kontrollgruppenvergleiche durchgeführt werden, ohne von Anfang an Daten in der Kontrollgruppe zu erheben (KfW, 2021a, S. 28, 34), wie es bei einer rigorosen Impact-Evaluierung gemacht werden müsste.

Datenerhebung für die Ex-post-Evaluierung

Zur Datenerhebung wird meist eine etwa zweiwöchige Reise ins Projektland unternommen.¹⁹ Davor werden ein Interview-Fragenkatalog entwickelt und Personen identifiziert, die befragt werden sollen (KfW3, KfW4). Da die Projektgebiete und die Zielgruppe der Projekte oft groß sind, erfolgt eine Stichprobenauswahl der Gebiete (ggf. Dörfer) und der Vertreter*innen der Zielgruppen.

17 Diese umfassen z. B. Fortschrittsberichte, die Abschlusskontrolle, Consultingberichte, Berichte der Partner etc.

18 Rigorose Impact-Evaluierungen können Wirkungen bzw. Effekte *kausal* einem Entwicklungsprojekt zuordnen. Eine genaue Erklärung findet sich bei der KfW (2022a).

19 Teils werden auch Schreibtisch- bzw. Remote-Evaluierungen durchgeführt.

Pro DAC-Kriterium werden bestimmte Leitfragen untersucht, die sich u. a. auf die Prinzipien der Agenda 2030 beziehen (DEV1). Sie sind in der Evaluierungskonzeption durch die Orientierungslinie zum Umgang mit den OECD/DAC-Evaluierungskriterien (BMZ, 2020) vorgegeben und müssen sowohl bei der Konzipierung von Neuvorhaben als auch bei jeder Berichterstattung während der Projektlaufzeit berücksichtigt werden. Die Leitfragen dienen der Vergleichbarkeit.

Die Noten werden von den Evaluierenden anhand von quantitativen Daten und qualitativen Eindrücken aus den Interviews vergeben. Die Gesamtnote eines Projektes ergibt sich nicht aus dem Durchschnitt der Teilnoten der DAC-Kriterien (KfW, 2021a, S. 43), sondern es gibt drei sogenannte K.-o.-Kriterien: Effektivität, entwicklungspolitische Wirkung und Nachhaltigkeit. Tritt bei diesen ein schwerwiegender Mangel auf, so soll er nicht durch andere Kriterien mit einer positiven Bewertung ausgeglichen werden können, und das ganze Projekt soll nicht mehr als „insgesamt erfolgreich“ gelten (KfW, 2021a, S. 42). Somit wird es mit einer schlechteren Gesamtnote als 3 bewertet (DEV1).

Der Berichtsentwurf wird an die Projektverantwortlichen oder an den zuständigen operativen Bereich geschickt. Sie haben die Möglichkeit, Stellung zu nehmen, falls Fakten ihrer Ansicht nach nicht korrekt dargestellt wurden oder fehlen. Laut den für diese Studie Befragten ergeben sich meist keine bedeutenden Änderungen daraus, auch nicht an der Note (KfW2, KfW5, KfW6). Abschließend wird die fertige EPE über den Evaluierungsverteiler u. a. an das zuständige Team bzw. die zuständige Abteilung des Projektes und an das BMZ verschickt sowie im Dokumentenmanagementsystem abgelegt. Teil 1²⁰ wird auf der Webseite der KfW veröffentlicht und an den Partner bzw. Träger geschickt. Teil 2 ist weder für den Partner noch für die Öffentlichkeit zugänglich.

5 Herausforderungen in Ex-post-Evaluierungen: Was ergeben die Interviewdaten?

In der Theorie wirken die Prozesse und Abläufe der Entstehung einer EPE sinnvoll und sind schlüssig aus der Methode des Rapid Appraisals hergeleitet. In der Praxis werden EPE mit erheblichem Aufwand und großer Umsicht betrieben, jedoch berichten mit dem System vertraute Expert*innen auch von Limitierungen. Die Datenauswertung der Interviews ergibt potenzielle Herausforderungen zur Umsetzung der Methodik der EPE und des Ablaufs. Hier folgt ein Überblick über die bedeutsamsten von den Interviewten genannten Kritikpunkte der Methodik und deren theoretische Einordnung. Diese Kritikpunkte wurden induktiv aus den Interviewdaten hergeleitet und in Oberkategorien zusammengefasst, die im Folgenden einzeln beschrieben werden. Die Analyse der gesammelten Kritikpunkte erhebt aufgrund der nicht repräsentativen Stichprobe keinen Anspruch auf Vollständigkeit. Jedoch illustrieren die Daten insbesondere wiederholt benannte Herausforderungen. Eine theoretische Einordnung dieser Einschätzungen zeigt mögliche Verbesserungsmöglichkeiten auf.

20 Teil 1 der EPE beginnt mit einem Deckblatt, auf dem Daten, Ziele und Umsetzung des Projektes, die bedeutsamsten Erkenntnisse in Stichpunktform, die Gesamtbewertung in Worten und die Schlussfolgerungen abgebildet sind. Ab S. 1 werden die einzelnen DAC-Kriterien mit Text und Note aufgeführt. Teil 1 umfasst ca. sechs bis acht Seiten und ist auf der KfW-Webseite in deutscher und englischer Sprache öffentlich zugänglich. Teil 2 enthält u. a. die Methodik und nähere Erläuterungen zu den Aussagen in Teil 1.

Mögliche Verzerrungen bei Methodik und Datengenerierung

Alle EPE werden mit Rapid Appraisal bewertet, nach derselben Methode. Nur teilweise wird nach Größe und Umfang des Projektes unterschieden.²¹ Auch werden keine Counterfactuals konstruiert, also Szenarien, wie das Geld alternativ hätte eingesetzt werden können (IDOS1; vgl. hierzu auch White, 2009, S. 213). Die Qualität einer Rapid-Appraisal-Methode hängt von der Komplexität der Fragestellung ab: Je komplexer diese ist, desto schwieriger ist es, sie in kurzer Zeit zu beantworten. Ob Rapid Appraisal als Methode ausreicht, hängt dementsprechend stark vom Projekttyp ab (KFW4). Für eine statistisch repräsentative Zufallsstichprobe reicht die Zeit für die EPE nicht immer aus: KFW4 gibt an, dass Evaluierende nicht mehr als zehn bis zwölf Interviews an einem Tag führen können. Hat die Evaluierungsreise eine Dauer von maximal zwölf Tagen, so kann die/der Evaluierende keine repräsentative Stichprobe interviewen, wenn die Zielgruppe mehrere tausend Menschen umfasst und das Gebiet zudem sehr groß ist (KFW4). Um dieses Problem zu minimieren, setzt die KfW nach Möglichkeit lokale Gutachter*innen ein (KFW4). DEV1 bewertet die begrenzten Evaluierungstage vor Ort als für einige Fragestellungen ausreichend; auf der anderen Seite würden über die DAC-Kriterien Fragen an das Projekt gestellt, die die/der Evaluierende ggf. nicht komplett in der begrenzten Zeit beantworten kann. Nur wenn die Monitoringsysteme der KfW umfassend sind, ist auch Rapid Appraisal als Format angemessen. Aktuell sind die Monitoringsysteme aber eher schwach (DEV1; vgl. dazu Hartmann, Amine, Klier & Vorwerk, 2019, S. 48). KFW2 war jedoch überrascht, wie umfangreich und sorgfältig die Evaluierung mit Rapid Appraisal war.

Teilweise sucht der Träger die zu interviewenden Zielgruppenvertreter*innen oder die Dörfer und Regionen selbst aus, die von den Evaluierenden besucht werden sollen. Das kann zu einer Verzerrung der Ergebnisse führen, wenn die Evaluierenden die Stichprobenauswahl unkritisch übernehmen bzw. dahingehend nicht ausreichend geschult wurden. Es ist ihnen dennoch möglich, sich vom Vorschlag des Partners zu distanzieren und Standorte nach eigenem Ermessen auszuwählen (KFW4). Manchmal bitten die Evaluierenden die operativen Teams um Empfehlungen für Projektstandorte (KFW3). Es kommt vor, dass Mitarbeitende des Projektträgers die Evaluierenden zu den Projektstandorten oder auch regionale KfW-Mitarbeitende sie zu den Gesprächen mit dem Träger begleiten (KFW2). Dadurch könnten sich Befragte eingeschüchtert oder kontrolliert fühlen und ihr Antwortverhalten dementsprechend anpassen (KFW4, IDOS1). Dass z. B. Zielgruppenvertreter*innen befragt werden, die nah an der Straße leben, während abseits der Straße wohnende Personen nicht interviewt werden, kann ebenfalls verzerrend wirken (KFW4). Denn Projekte und ihre EPE finden in Ländern statt, in denen die Infrastruktur teilweise nicht gut ausgebaut ist und die Evaluierenden deswegen nur einen Teil der Projektstandorte besuchen können. Der Länderkontext muss daher als Verzerrungsgrund ernstgenommen werden (IDOS1).

Herausforderungen mit Indikatoren und Zielformulierungen

Unpräzise, schwache Zielformulierungen stellen eine weitere Problematik dar, mit der eine Evaluierung nur positiv ausfallen kann (Wilhelm, 2015, S. 10–11). Ziele werden entweder von Beginn an schwammig formuliert oder die Zielwerte niedrig gesetzt, und anspruchsvolle Zielformulierungen werden vermieden (Amine & Eulenburg, 2022, S. 3; Stockmann, 2007, S. 64). Eine weitere Herausforderung ist die unzureichende Datengrundlage, da Daten in sogenannten Entwicklungsländern teils schwierig zu ermitteln sind (Wilhelm, 2015, S. 10–11).²² Die Befragten geben

21 Ggf. wird Rapid Appraisal um quantitative Methoden ergänzt, z. B. um Satellitendatenauswertungen.

22 Mit dieser Verallgemeinerung muss vorsichtig umgegangen werden: Viele sogenannte Entwicklungsländer haben in den letzten Jahren umfangreiche Fortschritte im Ausbau ihrer nationalen Evaluierungssysteme gemacht und sind somit inzwischen selber fähig, Daten zu aggregieren und bereitzustellen. Es scheint sich jedoch abzubilden, dass mit den Systemen der Partner bisher kaum gearbeitet wird.

an, dass teilweise „vergessen“ wird, Baseline-Daten zu Beginn des Projektes zu erheben. Die in der Abschlusskontrolle bzw. für die EPE erhobenen Daten können daher nicht mit den Baseline-Daten abgeglichen werden, um eine Entwicklung festzustellen. In diesem Bereich wäre ggf. eine engere Zusammenarbeit mit dem Partner sinnvoll (Holzapfel & Römling, 2020, S. 35). Viele Indikatoren²³ können zwar Output und ggf. auch Outcome darstellen, über die breite Wirkung eines Projektes lassen sie jedoch keine Aussagen zu (Borrmann & Stockmann, 2009b, S. 256; Holzapfel & Römling, 2020, S. 31). Auch in dieser Untersuchung werden Problematiken mit Indikatoren festgestellt: Am Ende wird ein Projekt u. a. daran gemessen, ob die am Anfang gesetzten Indikatoren erfüllt sind. Im Biodiversitätssektor gibt es jedoch z. B. teilweise keine passenden Indikatoren, um die Wirkung zu messen, bzw. die Wirkungen konnten sich bis zur Messung nicht vollständig entfalten (KfW5). Dies wurde von mehreren Befragten übereinstimmend geschildert.

Das Abgeordnetensystem als möglicher Verzerrungsgrund

Laut Bundesrechnungshof (BRH) haben das BMZ sowie die KfW betont, dass die von Abgeordneten durchgeführten EPE durch die Betreuung der FZ-E den Qualitätsstandards entsprechen und die Abgeordneten den Anforderungen gewachsen sind (BRH, 2022, S. 10). Andererseits macht der BRH deutlich, dass die FZ-E „schwerwiegende methodische Defizite bei Evaluierenden [gemeint sind Abgeordnete] gegebenenfalls nicht vollständig und mit angemessenem Aufwand ausgleichen“ (BRH, 2022, S. 11) könne.

Die Evaluierenden, die eine EPE erstellen, brauchen *umfassende Kenntnisse* von Ländern, Sektoren und Projekttypen. Selbst wenn dieses Wissen vorhanden ist, ist es problematisch, es in die EPE zu übertragen (BMZ1). Abgeordnete evaluieren häufig Sektoren und Länder, in denen sie *nicht* arbeiten (KfW6); daher mangelt es ihnen ggf. am Hintergrundwissen zu Land und Sektor. Ob Länder eine „perfekte potemkinsche Inszenierung“ vorführen können oder nicht, muss die/der Evaluierende mitbedenken (KfW4). Teilweise ist die KfW in Ländern aktiv, in denen der Zugang zu lokalen Akteuren problematisch ist oder in denen diese nicht frei reden können (IDOS1). Können die Evaluierenden diese Gegebenheiten nicht (richtig) einschätzen, so kann es zu Verzerrungen kommen.

Abgeordnete kennen die Prozesse und Abläufe in der FZ und der KfW besser als externe Evaluierende und können sie daher besser beurteilen, haben aber ggf. methodische Defizite. Externe Evaluierungsconsultants sind jedoch auch nicht zwingend ein Qualitätskriterium (KfW2), da es bei ihnen vorkommen kann, dass Bewertungen ins Positive verzerrt werden, um Folgeaufträge zu erhalten (KfW4). Die Schlussfolgerung aus den Ansichten der Befragten und der Literatur ist demnach, dass, wenn es um Verzerrungen geht, weder die Abgeordneten noch die Evaluierungsconsultants eine umfassende Lösung darstellen. Daher scheint die Kombination aus Abgeordneten, Evaluierungsconsultants und FZ-E-Mitarbeitenden angemessen, auch wenn Abgeordnete ggf. besser geschult werden sollten.

Objektive und subjektive Bewertung in der Ex-post-Evaluierung

Die EPE sollte eine möglichst objektive Darstellung sein.²⁴ Unter Objektivität wird verstanden, dass unabhängige Forschende bei denselben Untersuchungsinstrumenten zu denselben statistischen Befunden kommen. Das stellt die Vergleichbarkeit sicher, die vom BMZ gefordert wird.

23 Indikatoren sind – nicht nur in der deutschen EZ – ein umfangreiches Thema und können mit dieser Arbeit nicht allumfassend abgedeckt werden. Literatur zu Indikatoren, die von der KfW verwendet werden, findet sich in KfW 2021c; zur Diskussion um Indikatoren (in der EZ) äußern sich z. B. Armytage (2011, S. 267–268), Bartl, Papilloud und Terracher-Lipinki (2019), Goodwin (2017), Rottenburg, Merry, Park und Mugler (2015) sowie Sabbi und Stroh (2020).

24 Die Objektivität ist eines der Gütekriterien der empirischen Sozialforschung (siehe Lienert & Raatz, 1998).

Sie soll laut BMZ mit den gegenstandsangepassten Evaluierungsfragen und den einheitlichen Bewertungsmaßstäben – den Noten – gewährleistet werden (BMZ, 2021a, S. 28). Trotzdem sind Ergebnisse einer Evaluierung beobachtungsabhängig. Evaluierende nehmen eine bestimmte Sichtweise ein und können daher nicht völlig unabhängig oder objektiv bewerten (Wilhelm, 2015, S. 59–60; zur eingeschränkten Erreichbarkeit der Objektivität in der Wissenschaft siehe auch Weber, 1977). Viele Interviewte erwähnen die Subjektivität in der EPE: KFW2, KFW4, KFW6, BMZ1 und IDOS1 empfinden EPE ausdrücklich als (sehr) subjektiv, da sie auf dem Erfahrungsschatz der Evaluierenden basieren, diese ihren eigenen Fokus setzen können und ein Ermessensspielraum vorhanden ist. Die DAC-Kriterien zwingen jedoch alle Evaluierenden zur „Objektivität“ und Herleitung (KFW2). KFW5 setzt dagegen, dass die DAC-Kriterien von unterschiedlichen Evaluierenden sehr individuell befüllt werden können. Es werde versucht, objektive wissenschaftliche Kriterien zu erfüllen, eine tatsächliche Objektivität gebe es in der EPE jedoch nicht (IDOS1). Dabei soll jedoch auch dargelegt werden, wenn die Erkenntnisse andere Schlussfolgerungen oder Bewertungen nahelegen (KFW4).

Subjektivität betrifft nicht nur die Interpretation der Antworten, sondern auch das Antwortverhalten der Befragten an sich – teils aufgrund der eigenen subjektiven Einschätzung, teils aus strategischer Motivation. Eine völlige Objektivität kann somit nie erreicht werden – auch wenn die FZ-E versucht, diese anzustreben. Trotzdem muss eine Vergleichbarkeit zwischen den EPE vorhanden sein, da sonst die Daten z. B. nicht für eine Erfolgsquote genutzt werden können.

Notensystem der Ex-post-Evaluierung: objektiv oder subjektiv?

Ein weiterer Hauptbestandteil der EPE ist das System der Notenvergabe. Jedes DAC-Kriterium bekommt in der EPE eine eigene Note von den Evaluierenden, woraus die Gesamtnote berechnet wird. Diese soll die Feststellungen des Textes in eine Bewertung nach transparenten Maßstäben übersetzen. Sie sollte jedoch nicht isoliert betrachtet werden, sondern im Zusammenhang mit dem hinterlegten Text, der sie begründet. Für KFW2 und KFW4 ist dieser Text bedeutsamer als die Note selbst. Auch wenn diese Einschätzung von zwei Befragten geteilt wird, werden trotzdem besonders gute oder schlechte Noten hervorgehoben und diskutiert.

Laut Porter sind Noten in gewisser Weise immer subjektiv (Porter, 2015, S. 36). Eine Objektivität ist demnach kaum möglich, und auch standardisierte Fragenkataloge können die Subjektivität nicht vollständig eliminieren. KFW4 ist der Ansicht, dass eine Bewertung auch durch Formalisierung nicht glaubwürdiger wird. Die Notenvergabe erfolgt neben dem „gesunden Menschenverstand“ und der Erfahrung auch nach eigenem Ermessen und bietet einen „Ermessensspielraum“ (KFW4). Auch KFW6 hält Noten für subjektiv, da die Faktenlage nicht immer eindeutig sei. DEV1 stellt dar, dass Evaluierende allen Befragten die gleichen Fragen stellen und, wie all ihre Kolleg*innen, zur gleichen Faktenlage Daten erheben und auswerten. Trotzdem ergeben sich unterschiedliche Bewertungen, insbesondere bei Abgeordneten, die wenig Erfahrung mit EPE haben. Die FZ-E versucht das auszugleichen, indem die Noten im Berichtsentwurf zusätzlich mit früheren (aus den letzten 10 bis 20 Jahren) oder gleichzeitig durchgeführten EPE von ähnlichen Projekten verglichen werden und überprüft wird, ob diese EPE sich als Ausreißer darstellt – ob demnach systematisch positiver oder negativer bewertet wurde als in den letzten 20 Jahren. Dieser Vergleich wird innerhalb der FZ-E diskutiert und im Zweifel nachgearbeitet (DEV1). Dadurch kann jedoch eine gewisse Pfadabhängigkeit entstehen.

Notensystem der Ex-post-Evaluierung: Anreize und Verzerrungen

Gegen Ende der Erstellung erhält das operative Team des jeweiligen Projektes die EPE, um Feedback geben zu können. Interviewte berichten über (häufige) Diskussionen zwischen Evaluierenden und dem operativen Bereich während dieser Feedbackprozesse (KFW2, KFW4, KFW5). Bei manchen Projekten ist die Note für den operativen Bereich jedoch irrelevant und es entstehen keine Diskussionen, z. B. bei „konzeptionellen Ladenhütern“: „Das ist den Operativen

egal, solange es keine 4 oder schlechter ist. Bei 4 oder schlechter bei den Ladenhütern sagt man: Können die damit recht haben?“ (KFW4). KFW4 sagt zudem, dass die Noten 2 und 3 kaum Interesse wecken. Wie in Kapitel 6.3 dargelegt wird, werden diese Noten aber am häufigsten vergeben.

Wie bereits betont wurde, weckt der Text geringeres Interesse als die Noten. KFW3 und KFW6 geben an, dass Noten mit einem gewissen Stolz verknüpft sind. Sobald Zahlen im Umlauf sind, werden die Produkte, mit denen sie in Verbindung stehen, objektiviert und neutralisiert und neigen dazu, zu einer neuen „Realität“ zu werden (Desrosières, 2015, S. 334). Noten können leichter von den Mitarbeitenden aufgenommen und verarbeitet werden. Eine gute Note in der EPE ist laut Befragten, die längerfristig an Projekten arbeiten, mit Stolz verknüpft. Das deutet darauf hin, dass die EPE ernst genommen wird (was auf die Unabhängigkeit und Genauigkeit der Beurteilung hinweist) und Eigenmotivation vorhanden ist. Das kann aber auch zu Verzerrungen der Note ins Positive führen, da Mitwirkende an der EPE aus den oben dargelegten Gründen ein Interesse haben, das Projekt besser zu bewerten als es eigentlich ist. Dass Noten durchaus eine wichtige Rolle spielen, zeigt auch die Diskussion darum im Feedbackprozess der EPE. Hinzu kommt, dass Befragte angeben, den Text unter guten Noten nicht zu lesen bzw. zu hinterfragen; auch dies ist ein Hinweis auf Stolz. Gäbe es ein reines Erkenntnisinteresse, dann sollten auch die guten Noten im Hinblick auf mögliche Fehler betrachtet werden.

Auch das politische Umfeld kann einen verzerrenden Effekt auf Evaluierungen (Henry & Mark, 2003, S. 301–302) bzw. auf die Notenvergabe haben. Als Beispiel soll hier eine Meta-Evaluierung von Kirsch und Wilson über EPE von KfW und GIZ aus Afghanistan dienen: Wegen des hohen internationalen Interesses und der umstrittenen Militärmission stieg der politische Druck, dort schnelle Ergebnisse in der EZ zu erzielen (Kirsch & Wilson, 2014, S. 27). Die Informationen aus den Evaluierungen wurden verwendet, um bereits getroffene politische Entscheidungen zu stützen, anstatt aufzuzeigen, was vor Ort (nicht) funktionierte. Keines der Projekte erhielt eine schlechtere Note als 3, obwohl Outcome und Impact kaum bestimmt wurden (Kirsch & Wilson, 2014, S. 33–34). Diese unzureichende Bestimmung kommentierte das BMZ in seiner offiziellen Stellungnahme zu Kirsch und Wilson ebenfalls kritisch (BMZ, 2014). Dieses Beispiel zeigt, wie Noten auch von äußeren Umständen und politischen Interessen beeinflusst werden können.

6 Ex-post-Evaluierungen und ihr Nutzen: Wie werden Ergebnisse verwendet?

6.1 Wer liest die Ex-post-Evaluierung?

Um zu verstehen, welchen Nutzen EPE haben, muss ermittelt werden, wer die Berichte liest, was und warum (nicht) gelesen wird und wie Evaluierungsergebnisse insgesamt genutzt werden, denn das Lernen aus Evaluierungen erfolgt nicht ausschließlich über das Lesen der Berichte. Jährlich entstehen unzählige Dokumente zum Thema EZ. Dieses Wissen kann jedoch keinen Nutzen haben, wenn es nicht gelesen (Yanguas, 2021, S. 9) oder kaum in formellen Verfahren genutzt wird (Krämer et al., 2021, S. 38).

Die *primäre Zielgruppe* der EPE sind die Referent*innen im BMZ (DEV1). Die Interviews aus dieser Forschung legen nahe, dass EPE im BMZ nur in geringem Maße gelesen werden. Das BMZ besteht zwar auf einer Berichterstattung aller EZ-Organisationen über die Projekte (BMZ, 2021a, S. 12). Das führt aber dazu, dass die Berichterstattung an die BMZ-Referate viele Seiten umfasst, die (mutmaßlich aus Zeitmangel) nicht alle gelesen werden können. Die EPE konkurriert daher mit einer großen Menge anderer Berichte und Tätigkeiten der Länderreferate (KFW2; Hartmann et al., 2019, S. 7).

In der KfW sind die EPE (Teile 1 und 2) hausintern für alle Mitarbeitenden verfügbar (Borrmann & Stockmann, 2009b, S. 262–263). Das Leseverhalten des operativen Bereichs weicht nicht merklich von dem des BMZ ab: Von den Interviewten werden vorrangig die EPE der eigenen Projekte gelesen, dabei aber häufig nur das Deckblatt und maximal Teil 1. Wenn die eigenen Projekte evaluiert werden, liest das operative Team (falls noch vorhanden) die EPE aufgrund des Feedbackprozesses (KFW2). KFW6 liest Evaluierungen seiner eigenen Projekte, auch um in der Folgephase etwas aus der Vorgängerphase nachzuprüfen. Thematisch einander naheliegende EPE werden bei der Konzipierung von Neuvorhaben gelesen, was in der KfW keine Verpflichtung ist und laut den Interviewten auch selten gemacht wird. Zusätzlich gibt es Fortbildungstermine der FZ-E, in denen über die aktuellen Ergebnisse berichtet wird. Alle zwei Jahre wird ein Synthesenbericht von der FZ-E veröffentlicht, mit den bedeutsamsten übergreifenden Erkenntnissen und Kurzfassungen von interessanten EPE der letzten zwei Jahre. Die meisten Befragten aus dem operativen Bereich geben an, diesen Synthesenbericht zu lesen (KFW2, KFW3, KFW4, KFW5, KFW6).

Der Zeitfaktor scheint vorrangig dazu beizutragen, dass nicht oder nur wenig gelesen wird: Bei Cracknell (2002, S. 189) und in der Studie von Krämer et al. (2021, S. 42) geben Mitarbeitende von EZ-Organisationen an, keine Zeit zum Lesen von Evaluierungsberichten zu haben – ebenso die für diese Untersuchung Interviewten. Sie haben während ihrer täglichen Arbeit eine große Menge an Lesestoff zu bewältigen. Zudem ist der operative Bereich mit den eigenen Projekten beschäftigt, und es fehlt die Zeit, um sich auch noch mit EPE anderer Projekte zu befassen, auch wenn sie interessant sein mögen (KFW2, KFW6). KFW4 sagt zudem, die EPE interessiere niemanden und werde daher auch nicht gelesen, wenn es keine übergeordneten Schlussfolgerungen oder Folgeprojekte gebe, da die Arbeitsbelastung bereits anderweitig hoch sei. Demnach scheinen auch nur wenige neue Erkenntnisse aus den EPE für die operativen Bereiche ableitbar zu sein.

6.2 Wie nutzt die KfW ihre Ex-post-Evaluierungen?

Nutzen für die Mitarbeitenden

Laut Befragten gibt es für die EPE verschiedene Verwendungen, die die Mitarbeitenden der KfW direkt betreffen. Z. B. spiegelt die EPE die eigene Arbeit für die Projektbeteiligten in der KfW wider (KFW1, KFW2). Sie sorgt dafür, dass die Mitarbeitenden sich kritisch mit der eigenen Arbeit auseinandersetzen (KFW4) und die eigene Leistung einschätzen können (KFW6) – vorausgesetzt, die EPE wird gelesen. Sie hat jedoch nur einen geringen Einfluss auf das Design und die Umsetzung aktueller Projekte (KFW6). Wenn KFW5 die EPE von innovativen oder erfolgreichen Projekten liest bzw. diese vorgestellt bekommt, kann KFW5 daraus Inspiration ziehen. KFW5 hofft, dass sich die Qualität der eigenen Projekte durch die EPE verbessert.

Die EPE dient aufgrund der Personalrotation in der KfW (und im BMZ) auch dem Wissensmanagement (BMZ1). Auch für KFW5 ist dieses relevant, da in der KfW die PM regelmäßig rotieren und damit die Möglichkeit haben, die Arbeit ihrer Vorgänger*innen nachzuvollziehen. Es findet ein Wissenstransfer statt (KFW3), der dem operativen Bereich hilft.²⁵ Direkte interne Auswirkungen von schlechten Noten auf persönliche Karrieren oder das Ansehen einer Abteilung der KfW konnten nicht festgestellt werden. Dafür rotieren die Mitarbeitenden laut den Befragten zu häufig.

25 Weitere Lernerfahrungen des operativen Bereichs werden in Kapitel 6.5 dargestellt.

Glaubwürdigkeit durch Ex-post-Evaluierungen gegenüber dem BMZ

EPE können vertiefte Einblicke und somit Reflexionsprozesse ermöglichen, die die Legitimität der FZ-Arbeit gegenüber ihrem Auftraggeber BMZ beleuchten und stärken können. Es besteht jedoch die Voraussetzung, dass die Bewertung transparent und unabhängig ist. Für die FZ-E ist die Bewertung ein Balanceakt, denn den Stand der Unabhängigkeit muss sie nach außen vertreten und verteidigen. „Den Ruf des Gefälligkeits-Gutachtens, den sollte man auf jeden Fall vermeiden“ (KFW4). KFW6 erklärt, dass die Unabhängigkeit der FZ-E und des Abgeordneten-systems außerhalb der KfW umfangreich diskutiert werde. KFW6 sieht eine geringe externe Anerkennung von guten Noten, weil eine interne Evaluierung teilweise nicht als völlig unabhängig gilt.

Interviewte mit geringeren Kenntnissen der EPE und des Evaluierungssystems der KfW scheinen sich negativer über Unabhängigkeit und Methodik zu äußern (in diesem Fall BMZ1 und IDOS1) als solche, die häufig mit EPE der KfW arbeiten. BMZ1 sagt z. B., dass sich die Durchführungsorganisation angreifbar mache, wenn EPE nicht extern durchgeführt würden, während IDOS1 darstellt, dass nur externe Evaluierende Unabhängigkeit bedeuten. Die Aussagen von folgenden KfW-Mitarbeitenden können dem entgegengestellt werden: KFW3 gibt an, dass es vorkam, dass sie Abgeordnete kannte; trotzdem war die Unabhängigkeit für sie gewährleistet. KFW3 stellt dar, dass interne Evaluierende mehr Fragen stellten und tiefer griffen als z. B. externe Evaluierende. Dem stimmt KFW6 zu. KFW4 ergänzt, dass die FZ-E mindestens so unabhängig wie externe Evaluierungsconsultants sei, denn diese seien auf Folgeaufträge angewiesen. Daher könne es auch bei ihnen Bedenken bezüglich Verzerrungen hin zu einer besseren Note geben (KFW4).

Rechenschaftslegung / Transparenz

Rechenschaftspflicht und Transparenz bedeuten, die Pflicht der KfW gegenüber dem BMZ, Aktionen und Entscheidungen zu rechtfertigen. Theoretisch ist Rechenschaftspflicht mit Sanktionen verbunden und beruht auf Anreizen (Klingebiel, 2013, S. 207). Mit schlechten Bewertungen in der EPE sind jedoch, laut den Interviewten, kaum Sanktionen verbunden.

Um Transparenz zu gewährleisten, muss in der EPE nachvollziehbar vorgegangen werden. Laut DEV1 muss es für eine transparente EPE eine klare Fragestellung geben; die Datenerhebungsmethodik und das Analyseverfahren müssen darin aufgezeigt werden (DEV1). Einige Informationen, wie die in Teil 2 enthaltene Methodik, sind jedoch für die Öffentlichkeit intransparent. Auch der Partner bzw. Projektträger hat nicht den vollständigen Bericht vorliegen, sondern nur Teil 1. Noltze et al. empfehlen, im Sinne der Transparenz die Evaluierungsberichte vollständig zu veröffentlichen (Noltze, Euler & Verspohl, 2018a, S. 47). Datenschutz und Transparenz bilden hier einen Zielkonflikt, der aber ggf. vorgeschoben ist. Transparenz könnte demnach problemlos gewährleistet werden. Sollten politisch hochsensible Informationen in der EPE vorhanden oder der Datenschutz ausdrücklich gefährdet sein, könnte dieser (kleine) Bereich bei der Veröffentlichung auch geschwärzt werden.

Es besteht zudem ein Zielkonflikt zwischen der Rechenschaftslegungs- und der Lernfunktion, wie bereits in Kapitel 3 dargelegt. Im Jahr 2001 schreibt Borrmann, dass EPE und die damit verbundenen Wirkungsuntersuchungen in der deutschen EZ im Wesentlichen als Mittel der Rechenschaftslegung und nicht als Mittel zum Lernen verstanden würden. EPE dienen ihm zufolge selten als Hilfestellung bei der Planung von Neuprojekten in der EZ (Borrmann, 2001, S. 17). Im Jahr 2009 hingegen stellen Borrmann und Stockmann eine gegenläufige Tendenz fest: Evaluierungen würden überwiegend zum Lernen und seltener für Rechenschaftslegung genutzt (Borrmann & Stockmann, 2009a, S. 152). In einer Meta-Evaluierung von Kirsch und Wilson zu Projekten in Afghanistan heißt es dagegen: „Evaluierungsergebnisse dienen vornehmlich der Rechenschaftslegung für geleistete Beiträge und weniger zur Generierung von Erkenntnissen, die das

kritische Lernen aus Erfahrung fördern und die Portfoliosteuerung unterstützen“ (Kirsch & Wilson, 2014, S. viii). Die OECD (2021b, S. 99–100) kommt für die deutsche EZ zu einem ähnlichen Ergebnis. Auch international wird diese Beobachtung gemacht: Yanguas (2021, S. 3–4), Forestieri (2020, S. 72) und Reinertsen et al. (2022, S. 366) stellen dar, dass Evaluierungen in den von ihnen untersuchten EZ-Organisationen primär der Rechenschaftslegung dienen.²⁶

Bei der Untersuchung der EPE der KfW wird zwar von allen Interviewten wie auch auf der KfW-Webseite bestätigt, dass EPE primär dem Lernen dienen. Bei genauerem Hinsehen zeigt sich jedoch, dass der Rechenschaftslegung mehr durch die Zufallsstichprobe gedient ist sowie auch dadurch, dass kaum darin gelesen und – zumindest institutionell – wenig daraus gelernt wird.²⁷ Trotzdem versucht die FZ-E hier einen Spagat, indem interessante Projekte abseits der Zufallsstichproben evaluiert werden. Da in den EPE beide Funktionen angestrebt werden, ist dieser Zielkonflikt zwischen den beiden Funktionen ein zusätzlicher Faktor, warum aus den EPE ggf. nicht gelernt wird.

6.3 Die Frage nach der politischen Wirkung einer hohen Erfolgsquote

Die KfW erhebt seit 2007 aus allen Gesamtnoten der durchgeführten EPE eine Erfolgsquote (KfW, 2014, S. 48). Um zu bewerten, wie sinnvoll die Nutzung von Erfolgsquoten ist, muss unter anderem hinterfragt werden, ob EPE und die darin enthaltenen Noten vergleichbar sind, um eine aggregierte Quote bilden zu können. Denn wie in Kapitel 5 dargelegt, sind EPE teilweise subjektive Bewertungen. Einige Sektoren haben eine schlechtere Durchschnittsnote als andere, z. B. aufgrund des höheren Risikos (KfW6). Daher stellt sich auch hier die Frage nach der Zusammenfassung aller Sektoren zu einer Erfolgsquote. Zusätzlich sollte beachtet werden, dass Erfolgsquoten eine – oftmals klar sichtbare – Dimension zur Bewertung der Vorhaben darstellen, denn Noten lassen sich einfach kommunizieren. Wichtige Lernprozesse, die in den EPE dokumentiert sind, können jedoch primär in einzelnen Vorhaben mit *negativer* Bewertung vorangetrieben werden. Die Sichtbarkeit dieser Lernprozesse ist jedoch durch die komplexere Kommunikation von negativen Bewertungen eingeschränkt.

Die KfW gibt an, dass sich aufgrund der Stichprobengröße eine zuverlässige Aussage über die Erfolgsquote treffen lässt und das Ergebnis aussagekräftig für alle abgeschlossenen Projekte ist (KfW, 2014, S. 48). Sie warnt jedoch vor Fehlinterpretation: Über die tatsächliche Qualität der FZ könne die Erfolgsquote nur begrenzt Auskunft geben, denn auch externe Einflüsse müssten bei Erfolg oder Misserfolg berücksichtigt werden (KfW, 2014, S. 41). Auch Terberger beschreibt, dass Erfolgsquoten mit Vorsicht zu interpretieren sind. Zum einen könnten sie zwischen unterschiedlichen Institutionen nicht verglichen werden, da ggf. unterschiedliche Evaluierungssysteme dahinterstünden. Auch Vergleiche von EPE aus verschiedenen Regionen und Sektoren innerhalb einer Institution haben nur eine begrenzte Aussagekraft (Terberger, 2011, S. 233–234).

KfW6 lehnt eine Vergleichbarkeit der Noten ab. Gerade zwischen den verschiedenen Sektoren sollten sie nicht miteinander verglichen, sondern nur dazu genutzt werden, die Schwachstellen zu finden und konkrete Projekte zu verbessern (KfW6). Diese Einschätzung von KfW6 wurde von anderen interviewten Personen geteilt. Trotz gleicher Fragestellung sowie ähnlicher Datenerhebung und -auswertung kann es gerade für Abgeordnete mit geringerer Erfahrung eine Herausforderung sein, die eigene, subjektive Perspektive herauszuhalten.

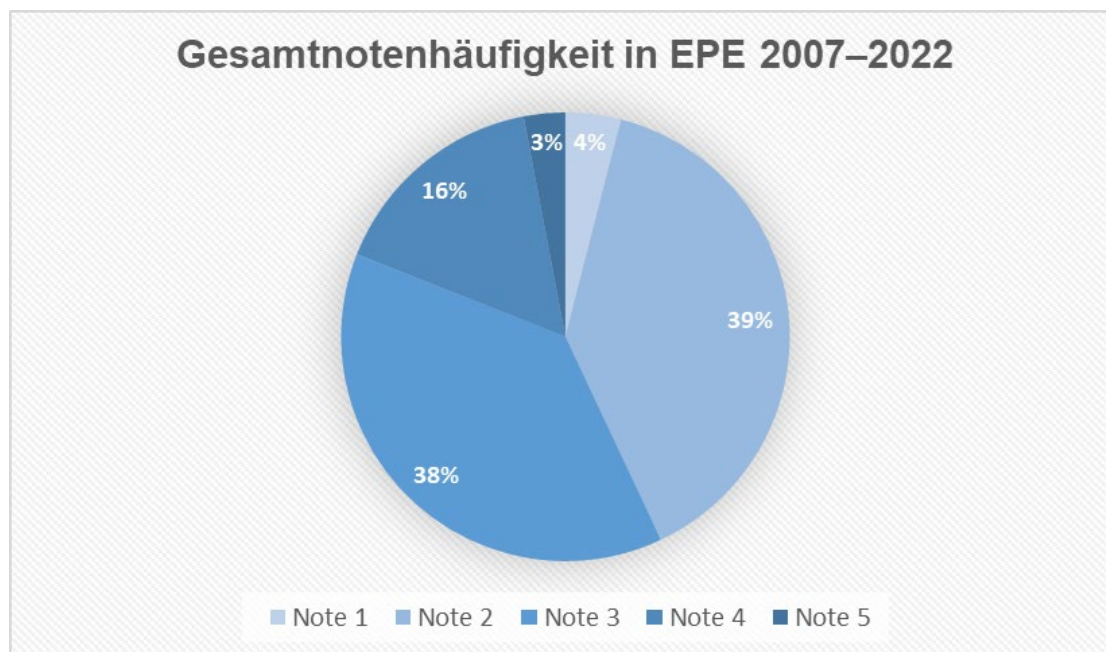
26 Yanguas (2020) beschäftigte sich in seiner Studie mit der britischen EZ und der Weltbank, Forestieri (2020) mit der italienischen EZ und Reinertsen et al. (2022) mit der schwedischen und norwegischen EZ.

27 Dies wird ausführlich in Kapitel 6.5 erläutert.

Bei der KfW-Notenverteilung fällt auf, dass eine Häufung der Noten 2 und 3 vorliegt, jedoch nur eine geringe Anzahl an Einsen existiert (KfW6). DEV1 ergänzt, dass nur wenige Einsen oder Fünfen und niemals eine 6 vergeben werden. DEV1 erklärt, dass die EZ in Kontexten mit hohem Risiko arbeitet. Daher stellt DEV1 die hohe Erfolgsquote infrage. Andererseits sieht DEV1 hier eine Pfadabhängigkeit: Da über einen langen Zeitraum hinweg sehr positiv bewertet wurde, sei es problematisch, die Noten zum jetzigen Zeitpunkt noch herunterzukorrigieren (DEV1).

Wie in Abbildung 1 zu erkennen ist, sammelt sich eine Notenhäufigkeit um die Noten 2 und 3. Ab der Note 4 gilt das Projekt als nicht mehr erfolgreich. Damit wurden 81 % der Projekte seit 2007 als erfolgreich evaluiert, 19 % galten als nicht erfolgreich. Seit der Ersterhebung 2007 schwankt die jährliche Erfolgsquote zwischen 77 % und 87 %, wobei sich die Konfidenzintervalle im Bereich 81 % überlappen (KfW, 2021a, S. 42–45). Hier muss betont werden, dass „erfolgreich“ in der Bewertung der DAC-Kriterien nicht mit langfristiger Wirksamkeit eines Projektes gleichgesetzt werden kann. Um eine solche nachzuweisen, wäre die Durchführung von rigorosen Impact-Evaluierungen notwendig, denn für die DAC-Kriterien „entwicklungspolitische Wirkung“ und „Nachhaltigkeit“ können – wie bereits beschrieben – häufig nur Annahmen getroffen werden.

Abbildung 1: Gesamtnotenhäufigkeit in EPE 2007–2022 in der KfW



Quelle: eigene Darstellung; Daten: KfW, 2022d

Die Mittelwerte der Erfolgsquote haben sich über die Jahre nur geringfügig verändert. Auf der Mikro-Ebene werden Erfolge verzeichnet, auf der Makro-Ebene finden sich diese aber kaum wieder (IDOS1). Diese Erkenntnis lässt sich dem von Mosley (1986) identifizierten Mikro-Makro-Paradoxon zuordnen, nach dem einzelne Entwicklungsprojekte eine hohe Erfolgsquote verzeichnen, auf der Makro-Ebene jedoch kaum Wirkungen der EZ in den Partnerländern sichtbar sind. Dafür wird die zu ungenaue Nachfrage (in Evaluierungen) nach der Wirkung auf der Mikro-Ebene verantwortlich gemacht. Caspari und Barbu (2008, S. 2) bezeichnen das als „Evaluation Gap“, da die Methoden in Evaluierungen nicht ausreichen, um Outcome und Impact von Entwicklungsprojekten präzise nachzuweisen. Terberger gibt zu bedenken, dass eine gute Evaluierung nicht ausschließlich aus der Messung der Wirkung bestehe, sondern auch daraus, die richtigen Fragen zu stellen und voreilige Schlüsse zu vermeiden (Terberger, 2011, S. 237). Bedeutsamer als die Erfolgsquote zu berechnen sei es zu ermitteln, was Erfolg oder Misserfolg in einem Projekt

verursacht (Terberger, 2011, S. 235). Trotzdem soll aber insbesondere die Wirksamkeit von Projekten in der EPE bestimmt werden, denn sie soll am Ende verbessert werden.

DEV1 ordnet die hohe Erfolgsquote wie folgt ein:

Heißt ja im Prinzip, dass 85 % aller FZ-Vorhaben erfolgreich sind. Und das kann man ja schön finden. Man muss sich aber auch immer fragen, jetzt mal im Vergleich: Wenn man jetzt so ein Start-up-Unternehmen im Silicon Valley nimmt: Die haben hervorragende Bedingungen, erfolgreich zu sein [...]. Und die EZ arbeitet in Kontexten, die schlechter eigentlich nicht sein könnten. [...] und ausgerechnet die haben eine viel höhere Erfolgsquote als die Tech-Start-ups im Silicon Valley. Also da... da ist im Prinzip irgendwo ein Verhau drin [...]. Weil wenn die [Erfolgsquote] jetzt runtergehen sollte, dann fragt natürlich jeder nach: ‚Warum werdet ihr denn da schlechter?‘ Aber ich glaube, die Nivellierung, irgendwann müsste man da schon nochmal drüber nachdenken, wie glaubwürdig ist das eigentlich, dass da 85 % der Vorhaben als erfolgreich gelten. Wobei natürlich auch bei einer 3, das muss man auch sagen, auch nicht alles gut ist. [...] Es gibt ja auch weniger Einser. Aber man muss sich auch fragen, warum gibt es denn niemals eine 6 und ganz wenige Fünfer. Also irgendwie stimmt was mit der Skala nicht. (DEV1)

Die Erfolgsquote prägt den Blick der Öffentlichkeit auf die EZ und im Falle der KfW auf die FZ (IDOS1). Jedoch sollte sie kritisch betrachtet werden, denn sie hat weit geringere Aussagekraft als auf den ersten Blick erscheinen mag. Über entwicklungspolitische Wirksamkeit sagt die Erfolgsquote nichts aus. Trotz teilweise mangelnder Vergleichbarkeit wird sie gezogen, auch wenn sie laut den für diese Forschung interviewten Personen ein etwas zu positives Bild vermittelt und eine Vergleichbarkeit zwischen Ländern und Sektoren suggeriert, die nicht miteinander vergleichbar sind. Dieses positive Bild hängt mit einer gewissen Pfadabhängigkeit zusammen, denn nach 15 Jahren konsequenter Vergabe guter Noten wäre eine Anpassung nicht einfach vor dem BMZ oder der Öffentlichkeit zu rechtfertigen.

6.4 Politische Steuerungsrelevanz: Nutzt das BMZ die Ex-post-Evaluierung?

Die Dexis Consulting Group empfiehlt, dass sich die Entwicklungszusammenarbeit einer mehrheitlich evidenzbasierten Entscheidungspolitik öffnen solle. Doch es gibt Schwierigkeiten, wirksame Strategien zu entwickeln, die die Integration von Evidenz in der Politik sicherstellen (Dexis Consulting Group, 2020, S. 39). Auch die Untersuchung von Krämer et al. zeigt, dass Evidenz häufig nicht zur strategischen Entscheidungsfindung im BMZ genutzt wird (Krämer et al., 2021, S. vi). Sie machen dafür die Unterteilung der deutschen EZ in politische Steuerung und Durchführungsorganisationen verantwortlich. Sie verhindert, dass die Evidenz von der Projektebene in die strategische Entscheidungsfindung einfließt (Krämer et al., 2021, S. 42). Der Bundesrechnungshof kritisiert, dass es durch die Berichterstattung auf Projektebene kaum möglich ist, Relevanz und Wirkung auf der Programmebene zu bestimmen. Eine übergreifende Analyse von Ergebnissen einzelner Projektevaluierungen wird im BMZ selten vorgenommen (BRH, 2021, S. 7). Daher hat das BMZ nur wenige Möglichkeiten, eine evidenzbasierte Politiksteuerung auf der Programmebene zu betreiben (BRH, 2021, S. 24).

Das BMZ hat laut eigener Aussage das Ziel, die politisch-strategische Steuerung mit aggregierten Ergebnissen von Projektevaluierungen zu verbessern (BMZ, 2021a, S. 40). Laut den für diese Studie geführten Interviews sind Projektevaluierungen zur Ausgestaltung des Programms oder Länderportfolios für das BMZ zu kleinteilig. Doch auch aggregierte Evidenz scheint nur in geringem Maße in politische Entscheidungen einzufließen und besitzt womöglich eine geringe Priorität im BMZ. Evaluierungen werden dort demnach als Verwaltungsaufgaben und formale Rechenschaftslegung angesehen, aber nicht als Lerninstrument (Krämer et al., 2021, S. 41; vgl.

dazu OECD, 2021b). Auch die OECD (2001, S. 72) kommt zu dem Schluss, dass die Erkenntnisse aus Evaluierungen nur eine kleine Komponente sind, die in politische Entscheidungen einfließt. Trotz noch so systematischer Lern- und Feedbackschleifen hat eine Reihe anderer Faktoren – wie (politische) Machtkämpfe, ungünstige politische Situationen im Partnerland, Ausgabendruck oder das Eigeninteresse von Entscheidungstragenden – einen größeren Einfluss auf die politische Steuerung von Programmen in der EZ (OECD, 2001, S. 72).

Mangelnde zeitliche Abstimmungen zwischen Projektzyklen und Evaluierungen sind ebenfalls ein Hindernis für die Übernahme von Ergebnissen im BMZ (Krämer et al., 2021, S. 41), da diese – wie die der EPE – erst (lange) nach Abschluss des Projektes vorliegen. BMZ1 bestätigt, dass die Ergebnisse für das BMZ häufig zum falschen Zeitpunkt kommen, wenn keine aktuellen Verhandlungen mit dem betroffenen Partner bevorstehen. Daher würden die EPE abgelegt bzw. nur überflogen. Sie sollen BMZ1 zufolge (aufgrund der zeitverzögerten Durchführung) den Eindruck bestätigen, der bereits bei Abschluss des Projektes vorlag und den auch die sonstige Berichterstattung über das Projekt wiedergibt. Daher erwartet BMZ1 auch keine entscheidenden neuen Erkenntnisse aus der EPE.

Was das Verfahren anbelangt, werden die Evaluierungsberichte (wie auch alle anderen) im Länderreferat „zur Kenntnis genommen“. Diese Berichte können in Summe Konsequenzen für die weitere Ausgestaltung des Portfolios eines Landes herbeiführen, z. B. ob es in diesem Bereich zu weiteren Engagements kommt bzw. ob gewisse Dinge angepasst werden müssen (BMZ1). Die EPE sei zwar bedeutsam, wenn die nächsten Verhandlungsgespräche zwischen BMZ und dem Partner geführt werden, aber BMZ1 merkt an, dass sie eigentlich nicht an einer Rückschau interessiert sei, sondern daran, was zukünftig besser gemacht werden könne.

Es zeigt sich, dass verschiedene Evaluierungsergebnisse auf verschiedenen Ebenen relevant sind, z. B. für politische Steuerung oder für die technische Verbesserung der Durchführung. Die EPE scheint die politische Steuerung nur in geringem Maße zu beeinflussen. Bemerkenswert ist, dass BMZ1 angibt, die EPE sei primär relevant für die Durchführungsorganisationen, wohingegen einige operative Befragte aus der KfW angeben, dass sie die EPE hauptsächlich für das BMZ relevant halten.

6.5 Wer lernt nun aus der Ex-post-Evaluierung?

Die KfW betont insbesondere die Lernfunktion der EPE, wie in Kapitel 3 dargelegt. Yanguas unterscheidet zwischen operativem und strategischem Lernen. Bei Ersterem geht es um den Prozess der Projektumsetzung. Strategisches Lernen ermöglicht es EZ-Organisationen, ihre Prozesse, Strukturen und Praktiken neu auszurichten, um Organisationsziele effektiver zu erreichen (Yanguas, 2021, S. 4–5). Die beiden Lernarten bedingen sich nicht zwingend gegenseitig. Nur weil auf der Mikro-Ebene operativ gelernt wird, kommt es auf der Makro-Ebene selten zu einer strategischen Änderung der Politik (Yanguas, 2021, S. 5). Strategisches Lernen kann mit institutionellem Lernen gleichgesetzt werden, wohingegen operatives Lernen im Folgenden als „Lernen des operativen Bereichs“ beschrieben wird. An dieser Stelle soll zudem die Ebene des individuellen bzw. persönlichen Lernens der Mitarbeitenden der KfW durch das Abgeordnetensystem unterschieden werden.

Institutionelles Lernen der KfW

Die KfW gibt an, dass institutionell gelernte Lektionen für zukünftige Vorhaben abgeleitet werden sollen (KfW, 2021a, S. 1). In der Zufallsstichprobe werden alle Projekte evaluiert, auch wenn teilweise solche gezogen werden, bei denen sich keine Lernerfahrungen mehr einstellen, da ähnliche Projektkonzepte bereits evaluiert wurden oder diese Art von Projekten nicht mehr durchgeführt wird (KfW4).

Institutionell gelernt wird laut KfW (2021a, S. 1), indem die Ergebnisse der Evaluierung zielgruppengerecht und effizient nutzbar aufbereitet werden. Auch wenn Projekte keine Folgephasen haben, sollte die Transfermöglichkeit geschaffen werden, gelernte Dinge in ähnliche Kontexte zu übertragen (DEV1). Querschnittsauswertungen aus den Evaluierungen werden – geschichtet nach Sektor oder Land – von der FZ-E erarbeitet, die teilweise von Consultants oder Universitäten unterstützt werden. KfW3 erklärt, dass es einen teilweisen Wissenstransfer gebe: So würden die Erfahrungen aus Projekten in Asien jetzt in Afrika angewandt. Obgleich manche Befragte Beispiele von institutionellen Lernerfahrungen aus EPE nennen, schildern sie dennoch, dass diese selten eintreten, auch aufgrund der Problematik, sie in andere Kontexte zu übertragen.

Lernen des operativen Bereichs

Die Interviews in der Studie von Krämer et al. zum Thema rigorose Impact-Evaluierungen deuten darauf hin, dass Erkenntnisse daraus kaum zum operativen Lernen genutzt werden, obwohl dies der häufigste angegebene Grund zur Nutzung rigoroser Erkenntnisse war (Krämer et al., 2021, S. 35–36). Eine ähnliche Beobachtung für EPE wurde auch in der Untersuchung für dieses Paper gemacht: Aus den hierfür geführten Interviews lässt sich ableiten, dass die Bedeutung der EPE für die operative Arbeit eher gering ist. Die meisten Aspekte sind nicht neu für die beteiligten PM und TSV (KfW4). Sie kennen die Empfehlungen und Kritikpunkte bereits (KfW6) und ziehen daher kaum eine Lernerfahrung aus der EPE. Stockmann bestätigt das Argument, dass Evaluierungsergebnisse für den operativen Bereich nichts Neues beinhalten (Stockmann, 2007, S. 180).

Yanguas (2021, S. 15) beantwortet die Frage, warum operativ nur geringfügig gelernt wird, mit vier Argumenten: Erstens lassen sich verwendete Instrumente und Methoden nicht ohne Weiteres auf andere Kontexte übertragen. Dieses Argument wird auch von den Interviewten genannt: KfW1 vermutet, dass in anderen Evaluierungen Punkte enthalten sein könnten, die im eigenen Projekt angewandt werden können. KfW1 denkt jedoch, dass es Komplikationen mit sich bringt, Erkenntnisse in andere Kontexte zu übertragen. Es gibt zwar Fehler, die sich wiederholen, sie wirken sich jedoch in unterschiedlichen Kontexten unterschiedlich aus. Nicht alle Fehler können vermieden werden, da mit einem Partner zusammengearbeitet wird, der eine große Autonomie hat bzw. haben sollte. Auch die KfW gibt in ihrem zweijährigen Synthesenbericht an, dass sich die Aussagen und Empfehlungen von Evaluierungen nicht ohne Weiteres auf andere Projekte übertragen lassen (KfW, 2019, S. 61).

Zweitens haben die Durchführungsorganisationen einen Anreiz, Wissen zu horten, da sie miteinander um Aufträge und Zuschüsse konkurrieren. In der deutschen EZ trifft diese Argument eher nicht zu: Technische Zusammenarbeit und FZ unterscheiden sich in der Art von Projekttypen und in dem dafür erforderlichen Fachwissen, daher konkurriert die KfW nicht um Mittel. Bei ihr kann sich das Horten von Wissen auf den Mangel an Transparenz in der EPE zurückführen lassen. Drittens besteht der Anreiz, Komplikationen in Projekten vor Ort (und damit deren Lösung) nicht zu veröffentlichen, um die breite Öffentlichkeit nicht darauf aufmerksam zu machen. Als letztes Argument nennt Yanguas, dass die gesamte Berichterstattung – und somit auch Evaluierungen – nicht systematisch oder auf operationell relevante Weise in die operativen Bereiche zurückgespielt wird (Yanguas, 2021, S. 15). Dies geschieht bei der KfW zwar, jedoch werden die Berichte selten gelesen. Ramalingam (2005, S. 30) ergänzt, dass das Lernen im Wettbewerb mit anderen Aufgaben der Mitarbeitenden steht. Dieses Problem spiegelt sich auch in den Interviews mit dem operativen Bereich der KfW wider.

Hat ein evaluiertes Projekt Folgephasen, so kann von der EPE der früheren Phasen gelernt werden (was z. B. Indikatoren und Zielsetzung angeht). Bei KfW1 und KfW2 z. B. hatte die EPE Einfluss auf die Indikatoren und die Wirkungsmatrix ihres aktuellen Folgeprojektes (KfW1, KfW2). Auch KfW4 bestätigt, dass eine Evaluierung zu Kurskorrekturen im noch laufenden Projekt führen kann. Wenn es Folgeprojekte bzw. eine weitere Zusammenarbeit im Sektor gibt,

sind Partner wie PM interessierter an der EPE (KFW4). Läuft die Zusammenarbeit aus, so bleibt das Interesse aus und die Lernerfahrung ist gering (KFW4). Wenn es keine Folgeprojekte gibt, ist die EPE laut KFW2 uninteressant.

Wenn PM der KfW neue Projekte planen, können sie seit kurzem auf eine Software namens „Quick Evaluation Results“ zugreifen, in der alle EPE abgelegt und nach Ländern, Sektoren und Lessons Learned sortiert sind. Bei der Neukonzipierung von Projekten besteht jedoch keine Pflicht, alte EPE zurate zu ziehen. Das stellt auch keine gängige Praxis der meisten Interviewten aus dem operativen Bereich dar: KFW2 und KFW6 geben an, dass ältere EPE oft nur geringen Einfluss auf das Design oder die Umsetzung aktueller Projekte hätten, weil sie daraus keine neuen Erkenntnisse mehr ziehen können. Einzig KFW5 gibt an, bei der Projektvorbereitung aktiv nach früheren EPE über Projektträger, Sektor und Land zu suchen und daraus Informationen zu ziehen.

Um Lernerfahrungen innerhalb der KfW zu verbreiten, nehmen FZ-E-Mitarbeitende an Sektorseminaren des operativen Bereichs teil und stellen dort die aktuellen Evaluierungsergebnisse vor (KFW4, KFW5), z. B. in Bezug auf Gestaltung, Planung oder Design zukünftiger Projekte (KFW3). Für das BMZ gibt es solch ein Format von KfW-Seite nicht, genauso wenig wie für die Partner bzw. Projektträger. Auch wird aus dem alle zwei Jahre erscheinenden Synthesenbericht gelernt. Die FZ-E disaggregiert darin die Ergebnisse der EPE nach Ländern und Sektoren und veröffentlicht sie (KFW2).

Lernen der Abgeordneten

Die Analyse der Interviewdaten zeigt, dass aus den EPE primär über das Abgeordnetensystem gelernt wird. Dessen Idee ist es, die Erfahrungen aus der Durchführung einer EPE mit in die tägliche operative Arbeit der/des Abgeordneten zu nehmen und sie/ihn für das Thema „Wirkung, Wirkungsmessung, Wirkungsnachweis“ zu sensibilisieren (KfW, 2021a, S. 3). Die für dieses Paper interviewten Personen bestätigen diesen Lerneffekt bei der Abgeordnetentätigkeit. Die Abgeordneten vergrößern ihre Sektor- und Länderkenntnisse, und die Durchführung der EPE dient der internen Weiterbildung (KFW4). Gerade z. B. das Dazulernen bei der Wirkungsmatrix ist bereichernd für die tägliche Arbeit der PM, da sie dieses Wissen bei der Neukonzipierung von Projekten verwenden können. Auch in den Interviews von Borrmann und Stockmann mit der KfW wurde der Lernaspekt der Abgeordneten als besonders interessant hervorgehoben, was dem großen Interesse der Mitarbeitenden entspricht, die an einer Abordnung teilnehmen (Borrmann & Stockmann, 2009b, S. 261). Der Bundesrechnungshof würdigt das Abgeordnetensystem in Bezug auf die Lernerfahrungen der Abgeordneten ebenfalls als sinnvoll (BRH, 2022, S. 10).

„Learning by doing“ ist der effektivste Weg, um zu lernen (OECD, 2001, S. 27). Daher ist es sinnvoll, die wichtigsten Interessengruppen in den Evaluierungsprozess einzubinden, denn sie haben den besten Zugang zu spezifischem Lernen, können diese Lehren verinnerlichen und in die nächste Planungsphase miteinfließen lassen (OECD, 2001, S. 27). Nicht nur der Bericht dient dem Lernen, sondern auch der Prozess der Erstellung einer Evaluierung führt dazu, dass die beteiligten Akteure lernen (Borrmann & Stockmann, 2009a, S. 133). Nichtsdestotrotz muss festgehalten werden, dass die Abgeordneten das größte Potenzial haben, die Ergebnisse der EPE durch ihre methodischen Defizite zu verzerren (siehe dazu BRH, 2022, S. 11). Auch hier entsteht ein Zielkonflikt zwischen den einzigen Akteuren, die aus der EPE lernen – den Abgeordneten –, und der Glaubwürdigkeit bzw. Unabhängigkeit der Bewertung.

Zusammenfassend gibt es verschiedene Gründe, warum aus den EPE nicht oder nur geringfügig gelernt wird: Bei Yanguas (2021) wird beschrieben, dass das erarbeitete Wissen nicht oder nur selten genutzt wird. Aus den gesammelten Informationen der EPE kann gelernt werden, jedoch ist häufig der Zyklus unterbrochen und endet mit der Evaluierung, weil der Bericht nicht

gelesen wird. Deswegen wird eben nicht aus den Evaluierungen gelernt (IDOS1). In Entwicklungsorganisationen werden keine Anreize gesetzt, um zu lernen: “Part of the challenge lies in convincing staff that knowledge sharing means smarter work rather than more work” (King & McGrath, 2003, S. 15).

Der hauptsächliche Grund, warum die EPE zu geringen Lerneffekten bei den Projekten führt, ist die große Zeitverzögerung. KfW4 führt an, dass das Handwerkszeug des operativen Bereichs zum Zeitpunkt der EPE deshalb bereits vorangeschritten ist: „Das ist zumindest die Rückmeldung, die häufig in meine FZ-E-Zeit kam: So eine Krücke würden wir heutzutage nie mehr anpacken, nach dem Motto: ‚Da haben wir selbst schon genügend Lehrgeld gezahlt‘“ (KfW4). Die KfW ist sich bewusst, dass mit dem zeitlichen Verzug Hindernisse einhergehen: Für die Ausgestaltung zukünftiger Projekte können keine kurzfristigen Erkenntnisse abgeleitet werden (KfW, 2021c, S. 4). Wird jedoch nicht abgewartet, so kann – wie erwähnt – die Frage nach Nachhaltigkeit und entwicklungspolitischer Wirkung nicht beantwortet werden.

Schlussendlich legen Interviews und Literatur nahe, dass operativ wie institutionell wenig aus den EPE gelernt wird, insbesondere nicht durch das Lesen des Berichts. Nur die Abgeordneten scheinen aus den von ihnen selbst durchgeführten EPE zu lernen. Ggf. führen EPE dazu, dass Folgephasen oder Folgeprojekte verbessert werden, wenn es welche gibt. Es ist nicht sinnvoll, z. B. das Problem des zeitverzögerten Erscheinens der Ergebnisse anzugehen, da hier der Zielkonflikt auftritt, dass Nachhaltigkeit und Wirkung eines Projektes nicht ausreichend bewertet werden können. Es scheint daher sinnvoller, EPE weiterhin zur Rechenschaftslegung und als Rohmaterial für weitere Forschungen und Meta-Evaluierungen zu nutzen.

7 Wie Ex-post-Evaluierungen in Zukunft genutzt werden können

Insgesamt zeigt die Analyse, dass die EPE trotz aller Herausforderungen ein wichtiges Instrument für die Rechenschaftslegung ist und zur Legitimität der FZ beitragen kann. Dabei ist hervorzuheben, dass die EPE in der KfW sorgfältig betrieben zu werden scheint und von den interviewten Mitarbeitenden ernstgenommen wird. Zur Einschätzung, dass die EPE *grundsätzlich wichtig* ist, kommen alle Interviewten aus der Untersuchung für dieses Paper.

Jedoch gehen Herausforderungen mit der Methodik der EPE einher. Sie hängen überwiegend von der evaluierenden Person und ihrem Umgang damit ab. Haben evaluierende Personen geringere Erfahrung mit der Anwendung von Evaluierungstechniken oder fehlt Vorwissen zu Sektoren und Ländern, so kann dies Evaluierungsergebnisse verzerren. Die Subjektivität der Evaluierenden stellt eine weitere Dimension dar: Abgeordnete können einzelne Projekte unabhängig bewerten, sind als KfW-Mitarbeitende jedoch Teil des Gesamtsystems der Projektkonzipierung und Umsetzung. Dieses System könnte verlässlicher und mit größerer Distanz von externen Evaluierenden hinterfragt werden. Aus diesen Gründen kann die Vergleichbarkeit der einzelnen EPE eine Herausforderung darstellen.

Einen Zielkonflikt bilden Transparenz und Datenschutz: Aufgrund des hohen Stellenwerts, den die KfW dem Datenschutz in der EPE zumisst, schränkt sie deren Transparenz und damit ihre Glaubwürdigkeit gegenüber der Öffentlichkeit ein. Der Datenschutz gewährleistet andererseits Sicherheit für den Partner, der ggf. durch in der EPE enthaltene Informationen angreifbar würde. Nichtsdestotrotz sollte mit Steuergeldern transparent umgegangen werden, daher ist eine transparente Bewertung von Projekten, die die EPE leisten könnte, empfehlenswert.

Auch wenn diese Studie nicht repräsentativ ist, weisen die Aussagen der Interviewten darauf hin, dass die EPE im operativen Bereich der KfW nur selten gelesen wird, weshalb sie dort auch nur einen geringen Nutzen haben kann. Denn das Lesen sollte eine Grundvoraussetzung für

alle darauffolgenden Lernprozesse sein. Alleiniges Lesen führt jedoch nicht zwingend zum Lernen, wenn die daraus gezogenen Erkenntnisse nicht angewandt werden. Würden Inhalte der EPE von den operativen Bereichen als für die Arbeit relevant angesehen, so würden die Berichte vermutlich auch ausführlicher gelesen. In der Tat gelten EPE im operativen Bereich laut den Befragten als größtenteils irrelevant für die eigene Arbeit.

Rund 19 % der evaluierten Projekte erhalten die Gesamtnote 4 oder schlechter, wodurch sie als „nicht erfolgreich“ gelten. Soweit hier festgestellt werden konnte, haben diese „nicht erfolgreichen“ Projekte geringe Auswirkungen innerhalb der KfW in Bezug auf persönliche Karrieren, das Ansehen einer Abteilung der KfW, die Kürzung der Mittel oder die Beendigung der Zusammenarbeit mit einem Partner. Wenn keine Konsequenzen aus schlechten Evaluierungsergebnissen gezogen werden, gibt es auch keinen Anreiz zu ihrer positiven Manipulation. Das dient der Rechenschaftsfunktion der EPE. Andererseits gibt es damit jedoch nur einen geringen Anreiz, Kritikpunkte anzugehen; die Lernfunktion wird somit kaum adressiert.

Die beiden wichtigsten Funktionen der EPE sind die Rechenschaftslegung und das Lernen, die, wie oben gezeigt, ebenfalls einen Zielkonflikt darstellen. Beide können nicht gleichzeitig mit derselben Methodik abgebildet werden: Eine Stichprobenziehung z. B. garantiert die Repräsentativität der Ergebnisse und trägt somit zur Rechenschaftslegung bei. Damit werden jedoch häufig Projekttypen evaluiert, die bereits öfter bewertet wurden oder deren Ansätze nicht innovativ sind. Lernerfahrungen stellen sich laut den Befragten daher seltener ein. Auch das Abgeordnetensystem der KfW hängt damit zusammen, da für die Lernfunktion möglichst interne Evaluierende eingesetzt werden sollten, um die gelernten Inhalte besser in die Praxis umsetzen zu können. Externe und damit unabhängige Evaluierende eignen sich wiederum besser für die Funktion der Rechenschaftslegung (vgl. hierzu auch Reinertsen et al., 2022, S. 364).

Die EPE regt, laut den Befragten, aufgrund der Zeitverzögerung nicht zum Lernen im operativen Bereich an. Ohne die Zeitverzögerung könnten Nachhaltigkeit und entwicklungspolitische Wirkung jedoch nicht ausreichend bewertet werden. Der größte Lerneffekt scheint bei den Abgeordneten während der Durchführung der EPE aufzutreten. Sie können aber Verzerrungen in die EPE bringen. Daher empfiehlt der BRH zusätzliche Schulungen für Abgeordnete (BRH, 2022, S. 11). Die Beauftragung externer Evaluierungsconsultants scheint keine bessere Lösung darzustellen, da auch sie einen Anreiz haben können, Noten ins Positive zu verzerren.

In der KfW und im BMZ kann nicht aus EPE gelernt werden, wenn diese nicht gelesen bzw. ihre Ergebnisse nicht anderweitig (z. B. akkumuliert) registriert werden. In den letzten Jahren versucht die FZ-E, diese Problematik innerhalb der KfW mit innovativen Formaten auszugleichen. Jedoch wird die Verbreitung von Evaluierungsergebnissen noch zu selten als weitere wichtige Evaluierungsphase wahrgenommen. Gibt es Folgephasen oder Folgeprojekte, so profitieren die Befragten aus dem operativen Bereich der KfW am ehesten von der EPE, da deren Kritikpunkte sofort auf das Projekt angewandt werden können. Insbesondere wenn Querschnittsauswertungen aus den EPE erstellt werden, können diese zu projektübergreifenden Lernerfahrungen führen.

Diese Analyse deutet darauf hin, dass EPE nur ein eingeschränktes Potenzial zur Verbesserung zukünftiger FZ-Projekte haben. Stattdessen beschränkt sich ihr Nutzen hauptsächlich auf die Rechenschaftslegung. Auch beim BMZ sind der Nutzen und die Nutzung der EPE offenbar recht begrenzt. Trotzdem scheint ihre vollständige Abschaffung keine Lösung darzustellen, denn sie sind das einzige Instrument, welches in der KfW für eine repräsentative Projektzahl relativ kostengünstig Nachhaltigkeit und entwicklungspolitische Wirkung bewertet und damit die Basis für die Rechenschaftsfunktion der FZ bilden kann. EPE erfüllen nur einen geringen Anteil des ursprünglich vom BMZ angedachten Zwecks. Folglich stellt sich die Frage, ob die Zwecke, die eine EPE erreichen soll, überdacht werden sollten.

Da einige Zielkonflikte vermutlich nicht lösbar sind, bleibt die Problematik der Umsetzung vieler Ziele voraussichtlich bestehen. Umsetzbares Verbesserungspotenzial besteht jedoch bei der

Steigerung des Lerneffekts aus der EPE: Die Forschende schließt sich hier dem BRH und auch der OECD an, die (der KfW) empfehlen, häufiger projektübergreifende Schlussfolgerungen aus den EPE in Querschnittsauswertungen zusammenzustellen (BRH, 2022, S. 6; OECD, 2001, S. 27–28), um institutionelles Lernen bei der KfW zu verbessern. Primär diese Auswertungen sollten im alle zwei Jahre erscheinenden Synthesenbericht thematisiert werden. Der operative Bereich lernt laut Befragten auf anderem Wege (z. B. aus besagtem Synthesenbericht oder von Informationsveranstaltungen) und nicht aus der einzelnen EPE.

Um ein lernendes System weiter auszubauen, wäre es zudem sinnvoll, erstens den Methodenteil jeder EPE öffentlich zur Verfügung zu stellen – wenn nicht sogar den vollständigen Teil 2, da damit auch andere Forschungsinstitute oder Organisationen unbürokratisch Zugriff darauf hätten und unabhängig Meta-Evaluierungen bzw. Querschnittsauswertungen erstellen könnten. Zweitens scheint es sinnvoll, auch für die Länder- und Sektorreferate im BMZ Querschnittsauswertungen aus EPE zu erstellen, da es bisher auf dieser Ebene keine standardisierten Verfahren zur Nutzung der EPE-Ergebnisse gibt (BRH, 2021, S. 20). Dazu schlagen z. B. Krämer et al. zusätzliche Plattformen für den Austausch zwischen politischen Entscheidungstragenden, Entwicklungspraktiker*innen sowie Forschenden und Evaluierenden vor, um sicherzustellen, dass sich die Schaffung von Evidenz am tatsächlichen Bedarf orientiert (Krämer et al., 2021, S. 59). Ein zusätzlicher Vorschlag ist, auch die Partner bzw. Träger der KfW in diesen Austausch miteinzubeziehen, denn so könnten auch diese aus den EPE oder anderen Evaluierungsformaten lernen. Darüber hinaus soll eine Unterstützungsstruktur für Praktiker*innen und Entscheidungstragende während des gesamten Zyklus von Evidenzgenerierung und -übernahme aufgebaut werden, die teilweise dezentral bei den Durchführungsorganisationen, teils zentral angesiedelt wäre, um gleichzeitig nahe am praktischen Bereich arbeiten wie auch organisationsübergreifendes Lernen und Beratung anbieten zu können (Krämer et al., 2021, S. 60). Diese Unterstützungsstruktur könnte auch dem Wunsch von BMZ1 entsprechen, dass Präsentationen der Ergebnisse von EPE (oder von übergeordneten Schlussfolgerungen) stattfinden sollten, um die Ergebnisse in den Länder- und Sektorreferaten im BMZ besser einordnen zu können.

Literaturverzeichnis

- Amine, M., & Eulenburg, A. G. zu (2022). *Wirkungsorientierte bilaterale Entwicklungszusammenarbeit? Der Dreiklang aus Länderstrategie, EZ-Programm und Modul* (DEval Policy Brief 7). Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval).
- Armytage, L. (2011). Evaluating aid: An adolescent domain of practice. *Evaluation*, 17(3), 261–276. <https://doi.org/10.1177/1356389011410518>
- Bartl, W., Papilloud, C., & Terracher-Lipinki, A. (2019). Governing by numbers – key indicators and the politics of expectations: An introduction. *Historical Social Research*, 44(2), 7–43. <https://doi.org/10.12759/HSR.44.2019.2.7-43>
- Beebe, J. (1995). Basic Concepts and Techniques of Rapid Appraisal. *Human Organization*, 54(1), 42–51. <https://doi.org/10.17730/humo.54.1.k84tv883mr2756l3>
- BMZ (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung). (2014). *BMZ-Stellungnahme zum DEval-Bericht: „Ein Review der Evaluierungsarbeit zur deutschen Entwicklungszusammenarbeit in Afghanistan“*.
- BMZ. (2020). *Evaluierungskriterien für die deutsche bilaterale Entwicklungszusammenarbeit. BMZ-Orientierungslinie zum Umgang mit den OECD-DAC-Evaluierungskriterien in Evaluierungen der deutschen bilateralen Entwicklungszusammenarbeit*.
- BMZ. (2021a). *Evaluierung der Entwicklungszusammenarbeit. Leitlinien des BMZ (BMZ-Strategien 4)*.
- BMZ. (2021b). *Glossar der Schlüsselbegriffe im Bereich Evaluierung der Entwicklungszusammenarbeit*.
- BMZ. (2021c). *Leitlinien für die bilaterale Finanzielle und Technische Zusammenarbeit mit Kooperationspartnern der deutschen Entwicklungszusammenarbeit*. BMZ.
- BMZ. (2022). *Deutsche ODA*. Abgerufen von <https://www.bmz.de/de/ministerium/zahlen-fakten/oda-zahlen/deutsche-oda-leistungen-19220>
- Borrmann, A. (2001). *Reform der Erfolgskontrolle in der deutschen Entwicklungszusammenarbeit. Eine Zwischenbilanz: Studie im Auftrag des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung* (HWWA-Studien, Bd. 63, 1. Aufl.). Baden-Baden: Nomos-Verl.-Ges. Abgerufen von <http://www.hwwa.de>
- Borrmann, A., & Stockmann, R. (2009a). *Evaluation in der deutschen Entwicklungszusammenarbeit. Band 1 – Systemanalyse* (Sozialwissenschaftliche Evaluationsforschung). Münster: Waxmann.
- Borrmann, A., & Stockmann, R. (2009b). *Evaluation in der deutschen Entwicklungszusammenarbeit. Band 2 – Fallstudien* (Sozialwissenschaftliche Evaluationsforschung, 8,1). Münster: Waxmann. Abgerufen von <http://www.socialnet.de/rezensionen/isbn.php?isbn=978-3-8309-2125-7>
- BRH (Bundesrechnungshof). (2021). *Abschließende Mitteilung an das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung über die Prüfung „Evaluierung von Maßnahmen der Entwicklungszusammenarbeit. Teil I – Steuerung im BMZ“*.
- BRH. (2022). *Abschließende Mitteilung an das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung über die Prüfung „Evaluierung von Maßnahmen der Entwicklungszusammenarbeit. Teil III – Maßnahmen der bilateralen Finanziellen Zusammenarbeit“*.
- Caspari, A., & Barbu, R. (2008). *Wirkungsevaluierungen: Zum Stand der internationalen Diskussion und dessen Relevanz für die Evaluierung der deutschen Entwicklungszusammenarbeit* (Evaluation Working Papers). Bonn: BMZ Evaluation Division.
- Chambers, R. (1981). Rapid rural appraisal: rationale and repertoire. *Public Administration and Development*, 1, 95–106.
- Cracknell, B. E. (2002). *Evaluating development aid. Issues, problems and solutions* (3rd printing). London: Sage.

- Czarniawska, B. (2007). *Shadowing. And other techniques for doing fieldwork in modern societies*. Koege: Copenhagen Business School Press – Universitetsforlage.
- Desrosières, A. (2015). Retroaction: How indicators feed back. In R. Rottenburg, S. E. Merry, S. J. Park & J. Mugler (Eds.), *The world of indicators. The making of governmental knowledge through quantification* (Cambridge studies in law and society, S. 329–353). Cambridge: Cambridge University Press.
- Dexis Consulting Group. (2020). *Evidence base for collaborating, learning and adapting. Summary of the literature review*. USAID.
- Faust, J. (2020). Rigorose Wirkungsevaluierung: Genese, Debatte und Nutzung in der Entwicklungszusammenarbeit. *der moderne staat: Zeitschrift für Public Policy, Recht und Management*, 13(1), 61–80. <https://doi.org/10.3224/dms.v13i1.08>
- Ferguson, J., Mchombu, K., & Cummings, S. (2008). *Management of knowledge for development: Meta-review and scoping study* (IKM Working Paper, 1).
- Forestieri, M. (2020). Equity Implications Evaluating Development Aid: The Italian Case. *Journal of Multi-Disciplinary Evaluation*, 16(34), 65–90.
- GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit). (2012). Wirkungsanalyse und Wirkungsmessung in Gesundheitsvorhaben der deutschen Entwicklungszusammenarbeit. *Zukunftsentwickler. Wir machen Zukunft. Machen Sie mit*.
- Glaser, B. G., & Strauss, A. L. (2010). *Grounded theory. Strategien qualitativer Forschung* (Programmbe- reich Gesundheit, 3. Aufl.). Bern: Huber.
- Goodwin, M. (2017). The poverty of numbers: reflections on the legitimacy of global development indicators. *International Journal of Law and Context*, 13(4), 485–497. <https://doi.org/10.1017/S1744552317000404>
- Hartmann, C., Amine, M., Klier, S., & Vorwerk, K. (Hrsg.) (2019). *Länderportfolioreviews. Ein Analyseinstrument für die deutsche Entwicklungszusammenarbeit*. Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit. Abgerufen von <https://nbn-resolving.org/urn:nbn:de:101:1-2019091913032369399229>
- Henry, G. T., & Mark, M. M. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation*, 24(3), 293–314. <https://doi.org/10.1177/109821400302400302>
- Hoey, L. (2015). Show me the numbers: Examining the dynamics between evaluation and government performance in developing countries. *World Development*, 70, 1–12. doi:10.1016/j.worlddev.2014.12.019
- Holzapfel, S., & Römling, C. (2020). *Monitoring in German bilateral development cooperation. A case study of agricultural, rural development and food security projects* (Discussion Paper 18/2020). Bonn: Deutsches Institut für Entwicklungspolitik (DIE).
- Hovland, I. (2003). *Knowledge management and organisational learning: An international development perspective* (Working Paper 224). Abgerufen von <https://www.files.ethz.ch/isn/95711/wp224.pdf>
- Kalton, G., & Anderson, D. W. (1986). Sampling Rare Populations. *Journal of the Royal Statistical Society, Ser. A*, 149, 65–82.
- KfW (Kreditanstalt für Wiederaufbau). (2014). 13. *Evaluierungsbericht 2013–2014. Wirkung im ländlichen Raum: Tragfähige Ansätze für Mensch und Natur*.
- KfW. (2019). 15. *Evaluierungsbericht 2017–2018. Zu größerer Wirkung in kleineren Städten*.
- KfW. (2021a). 16. *Evaluierungsbericht 2019–2020. Begleiten. Bewerten. Lernen*.
- KfW. (2021b). *KfW Entwicklungsbank: Zahlen und Fakten 2020*.
- KfW. (2021c). *Das Wirkungsmanagement der KfW-Bankengruppe. Wirkungsverständnis, -kategorien, -indikatoren und Zusammenspiel der Daten*.
- KfW. (2022a). *Rolle/Wichtigkeit von RIE*. Abgerufen von kfw-entwicklungsbank.de/s/dezBLLbP
- KfW. (2022b). *Unsere Arbeitsweise*. Abgerufen von kfw-entwicklungsbank.de/s/dezRK8Y
- KfW. (2022c). *Unsere Wirkungen*. Abgerufen von kfw-entwicklungsbank.de/s/dezqUw9

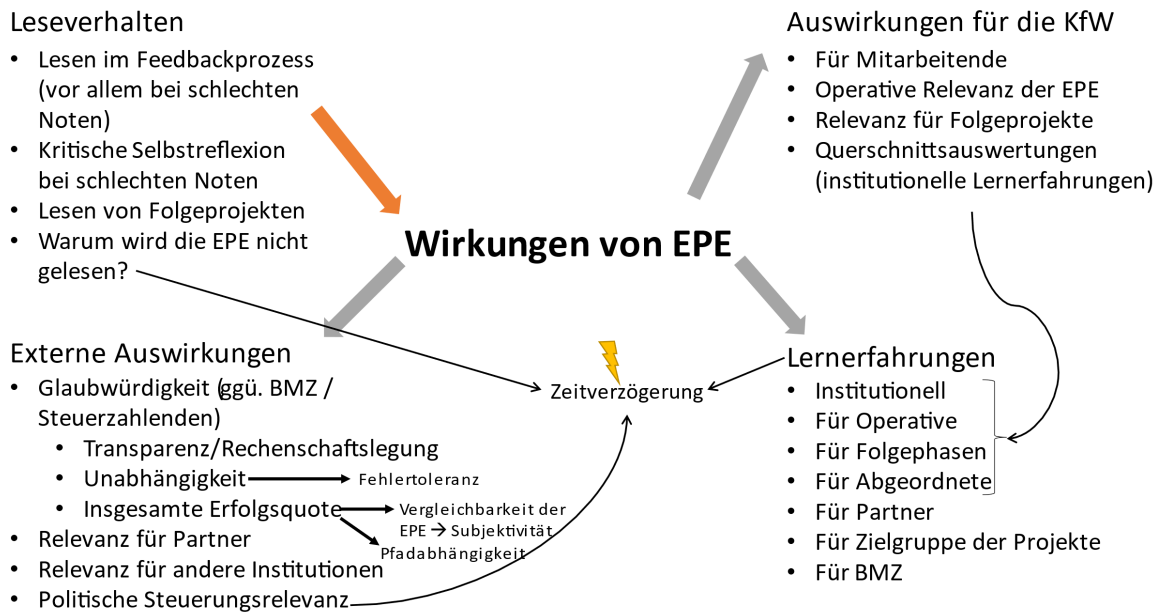
- KfW. (2022d). *Weltweites Engagement*. Abgerufen von <https://www.kfw.de/microsites/Microsite/transparenz.kfw.de/#/start>
- King, K., & McGrath, S. (2003). *Knowledge sharing in development agencies: Lessons from four cases*. Washington, DC: World Bank.
- Kirsch, R., & Wilson, M. B. (2014). *Ein Review der Evaluierungsarbeit zur deutschen Entwicklungszusammenarbeit in Afghanistan*. Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval). Abgerufen von <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-55213-3>
- Klingebiel, S. (2013). Ergebnisbasierte Ansätze in der Entwicklungszusammenarbeit: Möglichkeiten und Grenzen neuer Ansätze. In R. Öhlschläger & H. Sangmeister (Hrsg.), *Von der Entwicklungshilfe zur internationalen Zusammenarbeit* (S. 201–210). Nomos Verlagsgesellschaft mbH & Co. KG. <https://doi.org/10.5771/9783845245676-201>
- Kogen, L. (2018). What have we learned here? Questioning accountability in aid policy and practice. *Evaluation*, 24, 98–112. doi:10.1177/1356389017750195
- Krämer, M., Jechel, L., Kretschmer, T., & Schneider, E. (2021). *Rigorous impact evaluation: Evidence generation and take-up in German development cooperation*. Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval).
- Krohwinkel-Karlsson, A. (2007). Knowledge and learning in aid organisations: A literature review with suggestions for further studies. *SADEV Working Paper*, 1.
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* (Grundlagen Psychologie, 6. Aufl.). Weinheim: Beltz.
- Mayring, P. (2010). Qualitative Inhaltsanalyse. In G. Mey & K. Mruck (Hrsg.), *Handbuch qualitative Forschung in der Psychologie* (1. Auflage, 2010, S. 601–613). Wiesbaden: VS Verlag. https://doi.org/10.1007/978-3-531-92052-8_42
- Merton, R. K. (1949). Patterns of influence: a study of interpersonal influence and of communications behavior in a local community. In P. F. Lazarsfeld & F. N. Stanton (Hrsg.), *Communications research 1948-49* (S. 180–219). New York: Duell, Sloan & Pearce.
- Meuser, M., & Nagel, U. (1991). ExpertInneninterviews – vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In D. Garz & K. Kraimer (Hrsg.), *Qualitativ-empirische Sozialforschung. Konzepte, Methoden, Analysen* (S. 441–471). Wiesbaden: Springer Fachmedien.
- Meyer, W., Bär, T., Faust, J., Jan, S. von, Silvestrini, S., & Wein, S. (2019). Die DeGEval-Standards in der deutschen bilateralen Entwicklungszusammenarbeit. In J. U. Hense, W. Böttcher, M. Kalman & W. Meyer (Hrsg.), *Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis?* (1. Aufl., S. 165–181). Münster: Waxmann.
- Mosley, P. (1986). Aid-effectiveness: The micro-macro paradox. *IDS Bulletin*, 17(2), 22–27. <https://doi.org/10.1111/j.1759-5436.1986.mp17002004.x>
- Noltze, M., Euler, M., & Verspohl, I. (2018a). *Nachhaltigkeit in der deutschen Entwicklungszusammenarbeit. Meta-Evaluierung*. Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval).
- Noltze, M., Euler, M., & Verspohl, I. (2018b). *Nachhaltigkeit in der deutschen Entwicklungszusammenarbeit. Meta-Evaluierung, Zusammenfassung*. Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval).
- OECD (Organization for Economic Co-operation and Development). (1991). *Principles for evaluation of development assistance*. Paris: OECD Publishing. Abgerufen von <https://www.oecd.org/development/evaluation/2755284.pdf>
- OECD. (2001). *Evaluation feedback for effective learning and accountability* (Evaluation and aid effectiveness, Bd. 5). Paris: OECD Publishing.
- OECD. (2010). *Qualitätsstandards für die Entwicklungsevaluierung* (DAC-Reihe Leitlinien und Grundsatztexte). Paris: OECD Publishing. <https://doi.org/10.1787/9789264085183-de>

- OECD. (2020). *Better criteria for better evaluation: Revised and updated evaluation criteria*. Paris: OECD Publishing.
- OECD. (2021a). *Applying evaluation criteria thoughtfully*. Paris: OECD Publishing. <https://doi.org/10.1787/543e84ed-en>
- OECD. (2021b). *DAC-Prüfbericht über die Entwicklungszusammenarbeit: Deutschland 2021. Wichtigste Ergebnisse und Empfehlungen*. Paris: OECD Publishing. <https://doi.org/10.1787/83f90077-de>
- OECD/DAC Network on Development Evaluation. (2019). *Better criteria for better evaluation: Revised evaluation criteria definitions and principles for use*. Paris: OECD Publishing. Abgerufen von <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf>
- Porter, T. M. (2015). The Flight of the Indicator. In R. Rottenburg, S. E. Merry, S. J. Park & J. Mugler (Eds.), *The world of indicators: The making of governmental knowledge through quantification* (Cambridge studies in law and society, S. 34–55). Cambridge: Cambridge University Press.
- Ramalingam, B. (2005). *Implementing knowledge strategies: Lessons from international development agencies* (RAPID Working Paper 244, Vol. 244). London.
- Reinertsen, H., Bjørkdahl, K., & McNeill, D. (2022). Accountability versus learning in aid evaluation: A practice-oriented exploration of persistent dilemmas. *Evaluation*, 28(3), 356–378. <https://doi.org/10.1177/13563890221100848>
- Rottenburg, R., Merry, S. E., Park, S. J., & Mugler, J. (Eds.) (2015). *The world of indicators: The making of governmental knowledge through quantification* (Cambridge studies in law and society). Cambridge: Cambridge University Press.
- Sabbi, M., & Stroh, A. (2020). The “Numbers Game”: Strategic reactions to results-based development assistance in Ghana. *Studies in Comparative International Development*, 55(1), 77–98. <https://doi.org/10.1007/s12116-019-09296-z>
- Schönhuth, M., & Jerrentrup, M. T. (2019). *Partizipation und nachhaltige Entwicklung: Ein Überblick*. Wiesbaden: Springer Fachmedien.
- Shallwani, S., & Dossa, S. (2023). Evaluation and the White Gaze in International Development. In T. Khan, K. Dickson & M. Sondarjee (Hrsg.), *White saviorism in international development. Theories, practices and lived experiences* (S. 42–62). Wakefield, Quebec, Kanada: Daraja Press.
- Stockmann, R. (2004). *Was ist eine gute Evaluation? Einführung zu Funktionen und Methoden von Evaluationsverfahren* (CEval-Arbeitspapiere 9). Saarbrücken: Centrum für Evaluation.
- Stockmann, R. (2007). *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (Sozialwissenschaftliche Evaluationsforschung, Bd. 6). Münster, München, Berlin: Waxmann. Abgerufen von <http://www.socialnet.de/rezensionen/isbn.php?isbn=978-3-8309-1766-3>
- Terberger, E. (2011). Evaluierung in der Entwicklungszusammenarbeit. Das Beispiel der Finanziellen Zusammenarbeit. Theoretische Konzepte, strukturelle Herausforderungen und praktische Umsetzung. In J. König & J. Thema (Hrsg.), *Nachhaltigkeit in der Entwicklungszusammenarbeit. Theoretische Konzepte, strukturelle Herausforderungen und praktische Umsetzung* (Globale Gesellschaft und internationale Beziehungen, 1. Aufl., S. 219–238). Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-93091-6>
- Vindrola-Padros, C., & Johnson, G. A. (2020). Rapid Techniques in Qualitative Research: A Critical Review of the Literature. *Qualitative Health Research*, 30(10), 1596–1604. <https://doi.org/10.1177/1049732320921835>
- Weber, M. (1977). Objectivity of Social Science and Social Policy. In F. R. Dallmayr & T. A. McCarthy (Hrsg.), *Understanding and Social Inquiry* (S. 24–37). Notre Dame: University of Notre Dame Press.
- Wilhelm, J. L. (Hrsg.). (2015). *Evaluation komplexer Systeme. Systemische Evaluationsansätze in der deutschen Entwicklungszusammenarbeit* (Potsdamer Geographische Praxis, Bd. 10). Potsdam: Univ.-Verl.

- Yanguas, P. (2021). *What have we learned about learning? Unpacking the relationship between knowledge and organisational change in development agencies* (Discussion Paper 9/2021). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE). <https://doi.org/10.23661/DP9.2021>
- Zintl, M. (2009). Evaluierung in der deutschen Entwicklungszusammenarbeit. In T. Widmer, W. Beywl & C. Fabian (Hrsg.), *Evaluation*. VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91468-8_24

Anhang

Anhang 1: Kleines Kodierungssystem



Quelle: Eigene Darstellung

Anhang 2: Informationen zu den Interviewten

Befragungsperson	Hintergrund	Hat EPE in KfW erstellt
KFW1	KfW – operativer Bereich	nein
KFW2	KfW – früher operativer Bereich; jetzt Kompetenzzentrum	ja
KFW3	KfW – operativer Bereich	nein
KFW4	KfW – früher FZ-E; jetzt operativer Bereich	ja
KFW5	KfW – operativer Bereich	nein
KFW6	KfW – operativer Bereich	nein
BMZ1	BMZ – Länderreferent*in	-
DEV1	DEval – Kompetenzzentrum Methoden	-
IDOS1	IDOS – Forschung zur Entwicklungszusammenarbeit	-