



META-EVALUIERUNG ZUR QUALITÄT VON (PROJEKT-) EVALUIERUNGEN IN DER DEUTSCHEN ENTWICKLUNGS- ZUSAMMENARBEIT

Zusammenfassung
2022

IMPRESSUM

Verfasst von

Dr. Kerstin Guffler
Laura Kunert
Marian Wittenberg
Dr. Nico Herforth

Verantwortlich

Amélie Gräfin zu Eulenburg

Gestaltung Umschlag und Grafiken

Katharina Mayer, DEval

Lektorat

Marcus Klein, PhD

Bildnachweis

Titelseite: VektorMine, Shutterstock

Bibliografische Angabe

Guffler, K., L. Kunert, M. Wittenberg und N. Herforth (2022), *Meta-Evaluierung zur Qualität von (Projekt-)Evaluierungen in der deutschen Entwicklungszusammenarbeit*, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

Druck

Bonifatius, Paderborn

© Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), 2022

ISBN 978-3-96126-171-0 (gebundene Ausgabe)

ISBN 978-3-96126-172-7 (PDF)

Herausgeber

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

Tel: +49 (0)228 33 69 07-0

E-Mail: info@DEval.org

www.DEval.org

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

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

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

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

Eine Stellungnahme des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) findet sich unter: <https://www.bmz.de/de/ministerium/evaluierung/bmz-stellungnahmen-19404>

Dies ist die deutsche Zusammenfassung des DEval-Berichts "Meta-Evaluierung zur Qualität von (Projekt-)Evaluierungen in der deutschen Entwicklungszusammenarbeit".

Der vollständige Bericht kann hier heruntergeladen werden: <https://www.deval.org/de/evaluierungen/laufende-und-abgeschlossene-evaluierungen/metaevaluierung-zur-qualitaet-von-projekt-evaluierungen-in-der-deutschen-ez>

ZUSAMMENFASSUNG

Einleitung

Meta-Evaluierungen, die als Evaluierung von Evaluierungen bezeichnet werden können, gewinnen in der EZ zunehmend an Bedeutung. Laut Caracelli und Cooksy (2009, S. 2 f., eigene Übersetzung) werden Meta-Evaluierungen „durchgeführt, um den Evaluierungsprozess zu verbessern, die Stärken und Schwächen einer Evaluierung systematisch zu reflektieren, die zukünftige Evaluierungsarbeit zu verbessern oder Informationen zur Glaubwürdigkeit der Ergebnisse für die Nutzenden zur Verfügung zu stellen“. Diesem Verständnis folgt auch die vorliegende Meta-Evaluierung. In der internationalen Entwicklungszusammenarbeit (EZ) besteht inzwischen eine größere Anzahl an Meta-Evaluierungen, aber auch in der deutschen EZ werden organisationsinterne und -übergreifende Meta-Evaluierungen durchgeführt.

Die vorliegende Meta-Evaluierung, die zeitweise parallel zur Erstellung der Leitlinien Evaluierung des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) lief, schließt inhaltlich unter anderem an frühere Systemprüfungen des deutschen EZ-Systems, das Umsetzungsmonitoring einer Systemprüfung und die Meta-Evaluierung Nachhaltigkeit (Noltze et al., 2018) an. Im Jahr 1999 wurde im Auftrag des BMZ eine erste systematische Untersuchung der Evaluierungspraxis in der deutschen EZ vorgenommen (Borrmann et al., 1999) und 2009 eine zweite Systemprüfung abgeschlossen (Borrmann und Stockmann, 2009). Auf Letzterer aufbauend führte das DEval im Jahr 2015 ein Umsetzungsmonitoring zu den Ergebnissen und Empfehlungen durch (Lücking et al., 2015). Im Jahr 2018 wurde eine organisationsübergreifende Meta-Evaluierung¹ zur Analyse der Qualität von Projektevaluierungen der Deutschen Gesellschaft für Internationale Zusammenarbeit (GIZ) und der KfW Entwicklungsbank (KfW) vom DEval veröffentlicht (Noltze et al., 2018).

Vor Abschluss der Meta-Evaluierung wurden als bisher letzter Meilenstein die BMZ-Leitlinien zur Evaluierung in der EZ (BMZ, 2021) veröffentlicht, unter anderem um das Qualitätsverständnis in der deutschen EZ weiter zu festigen. Darin sind insbesondere die Standards des Entwicklungsausschusses (Development Assistance Committee, DAC) der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (Organisation for Economic Co-operation and Development, OECD), aber auch der Gesellschaft für Evaluation (DeGEval) als verbindlich für die Durchführungsorganisationen beziehungsweise als Orientierung für die nichtstaatlichen Organisationen verschriftlicht worden. Da der Untersuchungszeitraum der Meta-Evaluierung vor Verabschiedung der Leitlinien lag, konnten diese nicht in die Analyse einbezogen werden. Gleichwohl wurden wichtige Erkenntnisse für die zukünftige Ausgestaltung der Leitlinien generiert.

Eine Besonderheit der vorliegenden Meta-Evaluierung liegt darin, dass die Anwendung der Qualitätsstandards in Evaluierungen bei einer Vielzahl staatlicher und nichtstaatlicher Organisationen untersucht wurde. Die nachfolgenden Organisationen waren an der Meta-Evaluierung beteiligt: Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), CARE Deutschland e. V. (CARE), Deutscher Volkshochschul-Verband International (DVV), Deutsches Rotes Kreuz (DRK), Evangelisches Werk für Diakonie und Entwicklung e. V. (EWDE), GIZ, Heinrich-Böll-Stiftung (hbs), Konrad-Adenauer-Stiftung (KAS), KfW, MISEREOR und Physikalisch-Technische Bundesanstalt (PTB). Untersucht wurden in Deutschland (mit-)verantwortete und zwischen Oktober 2016 und Dezember 2020 umgesetzte Projektevaluierungen. Dabei wurden staatliche Organisationen vollumfänglich und nichtstaatliche entlang von Kriterien bezüglich ihrer strukturellen Heterogenität ausgewählt, sodass die Ergebnisse ein möglichst breites Spektrum an Erfahrungen mit der Anwendung von Qualitätsstandards abbilden konnten.

¹ In den BMZ-Leitlinien zur Evaluierung in der EZ werden die DEval-Meta-Evaluierungen als ein Teil der Qualitätssicherung des Evaluierungssystems benannt (BMZ, 2021).

Die Anwendung der Qualitätskriterien wurde entlang der Verpflichtungsgrundlage der Organisationen zu den Standarddokumenten der OECD-DAC- und/oder der DeGEval- und organisationsspezifischer Qualitätskriterien untersucht sowie für GIZ und KfW entlang der Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit.

Mit der organisationsübergreifenden Meta-Evaluierung sollen Erkenntnisse zum Qualitätsverständnis von Evaluierungen bei den beteiligten Organisationen sowie zu den Stärken und Schwächen in der Anwendung der Qualitätsstandards geliefert werden. Darüber hinaus wurden Faktoren identifiziert und untersucht, die mit der Anwendung der Qualitätsstandards zusammenhängen. Im Sinne des zukünftigen Lernens wurde in dieser Meta-Evaluierung auch das Ziel verfolgt, Erklärungen für die Nichtanwendung von Qualitätsstandards aufzuzeigen. Die nachfolgenden Evaluierungsfragen wurden dabei untersucht:

Evaluierungsfragen

1. Wie ist das Qualitätsverständnis von Evaluierungen bei den beteiligten Organisationen in der deutschen EZ?
2. Inwieweit werden Qualitätsstandards bei Evaluierungen der beteiligten Organisationen in der deutschen EZ angewandt?
 - a) Inwieweit zeigen sich Stärken und Schwächen bei der Anwendung der OECD-DAC- und der DeGEval-Standards in den Evaluierungen der beteiligten deutschen EZ-Organisationen?
 - b) Inwieweit zeigen sich Stärken und Schwächen bei der Anwendung der organisationsspezifischen Qualitätsstandards in den Evaluierungen der beteiligten deutschen EZ-Organisationen?
 - c) Inwieweit zeigen sich Stärken und Schwächen bei der Anwendung der Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit in den Evaluierungen von GIZ und KfW?
3. Inwieweit hängen länderkontext-, evaluierungs- und organisationsspezifische Faktoren mit der Anwendung der Qualitätsstandards zusammen?

Theoretische und empirische Herleitung

Qualitätsverständnis

In der vorliegenden Meta-Evaluierung wird die Evaluierungsqualität mit der Anwendung der einschlägigen beziehungsweise der für die Organisationen verpflichtenden Qualitätsstandards gleichgesetzt und dementsprechend untersucht. Der Qualitätsbegriff schließt vor allem – aber nicht ausschließlich – die Inhalte der OECD-DAC- und der DeGEval-Standarddokumente mit ein. Die Grundlage für die Identifizierung von Belegen guter Evaluierungen bilden die Qualitätsstandards des OECD DAC und der DeGEval aufgrund ihres international geltenden Charakters, ihres Bezugs zur EZ und ihrer Relevanz für deutsche EZ-Organisationen. Der Begriff „Anwendung der Qualitätsstandards“ wurde im Konsens mit der Referenzgruppe² gewählt und beschreibt, inwieweit der Nachweis erbracht werden kann (das heißt schriftlich dokumentiert war oder schriftlich auf Nachfrage zurückgemeldet wurde), ob und wie die Qualitätskriterien in den untersuchten Evaluierungen berücksichtigt wurden. Bei den OECD-DAC- und den DeGEval-Standards handelt es sich um Maximalstandards. Das bedeutet, dass die beteiligten Organisationen nicht alle Qualitätsstandards in allen Evaluierungen anwenden müssen. Das Qualitätsverständnis wurden im Weiteren um organisationsspezifische Qualitätsstandards und die Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit ergänzt.

² Die Referenzgruppe bestand aus Vertreter*innen der beteiligten Organisationen und VENRO sowie aus Referent*innen des BMZ-Referats GS 22 „Evaluierung und Ressortforschung, DEval, IDOS“. Ihre Mitglieder begleiteten den Prozess der Evaluierung in allen Evaluierungsphasen (zum Beispiel über virtuelle Treffen oder Kommentierungen von Evaluierungsdokumenten; DEval, 2021a).

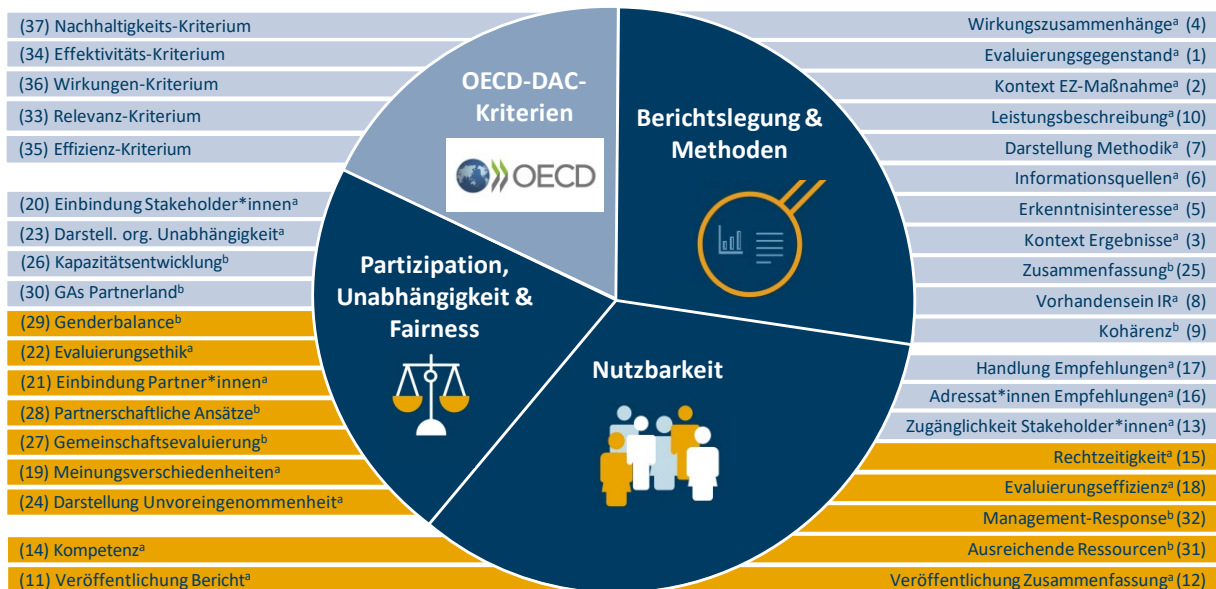
Analyseraster

Das Analyseraster beinhaltet Qualitätskriterien, die aus den OECD-DAC- und den DeGEval-Standarddokumenten abgeleitet wurden, organisationspezifische Qualitätskriterien und Qualitätskriterien aus der Meta-Evaluierung Nachhaltigkeit (Noltze et al., 2018). Die OECD-DAC- und die DeGEval-Qualitätskriterien lassen sich drei Bereichen zuordnen: 1) der Überschneidung zwischen den OECD-DAC- und den DeGEval-Standarddokumenten, 2) dem OECD-DAC-Standarddokument ohne Überschneidung mit dem DeGEval-Standarddokument (OECD DAC only) und 3) den OECD-DAC-Kriterien. Da alle Organisationen im Untersuchungszeitraum den „OECD-DAC-Kriterien“ (BMZ, 2006) verpflichtet waren, wird dies als separater Bereich aufgeführt, obwohl ihre Anwendung Teil der OECD-DAC-Standards ist (Qualitätsstandard 2.8; OECD DAC, 2010). Für DRK, EWDE, GIZ und hbs umfasste das Analyseraster darüber hinaus weitere organisationspezifische Qualitätskriterien³ und für GIZ und KfW Qualitätskriterien, die bereits in der vorangegangenen Meta-Evaluierung Nachhaltigkeit erhoben worden waren.

Standardcluster

Mit Ausnahme der fünf OECD-DAC-Kriterien wurden die Qualitätskriterien drei inhaltlichen Standardclustern – „Berichtslegung und Methoden“, „Partizipation, Unabhängigkeit und Fairness“ und „Nutzbarkeit“ – zugeordnet.⁴ Das Standardcluster „Berichtslegung und Methoden“ umfasst vor allem Qualitätskriterien, die sich auf die Darstellung von Informationen zur Methodik der Evaluierung beziehen oder auf das Vorliegen beziehungsweise den Informationsgehalt ausgewählter Evaluierungsdokumente (Abbildung 1). Dem Standardcluster „Partizipation, Unabhängigkeit und Fairness“ sind vor allem Qualitätskriterien zugeordnet, die sich mit der Berücksichtigung unterschiedlicher Personengruppen in der Evaluierung auseinandersetzen. Das Standardcluster „Nutzbarkeit“ fokussiert insbesondere auf die Nützlichkeit der Evaluierungen, während die aktive Nutzung nur eine kleine Rolle spielt und der Nutzen nicht untersucht wird.⁵

Abbildung 1 Zuordnung der 37 Qualitätskriterien zu den Standardclustern und OECD-DAC-Kriterien



Quelle: DEval, eigene Darstellung

Anmerkung: blauer Balken = Qualitätskriterium wurden je Evaluierung untersucht; gelber Balken = Qualitätskriterium wurde auf Organisationsebene über alle Evaluierungen hinweg untersucht; GAs = Gutachtende; Darstell. = Darstellung; org. = organisationale
^a Qualitätskriterium kommt aus der Überschneidung von OECD-DAC- und DeGEval-Standards; ^b Qualitätskriterium kommt aus dem Bereich „OECD DAC only“.

³ Organisationspezifische Qualitätskriterien wurden definiert als Anforderungen, die unabhängig von den OECD-DAC- oder den DeGEval-Standards eine große Bedeutung für eine Organisation hinsichtlich der Qualität ihrer Evaluierungen haben.

⁴ Die Benennungen der drei gebildeten Standardcluster zeigen Ähnlichkeiten zu den Benennungen der DeGEval-Standardgruppen (1. Nützlichkeit, 2. Durchführbarkeit, 3. Fairness und 4. Genauigkeit). Da die identifizierten Qualitätskriterien allerdings zum Teil die Überschneidung zwischen den OECD-DAC- und den DeGEval-Standards darstellen, war eine identische Benennung inhaltlich nicht zutreffend. Die OECD-DAC-Standards sind überwiegend entlang von Evaluierungsphasen strukturiert, sodass diese Benennungen nicht berücksichtigt wurden.

⁵ Weitere Details zu den Begriffsdefinitionen finden sich im Haupttext des Evaluierungsberichts in Abschnitt 2.1.

Faktoren für die Anwendung der Qualitätsstandards

Die nachfolgend beschriebenen Faktoren wurden in den Analysen untersucht, wenn sie drei Merkmale erfüllten, nämlich 1) eine eindeutige organisationsübergreifende Definition aufwiesen, 2) klare Wirkungszusammenhänge mit ausgewählten Qualitätsstandards beschrieben werden konnten und 3) Daten bei den Organisationen oder in Sekundärdatenbanken verfügbar waren. Um die Faktoren zu identifizieren, wurden drei Fokusgruppendifkussionen mit den Verantwortlichen der beteiligten Organisationen durchgeführt sowie wissenschaftliche und empirische Literatur gesichtet. Die Faktoren wurden anschließend entlang der 1) länderkontext-, 2) evaluierungs- und 3) organisationspezifischen Einflussdimensionen systematisiert.

Methodisches Vorgehen

Datengrundlage und -analyse

Nachdem die Organisationen in die Meta-Evaluierung aufgenommen waren, wurde eine geschichtete Stichprobe von insgesamt 296 Evaluierungen für die Untersuchung der Anwendung der Qualitätsstandards gezogen. Die Organisationen wurden anhand von vier Kriterien ausgewählt, um eine möglichst große strukturelle Heterogenität abzudecken und je Organisation ausreichend Evaluierungen untersuchen zu können. Insgesamt haben die Evaluierungseinheiten/-stellen der Organisationen im Untersuchungszeitraum von Oktober 2016 bis Dezember 2020 839 Evaluierungen in Deutschland (mit-)verantwortet. In die Grundgesamtheit wurden daraufhin die 576 Evaluierungen aufgenommen, die das BMZ entweder (mit-)gefördert hatte oder in denen eine vom BMZ (mit-)geförderte EZ-Maßnahme untersucht wurde. Die daraus gezogene Stichprobe umfasste 296 Evaluierungen.

Qualitätsverständnis (Evaluierungsfrage 1): Für die Erarbeitung des Qualitätsverständnisses beziehungsweise der Verpflichtungsgrundlage der beteiligten Organisationen wurden Organisationsdokumente sowie BMZ-Vorgaben mithilfe einer qualitativen Inhaltsanalyse ausgewertet.

OECD-DAC- und/oder DeGEval-Standards (Evaluierungsfrage 2a): Um die Anwendung der Qualitätskriterien in den 296 Evaluierungen zu untersuchen, wurden Evaluierungsdokumente (Evaluierungsberichte und -anhänge, Leistungsbeschreibungen, Inception Reports) und Organisationsdokumente (zum Beispiel Evaluierungskonzepte, Leitfäden und Handreichungen zur Durchführung von Evaluierungen) herangezogen. Für 14 der 37 OECD-DAC- und DeGEval-Qualitätskriterien konnten in der Inter-Kodierenden-Phase keine oder nur sehr wenig Informationen in den von den Organisationen bereitgestellten Evaluierungsdokumenten kodiert werden. Um keine fehlerhaften Rückschlüsse auf eine Nichtanwendung zu ziehen, wurden die Verantwortlichen der Evaluierungseinheiten/-stellen in einem weiteren Schritt zur Anwendung dieser Qualitätskriterien in ihrer Organisation befragt.⁶ Die im Analyseraster festgelegten Qualitätskriterien wurden entlang von ordinalen oder binären Bewertungsstufen kodiert. Den von den Verantwortlichen der Evaluierungseinheiten/-stellen in der Onlinebefragung zurückgemeldeten durchschnittlichen Häufigkeiten der Anwendung der Qualitätskriterien wurden ebenfalls Werte zugeordnet. Für jedes Qualitätskriterium wurden je Organisation verschiedene Kennzahlen (zum Beispiel Mittelwerte) berechnet. Die Kennzahlen wurden nachfolgend in Prozentwerte umgewandelt und den vorab festgelegten Schwellenwerten des Anspruchsniveaus zugeordnet. Anschließend wurde ein Mittelwert über alle Organisationen hinweg ermittelt und mit Verfahren der deskriptiven Statistik ausgewertet.

Organisationsspezifische Qualitätsstandards (Evaluierungsfrage 2b): Für die Untersuchung der Anwendung der elf organisationsspezifischen Qualitätskriterien wurden diese in den Evaluierungsdokumenten der vier betroffenen Organisationen kodiert und anschließend wie die OECD/DAC- und die DeGEval-Qualitätskriterien berechnet und ausgewertet.

⁶ Da die Qualitätskriterien der Onlinebefragung durchschnittlich rund 6 Prozent weniger angewandt wurden als die Qualitätskriterien der Dokumentenanalyse, gab es keinen Anlass anzunehmen, dass sich die Organisationen systematisch besser bewerteten, als sie durch die objektive Kodierung bewertet worden wären.

Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit (Evaluierungsfrage 2c): Für die Analyse der Anwendung beziehungsweise der erneuten Anwendung der Qualitätskriterien aus der Meta-Evaluierung Nachhaltigkeit wurden 15 Qualitätskriterien aus der vorangegangenen Meta-Evaluierung von Noltze et al. (2018) herangezogen. Acht Qualitätskriterien waren bereits als Qualitätskriterien im OECD-DAC- und im DeGEval-Analyseraster berücksichtigt und wurden lediglich transformiert, die übrigen sieben wurden in den aktuellen Evaluierungen erneut kodiert. Für die Auswertung wurden deskriptive Statistiken und für die Differenz der Ergebnisse zwischen der Meta-Evaluierung Nachhaltigkeit und der vorliegenden Meta-Evaluierung Strukturgleichungsmodelle berechnet (Weiber und Mühlhaus, 2010).

Um die Zusammenhänge zwischen ausgewählten Faktoren und der Anwendung der Qualitätskriterien zu untersuchen, wurden multivariate Regressionsanalysen geschätzt (Evaluierungsfrage 3). Regressionsanalysen erlauben die Identifikation von statistischen Zusammenhängen zwischen den Faktoren (unabhängige Variablen) und den Qualitätskriterien sowie dem Standardcluster „Berichtslegung und Methoden“ (abhängige Variablen; Backhaus et al., 2011). Konkret wurden Faktoren in 1) der Länderkontext-, 2) der Evaluierungs- und 3) der Organisationsdimension untersucht. Die Informationen für die einzelnen Faktoren wurden mithilfe von Daten gewonnen, die von den Organisationen übermittelt worden waren, sowie über Sekundärdatenbanken.

Bewertung der Anwendung der Qualitätsstandards

Die Anwendung der Qualitätskriterien wurde bei Organisationen mit Verpflichtungsgrundlage zur Anwendung der Qualitätskriterien sowohl untersucht als auch bewertet (Gruppe 1), bei Organisationen ohne Verpflichtungsgrundlage ausschließlich untersucht (Gruppe 2). Die Schwellenwerte des Anspruchsniveaus für die Anwendung der Qualitätskriterien wurden im Austausch mit der Referenzgruppe festgelegt und dienten als Basis für die Bewertung. Das Anspruchsniveau stellte die ex ante festgelegte Einschätzung dar, ab wann ein Qualitätskriterium in einer Evaluierung als kaum, teilweise, größtenteils und vollständig angewandt gilt. Bei der Festlegung der Schwellenwerte in 25-Prozent-Schritten ($0 \leq 25$ Prozent = „kaum angewandt“, $25 < 50$ Prozent = „teilweise angewandt“, $50 \leq 75$ Prozent = „größtenteils angewandt“, $75 < 100$ Prozent = „vollständig angewandt“) wurde berücksichtigt, dass die Qualitätsstandards als Maximalstandards zu verstehen sind. Die Bewertung wurde entlang der Schwellenwerte des Anspruchsniveaus mit Hinzunahme der Extremwerte 0 (verfehlt) und 100 (übertroffen) für Gruppe 1 vorgenommen. Da eine erneute Untersuchung der Anwendung der Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit stattfand, wurde an dieser Stelle der Anspruch gestellt, dass sich die Anwendung der Qualitätskriterien seit der Meta-Evaluierung Nachhaltigkeit verbessert hatte. Bei den Qualitätskriterien der OECD-DAC- und der DeGEval-Standards wurden beide Gruppen hinsichtlich ihres Grads der Anwendung analysiert, Gruppe 1 wurde zusätzlich bewertet. Bei den OECD-DAC-Kriterien (BMZ, 2006), den organisationspezifischen Qualitätskriterien und den Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit wurden ausschließlich Organisationen der Gruppe 1 untersucht und bewertet.

Stärken und Herausforderungen des methodischen Vorgehens

Aufgrund der Auswahl der Organisationen entlang ihrer strukturellen Heterogenität wurde eine große Bandbreite in der Anwendung einzelner Qualitätskriterien untersucht und dargestellt. Die Erkenntnisse der Meta-Evaluierung gelten für die beteiligten Organisationen. Die staatlichen Organisationen wurden vollständig abgebildet. Die Auswahl der beteiligten nichtstaatlichen Organisationen ist nicht repräsentativ für alle Organisationen der nichtstaatlichen Entwicklungszusammenarbeit. Entsprechend der Auswahl der Organisationen entlang ihrer strukturellen Heterogenität, können sich nicht beteiligte nichtstaatliche Organisationen innerhalb dieser Bandbreite verorten und Erkenntnisse der Meta-Evaluierung für sich nutzen. Eine Übertragbarkeit der Erkenntnisse auf die Grundgesamtheit der Evaluierungen je Organisation war im Rahmen der gewählten statistischen Kennwerte für die Stichprobenziehung gewährleistet, eine Übertragbarkeit auf andere Evaluierungstypen einer Organisation nicht.

Aufgrund der systematischen Ableitung des Analyserasters entlang der OECD-DAC- und der DeGEval-Standards kann es auch von anderen Organisationen genutzt werden. Das Analyseraster der Meta-Evaluierung kann somit zukünftig als Grundlage für die Erstellung eines Analyserasters auf der Basis der BMZ-Leitlinien Evaluierung herangezogen werden.

Bei der Analyse der Qualitätskriterien aus der Onlinebefragung bestanden Einschränkungen hinsichtlich der Methoden-Triangulation. Die Hinzunahme einer weiteren Datenerhebungsmethode hätte allerdings angesichts des Aufwand-Nutzen-Verhältnisses in keiner angemessenen Relation gestanden. Neben der Befragung der Verantwortlichen der Evaluierungseinheiten/-stellen hätte zur Triangulation der Daten zum Beispiel auch die Bewertung der Gutachtenden der jeweiligen Evaluierungen herangezogen werden können. Da in der vorliegenden Meta-Evaluierung aber eine große Anzahl an Evaluierungen untersucht wurde, lag es außerhalb des Umfangs dieser Untersuchung, anstatt oder in Ergänzung zu den Evaluierungsverantwortlichen ehemalige Gutachtende der Evaluierungen zu befragen. Darüber hinaus hätten aufgrund von Personalfluktuation die Gutachtenden zum Teil nicht mehr ausfindig gemacht und befragt werden können. Die Stichprobengröße wäre somit gemindert gewesen.

Generell bestanden Grenzen in der Messung einiger Qualitätskriterien. Bei bestimmten Qualitätskriterien bedürfte es eines hohen Aufwands, um eine „gute“ Anwendung zu untersuchen. Es gibt Qualitätskriterien, die nur mit viel Aufwand in der Tiefe untersucht werden können. Zum Beispiel ist beim Qualitätskriterium „Einbindung der Stakeholder*innen“ und „Zugänglichkeit für Stakeholder*innen“ sowohl die angemessene Anzahl der Stakeholder*innen, die eingebunden werden können, als auch die Intensität der Einbindung in den verschiedenen Evaluierungsphasen schwer ermittelbar.

Die Operationalisierungen der Qualitätskriterien wurden über alle Organisationen hinweg entwickelt. Einige dieser Operationalisierungen trafen nicht auf die Evaluierungspraxis aller Organisationen zu, sodass die Bewertung der Anwendung für diese Organisationen niedriger war, als sie entlang alternativer Operationalisierungen ausgefallen wäre. Es gibt Qualitätskriterien, bei denen mehr Spielraum für die Operationalisierung bestand (beispielsweise bei den „Qualitätssicherungsprozessen“). Hier besteht entsprechend ein Zielkonflikt zwischen dem Erkenntnisinteresse der Anwendung ausgewählter Qualitätskriterien über Organisationen hinweg und der Heterogenität der Anwendung der Qualitätskriterien.

Aufgrund eines uneinheitlichen organisationsübergreifenden Verständnisses der Messung der „Kosten der Evaluierung“ konnten Analysen zu Erklärungen der (Nicht-)Anwendung von Qualitätsstandards zum Teil nur eingeschränkt durchgeführt werden.

Die wiederholte Untersuchung der Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit ermöglichte eine Untersuchung der Differenz in der Anwendung der Qualitätskriterien von GIZ und KfW über die Zeit. Sie liefert somit Hinweise, inwieweit organisationsinterne Reformen und die Unterstützung externer Akteure (BMZ und DEval) die Anwendung der Qualitätskriterien verbessern konnten. Sie zeigte darüber hinaus Herausforderungen auf, die mit einer längsschnittlichen Untersuchung einhergehen (zum Beispiel die Anhebung der Schwellenwerte des Anspruchsniveaus und die gegebenenfalls angemessenen Anpassungen von Qualitätskriterien über die Zeit).

Ergebnisse

Evaluierungsfrage 1: Wie ist das Qualitätsverständnis von Evaluierungen bei den beteiligten Organisationen in der deutschen EZ?

Das Qualitätsverständnis der beteiligten Organisationen beruhte überwiegend auf den OECD-DAC- und/oder den DeGEval- und gegebenenfalls organisationsspezifischen Qualitätsstandards. Mit diesen Qualitätsstandards hatten sich die beteiligten Organisationen zu Beginn der Meta-Evaluierung zum Teil wenig systematisch auseinandergesetzt. Darüber hinaus variierten die BMZ-Vorgaben zur Anwendung der Qualitätsstandards im Untersuchungszeitraum in den ausgewählten Haushaltstiteln – zum Teil wurden die OECD-DAC-Standards als verpflichtend gekennzeichnet, zum Teil bestanden keine Vorgaben.

Evaluierungsfrage 2a: Inwieweit zeigen sich Stärken und Schwächen bei der Anwendung der OECD-DAC- und der DeGEval-Standards in den Evaluierungen der beteiligten deutschen EZ-Organisationen?

Insgesamt zeichnete sich ein positives Bild in Bezug auf die Anwendung der OECD-DAC- und der DeGEval-Qualitätsstandards ab. Die beteiligten Organisationen der deutschen EZ wandten die Qualitätsstandards bei circa zwei Drittel ihrer Evaluierungen an. Dies traf auch – zu einem etwas niedrigeren Grad – auf Organisationen ohne Verpflichtungsgrundlage zu. Die Anwendung der Qualitätsstandards zwischen den Organisationen wich dabei zum Teil deutlich voneinander ab. Dies war aufgrund der gewählten Auswahlkriterien für die Aufnahme der beteiligten Organisationen in die Stichprobe zu erwarten und ermöglichte somit ein heterogenes Bild über die unterschiedlichen Anwendungsgrade hinweg.

Jedoch zeigte sich, dass die Qualitätsstandards in den Organisationsdokumenten überwiegend noch nicht vollumfänglich von den Organisationen identifiziert worden waren und ihre (Nicht-)Anwendung nicht systematisch verschriftlicht worden ist. Dies trifft ebenfalls auf die Nachvollziehbarkeit der Anwendung und Nichtanwendung bei einigen ausgewählten Qualitätsstandards auf Ebene der einzelnen Evaluierung zu. Es ist anzumerken, dass die Anwendung einiger Qualitätsstandards aus verschiedenen Gründen nicht auf Evaluierungsebene, sondern auf Ebene der Organisation erfasst wurden. Dies könnte an der gewählten Operationalisierung einiger Qualitätskriterien in der Meta-Evaluierung liegen, durch eine fehlende Dokumentation der (Nicht-)Anwendung bedingt sein oder auch dadurch, dass die Dokumentation der Anwendung ausschließlich auf Organisationsebene und nicht auf Ebene der einzelnen Evaluierung erfolgte. Hier besteht deutlicher Verbesserungsbedarf, da eine externe Untersuchung ohne Informationen zu einer (begründeten) (Nicht-)Anwendung der Qualitätsstandards auf Ebene der Evaluierung nur eingeschränkt möglich war. Es war somit nicht nachvollziehbar, ob ein Qualitätsstandard (mit oder ohne Begründung) nicht angewandt oder angewandt, aber nicht dokumentiert wurde.

Die OECD-DAC-Kriterien wurden zu einem sehr großen Teil von den beteiligten Organisationen in der deutschen EZ erfüllt. Allerdings ist explizit darauf hinzuweisen, dass die Operationalisierung entlang der OECD-DAC-Standards und nicht der OECD-DAC-Kriterien erfolgte und eine Erfüllung somit leicht möglich war. Bei der Anwendung der OECD-DAC-Kriterien bestehen darüber hinaus erste Dokumentationen einer Nichtanwendung auf Organisations- und Evaluierungsebene. Hierin unterscheidet sich die Anwendung der OECD-DAC-Kriterien bereits – wenn auch nur in kleinerem Umfang – von der Anwendung der meisten anderen Qualitätskriterien. Es ist anzunehmen, dass in zukünftigen Evaluierungen die Dokumentation der Nichtanwendung der OECD-DAC-Kriterien weiter ansteigen wird, da ab 2020 (BMZ, 2020) die aktualisierte BMZ-Orientierungslinie zu den Evaluierungskriterien eine begründete und transparente Schwerpunktsetzung vorsieht.

Im Berichtsanhang im Abschnitt 7.1 werden die Erkenntnisse für die vier staatlichen Organisationen BGR, GIZ, KfW und PTB zusätzlich auf Ebene der einzelnen Organisation dargestellt und eingeordnet.

Evaluierungsfrage 2b: Inwieweit zeigen sich Stärken und Schwächen bei der Anwendung der organisationspezifischen Qualitätsstandards in den Evaluierungen der beteiligten deutschen EZ-Organisationen?

Es zeigt sich ebenfalls ein positives Bild für die Anwendung der organisationspezifischen Qualitätsstandards durch DRK, EWDE, GIZ und hbs. Sie wurden durchschnittlich „größtenteils angewandt“. Verbesserungspotenzial zeigen sich erneut in der Nachvollziehbarkeit der Nichtanwendung auf Ebene der Evaluierung.

Evaluierungsfrage 2c: Inwieweit zeigen sich Stärken und Schwächen bei der Anwendung der Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit in den Evaluierungen von GIZ und KfW?

Hinsichtlich der Anwendung der Qualitätskriterien zeichnet sich – mit wenigen Ausnahmen – ein positives Bild ab. Konkret wurden die Qualitätskriterien durchschnittlich zu circa 75 Prozent erfüllt. Dies entspricht einer etwas höheren Ausprägung der Anwendung als bei den OECD-DAC- und den DeGEval-Standards. Herausforderungen bestanden allerdings weiterhin in der Anwendung der Qualitätskriterien „Auswahlverfahren der Gesprächspartner beschrieben“ und „Kontroll-/Vergleichsgruppe einbezogen“. Darüber hinaus ist anzumerken, dass sich die Anwendung der Qualitätskriterien seit der Meta-Evaluierung Nachhaltigkeit in allen Qualitätskriterien – zum Teil bedeutsam – verbesserte. Insgesamt zeigt sich ein

durchschnittlicher Unterschied von 36 Prozent. Die Veränderungen deuten darauf hin, dass die nach der Meta-Evaluierung Nachhaltigkeit durchgeführten Maßnahmen zur Verbesserung der Evaluierungspraxis von GIZ und KfW einen Einfluss auf die Anwendung gehabt haben könnten. Dies ist vor dem Hintergrund der damit einhergehenden umfassenden Anstrengungen einer Vielzahl an Akteuren ein sehr positives Ergebnis. Anzumerken ist aber, dass Alternativerklärungen nicht ausgeschlossen werden können (beispielsweise relativ leicht zu erfüllende Operationalisierungen der Qualitätskriterien oder veränderte Dokumentationsweisen).

Evaluierungsfrage 3: Inwieweit hängen länderkontext-, evaluierungs- und organisationspezifische Faktoren mit der Anwendung der Qualitätsstandards zusammen?

Insgesamt geben die Ergebnisse hinsichtlich der in der Literatur und aus den Fokusgruppendifkussionen identifizierten Faktoren sowie der Anwendung der Qualitätskriterien wenig Hinweise auf signifikante Zusammenhänge. Dies sind insbesondere Faktoren aus der evaluierungsspezifischen Dimension „Anzahl der internen und externen Gutachtenden“ und die Umsetzung verschiedener „Qualitätssicherungsprozesse“.

Schlussfolgerungen und Empfehlungen

Die Empfehlungen der Meta-Evaluierung leiten sich vor allem aus den Evaluierungsfragen zur Anwendung ausgewählter Qualitätsstandards und -kriterien (OECD-DAC-, DeGEval- und/oder organisationspezifischer Standards als auch Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit, Evaluierungsfragen 2a bis 2c) ab. Die Empfehlungen sind allgemein formuliert, sodass alle beteiligten Organisationen sich entlang ihrer organisationspezifischen Ergebnisse selbst innerhalb der Empfehlungen verorten müssen, da der organisationsübergreifende Mittelwert der Ergebnisse für die Bewertung der Anwendung der Qualitätsstandards einzelner Organisationen nicht ausreichend aussagekräftig ist. Da in der kriterienbasierten Auswahl der nicht-staatlichen Organisationen ein Fokus auf ihre strukturelle Heterogenität gelegt und somit die Bandbreite möglicher Anwendungsgrade und -formen für unterschiedliche Organisationen abgebildet wurde, können sich auch nicht beteiligte nichtstaatliche Organisationen in den Ergebnissen verorten und damit an den Schlussfolgerungen und Empfehlung orientieren. Die an das BMZ gerichteten Empfehlungen beziehen sich auf das BMZ-Referat für Evaluierung.

Identifikation (nicht) relevanter Qualitätsstandards und Verschriftlichung dieser in Organisationsdokumenten

Empfehlung 1

- a) Die Evaluierungseinheiten/-stellen von BGR, CARE, DRK, DVV, EWDE, GIZ, hbs, KAS, KfW, MISEREOR und PTB sollten im Rahmen einer Revision ihrer Evaluierungspraxis – wenn noch nicht vorhanden – die für ihre Organisation verpflichtenden Qualitätsstandards identifizieren, in Organisationsdokumenten explizit benennen und ihre Anwendung in Evaluierungsprozessen festlegen. Die Identifikation und die systematische Verschriftlichung der Qualitätsstandards sollten in regelmäßigen Abständen überprüft werden. Dabei sollten die Organisationen ihren Anspruch an den Grad der Anwendung der einzelnen Qualitätsstandards konkret bestimmen. (Ergebnis: 2, 5.1, 9.1)
- b) Das BMZ sollte im Rahmen anstehender Aktualisierungen von Förderrichtlinien oder Nebenbestimmungen für einzelne Haushaltstitel einen Beitrag dazu leisten, die BMZ-Leitlinien Evaluierung als Referenzdokument für Evaluierungen zu stärken. Im Rahmen der Aktualisierungen sollte das BMZ gemeinsam mit den betroffenen nichtstaatlichen Organisationen organisationale Besonderheiten (beispielsweise wie bei den Förderrichtlinien der politischen Stiftungen) festhalten und verschriftlichen. Das Maximalstandardprinzip sollte dabei erhalten bleiben. (Ergebnis: 3)
- c) Das BMZ sollte basierend auf den BMZ-Leitlinien Evaluierung und im Austausch mit staatlichen und nichtstaatlichen Organisationen sowie unter Berücksichtigung des Analyserasters der vorliegenden Meta-Evaluierung ein Analyseraster für die Anwendung der Qualitätsstandards erarbeiten und den staatlichen und nichtstaatlichen Organisationen bereitstellen.

Empfehlung 2

- a) Die Evaluierungseinheiten/-stellen von BGR, CARE, DRK, DVV, EWDE, GIZ, hbs, KAS, KfW, MISEREOR und PTB sollten in Organisationsdokumenten die generelle Nichtanwendung einzelner für sie verpflichtender Qualitätsstandards begründen und verschriftlichen. (Ergebnis: 5.1, 6.2)
- b) Das BMZ sollte sich mit den staatlichen Organisationen bezüglich einer Anwendung und (begründeten) Nichtanwendung der in den BMZ-Leitlinien Evaluierung beschriebenen Qualitätsstandards verständigen, um eine Nichtanwendung auf Organisationsebene gemeinsam festzulegen oder Unstimmigkeiten zu dokumentieren.

Sicherstellung der Anwendung und Nachvollziehbarkeit der (Nicht-)Anwendung relevanter Qualitätsstandards auf Ebene der Evaluierung

Empfehlung 3

- a) Die Evaluierungseinheiten/-stellen von BGR, CARE, DRK, DVV, EWDE, GIZ, hbs, KAS, KfW, MISEREOR und PTB sollten – wenn noch nicht vorhanden – die Anwendung der auf Organisationsebene festgelegten Qualitätsstandards (Empfehlung 1) in den einzelnen Evaluierungen weiter verbessern, insbesondere die kaum oder teilweise angewandten Qualitätsstandards. Darüber hinaus sollte eine Anwendung oder (begründete) Nichtanwendung aller Qualitätsstandards auf Ebene jeder Evaluierung nachvollziehbar sein und von den Organisationen regelmäßig untersucht werden. (Ergebnis: 4.1, 4.2, 5.2, 5.3, 6.1, 7.1, 8.1, 9.2, 9.3)
- b) Das BMZ sollte die staatlichen Organisationen zur Sicherstellung der Anwendung und Nachvollziehbarkeit der (Nicht-)Anwendung der relevanten Qualitätsstandards auf Ebene der Evaluierung anhalten.

Gemeinsames Lernen

Empfehlung 4

- a) Die Evaluierungseinheiten/-stellen von BGR, CARE, DRK, DVV, EWDE, GIZ, hbs, KAS, KfW, MISEREOR und PTB sowie Vertreter*innen von VENRO sollten ihre unterschiedlichen Erfahrungen in der Identifikation, Verschriftlichung, Sicherstellung und Nachvollziehbarkeit der (Nicht-)Anwendung aller Qualitätsstandards regelmäßig untereinander austauschen. Der Austausch sollte auch nicht beteiligte Organisationen integrieren und weitere Evaluierungstypen beinhalten – zum Beispiel dezentrale Evaluierungen –, um die Anwendung der Qualitätsstandards weiter zu verbessern. (Ergebnis: 4.2, 4.3)
- b) Das BMZ sollte den Austausch zur Identifikation, Verschriftlichung, Sicherstellung und Nachvollziehbarkeit der (Nicht-)Anwendung der Qualitätsstandards mit und zwischen den Organisationen finanziell unterstützen.

Sicherstellung der Anwendung und Nachvollziehbarkeit der (Nicht-)Anwendung der Qualitätskriterien der Meta-Evaluierung Nachhaltigkeit

Empfehlung 5

- a) Das BMZ sollte im Zuge der Erarbeitung des Analyserasters für die in den BMZ-Leitlinien Evaluierung beschriebenen Qualitätsstandards (Empfehlung 1) die Übernahme der Qualitätskriterien aus der Meta-Evaluierung Nachhaltigkeit überprüfen und diese gegebenenfalls in das Analyseraster aufnehmen.
- b) GIZ und KfW sollten (angelehnt an Empfehlung 5a) die Anwendung und (Nicht-)Anwendung der Qualitätskriterien aus der Meta-Evaluierung Nachhaltigkeit, die in ein BMZ-Analyseraster übernommen wurden, sicherstellen beziehungsweise verbessern und die Nachvollziehbarkeit der (begründeten) (Nicht-)Anwendung je Evaluierung gewährleisten. (Ergebnis: 10.1, 10.2, 11.1)

Dies ist die deutsche Zusammenfassung des DEval-Berichts "Meta-Evaluierung zur Qualität von (Projekt-)Evaluierungen in der deutschen Entwicklungszusammenarbeit".

Der vollständige Bericht kann hier heruntergeladen werden: <https://www.deval.org/de/evaluierungen/laufende-und-abgeschlossene-evaluierungen/metaevaluierung-zur-qualitaet-von-projekt-evaluierungen-in-der-deutschen-ez>