

# AUSGEWÄHLTE ERKENNTNISSE ZU METHODISCHEN QUALITÄTSSTANDARDS IN PROJEKTEVALUIERUNGEN

## *Anwendungsbeispiele aus der Evaluierungspraxis von elf Organisationen der deutschen Entwicklungszusammenarbeit*

### Zusammenfassung

Die Anwendung von Qualitätsstandards ist von großer Bedeutung für Evaluierungen in der Entwicklungszusammenarbeit (EZ), damit sie ihren drei Funktionen gerecht werden können: praxisrelevante Erkenntnisse zu generieren, Lernen zu fördern und die Rechenschaftslegung zu stärken. Die Leitlinien des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) zu Evaluierungen betrachten dabei die Qualitätsstandards des Entwicklungsausschusses der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD-DAC) und der Gesellschaft für Evaluation (DeGEval) als wesentlich für die Güte von Evaluierungen in der deutschen Entwicklungszusammenarbeit (BMZ, 2021).

Um die Qualität von Evaluierungen zu untersuchen, hat das Deutsche Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) in einer Meta-Evaluierung die Anwendung von Qualitätsstandards anhand von Qualitätskriterien analysiert, die aus den OECD-DAC- und DeGEval-Standards abgeleitet wurden. Dafür wurden 296 Evaluierungen von elf staatlichen und nichtstaatlichen Organisationen der deutschen EZ ausgewählt und für den Zeitraum 2016 bis 2020 untersucht (Guffler et al., 2022). Darüber hinaus wurden weitere Qualitätskriterien (diese sind zur Unterscheidung im Folgenden mit einem \* markiert) aus einer vorangegangenen Meta-Evaluierung analysiert, um die Entwicklung der Anwendung dieser Qualitätskriterien für zwei der elf Organisationen, Deutsche Gesellschaft für internationale Zusammenarbeit (GIZ) und KfW Entwicklungsbank (KfW), seit 2018 abbilden zu können.

Dieser Policy Brief konzentriert sich auf ausgewählte Qualitätskriterien aus dem Bereich „Berichtslegung und Methoden“.

Methodischen Qualitätsstandards wird eine besondere Relevanz im Hinblick auf die Belastbarkeit und Nachvollziehbarkeit von Evaluierungsergebnissen zugeschrieben. Um einen Beitrag zur Verbesserung der Anwendung dieser Qualitätskriterien zu leisten, greift der vorliegende Policy Brief Qualitätskriterien heraus, zu denen in der Meta-Evaluierung Beispiele guter Praxis identifiziert werden konnten. Aus den Ergebnissen lassen sich organisationsübergreifend die folgenden Aussagen ableiten:

- Die dargestellten OECD-DAC und DeGEval-Qualitätskriterien werden größtenteils angewendet, jedoch nicht systematisch in den Evaluierungsdokumenten verschriftlicht.
- Stärken zeigen sich in der Nachvollziehbarkeit der Informationsquellen und der Beschreibung des Kontexts einer Entwicklungsmaßnahme.
- Schwächen weisen die Darstellung sowohl der Angemessenheit der Methodik als auch der Wirkungszusammenhänge auf.
- Die zusätzlich untersuchten Qualitätskriterien für GIZ und KfW zeigen Stärken in der Durchführung eines Vorher-Nachher-Vergleichs\* und Schwächen in der Anwendung von Kontroll- und Vergleichsgruppen\*.

Es wird empfohlen, methodische Qualitätsstandards, die für eine Organisation relevant sind, zu identifizieren und zu verschriftlichen sowie die Nachvollziehbarkeit ihrer Anwendung auf Evaluierungsebene sicherzustellen. Die Anwendung von Qualitätsstandards sollte generell weiter verbessert und regelmäßig von den Organisationen untersucht werden. Zudem sollten sich Evaluierungseinheiten/-stellen der deutschen EZ-Organisationen untereinander systematisch zur Anwendung der Qualitätsstandards austauschen.

## Abbildung 1 Ausgewählte Qualitätskriterien aus dem Bereich „Berichtslegung und Methoden“

Einbindung des Kontexts	Auswahlverfahren der Gesprächspartner beschrieben*
Darstellung der Wirkungszusammenhänge	Vorher-Nachher-Vergleich*
Nachvollziehbarkeit der Informationsquellen	Kontroll- und Vergleichsgruppe einbezogen*
Darstellung der Angemessenheit des methodischen Vorgehens	Methoden-Triangulation*
Kohärenz von Daten–Ergebnissen–Schlussfolgerungen	

Quelle: DEval, eigene Darstellung. Die Qualitätskriterien, die mit einem \* versehen sind, wurden zusätzlich für einen zeitlichen Vergleich der Ergebnisse zur Meta-Evaluierung Nachhaltigkeit (Noltze et al., 2018) für GIZ und KfW erhoben und lassen sich thematisch dem Bereich „Berichtslegung und Methoden“ zuordnen.

### Hintergrund und methodisches Vorgehen

Die in der Meta-Evaluierung analysierten Organisationen wurden anhand von mehreren Kriterien ausgewählt (zum Beispiel in Bezug zu Mittelzuwendungen und Evaluierungstätigkeit), um die strukturelle Heterogenität der Akteure abzudecken. Insgesamt wurden elf Organisationen<sup>1</sup> mit einer Grundgesamtheit von 576 (Projekt-)Evaluierungen aufgenommen, die BMZ-(mit-)finanziert waren und zwischen 2016 und 2020 in Deutschland (mit-)verantwortet wurden. Daraus wurde eine für jede Organisation repräsentative Stichprobe gezogen oder eine Vollerhebung durchgeführt. Insgesamt waren dies 296 Evaluierungen. Anschließend wurden aus inhaltlichen Überschneidungen der OECD-DAC- und DeGEval-Standards Qualitätskriterien abgeleitet und den drei Standardclustern „Berichtslegung und Methoden“, „Nutzbarkeit“ und „Partizipation, Unabhängigkeit und Fairness“ zugeordnet. Außerdem wurden für die Untersuchung von KfW und GIZ weitere Qualitätskriterien aus einer vorangegangenen Meta-Evaluierung des DEval aufgenommen, um einen zeitlichen Vergleich ihrer Anwendung zu ermöglichen (Noltze et al., 2018). Diese zusätzlichen Qualitätskriterien können dem Standardcluster „Berichtslegung und Methoden“ zugeordnet werden und sind mit der Markierung \* versehen, da sich die Erkenntnisse lediglich auf Evaluierungen dieser beiden Organisationen beziehen (Abbildung 1). Die Anwendung der Qualitätskriterien wurde anhand von Evaluierungs- und Organisationsdokumenten, sowie einer Online-Befragung untersucht. Die Qualitätskriterien im Standardcluster „Berichtslegung und Methoden“ konnten

ausschließlich anhand der Evaluierungsdokumente untersucht werden.

### Fokus „Berichtslegung und Methoden“

In der Meta-Evaluierung wurden insgesamt 37 Qualitätskriterien untersucht, unter anderem „Genderbalance“ und „Unabhängigkeit des Gutachter\*innen-Teams“, „Kapazitätsentwicklung in Partnerländern“, „Berücksichtigung von partnerschaftlichen Aktivitäten“ oder „Einbindung von Stakeholder\*innen“ im Evaluierungsprozess. Der Fokus in diesem Policy Brief liegt auf methodischen Qualitätskriterien, da ihnen eine besondere Relevanz für die Belastbarkeit und Nachvollziehbarkeit der Evaluierungsergebnisse zukommt. Evaluierungen werden unter einer Vielzahl von Restriktionen durchgeführt (beispielsweise in Bezug auf Zeit, Budget oder Datenverfügbarkeit). Daher kann sich die Auswahl der Evaluierungsmethoden nicht ausschließlich an der Wissenschaftlichkeit orientieren, sondern sollte auch die Effizienz und Durchführbarkeit der Evaluierung berücksichtigen. Diesbezügliche Entscheidungen sollten gemäß den Qualitätsstandards im Evaluierungsbericht transparent dargestellt und ihre Auswirkungen auf die Evaluierung erläutert werden.

Im Folgenden werden einzelne Qualitätskriterien des Standardclusters „Berichtslegung und Methoden“ vorgestellt, zu denen die Meta-Evaluierung Beispiele guter Praxis in einzelnen Organisationen identifiziert hat.

<sup>1</sup> 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), Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), Heinrich-Böll-Stiftung (hbs), Konrad-Adenauer-Stiftung (KAS), KfW Entwicklungsbank (KfW), MISEREOR und Physikalisch-Technische Bundesanstalt (PTB)

## Einbindung des Kontexts

---

Der Kontext der EZ-Maßnahme wird in 87 % der Evaluierungen ausreichend oder ausführlich beschrieben. Die Berücksichtigung des Kontexts bei den Ergebnissen der Evaluierungen ist jedoch in 44 % der Fälle nicht ausreichend. So werden hemmende oder förderliche Kontextfaktoren und ihr spezifischer Einfluss auf die Ergebnisse der Maßnahme oftmals nicht systematisch dargestellt. Organisationen erfüllen diese Qualitätskriterien besonders gut, wenn im Rahmen einer standardisierten Struktur des Evaluierungsberichts ein Abschnitt für die Beschreibung des Kontexts sowohl bei der Einführung des Evaluierungsgegenstands als auch bei der Beschreibung der Ergebnisse vorgesehen ist.

## Darstellung der Wirkungszusammenhänge

---

Die meisten Evaluierungen stellen die Wirkungsmodelle, die den Maßnahmen zugrunde liegen, dar, in 65 % der Evaluierungen sind diese allerdings lückenhaft. Das heißt, dass Informationen über spezifische Annahmen und Wirkungsbeziehungen zwischen den Ergebnisebenen (*Input, Output, Outcome, Impact*) fehlen, die wiederum essenziell für eine Bewertung des Erfolgs von Maßnahmen sein können. Besonders gut wird dieses Qualitätskriterium in Evaluierungen erfüllt, in denen in einem gesonderten Abschnitt die Wirkungslogik anhand einer Grafik und einer dazugehörigen textlichen Erläuterung der Wirkzusammenhänge dargestellt wird.

## Nachvollziehbarkeit der Informationsquellen

---

In 92 % der Evaluierungen werden Informationsquellen für Analysen einer Evaluierung mindestens hinreichend oder umfassend beschrieben. In den anderen 8 % wird nicht klar, welche Quellen für welche Analysen herangezogen werden. Tabellen, in denen die Quellen nach Evaluierungsfragen und zugehörigen Analysen genannt werden, bieten einen schnellen Überblick.

## Darstellung der Angemessenheit des methodischen Vorgehens

---

In 33 % der Evaluierungen werden weder Begründungen für die Auswahl der angewandten Methoden noch deren Limitationen genannt. In 40 % wird einer der beiden Aspekte beschrieben und in 27 % der Evaluierungen wird beides mindestens hinreichend (13 %) oder umfassend (14 %) begründet. Eine Begründung der Auswahl der Methoden und ihrer Limitationen ist wichtig, um das Vorgehen nachvollziehen und die Belastbarkeit der Ergebnisse

einschätzen zu können. Wenn etwa eine geplante Felderhebung aufgrund einer Konflikteskalation nicht stattfinden konnte und stattdessen eine Dokumentenstudie zur Einschätzung der Projekterfolge durchgeführt wurde, ist eine diesbezügliche Erläuterung der damit einhergehenden Limitationen essenziell. Beispielgebend sind Evaluierungen, in denen in spezifischen Abschnitten die Auswahl und Limitationen der verwendeten Methoden begründet werden.

## Kohärenz von Daten–Ergebnissen–Schlussfolgerungen

---

In 63 % der Berichte basieren mehr, in 37 % weniger als die Hälfte der Schlussfolgerungen einer Evaluierung auf beschriebenen Erhebungen und Analysen. Insbesondere diejenigen, die lediglich den Ergebnisteil lesen, können die Belastbarkeit von Aussagen der Evaluierung dadurch schlecht einschätzen oder nachvollziehen. Gute Praxis ist es hingegen, bei der Verschriftlichung der Schlussfolgerungen durch Referenzen (zum Beispiel durch Fußnoten) direkten Bezug zu den jeweiligen Ergebnissen zu nehmen.

## Auswahlverfahren beschrieben\*

---

In 43 % der Evaluierungen der GIZ und KfW werden die Kriterien beschrieben, anhand derer beispielsweise Personen für ein Interview oder Dokumente für eine Auswertung ausgewählt werden. In den verbleibenden 57 % wird zwar beschrieben, dass zum Beispiel eine bestimmte Anzahl an Personen in drei von sieben Dörfern der Projektregion befragt werden, aber nicht, warum diese Anzahl, diese Personen und diese Dörfer ausgewählt werden. Der Zugang zur Projektregion oder eine Zufallsstichprobe könnten solche Kriterien darstellen.

## Vorher-Nachher-Vergleich\*

---

81 % der Evaluierungen von GIZ und KfW ziehen einen Vorher-Nachher-Vergleich relevanter Indikatoren heran, um die Wirkungen der Maßnahme über einen bestimmten Zeitraum zu überprüfen. Insbesondere bei langjährigen, strukturellen Maßnahmen, wie etwa dem Aufbau einer Gesundheitsversorgung in einer ganzen Region, werden häufig externe Makroindikatoren, wie die Kindersterblichkeitsrate, für den Vergleich herangezogen. Dieser Vergleich ermöglicht es, einen Eindruck von der Entwicklung der Region zu gewinnen, aber nicht, welchen Beitrag die Maßnahme dazu geleistet hat. Eine Plausibilisierung des Einflusses eigener beziehungsweise Maßnahmen anderer Geber ist der Nachvollziehbarkeit und Belastbarkeit der Ergebnisse förderlich.

## Kontroll- und Vergleichsgruppen\*

Bei 18 % der Evaluierungen von GIZ und KfW werden Untersuchungen ähnlicher Maßnahmen zur Einschätzung der Wirkung der betrachteten Maßnahme herangezogen. Dieser Vergleich wäre besser nachzuvollziehen, wenn weitere Details der Maßnahmen, beispielsweise hinsichtlich der finanziellen und personellen Mittel, dargestellt wären. Die Verwendung einer randomisierten Kontrollgruppe produziert robustere Evaluationsergebnisse als nichtrandomisierte Vergleiche. In Fällen, in denen keine randomisierte Kontrollgruppe herangezogen werden kann, muss allerdings nicht gänzlich auf ein experimentelles Design verzichtet werden. Auch nichtrandomisierte Befragungen von einer geringen Anzahl an Personen, die eine Maßnahme nicht erhalten haben, können Erkenntnisse über die Wirkung der Maßnahme liefern. Dabei sollten die Einschränkungen dieses Vorgehens transparent dargelegt werden.

## Methoden-Triangulation\*

Fast alle Evaluierungen von GIZ und KfW wenden im Evaluierungsprozess mehrere Methoden an. Allerdings verwenden lediglich 63 % der Evaluierungen zwei oder mehr Methoden, um eine

## Literaturhinweise

**BMZ (2021)**, *Evaluierung der Entwicklungszusammenarbeit: Leitlinien des BMZ*, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn/Berlin.

**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.

**Noltze, M., M. Euler und I. Verspohl (2018)**, *Meta-Evaluierung von Nachhaltigkeit in der deutschen Entwicklungszusammenarbeit*, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), Bonn.

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 Ergebnisse der Entwicklungszusammenarbeit transparenter zu machen.

spezifische Aussage zu belegen. Ein expliziter Hinweis darauf, wie Triangulation in einer Evaluierung verstanden und angewendet wird, kann die Einschätzung der Belastbarkeit der Ergebnisse erleichtern.

## Fazit und Empfehlungen

Die Meta-Evaluierung kommt zu dem Schluss, dass die dargestellten OECD-DAC- und DeGEval-Qualitätsstandards im Standardcluster „Berichtslegung und Methoden“ durchschnittlich größtenteils erfüllt werden, wobei sich insbesondere Defizite in der „Darstellung der Wirkungszusammenhänge“ einer Entwicklungsmaßnahme und der „Darstellung der Angemessenheit des methodischen Vorgehens“ zeigen. Wenn Qualitätsstandards nicht erfüllt werden, wird dies in den Evaluierungsdokumenten selten nachvollziehbar begründet. Daher gehört die systematische schriftliche Verankerung der Qualitätsstandards in den Organisationsdokumenten sowie die Sicherstellung ihrer Anwendung und Nachvollziehbarkeit auf der Ebene einzelner Evaluierungen zu den zentralen Empfehlungen. Weiterhin wird der Austausch zur Umsetzung der Qualitätsstandards zwischen den Evaluationseinheiten/-stellen der deutschen EZ-Organisationen empfohlen. Die vorgestellten Beispiele guter Praxis bieten hierfür erste Anhaltspunkte.



**Marian Wittenberg**  
Evaluators



**Dr. Kerstin Guffler**  
Senior-Evaluators,  
Teamleitung



**Amélie zu Eulenburg**  
Abteilungsleiterin