

# Bericht

für das Bischöfliche Hilfswerk MISEREOR e.V.

Arbeitsbereich Evaluierung und Qualitätsmanagement (EQM)



## Meta-Evaluierung der Projektevaluierungen aus den Jahren 2014 – 2015

Eval.-Nr. 2089-Z1031-1089

29. September 2016

Durchgeführt von

**.Evaluation.Consulting**  
Jens Koy

Wehrhausweg 33, 53227 Bonn  
koy@koy-evaluation.de  
www.koy-evaluation.de

**Centrum für Evaluation (CEval)**

Im Stadtwald, 66123 Saarbrücken  
www.ceval.de

Dr. Susanne Johanna Väth

s.vaeth@ceval.de

Dr. Cornelia Römling

c.roemling@ceval.de

# Inhalt

Abkürzungsverzeichnis .....	3
Executive Summary .....	4
1 Einleitung.....	9
2 Ziele der Evaluation .....	10
3 Untersuchungsgegenstand – ELBE und EQM-Evaluierungen .....	11
4 Methodisches Vorgehen und Grenzen der Meta-Evaluierung .....	12
4.1 Methodisches Vorgehen.....	12
4.2 Grenzen der Meta-Evaluierung .....	16
5 Ergebnisse .....	18
5.1 Vollständigkeit und Angemessenheit der Dokumentation .....	18
5.2 Unabhängigkeit und Überparteilichkeit .....	24
5.3 Glaubwürdigkeit .....	28
5.4 Nützlichkeit.....	44
5.5 Konsistenz der TOR.....	47
5.6 Partizipation / Beteiligung .....	50
6 Schlussfolgerungen und Empfehlungen.....	53
6.1 Zusammenfassende Bewertung .....	53
6.2 Einschätzung nach Untersuchungskriterien .....	55
Literatur und Quellen .....	67
Anhang.....	68
Anlage 1: TOR der Meta-Evaluierung.....	69
Anlage 2: Analyseraster - Qualitäts-Basischeck (Stufe 1).....	75
Anlage 3: Die 70 ausgewählten Evaluierungsberichte des Qualitäts-Basischecks..... (separater Anlagenteil wg. Datenschutz)	78
Anlage 4: Analyseraster der vertieften Analyse (Stufe 2) .....	83
Anlage 5: Die 20 ausgewählten Evaluierungsberichte der vertieften Analyse (Stufe 2) .....	97
(separater Anlagenteil wg. Datenschutz)	
Anlage 6: Ergebnisse der vertieften Analysen (separater Anlagenteil wg. Datenschutz) ...	99
Anlage 7: Gute Praxis Beispiel: Methodenkapitel (Nr.301, Mexiko, ELBE) .....	100
Anlage 8: Kurzprofil der GutachterInnen der Metaevaluierung .....	103

**Quellen der Fotos der Titelseite:** von links nach rechts, Titelseiten der untersuchten Evaluierungen Nr. 195 (Uganda), Nr. 86 (Burkina Faso) und aus Annex Nr. 238 (Indien)

## Abkürzungsverzeichnis

<b>BMZ</b>	Bundesministerium für Wirtschaftliche Zusammenarbeit und Entwicklung
<b>DAC</b>	Development Assistance Committee
<b>DeGEval</b>	Gesellschaft für Evaluation (Österreich und Deutschland)
<b>ELBE</b>	Externe, lokal beauftragte Evaluierung
<b>EQM</b>	Arbeitsbereich Evaluierung und Qualitätsmanagement
<b>InPro</b>	Inhaltliche Projektbegleitung (in Kontinentalabteilungen von MISEREOR)
<b>MPV</b>	MISEREOR Projektverwaltung (Datenbank)
<b>SWOT</b>	Strengths, Weaknesses, Opportunities and Threats (Analysetool in dem Stärken, Schwächen, Chancen und Bedrohungen z.B. einer Gruppe untersucht werden)
<b>TOR</b>	Terms of Reference / Referenzrahmen

## Executive Summary

Das Bischöfliche Hilfswerk MISEREOR unterstützt weltweit lokale Partnerorganisationen in der Durchführung von Projekten und Programmen, die zur nachhaltigen Entwicklung und zur Armutsbekämpfung beitragen sollen. Für diese Arbeit stellt das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) Mittel zur Verfügung. In 2015 waren dies 147,59 Mio. €. Neben staatlichen Mitteln setzt MISEREOR auch Spendenmittel und Kirchensteuern für die Projektarbeit ein. Insgesamt wurden 2015 178,75 Mio. € in die Projektförderung investiert. Für die ziel- und wirkungsorientierte Erfolgskontrolle sind zwischen BMZ und der Katholischen Zentralstelle für Entwicklung (KZE), vertreten durch MISEREOR, Leitlinien vereinbart, die u.a. regelmäßige Einzelprojektevaluierungen und Meta-Evaluierungen vorsehen. Für die hier vorliegende Meta-Evaluierung bildeten 261 Evaluierungen aus den Jahren 2014/15 die Grundgesamtheit. Die Untersuchung wurde im Auftrag von MISEREOR durch die externen Beratungsbüros *.Evaluation.Consulting*, Bonn und *Centrum für Evaluation (CEval)*, Saarbrücken durchgeführt.

Ziel der Meta-Evaluierung ist, dass MISEREOR über aussagekräftige Rückmeldungen zu der Qualität sowohl der von Partnern beauftragten Evaluierungen als auch der selbst beauftragten Evaluierungen verfügt. Das methodische Vorgehen umfasste einen repräsentativen Qualitäts-Basischeck von 70 der 261 Evaluierungen. Dabei wurden 14 Evaluierungen des MISEREOR-Arbeitsbereichs *Evaluierung und Qualitätsmanagement (EQM)* und 56 von Partnerorganisationen beauftragte sog. *extern lokal beauftragte Evaluierungen (ELBE)* untersucht. Untersuchungskriterien waren *Vollständigkeit und Angemessenheit der Dokumentation, Unabhängigkeit und Überparteilichkeit, Glaubwürdigkeit und Nützlichkeit*. Die Prüfung erfolgte auf der Grundlage der Lektüre von Zusammenfassung, Einleitung, Methodenkapitel, Schlussfolgerungen und Empfehlungen sowie einer raschen Durchsicht des Berichts. In einem zweiten Schritt wurden 20 Evaluierungen einer vertieften Analyse unterzogen. Dabei wurden auch die Kriterien *Partizipation* und *Konsistenz der TOR* untersucht. Bei den Untersuchungen wurde jeweils der ganze Bericht mit Anlagen gelesen.

In der zusammenfassenden Qualitätsbewertung der Berichte wird insgesamt eine befriedigende Qualität festgestellt. Die Bewertung umfasst die Ergebnisse des Qualitäts-Basischecks und der vertieften Analyse.

**Unterschied ELBE und EQM-Evaluierungen.** EQM-Evaluierungsberichte wurden in der Regel besser eingestuft als ELBE. Die Erklärungen dafür liegen aus Sicht des Meta-Evaluierungsteams in folgenden Aspekten:

- a) Direkte Qualitätskontrolle der Berichte durch Evaluierungsfachleute von EQM, höheres methodisches Know-How/ Evaluierungs-Know-How
- b) Mehr Arbeitstage und mehr Finanzmitteleinsatz für die GutachterInnen pro EQM-Evaluierung im Vergleich zu ELBE; Einhaltung des Vieraugenprinzips (i.d.R. immer 2 GutachterInnen bei EQM-Evaluierungen, bei ELBE oft nur eine GutachterIn)
- c) Mehr Distanz der GutachterInnen zum Projekt sowie die EQM-Unterstützung bei der Äußerung von Kritik am Projekt

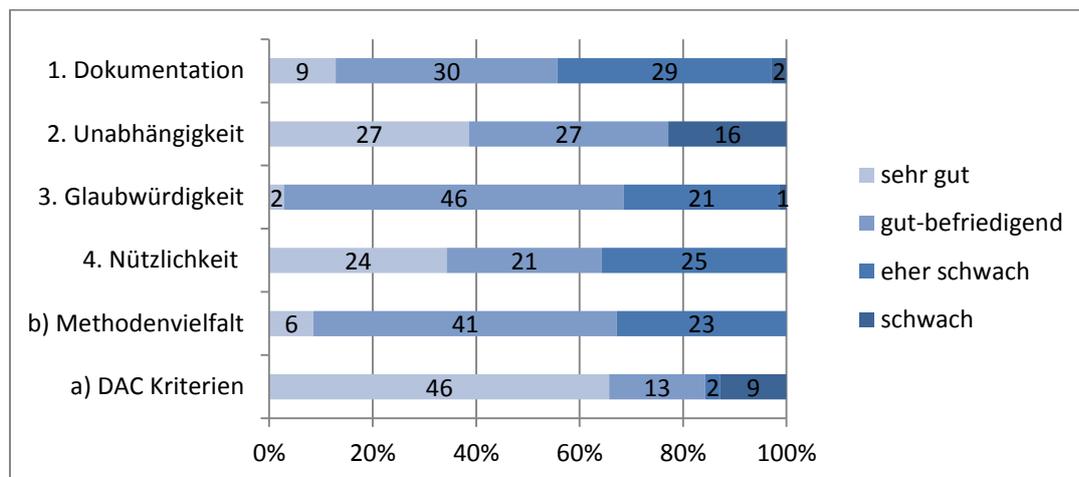
Im Bereich Partizipation welche nur in Stufe 2 untersucht wurde, lässt sich von der Tendenz ein vergleichbares und zum Teil besseres Abschneiden der ELBE finden. Dieses Ergebnis ist

nicht repräsentativ, zeigt jedoch, dass mögliche Vorteile in der Einbeziehung der Stakeholder in die Evaluierungen liegen.

### Ergebnisse und Schlussfolgerungen im Einzelnen

**Ergebnisse des Qualitäts-Basischecks:** 60-70% der Evaluierungsberichte befinden sich je Kriterium im „grünen“ Bereich, d.h. zwischen den Bewertungen sehr gut bis befriedigend. Der andere Teil von 30-40% wurde als eher schwach bis schwach eingestuft (siehe Abbildung).

Abbildung 1 Übersicht der Ergebnisse des Basischecks



### Dokumentation

<b>Bewertung der Vollständigkeit und Angemessenheit der Dokumentation:</b>	<b><u>Befriedigend bis eher schwach</u></b>
--	---

Anforderungen an die Dokumentation sind nicht ausreichend klar und meist zu komprimiert dargestellt, d.h. es sind nicht ausreichend praktische Umsetzungshilfen vorhanden. Zentraler Schwachpunkt fast aller Dokumentationen ist die praktisch nicht vorhandene Darstellung der Auswahl der Stichprobe und die unvollständige Dokumentation im Annex. Das führte in der Bewertung zum Zusatz „bis eher schwach“. Gliederungen sind in der Regel angemessen, jedoch fehlen in den schwächeren Evaluierungen wichtige Kapitel wie Zusammenfassung und Methodenkapitel. Feste Vorgaben sind erforderlich und würden Abhilfe schaffen.

### Unabhängigkeit und Überparteilichkeit

<b>Bewertung: Hinweise in den Berichten auf die Unabhängigkeit und Überparteilichkeit der EvaluatorInnen</b>	<b><u>gut bis befriedigend</u></b>
--	------------------------------------

Die Einschätzung der Unabhängigkeit und Glaubwürdigkeit ist über die reine Analyse von Evaluationsberichten nicht abschließend überprüfbar. Im vorliegenden Fall deuten ca. 3/4

(73%) der Berichte darauf hin, dass die GutachterInnen extern waren. Es ist davon auszugehen, dass diese Zahl höher ist, da es keine spezifischen Vorgaben durch MISEREOR oder auch durch maßgebliche internationale Institutionen wie OECD-DAC gibt, die die Dokumentation der Unabhängigkeit und Qualifikation der EvaluatorInnen im Bericht explizit einfordern. Daher wird davon ausgegangen, dass mehr externe EvaluatorInnen und GutachterInnen zum Einsatz kommen, als in den Berichten ausgewiesen wird.

### **Glaubwürdigkeit**

<b>Bewertung: Glaubwürdigkeit der Berichte</b>	<b><u>Befriedigend</u></b>
--	----------------------------

Ein wichtiger Beitrag zur Glaubwürdigkeit einer Evaluierung ist die Darlegung der Untersuchungsmethoden und die Information zu den Befragten. Gerade die Auswahlkriterien werden fast nie dargelegt und stellen somit die Glaubwürdigkeit in Frage. Eine Darlegung der Auswahlkriterien würde zudem eine bewusstere Auswahl unterstützen, da es einen Rechtfertigungsdruck gibt.

In der Auswertung der oft umfassenden Daten gibt es noch einige Verbesserungsmöglichkeiten. Sowohl die Quantifizierung der Aussagen als auch die Differenzierung der Zielgruppen findet kaum statt, obwohl beides explizit in den Handreichungen gefordert ist. Der Überprüfung auf Wirkungen wird sehr viel Raum in den Evaluierungen eingeräumt, jedoch ist fraglich, ob mit den angewandten Methoden sichergestellt werden kann, dass die identifizierten Wirkungen wirklich auf das Projekt zurückzuführen sind.

Weiterhin fällt auf, dass wichtige Bereiche des Projektmanagements oder der Untersuchung der konkreten inhaltlichen Ausgestaltung von Arbeitsabläufen in der Evaluierung nicht oder i.d.R. nur relativ kurz angesprochen werden. Eine Ursache dafür kann in der Fokussierung auf die Wirkungserfassung und Zielerreichung liegen. Möglicherweise führt die Praxis, das Ergebnis-Kapitel mit den DAC-Bewertungen zu kombinieren dazu, dass in vielen Fällen die Beschreibung der Situation anhand der erhobenen Daten ganz unterbleibt und sofort die Bewertung dargelegt wird. Die Kapitel heißen zwar oft „Ergebnisse und Analysen“, die puren Ergebnisse der Datenerhebung werden dort aber selten aufgezeigt.

### **Nützlichkeit**

<b>Bewertung: Nützlichkeit der Berichte</b>	<b><u>befriedigend</u></b>
---	----------------------------

Die Evaluierungsfragen werden in der Regel gut abgedeckt, jedoch wird durch die zum Teil große Anzahl und den sehr detaillierten Fragen der Überblick verloren und es fehlen Gesamteinschätzungen. Zum Teil werden so in der Beantwortung der DAC-Kriterien gewissenhaft alle Fragen zu dem Themenbereich beantwortet, ohne eine generelle Einschätzung zu geben wie relevant das Projekt einzuschätzen ist. Außerdem werden so auch nur die in den Fragen beschriebenen Themen in der Evaluierung eingebracht. Für die Vertiefung von zentralen Fragestellungen, die sich in der Feldphase ergeben können, fehlt dann oft der Raum. Die Erfahrung zeigt, dass die Nützlichkeit der Empfehlungen sehr stark verbessert werden kann, wenn für diese eine strukturierte Umsetzungsplanung vorgenommen wird. Dieser wichtige erste Schritt der Umsetzung von Empfehlungen wird

jedoch bei der ELBE-Handreichung nicht berücksichtigt. Bei der Umsetzung von Evaluierungsempfehlungen ist die Erfahrung der vergangenen Jahre zu berücksichtigen, dass Evaluierungen teilweise stark die Komplexität von Abläufen und Verfahren erhöhten und damit eine gewisse hemmende „Bürokratie“ unterstützten. Dies sollte bei der Umsetzung von neuen Verfahren wie z.B. Monitoring-Systemen überlegt und abgewogen werden: Welche Daten sind notwendig? Inwiefern führen auch einfachere Verfahren zu einem vergleichbaren Ergebnis?

Eine deutliche Trennung von Beschreibungen und Bewertungen konnte immer dann vorgenommen werden, wenn zunächst eine Beschreibung des Sachverhalts, möglichst mit Bezug auf die Quellen und dann eine Bewertung ggf. mit Begründung bzw. weiterer Erläuterung vorgenommen wurde. Es entspricht auch guter Praxis, dass Ergebnisse und Bewertungen auch in einem Kapitel zusammengefasst werden können, wenn die Inhalte deutlich voneinander getrennt sind.

### Partizipation

<b>Bewertung: Hinweise in den Berichten auf Partizipation von Partnerorganisation und Zielgruppen</b>	<b><u>gut bis befriedigend</u></b> <u>Partnerorganisation: gut; Zielgruppen: befriedigend</u>
---	--

Die explizit durch das MISEREOR-Evaluierungskonzept geförderte Partizipation von Partnerorganisationen und geförderten Personen steht im Spannungsfeld mit der Unabhängigkeit der Evaluierung. Im MISEREOR-Evaluierungskonzept wird diese Herausforderung nur indirekt thematisiert, indem der Anregung aus der Fachdiskussion gefolgt wird, dass Schlussfolgerungen und Empfehlungen sich aus den Ergebnissen nachvollziehbar im Evaluierungsbericht ergeben sollten.

MISEREOR und seine Partnerorganisationen haben teilweise direkten Zugang zu den Zielgruppen und können diesen noch weiter ausbauen und damit die Unterstützung für Benachteiligte noch effektiver machen.

### Konsistenz der TOR

<b>Bewertung: Konsistenz der TOR</b>	<b><u>befriedigend</u></b>
--------------------------------------	----------------------------

Die Aufteilung von Wirkungen auf der Outcome-Ebene in die Kapitel *Wirkungen* und *Effektivität* sollte aufgelöst werden und die Abgrenzung der Fragestellungen zwischen den Kapiteln *Relevanz* und *Wirkungen* klarer herausgearbeitet werden.

Die Erfahrung zeigt, dass die Entwicklung guter TOR Zeit und Know-how benötigt, um die Evaluierung wirklich auf die entscheidenden Fragen zu fokussieren. Daher wird es als angemessen angesehen, wenn Partnerorganisationen bei ELBE wenige zentrale Fragestellungen entwickeln und dann die Standardfragen der TOR einsetzen.

Gute TOR sind wichtig. Noch wichtiger ist für die Erstellung einer guten ELBE die Auswahl einer guten EvaluatorIn. Dazu ist es wichtig, dass die Vorgaben für das Evaluierungsbudget angemessen sind und die Partnerorganisation Zugang zu Netzwerken erfahrener EvaluatorInnen hat.

## **Die wichtigsten Empfehlungen**

Die zentralen Empfehlungen sind hier nach Adressaten und nach Priorität geordnet und in Kurzform dargestellt. Im Kapitel *Schlussfolgerungen und Empfehlungen* werden Erläuterungen und Ansatzpunkte für die Umsetzung benannt und alle 19 Empfehlungen ausgeführt.

### **An die Kontinentalabteilungen und EQM**

- 1)** Weitere Integration des Werkzeugs „Evaluierung“ in den Partnerdialog. Evaluierungen dienen der Rechenschaftslegung und des Lernens. Für die Stärkung des Lernens bedarf es des Vertrauens der Partnerorganisation in MISEREOR, dass auch gewisse Schwachpunkte nicht zum Abbruch der Förderungen führen (Empfehlung 2).
- 2)** Gesamtkonzept der ELBE und EQM-Evaluierungen überprüfen und ggf. unterschiedliche Schwerpunktsetzungen und inhaltliche Ausgestaltungen für ELBE und EQM-Evaluierungen entwickeln (Empfehlung 8).
- 3)** Wirkungen nicht methodisch aufwendiger prüfen, jedoch Methoden anwenden um diese plausibel auf das Projekt zurückzuführen. Zum Beispiel über verstärkte Diskussion/Hinterfragung der Wirkungslogik und stärkerer Triangulation der Informationen zu möglichen Wirkungen (Empfehlung 5).
- 4)** Zusätzlich zu den Wirkungen auch andere zentrale Faktoren des Projekterfolgs in den Blick der Evaluierungen nehmen. Zum Beispiel mehr auf konkrete Arbeitsabläufe, Projektmanagements (nicht nur Monitoring!) oder inhaltliche Ausgestaltung des Projektes untersuchen (Empfehlung 7).
- 5)** ELBE benötigen eine angemessene finanzielle Ausstattung, um qualifizierte Evaluierungsteams mit ausreichend Arbeitstagen unter Vertrag nehmen zu können (8000 - 15.000 €) (Empfehlung 18).
- 6)** Unterstützung von Partnerorganisationen bei der Suche nach geeigneten EvaluatorInnen für ELBE. Das könnte über EQM oder über BeraterInnen (BAZ) erfolgen (Empfehlung 19).

### **Weitere zentrale Empfehlungen für EQM:**

- 7)** Darstellungen der Wirkungen des Projekts in einem Kapitel „Outcome & Effektivität“ sowie „Impact“ oder ganz klassisch nach DAC in den Kriterien „Effektivität“ und „Impact“. Klarere Abgrenzung der Fragestellungen des Kapitel Relevanz zu denen der Kapitel Wirkungen und Effektivität. Darstellung der DAC-Kriterien in internationaler Reihenfolge, da chronologisch sinnvoll. Alternativ, Bewertung der DAC-Kriterien in einem separaten Kapitel und Untergliederung der Evaluierungsergebnisse beispielsweise auf Aktivitätenebene (Empfehlung 15).
- 8)** MISEREOR sollte die Anforderungen und die verbindlichen Inhalte für die Dokumentation in Evaluierungen klarer benennen (Empfehlung 1).
- 9)** Die Vorgaben für ELBE und EQM-Evaluierungen sollten die Wichtigkeit der Auswahl der Stichproben und deren Dokumentation verdeutlichen und dafür klare Vorgaben entwickeln (Empfehlung 3).

# 1 Einleitung

Das Bischöfliche Hilfswerk MISEREOR unterstützt weltweit lokale Partnerorganisationen in der Durchführung von Projekten und Programmen, die zur nachhaltigen Entwicklung und zur Armutsbekämpfung beitragen sollen. Für diese Arbeit stellt das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) Mittel zur Verfügung. In 2015 waren dies 147,59 Mio. €. Neben staatlichen Mitteln setzt MISEREOR auch Spendenmittel und Kirchensteuern für die Projektarbeit ein. Insgesamt wurden 2015 178,75 Mio. € in die Projektförderung investiert.<sup>1</sup>

Für die ziel- und wirkungsorientierte Erfolgskontrolle sind zwischen BMZ und der Katholischen Zentralstelle für Entwicklung (KZE) - vertreten durch MISEREOR - Leitlinien vereinbart, die u.a. regelmäßige Einzelprojektevaluierungen und Meta-Evaluierungen vorsehen. Dieses System der Evaluierung von Projekten und der Überprüfung der Qualität der Evaluierungen durch Meta-Evaluierungen nutzt MISEREOR auch für Projekte, die durch Spenden oder Kirchensteuern finanziert werden. Die Evaluierungen aus 2014/15, die innerhalb des Gesamtportfolios an Projektförderungen entstanden, bildeten mit 261 Evaluierungen die Grundgesamtheit der hier vorgelegten Meta-Evaluierung.

Für die Meta-Evaluierung, deren Vorgehen und Ergebnisse hier im Bericht dokumentiert werden, hat MISEREOR im Mai 2016 die externen Beratungsbüros *.Evaluation.Consulting*, Bonn und *Centrum für Evaluation (CEval)*, Saarbrücken beauftragt. Teamleiter Jens Koy brachte neben seiner über 10-jährigen Erfahrung im Bereich Evaluierung vor allem die Kenntnis von kirchlicher Entwicklungszusammenarbeit und die Vertrautheit mit den Herausforderungen von NRO-Arbeit bei Partnerorganisationen ein. Herr Koy war u.a. 5 Jahre Referent für Evaluierung beim Evangelischen Entwicklungsdienst (EED). Die Evaluatorinnen des Centrum für Evaluation, Dr. Susanne Johanna Väth und Dr. Cornelia Römling brachten Evaluierungsexpertise, methodisches Fachwissen und umfangreiche Erfahrungen im Bereich von Meta-Evaluierungen in das Team ein. Am CEval werden seit 2002 Evaluationen in verschiedenen Bereichen des zivilgesellschaftlichen und staatlichen Handelns durchgeführt, es berät öffentliche und private Organisationen bei der Entwicklung von Monitoring-Systemen und Evaluationskonzepten und bietet akademische und nicht-akademische Aus- und Weiterbildungsprogramme im Themenbereich Monitoring und Evaluation (kurz: M&E) an.

MISEREOR blickt auf eine lange Evaluierungspraxis zurück. 1968 wurden die ersten Gutachten durch eine unabhängige Evaluierungsgruppe in Auftrag gegeben.<sup>2</sup> Seither hat sich die Evaluierungspraxis stetig weiterentwickelt. Seit 2010 nutzt MISEREOR zwei Verfahren, um Einzelprojekte zu evaluieren. Dabei handelt es sich um sogenannte *externe, lokal beauftragte Evaluierungen* (ELBE), die von den jeweiligen Partnerorganisationen beauftragt, und um Evaluierungen die direkt durch den in der MISEREOR-Geschäftsstelle in Aachen angesiedelten *Arbeitsbereich Evaluierung und Qualitätsmanagement* (EQM) vergeben werden. Nach der Einführung der ELBE in 2010 wurde in 2013/14 eine erste Meta-Evaluierung durch ein externes Berliner Consulting durchgeführt, die zum einen die administrativen Prozesse von MISEREOR zur Systematisierung von Projektevaluierungen

---

<sup>1</sup> MISEREOR Jahresbericht (MISEREOR 2015a: 45)

<sup>2</sup> MISEREOR Jahresevaluierungsbericht 2015 (MISEREOR 2015b: 6)

untersuchte sowie eine Prüfung der Qualität von 20 ELBE Berichten vornahm.<sup>3</sup> MISEREOR konzentrierte sich bei der Umsetzung der Evaluierungs-Empfehlungen auf die administrativen Verbesserungsvorschläge in der MISEREOR-Geschäftsstelle in Aachen.<sup>4</sup>

Die Meta-Evaluierung 2013/14 zeigte: MISEREOR steht mit seinen rund 2.500 Partnerorganisationen vor einer Herkules-Aufgabe bei der Umsetzung eines systematischen, partizipativen und methodisch angemessenen Evaluierungsverfahrens. In 2015 wurden insgesamt 1.093 Projekte, darunter 461 mit staatlichen Mitteln geförderte Vorhaben in Lateinamerika/Karibik, Afrika/Naher Osten und Asien/Ozeanien bewilligt.<sup>5</sup> Darin setzt sich die Entwicklung der Projektbewilligungen der vergangenen Jahre fort.

Dieser Bericht gliedert sich wie folgt: Zunächst werden die Ziele der Evaluation kurz ausgeführt (Kapitel 2) und eine Einführung in den Evaluationsgegenstand (Kapitel 3) gegeben. In einem weiteren Schritt werden die methodische Vorgehensweise (Kapitel 4) und die Grenzen der Meta-Evaluierungen dargelegt. Dem schließt sich die Ergebnisse der Analysen an (Kapitel 5). Der Bericht endet mit Schlussfolgerungen und daraus abgeleiteten Empfehlungen (Kapitel 6).

## 2 Ziele der Evaluation

Ziel der Meta-Evaluierung ist, dass MISEREOR über aussagekräftige Rückmeldungen zu der Qualität sowohl der von Partnern beauftragten Evaluierungen als auch der selbst beauftragten Evaluierungen verfügt. Diese Rückmeldungen können bei Bedarf zusätzliche Anstrengungen im Bereich der Qualitätssicherung initiieren.

Auf Wunsch von MISEREOR ist die Untersuchung der methodischen Qualität der Evaluierungen der Schwerpunkt dieser Meta-Evaluierung, um damit die Resultate der intensivierten Wirkungsbeobachtung bei den von MISEREOR geförderten Projekten der vergangenen 10 Jahre zu untersuchen. Gleichfalls bestand der Wunsch, die Untersuchung aus Effizienzgründen auf die Qualität der Berichte zu fokussieren und keine zusätzlichen Interviews mit den für die Projektarbeit zuständigen KollegInnen in den MISEREOR-Kontinentalabteilungen oder mit Partnerorganisationen zu führen. Dieses Vorgehen entspricht der weit verbreiteten Praxis von Meta-Evaluierungen.<sup>6</sup>

Als Zielgerüst wurden von Seiten MISEREORs Evaluationsfragen zu verschiedenen Themenbereichen im Referenzrahmen verankert, die in der Meta-Evaluierung Aufschluss über die Qualität der einzelnen Berichte in den 6 folgenden Bereichen geben sollen:

- a) Vollständigkeit und Angemessenheit der Dokumentation
- b) Unabhängigkeit und Überparteilichkeit
- c) Glaubwürdigkeit
- d) Nützlichkeit

---

<sup>3</sup> Siehe Meta-Evaluierungsbericht der ELBE aus 2014 (MISEREOR 2014)

<sup>4</sup> Information von EQM

<sup>5</sup> MISEREOR Jahresbericht (MISEREOR 2015a: 30f) und <https://www.misereor.de/ueber-uns/staatliche-foerderung/>; Zugriff 14.09.2016

<sup>6</sup> Siehe Vortrag von Prof. Dr. Alexandra Caspari am Deutschen Evaluierungsinstitut der Entwicklungszusammenarbeit am 06. April 2016 in Bonn: Meta-Evaluationen – die Evaluation von Evaluationen

- e) Partizipation/Beteiligung
- f) Konsistenz der TOR

Diesen Bereichen ordnete MISEREOR 47 Evaluierungsfragen zu, die als Grundlage zur Entwicklung der Analyseraster dienen und anhand derer die Evaluierungen bewertet werden.

### 3 Untersuchungsgegenstand – ELBE und EQM-Evaluierungen

**ELBE.** *externe, lokal beauftragte Evaluierungen* (ELBE) werden zwischen der Partnerorganisation und MISEREOR verpflichtend im Projektvertrag festgehalten. Für ELBE entwickeln die Partnerorganisationen und bei Bedarf in Abstimmung mit den zuständigen MISEREOR-MitarbeiterInnen, den InPros (Inhaltliche Projektbegleitung) die Terms of Reference (TOR) der Evaluierung. Im Anschluss daran wählen die Partnerorganisationen die GutachterInnen selbst aus, begleiten den Evaluierungsprozess, nehmen den Evaluierungsbericht am Ende ab und leiten diesen an die InPros weiter.<sup>7</sup> Das Verfahren wurde 2010 bei MISEREOR eingeführt. In den ersten dreieinhalb Jahren (01.01.2010 bis 31.08.2013) wurden insgesamt 93 ELBE durchgeführt. Das war ein Durchschnitt von ca. 26 ELBE pro Jahr. Danach stieg die Zahl rasant an auf 100 ELBE in 2014 und 84 ELBE in 2015.

Für den Zeitraum 2010 bis 2013 (01.01.2010 bis 31.08.2013) liegen Ergebnisse und Analysen einer externen Meta-Evaluierung vor, die u.a. eine Auswahl von 20 Evaluierungen überprüfte. Nach Aussage des Evaluierungsteams sind die Ergebnisse nicht repräsentativ, zeigten aber Tendenzen auf. Folgende zentrale Aussagen lassen sich dem Bericht jener Meta-Evaluierung entnehmen:<sup>8</sup>

*„Die formalen Anforderungen (**Nützlichkeit der Berichte**) werden von den Evaluierungen insgesamt befriedigend erfüllt... Die fehlende innere Struktur der Berichte und die wenig aussagekräftigen Zusammenfassungen stellen die größten Defizite in dieser Kategorie dar... Die Bewertung der Kategorie „**Genauigkeit**“ fällt bei ELBE in Afrika im Vergleich zu den anderen Parametern deutlich schlechter aus. Hier treten insbesondere Schwächen bei der Erstellung einer Wirkungslogik und bei der konsequenten Anwendung der international gültigen DAC-Kriterien zu Tage. Der Anspruch der Anwendung der DAC-Kriterien durch die Gutachter/innen kann in den meisten Ländern Afrikas und Asiens nur in Ansätzen eingelöst werden.“*

Im Rahmen der Evaluierung 2013/14 wurden auch MISEREOR-MitarbeiterInnen interviewt, die Projekte begleiten. Die Rückmeldungen wurden damals im Bericht wie folgt dargestellt:

*„Die MitarbeiterInnen der KA<sup>9</sup>-Afrika und KA-Asien stehen der ELBE-Praxis eher skeptisch gegenüber, weil sie die verantwortlichen InPros in ihrer täglichen Arbeit vor erhebliche Herausforderungen stellt. Bei relativ schwachen, meist kirchlichen PO impliziert jeder Schritt einer ELBE-Durchführung (wirkungs- und DAC-bezogene TOR-Formulierung, GA-Auswahl, Nutzung der Ergebnisse für den Folgeantrag) hohen Kommunikationsaufwand, da es in den*

---

<sup>7</sup> Für die genaue Beschreibung des Vorgehens, siehe Handreichung für ELBE von MISEREOR (MISEREOR, (o. J.))

<sup>8</sup> Siehe Dokument (MISEREOR 2014: 6)

<sup>9</sup> KA steht für Kontinentalabteilung bei MISEREOR

*meisten Ländern dieser Kontinente bisher nur wenige qualifizierte lokale Beratungsstrukturen gäbe, die hier Entlastung bringen könnten...<sup>10</sup>*

Jedoch können ELBE auch von gewissen Vorteilen profitieren. Wenn Evaluierungen durch lokale Gutachter durchgeführt werden, sind einmal lokales Wissen und lokale Zusammenhänge bekannt, die für die Evaluierung nützlich sein können. Auch ein partizipatives Vorgehen kann durch größere kulturelle Nähe vereinfacht sein. Außerdem werden so Erwerbsmöglichkeiten vor Ort geschaffen und der Gedanke der Evaluierung im Partnerland weiter verankert. Deshalb sollte ein Augenmerk dieser Meta-Evaluierung darauf liegen, wie man die Qualität der ELBE verbessern könnte.

**EQM-Evaluierungen.** Die *MISEREOR-Arbeitseinheit Evaluierung und Qualitätsmanagement* (EQM) beauftragt in Absprache mit den zuständigen MISEREOR-ProjektbearbeiterInnen und den Partnerorganisationen Einzelprojektevaluierungen. Im Jahre 2014 wurden 47 und im Jahr 2015 28 EQM-Evaluierungen durchgeführt. Von Seiten MISEREORs werden diese als methodisch eher gut eingeschätzt. Es gilt zu überprüfen inwieweit die EQM-Evaluierungen den verschiedenen Fragen der Meta-Evaluierung gerecht werden und diese hohe Qualität bestätigen.

## 4 Methodisches Vorgehen und Grenzen der Meta-Evaluierung

### 4.1 Methodisches Vorgehen

In der Meta-Evaluierung wurde ein zweistufiges Verfahren angewendet, welches sich untergliedert in Stufe 1 (Qualitäts-Basischeck) und Stufe 2 (vertiefte Analysen). Durch dieses Vorgehen wurde sowohl eine große Anzahl von Berichten mit einem eingeschränkten Raster als auch eine kleinere Anzahl von Studien sehr intensiv untersucht. Gerade in der zweiten Stufe stand dabei das Lernen aus den Evaluierungen im Vordergrund, umso die Qualität der Evaluierungen zu erhöhen. Die erste Stufe legte den Fokus eher auf die Rechenschaftslegung, nichtsdestotrotz flossen die Ergebnisse dieser breiten Analyse gleichzeitig in die Diskussion über Lernen ein.

Aufgrund der seit mehreren Jahren geführten intensiven Fachdebatte zu Wirkungen und aufgrund des Wunsches von MISEREOR wurde der Schwerpunkt der Untersuchungen auf die Kategorie „Glaubwürdigkeit und Genauigkeit“ gelegt, die sich mit der angewandten Evaluierungsmethodik, der angemessenen Beantwortung der Evaluierungskriterien sowie der Nachvollziehbarkeit der Schlussfolgerungen beschäftigt. Ein weiterer Schwerpunkt in der Untersuchung wurde auf den Bereich Nützlichkeit der Evaluierungen gelegt, um zu bewerten inwieweit Evaluierungen als hilfreiche Instrumente für MISEREOR und die Partnerorganisationen zur Verbesserung, Weiterentwicklung und Anpassung der Projekte dienen können.

---

<sup>10</sup> Siehe Dokument (MISEREOR 2014: 6)

#### 4.1.1 Vorgehen im Qualitäts-Basischeck (Stufe 1)

In der Stufe 1 der Meta-Evaluierung wurden 70 aus insgesamt 261 Evaluierungen aus den Jahren 2014 und 2015 anhand eines Analyserasters untersucht. Die Stichprobe von 70 Evaluierungen wurde per Zufall aus der Gesamtheit gezogen. So ist ein Mindestmaß an Repräsentativität gegeben, die zu Aussagen über die Gesamtheit der Evaluierungen befähigt.

Die geplante Schichtung der Evaluierungen nach ELBE und EQM-Evaluierungen in der Stichprobenziehung, um in jeder der beiden Kategorien Repräsentativität zu erlangen, wurde nicht durchgeführt, da dafür eine wesentliche höhere Anzahl von Studien hätte gezogen werden müssen. So ist insbesondere die geringe Anzahl der EQM-Evaluierungen (14) für sich alleinstehend nur bedingt repräsentativ und kann durch einzelne Evaluierungen verzerrt sein.

Die Stichprobe wurde mit dem Konfidenzlevel von 95% und einem Konfidenzintervall von 10 berechnet. Diese statistischen Messwerte erklären die Aussagekraft der Stichprobe. Das Konfidenzintervall 10 steht dabei für eine mögliche Abweichung der gefundenen Ergebnisse von plus minus 10% vom realen Wert. Das Konfidenzlevel besagt wie sicher die gefundenen Ergebnisse sind. Zusammen mit dem Konfidenzlevel kann man also Aussagen treffen, dass zum Beispiel bei einem im Sample gefundenem Prozentanteil von 30% zu einem bestimmten Indikator, mit 95% Wahrscheinlichkeit der Wert in der Grundgesamtheit zwischen 20% und 40% liegt. Die Stichprobe lässt valide Aussagen über die Grundgesamtheit zu. Gewisse Abweichungen sind durch das relativ weite Konfidenzintervall jedoch möglich. Diese statistischen Rahmendaten wurden einmal den Limitierungen von finanziellen und zeitlichen Ressourcen gerecht und bieten gleichzeitig eine Stichprobe, die über eine angemessene Aussagekraft verfügt.

Bei einigen der gezogenen Projekte handelte es sich um Folgeprojekte, die gemeinsam untersucht wurden und zu denen es nur einen gemeinsamen Evaluierungsbericht gibt. Somit wurden diese schon mit anderen Evaluierungen abgedeckt. Um wieder eine Gesamtzahl von 70 Evaluierungsberichten zu untersuchen, wurden deshalb vier weitere als Ersatz gezogene Projekte zusätzlich miteinbezogen. Da in der Praxis immer mit unvorhergesehenen Problematiken bezüglich der Stichprobe zu rechnen ist, ist vorbeugend die mögliche Nutzung von Ersatzprojekten bei Ziehung der Stichprobe miteinkalkuliert worden.

Insgesamt waren 14 EQM-Evaluierungen und 56 ELBE Bestandteil der Analyse der Stufe 1. Dabei betrafen 22 Evaluierungen Projekte in Afrika, 10 Evaluierungen Projekte in Asien und 37 Evaluierungen Projekte in Lateinamerika. Dies spiegelt in etwa auch die Verteilung in der Gesamtheit der Projekte wieder. EQM-Evaluierungen wurden sowohl in der Gesamtzahl als auch in der Stichprobe überwiegend in Afrika und ELBE überwiegend in Lateinamerika durchgeführt. Nur Asien ist in dieser Stichprobe leicht unterrepräsentiert im Vergleich zu Gesamtpopulation.

**Tabelle 1 Übersicht der Evaluierungen im Qualitäts-Basischeck (Stufe 1)**

	Afrika		Asien		Lateinamerika		Summe	
	Popula- tion <sup>11</sup>	Sample <sup>12</sup>	Popula- tion	Sample	Popula- tion	Sample	Popula- tion	Sample
<b>ELBE</b>	44	12	33	8	119	35	196	56
	22%	21%	17%	14%	60%	63%		
<b>EQM</b>	35	10	18	2	12	2	65	14
	54%	71%	27%	14%	19%	14%		
<b>Gesamt</b>	79	22	51	10	131	37	261	70
	30%	31%	20%	14%	50%	53%		

Für den Qualitäts-Basischeck wurde ein Analyseraster, auf Grundlage der im Referenzrahmen zur Meta-Evaluierung von MISEREOR gegebenen Evaluierungsfragen, entwickelt. Diese Evaluierungsfragen wurden auf die Möglichkeit der sinnvollen Abfrage in einem geringen zeitlichen Rahmen und auf Aussagekraft der Ergebnisse aus einer Kurzanalyse hin untersucht. Dementsprechend lag der Fokus auf bestimmten Bereichen und Evaluierungsfragen, die sich angemessen in Stufe 1 prüfen ließen. Die Grundlage für die Untersuchungen im Qualitäts-Basischeck bildeten die vier Bereiche:

- a) Vollständigkeit und Angemessenheit der Dokumentation
- b) Unabhängigkeit und Überparteilichkeit
- c) Glaubwürdigkeit
- d) Nützlichkeit

Die in den TOR (IV, 5+6) aufgeführten Kategorien Partizipation und Konsistenz der TOR wurden zugunsten einer Vertiefung in der weiter unten aufgeführten Untersuchungsstufe 2 hier nicht analysiert.

Die Ergebnisse der Untersuchungen zu den 34 Aspekten (Fragestellungen) wurden per Viererskala bzw. Ja-Nein-Einstufung festgehalten. Die vier Kategorien in der Viererskala sind: sehr gut, gut-befriedigend, eher schwach und schwach. Die Viererskala ist eine bewährte Skalierung zur Vermeidung der Tendenz in der Mitte zu antworten. Mit dieser Skalierung gelangt man ausschließlich zu bewertenden Antworten. Das finale Analyseraster der Stufe 1 ist als Anlage 1 beigefügt.

In der Auswertung analysierte das Meta-Evaluierungsteam einmal die Ergebnisse der einzelnen Fragen direkt. In einem zweiten Schritt wurden Aggregationen zur übergeordneten

<sup>11</sup> Unter dem Begriff Population ist die Grundgesamtheit der Evaluierungen zu verstehen.

<sup>12</sup> Unter dem Begriff Sample wird hier die Anzahl der Evaluierungen aus Afrika verstanden, die per Zufallsauswahl gezogen wurden und in der Meta-Evaluierung untersucht werden.

Bewertung der Untersuchungsbereiche aus den Einzelfragen vorgenommen. Die Kategorien Vollständigkeit und Angemessenheit der Dokumentation, Unabhängigkeit, Glaubwürdigkeit und Nützlichkeit wurden so übergreifend für jeden Bericht einmal bewertet. Im Bereich Glaubwürdigkeit wurden dabei aufgrund der Vielzahl von Fragen und Themen weitere Unterkategorien gebildet. Diese sind die Bewertung der DAC-Kriterien, die Bewertung der Methodenvielfalt und das Stattfinden von Workshops.

In dieser Aggregation ist bei allen Fragen die Viererskala mit Zahlen von 1-4 hinterlegt und die Ja-Nein Antworten mit 1 und 4. In einem nächsten Schritt erhielt jede Frage eine Gewichtung, die je nach der methodischen Wichtigkeit für eine Evaluierung mit hoher Qualität und je nach Beitrag zur jeweiligen Oberkategorie bemessen wurde. Die sich daraus ergebenden Durchschnittswerte wurden der Viererskala von sehr gut-schwach zugeordnet.

#### 4.1.2 Vorgehen in der vertieften Analyse (Stufe 2)

In der vertieften Analyse wurden 20 Evaluierungsberichte auf der Grundlage der anwendbaren Fragestellungen der TOR sowie tiefergreifender Aspekte untersucht. Ziel der Stufe 2 war eine vertiefte Beschäftigung mit Evaluierungsberichten, um so Stärken und Schwächen detailliert zu analysieren und einen Beitrag zum Lernen leisten.

Bestandteil der 20 Evaluierungen der Stufe 2 waren jeweils 5 ELBE-Evaluierungen aus den drei Kontinentalabteilungen und 5 EQM-Evaluierungen mit Abdeckung der drei Regionen. Die Auswahl erfolgte überwiegend gerichtet, um sich auf Evaluierungen zu konzentrieren, die für das Lernen geeignet sind. Durch diese gerichtete Auswahl und auch wegen der geringeren Fallzahl sind in dieser Auswahl keine Aussagen über die Grundgesamtheit der Evaluierungen möglich.

Es sind dabei sowohl Evaluierungen, die in Stufe 1 durch Besonderheiten aufgefallen sind, also auch von MISEREOR vorgeschlagene Studien in die Auswahl eingeflossen. Insgesamt sind 8 Evaluierungen mit Aspekten „guter Praxis“, 6 Evaluierungen mit Schwächen und 5 Evaluierungen mit Stärken und Schwächen ausgewählt. Die 5 EQM-Evaluierungen umfassen 2 Evaluierungen aus Afrika, 2 Evaluierungen aus Asien und eine Evaluierung aus Lateinamerika.

Insgesamt wurden 14 Evaluierungen, die von MISEREOR vorgeschlagen wurden, durch das Evaluierungsteam ausgewählt. 6 weitere Evaluierungsberichte wurden durch das Evaluierungsteam auf Grundlage der Analysen der Stufe 1 ausgewählt. Die genaue Auflistung der Stufe 2-Evaluierungen kann der Anlage 5 entnommen werden.

Tabelle 2 Übersicht der Evaluierungen in der vertieften Analyse (Stufe 2)

	Afrika	Asien	Lateinamerika	Summe
ELBE	5	5	5	15
EQM	2	2	1	5
Gesamt	7	7	6	20

Das Analyseraster für die vertiefte Analyse bezieht Fragen mit ein, die einen größeren Zeitrahmen zur Beantwortung benötigen und zu denen die Auswertung detaillierter Informationen sinnvoll ist.<sup>13</sup> Grundlage für die Entwicklung sind die in den TOR (IV) dargestellten Evaluierungsfragen sowie die Vorgaben für ELBE und EQM von MISEREOR.<sup>14</sup> Ziel ist hier alle Fragen abzudecken, die nicht in Stufe 1 behandelt wurden. Das Raster umfasst alle Kategorien der Fragestellungen der TOR sowie 68 Unterfragen. Im Folgenden werden die Kategorien zusammen mit der Anzahl der Unterfragen angegeben:

- a) Vollständigkeit und Angemessenheit der Dokumentation (7)
- b) Unabhängigkeit und Überparteilichkeit (7)
- c) Glaubwürdigkeit und Genauigkeit (39)
- d) Nützlichkeit (10)
- e) Partizipation/Beteiligung (2)
- f) Konsistenz der TOR (3)

Außerdem hat jede Kategorie eine Oberfrage, in der nach einer Gesamtschätzung der Evaluierung zu diesem Thema gefragt ist. Wegen seiner Größe wurden im Kapitel Glaubwürdigkeit zu verschiedenen Bereichen einschätzende übergreifende Fragestellungen hinzugefügt, in denen eine Gesamtschätzung zu den Bereichen Wirkungsgefüge, Methoden, Auswahl der Stichprobe, Datenauswertung, Beantwortung der DAC-Kriterien und zu den Schlussfolgerungen gefragt ist. Das Kapitel Nützlichkeit erhält drei übergreifende Fragen zur Einschätzung der Zusammenfassung, der Beantwortung der Evaluierungsfragen und zur Qualität der Empfehlungen.

## 4.2 Grenzen der Meta-Evaluierung

Eine Meta-Evaluierung bietet die Möglichkeit umfassend Stärken und Schwächen einer Vielzahl von Berichten zu untersuchen. So kann ein Überblick über das derzeitige Evaluierungssystem gewonnen und gleichzeitig ein möglicher Optimierungsbedarf herausgearbeitet werden. Jedoch gibt es auch einige Einschränkungen, die zu beachten sind.

Diese Meta-Evaluierung wurde anhand der Berichte und der vorhandenen Dokumente durchgeführt. Wenn etwas nicht dokumentiert wurde, ist es für das Meta-Evaluierungsteam nicht sichtbar gewesen und konnte somit nicht honoriert oder bewertet werden. Dies bezieht sich einmal auf geschriebenen Text. Wenn keine Informationen zu bestimmten Auswahlverfahren oder Methoden gegeben wurden, so wurde dies als nicht vorhanden und somit negativ bewertet. Ob wirklich keine Auswahlkriterien angewendet wurden, ist dabei

---

<sup>13</sup> Siehe Anlage 4

<sup>14</sup> Siehe hierzu die MISEREOR-Dokumente: „Hinweise für Misereor-Partnerorganisationen zur Durchführung lokal beauftragter externer Evaluierungen“ (2014) und „Evaluierung in der partnerschaftlichen Entwicklungszusammenarbeit von MISEREOR“ (2014).

nicht ausschlaggebend. An zweiter Stelle betrifft es den Annex. Das Meta-Evaluationsteam konnte nur Dokumente bewerten, die bis zum Team gelangten. Bei 10 ELBE-Berichten der Stufe 1 und in 5 ELBE-Berichten der Stufe 2 fehlt in Berichten der Annex. Zum Teil wurde deutlich, dass zum Bericht ein Annex gehörte, dieser jedoch im Dokument oder als Zusatzdokumente nicht vorhanden waren. Nach Rücksprache mit MISEREOR wurde zum Teil deutlich, dass die Anlagen nicht in Aachen verfügbar waren. Aus Zeitgründen wurden weitere Nachfragen bei den Partnerorganisationen nicht vorgenommen. Zusammenfassend heißt dies: Die Analyse der Berichte kann nur prüfen, was niedergeschrieben ist und/oder dokumentiert wurde. Evaluierungen können unter Umständen besser durchgeführt sein, als die Dokumentation es zeigt. Jedoch ist eine gute Dokumentation der Schlüssel zur Glaubwürdigkeit und somit ist es legitim danach zu bewerten.

In den Untersuchungen der Stufe 1 wurde die Bewertungen auf Grundlage der Zusammenfassung, der Einleitung, des Methodenkapitels, der Schlussfolgerungen und der Empfehlungen gefällt. Was nicht in diesen Kapiteln oder beim Überfliegen des Textes zu erkennen war, konnte nicht gewertet werden. Somit können in Stufe 1 Aspekte im Gesamtbericht dokumentiert sein, die in dieser Stufe aber nicht bewertet wurden. Zum Beispiel wurden Methoden zum Teil im Methodenkapitel nicht genannt oder es gab kein Methodenkapitel. Ausführungen zu den Methoden flossen aber in den Ergebnissen mit ein. Dies konnte somit in Stufe 1 nicht entdeckt und dementsprechend korrekt beantwortet werden. Wenn Evaluierungen auch Bestandteil der Stufe 2 waren, ist dies dann aufgefallen und Bewertungen der Stufe 1 wurden dann angepasst oder es wurde in Stufe 2 die Bewertung kommentiert.

Bei gewissen Fragen muss die vorgenommene Bewertung vorsichtig interpretiert werden. Die Unabhängigkeit einer Evaluierung lässt sich durch eine Kurzdurchsicht des Berichts nur sehr schwer untersuchen. Es wurden 2 Evaluationsfragen in Stufe 1 gewählt, um so auch über die Breite der Evaluierungen Aussagen treffen zu können. Diese Fragen sind aber nur sehr bedingt aussagekräftig, da es sich nur um die dokumentierte Information bezüglich der EvaluatorInnen handelt.

In Stufe 2 konnte dies schon besser überprüft werden, da der Eindruck des gesamten Berichts in die Bewertung miteinbezogen werden konnte. Dies wurde unter anderem anhand der Darstellung von Stärken und Schwächen im Bericht bewertet. Jedoch lässt eine deutliche Vorhebung der Stärken nicht automatisch auf einen parteiischen EvaluatorIn schließen. Hier könnte es sich um ein besonders gutes Projekt handeln. Auch wenn jemand besonders kritisch analysiert, können nicht direkt Rückschlüsse auf die Unparteilichkeit und Unabhängigkeit gezogen werden. Wenn es sich um ein besonders gutes Vorzeigeprojekt oder um ein besonders zu kritisierendes Projekt handelt, dann ist die Evaluierung ausgewogen und unabhängig, wenn diese Extreme auch als solche dargestellt werden. Die vorgenommenen Bewertungen können also höchstens Hinweise auf eine fehlende Unabhängigkeit sein, aber für eine abschließende Einschätzung ist eine bessere Kenntnis des Projekts und der Umstände vor Ort notwendig.

Zur weiteren Überprüfung der Distanz zur Organisation wurde von Seiten des Meta-Evaluationsteam erwogen die CVs der Gutachter hinzuzuziehen. Dies wurde jedoch mit Blick auf die Verfügbarkeit und den zeitlichen Aufwand für die Prüfung wieder verworfen. Eine umfassendere Kenntnis des Projekts würde auch die Einschätzung der Empfehlungen verlässlicher machen. Hier ist der Projektkontext sehr ausschlaggebend dafür, inwieweit

Empfehlungen hilfreich und pragmatisch sind. Dies lässt sich als externe Person schwer bewerten.

## 5 Ergebnisse

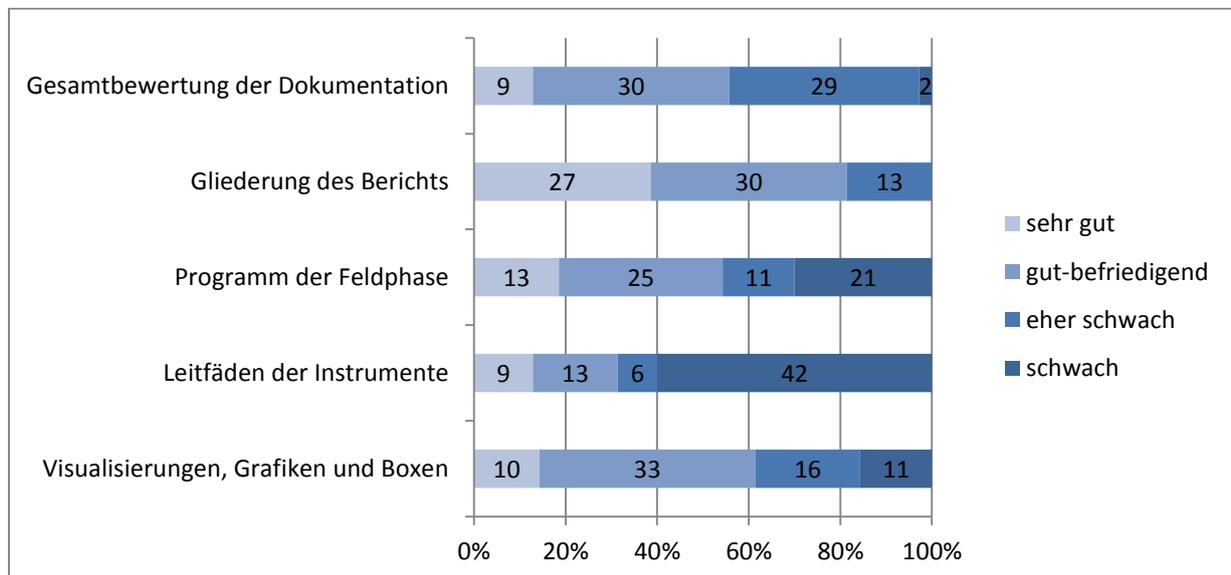
### 5.1 Vollständigkeit und Angemessenheit der Dokumentation

#### 1. Hintergrund

In der MISEREOR-Mustergliederung für den Referenzrahmen von ELBE und EQM-Evaluierungen werden im Abschnitt 6 (Bericht) Hinweise für eine Gliederung gegeben, die sich als nützlich erwiesen habe.<sup>15</sup> In der Darstellung werden die Anforderungen an die Zusammenfassung, das Methodenkapitel, Projektdarstellung und Kontext, Ergebnisse, Schlussfolgerungen und Empfehlungen dargelegt. Als Anhang werden erbeten: Referenzrahmen (TOR), chronologischer Verlauf des Einsatzes, Liste der Gesprächspartner und falls zutreffend die Liste der zugrunde gelegten Dokumente. In den Ausführungen werden keine Hinweise vorgenommen, inwiefern Visualisierungen von Ergebnissen im Text vorgenommen oder Gesprächsleitfäden und Protokolle von Abschlussworkshops als Annex beigefügt werden sollten.

#### 2. Ergebnisse des Qualitäts-Basischecks

Abbildung 2 Vollständigkeit und Angemessenheit der Dokumentation, Teil 1



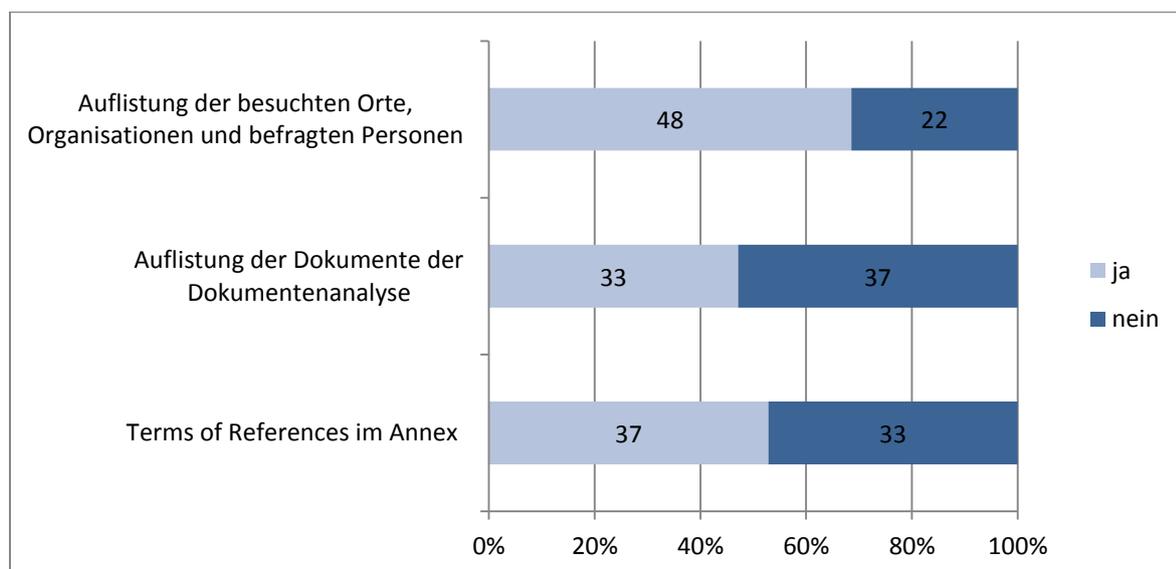
Wie aus Abbildung 2 ersichtlich ist die Gliederung der Berichte in keinem Fall als „schwach“ und überwiegend „gut-befriedigend“ und „sehr gut“ bewertet (57 (81%) der Berichte). 13 Berichte (19%) haben eine „eher schwache“ Gliederung. Das heißt in ihnen fehlen von den geforderten Kapiteln Zusammenfassung, Einleitung/Kontext, Methodenkapitel, Ergebnisse,

<sup>15</sup> Siehe Dokumente (MISEREOR o.J.: 3 (Anlage)) und (MISEREOR, EQM 2014: 25)

Schlussfolgerungen, Empfehlungen und der Unterteilung nach DAC-Kriterien mehr als zwei Unterteilungen.

Das von MISEREOR geforderte Programm der Feldphase der Evaluierung ist sehr unterschiedlich in den Bericht eingegliedert. Es kommt gleichermaßen im Annex wie auch im Bericht selbst vor. Beides wurde gewertet. Wichtig für eine hohe Wertung war hier die detaillierte Information über Ablauf, besuchte Orte, Organisationen und Methoden. Oftmals ließ sich hier das Stattfinden von Workshops ablesen. In 21 Berichten (30%) war kein Ablaufplan vorhanden. Ein Großteil der Berichte verfügt über einen Ablaufplan, der aber noch verbesserungswürdig ist. Dies wird an den 25 Berichten (36%) mit der Bewertung gut-befriedigend und 11 eher schwachen Berichten (16%) deutlich. Die Leitfäden der Interviews, Fokusgruppen oder Befragungen werden bei 42 Berichten (60%) überhaupt nicht dokumentiert. Jedoch verfügen auch 22 Berichte (31%) über eine sehr gute bis befriedigende Dokumentation der Leitfäden. Visualisierungen werden von einem Großteil der Berichte (43, 61%) angemessen eingesetzt. 11 Berichte (16%) nutzen keine Visualisierungen in den Berichten.

Abbildung 3 Vollständigkeit und Angemessenheit der Dokumentation Teil 2



Die drei abgefragten Inhalte der Abbildung 3 Vollständigkeit und Angemessenheit der Dokumentation Teil 2) sind wie oben beschrieben von MISEREOR gewünschte Dokumentationen. Etwas mehr als die Hälfte aller Berichte (37) hängt die TOR der Evaluierung an. Auch die Dokumente, die in der Dokumentenanalyse genutzt wurden, werden von fast 50% der Berichte (33) aufgezeigt. Fast 70% der Berichte (48) listen die besuchten Orte und Organisationen sowie die befragten Personen auf.

Alle EQM-Evaluierungen verfügen über einen relativ ausführlichen Annex. 16 Berichte (29%) der 56 ELBE haben keinen Annex. Bei 5 Berichten gibt es aber Hinweise auf einen Anhang im Text oder in der Gliederung, die darauf hindeuten, dass die Anhänge lediglich nicht bis zu MISEREOR gelangten. Somit werden 5 Berichte schlechter bewertet, weil man diesbezügliche Fragen nicht beantworten konnte. Auch die Qualität der Annexe der ELBE Berichte ist sehr unterschiedlich. Einige haben sehr ausführliche, sehr informative

Dokumente angehängt. Andere haben einen Annex, der zum Beispiel nur aus Fotos besteht. Dies ist insbesondere in Lateinamerika recht üblich.

Insgesamt schneiden ELBE-Berichte aus Lateinamerika und Asien in der Oberkategorie Vollständigkeit und Angemessenheit der Dokumentation schlechter ab als Afrika. Dies lässt sich bei genauerer Analyse der Einzelfragen nach Regionen auf mehrere Aspekte wie die besser bewertete Gliederung in Afrika als auch die Dokumentation von TOR und Dokumenten zurückführen.

**Tabelle 3 Vollständigkeit und Angemessenheit der Dokumentation**

	<b>ELBE (56)</b>	<b>EQM (14)</b>
<b>sehr gut</b>	4% (2)	50% (7)
<b>gut-befriedigend</b>	41% (23)	50% (7)
<b>eher schwach</b>	52% (29)	0
<b>schwach</b>	4% (2)	0

Tabelle 3 zeigt, dass EQM-Evaluierungen und ELBE sich deutlich unterscheiden.<sup>16</sup> Bei der Dokumentation sind 50% der EQM-Evaluierungen mit sehr gut bewertet und 50% mit gut-befriedigend. Bei den ELBE ist der Schwerpunkt eher in den Kategorien gut-befriedigend (42%, 23) und eher schwach (52%, 29).

### **3. Vertiefte Analyse**

In der vertieften Analyse der ausgewählten 20 Evaluierungen bestätigen sich die Ergebnisse aus dem Basischeck. Insbesondere als gutes Beispiel fallen Gliederungen auf (vgl. Abbildung 4), die alle gewünschten Kapitel abdecken und nicht überfrachtet sind mit von der EvaluatorIn neu eingebrachten Aspekten.<sup>17</sup>

---

<sup>16</sup> Dieses Ergebnis ist mit Blick auf die direkten Überprüfbarkeit dieser „Formalia“ durch den Arbeitsbereich EQM plausibel, es sollte hier jedoch die geringe Fallzahl von 14 EQM-Evaluierungen beachtet werden, die nur bedingt Rückschlüsse zulässt und verzerrt sein kein.

<sup>17</sup> Siehe Nr. 301 (Mexiko, ELBE)

Tabla de Contenido

I. Siglas y abreviaturas.....	3
II. Resumen Ejecutivo.....	4
III. Objetivos de la Evaluación.....	6
Objetivo General:.....	6
Objetivos Especificos.....	6
IV. Metodología.....	6
V. El Proyecto.....	9
VI. Resultados y Análisis de la Información.....	12
1.- Pertinencia del proyecto.....	12
2.- Efectos directos e impactos del proyecto.....	15
3.- Eficiencia.....	19
4.- Eficacia.....	20
5.- Sostenibilidad y futuro.....	21
VII. Conclusiones.....	22
VIII. Recomendaciones.....	23
IX. Anexos.....	25

Fehlende Kapitel sind in der vertieften Analyse zumeist die Zusammenfassung, das Methodenkapitel und die Unterteilung nach DAC-Kriterien. Gerade das Methodenkapitel wird oft nicht einzeln aufgeführt, so dass die Methoden oft zu knapp in der Einleitung oder in anderen einführenden Kapiteln eingebracht werden. Teilweise finden sich im Bericht aber auch gar keine Hinweise auf die Methoden der Evaluierung. Dies liegt nicht nur daran, dass EvaluatorInnen das Kapitel und die Erklärung weglassen, sondern dass auch die TOR der Partnerorganisation nicht auf eine Erklärung der Methoden und/oder einem Kapitel für Methoden hinweisen. Dies geschieht, obwohl von MISEREOR in der Handreichung dieses Kapitel empfohlen ist.

Oft wird die Bearbeitung der DAC-Kriterien und ggf. zusätzlicher Evaluierungsfragen im Ergebniskapitel eingegliedert, das dann wie im oben gezeigten Beispiel (Abbildung 4) zum Kapitel „Ergebnisse und Analysen“ wird. Die Konsequenz ist, dass in einem solchen Kapitel Bewertungen und Schlussfolgerungen vorgenommen werden und keine pure Ergebnisdarstellung erfolgt. Wenn in dem Kapitel klar zwischen Ergebnissen und Analyse zu unterscheiden ist, ist diese Herangehensweise durchaus auch als angemessen zu bewerten. Jedoch ist die Wahrscheinlichkeit einer Vermischung hoch, so dass nicht mehr deutlich ist was Bewertung der EvaluatorIn und was Ergebnis ist. Dadurch dass in den TOR die zu beantwortenden Fragen unter die DAC-Kriterien geordnet sind und so direkt miteinander verknüpft sind, könnte diese Vermischung gefördert werden.

Auch in der vertieften Analyse finden sich qualitativ sehr verschiedene Ablaufpläne der Evaluierung und es sind zum Teil keine solchen Pläne dokumentiert. Die chronologische Darstellung wurde insbesondere im ELBE-Bericht aus El Salvador (Nr.51) gut gelöst, wo auch Orte, Personen und Methoden direkt im Schema vermerkt wurden (vgl. Abbildung 5):

Abbildung 5 Beispiel Guter Praxis: Ablaufplan

<b>Martes, Octubre 7, 2014</b>	
mañana	FODA con equipo del proyecto actual Entrevista grupal con asistentes zonales del proyecto actual
tarde	Continuación Entrevista grupal con asistentes zonales del proyecto actual Visita de campo a Belén Zafra, Observación participativa en Escuela de Vida, entrevista con integrantes y familiares de la Escuela de Vida de Belén Zafra, Entrevista con el párroco de Belén Zafra
<b>Miércoles, Octubre 8, 2014</b>	
mañana	Entrevista con administradora de Pastoral Social, Nancy Mora Soto Entrevista con el coordinador de la delegación "Vida, Justicia y Paz" de la Pastoral Social, Padre Fabián Ortega Entrevista con el ex coordinador de proyectos (hasta 8/2014) de la Pastoral Social, Andrés Poveda (puesto ahora vacante)
tarde	Entrevista con el asistente zonal y ex No Matarás, Olimpo Herrera Entrevista con coordinadora del proyecto actual, Magaly Taborda
<b>Jueves, Octubre 9, 2014</b>	
mañana	Entrevista con Cristian Arias, ex asistente zonal del proyecto actual y beneficiario del proyecto anterior "Incidencia sociopolítica" Entrevista con ex coordinadora del proyecto actual, Marta Gonzales
tarde	Visita de campo a Peñitas, Observación participativa en Escuela de Vida, entrevista con integrantes y familiares de la Escuela de Vida de Peñitas

Auch bezüglich des Annexes bestätigt die vertiefte Analyse die Ergebnisse aus dem Basischeck. Alle untersuchten EQM-Evaluierungen verfügen über einen Annex. Bei den ELBE der Stufe 2 hat eine Studie in Lateinamerika nur ein Dokument angehängt und 5 Studien haben gar keinen Annex. Bei zwei Studien wird ersichtlich, dass dieser eigentlich vorhanden ist und entweder in der Gliederung oder im Text mit genauen Dokumenten aufgezählt wird.

Bei der vertieften Analyse wird sehr deutlich, um wie viel schwerer der Zugang und die Verständlichkeit einer Evaluierung ist, wenn Kapitel wie zum Beispiel die Zusammenfassung oder das Methodenkapitel fehlen. Wenn die TOR nicht angehängt sind, lässt sich aber schwer überprüfen, ob dies Vorgaben in den TOR waren und die EvaluatorIn diese dementsprechend umgesetzt hat.

Im Rahmen der Vollständigkeit und Angemessenheit der Dokumentation ist auch die Berichtslänge ein wichtiger Faktor, der sehr unterschiedlich gehandhabt wird. Zum größten Teil werden in den TOR Angaben zu den gewünschten Berichtslängen gemacht, die zwischen 30-50 Seiten liegen. EQM-Evaluierungen erscheinen in der Gänze sehr viel umfangreicher, was zum Teil aber auch an der wesentlich ausführlicheren und vollständigen Dokumentation liegt. Jedoch muss auch angemerkt werden, dass manche Berichte durch ihre Länge und Detailliertheit an Information nicht nur positiv ins Auge fallen. Es gibt ELBE-Berichte, die kurz

und knapp sind und dennoch alle wichtigen Informationen enthalten. Die mexikanische ELBE (Nr. 301) fällt hier sehr positiv auf. Auf 19 Seiten Bericht und 2 Seiten Zusammenfassung werden alle wichtigen Aspekte eingebracht. Auch etwas längere Evaluierungen sind zum Teil sehr gut und bringen viele Details über das Projekt ein.<sup>18</sup> Bei der Analyse der vertieften Studien hierzu stellt sich heraus, dass die EQM-Evaluierungen in einem Rahmen von 30-40 Seiten mit Tendenz zu 40 Seiten im Schnitt liegen. Die ELBE Berichte bestehen im Durchschnitt aus ca. 33 Seiten. Bei den ELBE sind jedoch die Extreme auch viel stärker ausgebildet. Es gibt einige sehr kurze, z.B. 8 Seiten, und einige sehr lange Berichte. Ein Bericht enthielt 95 Seiten im Hauptteil. In den Extremen sind die Berichte nicht mehr sinnvoll zu nutzen. Die zu Kurzen enthalten viel zu wenig und die zu Langen übermäßig viel Information.

Gerade in Hinblick auf den hohen Arbeitsdruck und auf knappe Zeitfenster zur Bearbeitung könnte es wichtig sein bündige Berichte zu liefern, aus denen auch die Partnerorganisation mit ihren knappen Ressourcen schnell herausfinden kann, was ein guter Follow-up Prozess nach der Evaluierung wäre. Mehrseitige Ausführungen über den historischen Kontext, die Geographie des Landes oder aber auch narrative Ergebnisdarlegung führen eher zu aufgeblasenen Berichten. Inhaltlich sind die Informationen oft durchaus von Interesse, jedoch sollte es die Aufgabe eines Evaluators sein, diese zusammenzufassen und nur die relevante Information im Bericht zu liefern.

#### **Fazit**

Die Gliederung eines Berichts ist der Schlüssel, um sich leicht in diesem zurechtzufinden und Informationen schnell zu finden. Die überwiegende Zahl der Evaluierungen wird hier mit gut-befriedigend oder sehr gut bewertet und erfüllt die Grundprinzipien einer guten Gliederung.

Die Dokumentation liefert dagegen ein sehr gemischtes Bild. Insbesondere der Ablaufplan und die Instrumente könnten noch besser dokumentiert werden. Die Berichte sind hier aber sehr gegensätzlich, etwa die Hälfte der Evaluierungen dokumentiert befriedigend bis sehr gut, die andere Hälfte eher schwach, was oftmals an einem nicht-existierenden Annex liegt. Oft wird die Ergebnisdarstellung nicht in einem separaten Kapitel vorgenommen, sondern mit der Darstellung von Analysen kombiniert. In den Fällen, in denen aus dem Text deutlich hervorgeht, was Ergebnisdarstellung und was Analyse ist, unterstützt dies die Nachvollziehbarkeit von Analysen aus den Ergebnissen.

Eine auf den Punkt gebrachte Berichterstattung trägt dazu bei, dass alle wichtigen Informationen bei der Partnerorganisation und MISEREOR ankommen. So geht keine Information unter und die Zielgruppen der Evaluierung können leicht den Bericht nutzen, um Änderungsprozesse zu initiieren.

---

<sup>18</sup> Siehe Nr. 211 (Kolumbien, EQM) und Nr. 51 (El Salvador, ELBE)

## 5.2 Unabhängigkeit und Überparteilichkeit

### 1. Hintergrund

MISEREOR setzt bei ELBE und EQM-Evaluierungen Mechanismen zur Einforderung und Sicherstellung von unabhängigen und überparteilichen Gutachterinnen und Gutachtern ein. Die zentralen Elemente für EQM-Evaluierungen sind:<sup>19</sup>

- a) Der Arbeitsbereich EQM ist selbst unabhängig von der Projektbearbeitung von MISEREOR und berichtet direkt an die Geschäftsführung
- b) EQM stellt sicher, dass für Evaluierungen unabhängige GutachterInnen beauftragt werden; unter anderem dürfen GutachterInnen pro Jahr maximal 100 Arbeitstage für MISEREOR eingesetzt werden.
- c) EQM koordiniert den gesamten Evaluierungsprozess und stellt sicher, dass in Vor- und Nachbereitungsgesprächen keine Einflussnahme auf die Evaluierungsergebnisse erfolgt.

Bei ELBE weist die Handreichung für MISEREOR-Partnerorganisationen zur Durchführung von ELBE darauf hin, dass für eine unabhängige Evaluierung die Perspektive einer unabhängigen Gutachterin bzw. eines unabhängigen Gutachters benötigt wird. Dafür werden drei „Regeln“ aufgestellt:<sup>20</sup>

- a) Die Person muss fachlich qualifiziert sein, das Projekt zu analysieren und zu bewerten.
- b) Die Person muss über das notwendige methodische Know-how verfügen, um eine Evaluierung durchzuführen.
- c) Die Person muss über ausreichend Distanz zu dem zu evaluierenden Projekt und seinem Projektträger verfügen (für diese Regel werden klare Ausschlusskriterien formuliert, wie Projektmitarbeiter und Verwandte sowie Personen, die das Projekt beraten haben).

Das Funktionieren dieser Mechanismen wird seitens MISEREORs unterschiedlich eingeschätzt. Eine in 2013/14 extern durchgeführte Auswertung der ELBE, die die Erfahrungen mit dem Instrument seit seiner Einführung in 2010 untersuchte, dokumentiert die Einschätzung von MitarbeiterInnen der Kontinentalabteilungen, dass EvaluiererInnen von EQM-Evaluierungen als unabhängiger angesehen würden, als die von ELBE. Die Gefahr von „Gefälligkeitsgutachten“ wird als sehr hoch und der Nutzen der Empfehlungen für die weitere Strategieentwicklung in den Vorhaben als gering eingeschätzt.<sup>21</sup>

Für die Dokumentation der o.g. Anforderungen an die GutachterInnen gibt MISEREOR keine Vorgaben für den Evaluierungsbericht.

---

<sup>19</sup> MISEREOR, EQM 2014: 10

<sup>20</sup> MISEREOR, o.J.: 1

<sup>21</sup> MISEREOR 2014: 6

## 2. Ergebnisse des Qualitäts-Basischecks

Im Basischeck wurde der Bericht auf Hinweise überprüft, die auf eine Evaluierung durch externe GutachterInnen hinweisen. Dabei wurde geprüft, ob auf dem Titelblatt, der Zusammenfassung, der Einleitung oder dem Methodenteil Hinweise auf den Status der GutachterIn vorliegen. Wenn z.B. auf der Titelseite der Name der AutorIn mit dem Hinweis „externe Evaluierung“ vermerkt wurde, dann wurde diese als externe EvaluatorIn eingestuft. Im Falle der Nennung von Kontaktdaten (Adresse, Internet) einer Consulting-Agentur oder eines Hinweises im Text der Evaluierung, der auf die Qualifikation oder den Hintergrund der Person einging, wurden diese als weitere Informationen zu den (externen) EvaluatorenInnen gewertet.

Abbildung 6 Unabhängigkeit der Evaluatoren

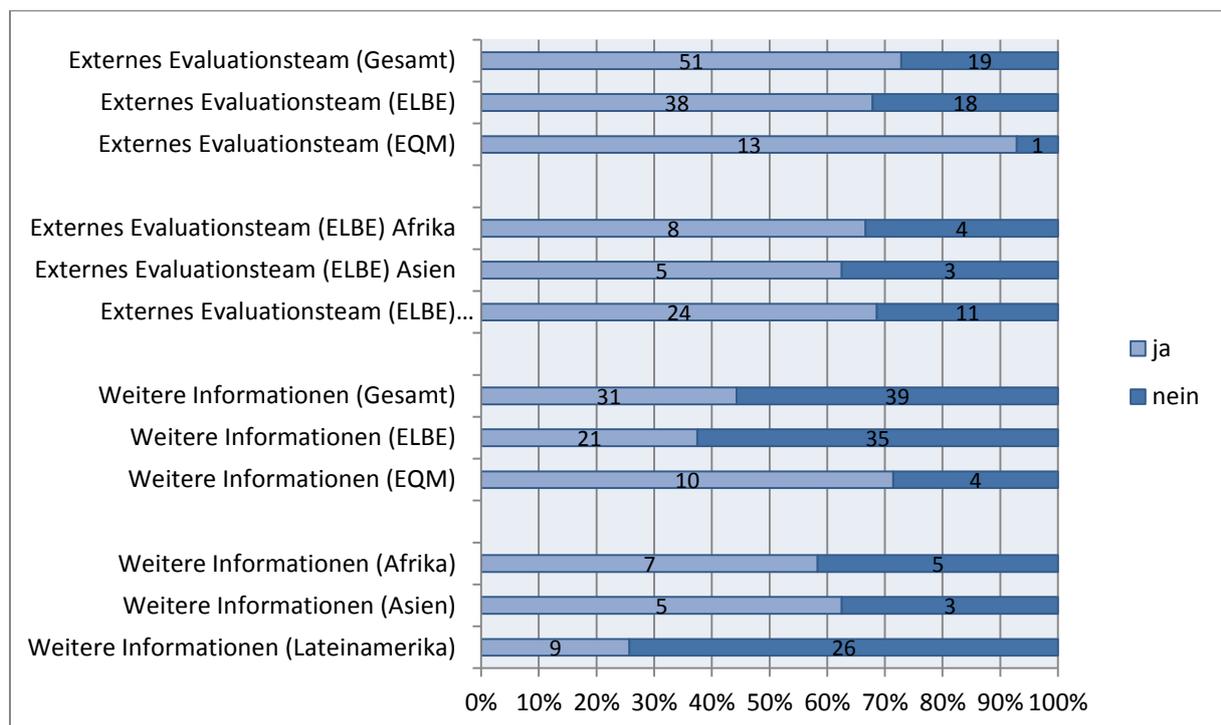


Abbildung 6 verdeutlicht, dass in 51 (73%) der 70 Evaluierungen die EvaluatorenInnen als Externe beschrieben sind. In 31 (44%) der Evaluierungen werden weitere Informationen zu institutionellen Zugehörigkeiten oder Arbeitsschwerpunkten der EvaluatorenInnen gegeben.

In den EQM-Evaluierungen liegt der Wert bei 72% (10 der 14) der Evaluierungen, die Informationen zu den Evaluatoren beifügen. In ELBE werden nur in 21 (38%) der 56 Evaluierungen weitere Informationen gegeben. Hier gibt es bei den ELBE auch Unterschiede zwischen den Kontinenten. In den ELBE von lateinamerikanischen Projekten sind weitaus weniger oft (nur bei 26% (9) der Evaluierungen) weitere Informationen gegeben als in Afrika (58%, 7) und in Asien (63%, 5).

## 3. Vertiefte Analyse

Eine Überprüfung der Unabhängigkeit und Überparteilichkeit der GutachterInnen in der von MISEREOR für ELBE und EQM-Evaluierungen eingeforderten Detailtiefe, konnte auch in der

vertieften Analysestufe (von 20 Evaluierungen) nicht durchgeführt werden. Als gute Praxis wurde hier die kurze Darstellung der fachlichen Schwerpunkte der Evaluatoren gewertet und der explizite Hinweis darauf, dass es sich um externe EvaluatorInnen handelt.<sup>22</sup>

Für die vertiefte Analyse von Unabhängigkeit und Überparteilichkeit wurde zusätzlich untersucht, inwiefern die Informationsquellen so gewählt wurden, dass eine Vielzahl von Perspektiven in die Datenerhebung, Datenauswertung und Darstellung der Ergebnisse einfluss. Die Ergebnisse sind wie folgt:<sup>23</sup>

- a) Die Perspektiven der Zielgruppen, in der Regel vertreten durch ZielgruppenvertreterInnen, nehmen in allen Evaluierungen einen wichtigen Stellenwert ein. Keine Evaluierung beschränkt sich auf Datenerhebungen bei der Partnerorganisation und Befragung von ExpertInnen z.B. in der Hauptstadt.
- b) Jedoch werden nicht immer bei der Datenauswertung beide Perspektiven ausgewogen dargestellt. So gibt es Evaluierungen, in denen vor allem aus Sicht der Partnerorganisation die Daten ausgewertet werden und welche in denen vor allem die Interessen der Zielgruppen im Fokus dargestellt werden.<sup>24</sup>
- c) Perspektiven von externen Fachleuten sind rar. Bei Projekten der ländlichen Entwicklung werden z.B. VertreterInnen von staatlichen Einrichtungen (z.B. Krankenhäuser, Gesundheitsbehörden, Veterinäreinrichtungen etc.) einbezogen, aber es wird praktisch keine externe Perspektive eingeholt, die das grundsätzliche Vorgehen des Projektes kritisch betrachtet.<sup>25</sup>
- d) Eine Einbeziehung der Fachdiskussion und der lessons learnt aus Evaluierungen vergleichbarer Projekte werden in den Berichten praktisch nicht dargestellt. Die Fachdiskussion und die lessons learnt fließen nur durch die Einschätzungen der beauftragten EvaluatorInnen in die Analysen ein.<sup>26</sup>
- e) In dieser kleinen nicht-repräsentativen Auswahl von Berichten sind bezüglich der Anzahl der Perspektiven keine systematischen Unterschiede zwischen ELBE und EQM-Evaluierungen zu finden.

Für die Ortung fehlender „Unabhängigkeit und Überparteilichkeit“ wurde in den Berichten auch die Art und Weise untersucht, wie Stärken und Schwächen des Berichts dargelegt wurden. Es wird davon ausgegangen, dass GutachterInnen mit mangelnder Distanz zur Partnerorganisation und zum Projekt diese eher positiv darstellen und dementsprechend positive Empfehlungen aussprechen, die nicht wirklich potentielle Herausforderungen wie

---

<sup>22</sup> Beispiele guter Praxis sind hier zum Beispiel: Nr. 152 (Kolumbien, ELBE) mit einer Kurzpräsentation des Evaluators im Annex oder auch absolut ausreichend wie in Nr. 208 (Brasilien, EQM) mit einer kleinen Vorstellung des Evaluationsteams im Bericht.

<sup>23</sup> Die Ergebnisse werden hier nicht quantifiziert (z.B. bei 4 von 20 Evaluierungen trifft das zu), da die Auswahl der 20 Evaluierungen von verschiedenen Akteuren mit verschiedenen Perspektiven vorgenommen wurde; siehe Informationen dazu im Methodenkapitel).

<sup>24</sup> Nr. 12 (Philippinen, ELBE) für Auswertung eher aus Partnerorganisationsperspektive und Nr. 51 (El Salvador ELBE) für Sicht aus Zielgruppenperspektive.

<sup>25</sup> Ein Beispiel guter Praxis ist Nr. 302 (Bolivien, ELBE) Befragung von Regierungsvertretern, Frauenorganisationen und Gewerkschaften

<sup>26</sup> Ein Beispiel mit externen Literaturquellen bei Projekthintergrund und Methoden: Nr.51 (El Salvador, ELBE), lessons learnt aus anderen Evaluierungen werden bei Nr. 217 (Burkina Faso, EQM) genutzt (Misserfolge bei Gruppenbildungen und Degradation von Weideland).

eine angemessene Konzeption der Projektinhalte und gutes Projektmanagement angehen. Die Ergebnisse der Analyse sind:

- a) Viele Berichte der vertieften Analyse stellen sowohl Stärken als auch Schwächen der Projekte ausgewogen dar.
- b) Es gibt in der vertieften Analyse ELBE- sowie EQM-Evaluierungsberichte, in denen die EvaluatorInnen vor allem die Stärken des Projekts unterstreichen.<sup>27</sup> Die besonders positiven Darstellungen könnten auf eine gewisse Parteilichkeit und nicht ausgewogene Nähe zum Projekt hindeuten oder können für ein sehr gutes Projekt stehen, was durchaus unabhängig evaluiert wurde. Dies ist aus Sicht der Meta-EvaluatorInnen nicht abschließend zu klären.
- c) In einigen Berichten folgen auf eine sehr positive Darstellung teilweise sehr kritische und umfangreiche Empfehlungen, die eindeutig auf Schwächen in substantziellen Bereichen des Projekts oder der Partnerorganisation hindeuten.<sup>28</sup>
- d) Insbesondere in den Fällen, wenn diese indirekte (konstruktive) Kritik erst sehr deutlich in den Empfehlungen geäußert wird, kommt diese Ausgewogenheit sehr spät zum Vorschein. Sie ist aber plausibel aus den Ergebnissen und Analysen ableitbar.<sup>29</sup>
- e) Es gibt Berichte, die dezidiert die Projektarbeit der Partnerorganisation analysieren und Empfehlungen aussprechen, Teilbereiche oder das Gesamtprojekt sofort zu stoppen bzw. neu zu konzipieren. Die Untersuchung identifizierte davon zwei EQM-Evaluierungen.<sup>30</sup>

### **Fazit**

Die Ergebnisse des Qualitäts-Basischecks zeigen, dass in ca. 3/4 der Evaluierungen (73%) unabhängige GutachterInnen explizit genannt werden und in etwa 30% der Evaluierungen weitere Informationen zum Evaluierungsteam gegeben werden. Die Perspektive der Partnerorganisationen und der Zielgruppen werden größtenteils tief in die Analysen mit eingebunden. Externe Perspektiven in der Form von externen Experten oder Fachdiskussion fließen explizit eher weniger mit ein. Die meisten der Evaluierungen legen sowohl Stärken als auch Schwächen des evaluierten Projekts dar. Dies geschieht aber in unterschiedlicher Ausgewogenheit und Detailliertheit. Zum Teil wird nur sehr positiv berichtet, was aber nicht unbedingt für Parteilichkeit und fehlende Unabhängigkeit steht.

---

<sup>27</sup> Ein Beispiel dafür ist: Nr. 301 (Mexico, ELBE)

<sup>28</sup> Beispiele hierfür sind die Berichte Nr. 228 (Madagaskar, EQM) und Nr. 187 (Südafrika, ELBE)

<sup>29</sup> Ein Beispiel ist 13 (Philippinen, ELBE) in dem 70 plausible Empfehlungen im Schlusskapitel entwickelt werden.

<sup>30</sup> Siehe Nr. 217 (Burkina Faso, EQM) und Nr. 211 (Kolumbien, EQM).

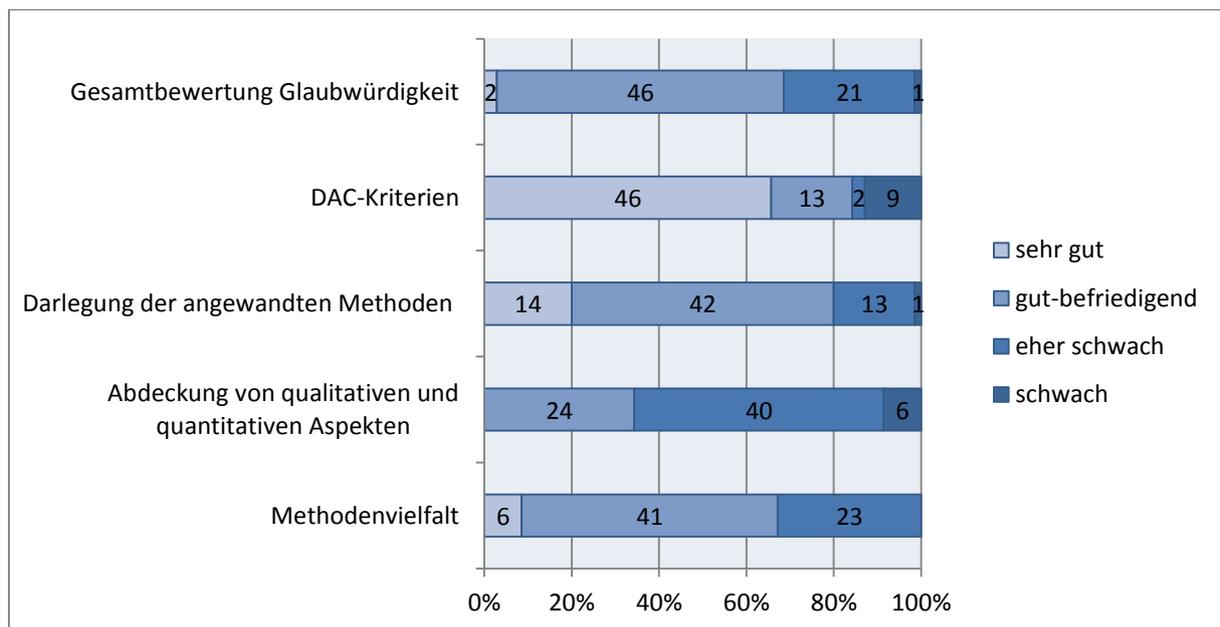
## 5.3 Glaubwürdigkeit

### 1. Hintergrund

Für EQM-Evaluierungen werden im MISEREOR-Evaluierungskonzept Qualitätskriterien für eine angemessene Evaluierungsmethodik definiert. Zwei der drei Kategorien beschäftigen sich mit „Genauigkeit und Glaubwürdigkeit“ sowie „Wirkungserfassung“. Die Anforderungen werden in „Mindeststandards“ und „Gute Praxis“ untergliedert.<sup>31</sup> Für ELBE gelten reduzierte Vorgaben, die zum einen die Standardfragen zu den DAC-Kriterien Relevanz, Wirkungen (Outcome und Impact), Effektivität, Effizienz und Nachhaltigkeit umfassen. Es wird zum anderen ein Mix aus Erhebungsmethoden empfohlen, um qualitative und quantitative Aspekte zu untersuchen. Hervorgehoben wird, dass insbesondere die ggf. unterschiedliche Wirkung auf verschiedene Zielgruppen untersucht werden soll.<sup>32</sup> MISEREOR empfiehlt, das DAC-Kriterium Impact, um den Bereich des Outcomes zu erweitern und als Kapitel „Wirkungen (direkte und indirekte)“ zu bezeichnen. Im Englischen wird dies mit „Outcome and Impact“ übersetzt.

### 2. Ergebnisse des Qualitäts-Basischecks

Abbildung 7 Glaubwürdigkeit Teil 1



Die Gesamtbewertung der Glaubwürdigkeit und Genauigkeit der Berichte liegt bei 46 Berichten (66%) im Bereich von „gut-befriedigend“, nur zwei Berichte sind mit „sehr gut“ bewertet (siehe Abbildung 7). 30% (21) der Berichte sind „eher schwach“ in Bezug auf die Glaubwürdigkeit zu bewerten, einer als „schwach“.

<sup>31</sup> MISEREOR, EQM 2014: 20f

<sup>32</sup> MISEREOR o.J.: 3 (Anhang Referenzrahmen)

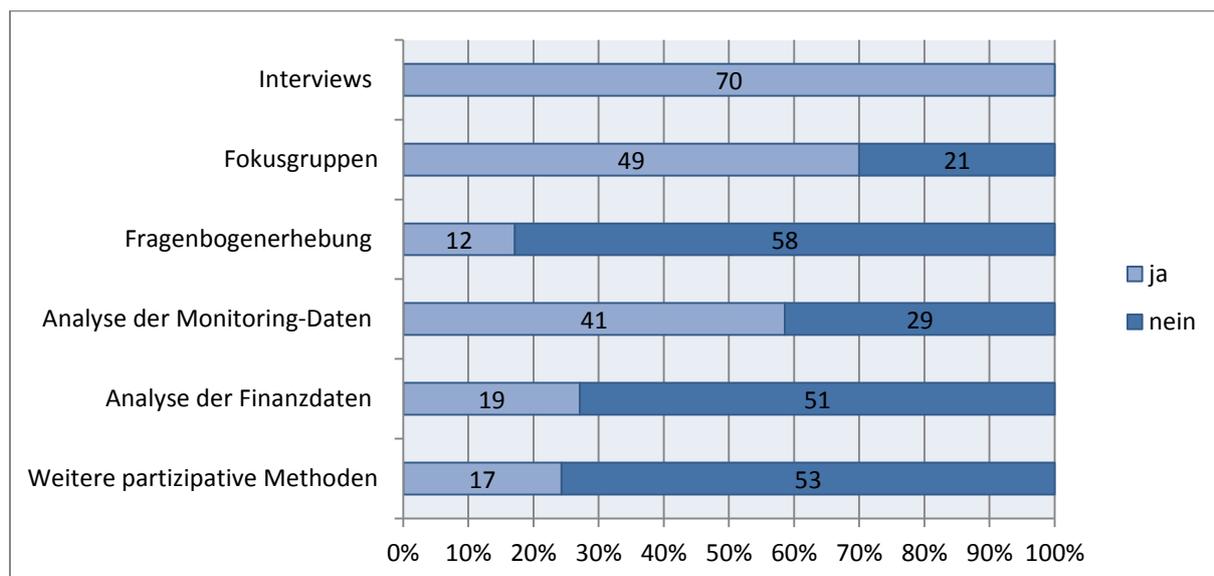
Wird nur die Behandlung der DAC-Kriterien betrachtet, so sind alle 5 Kriterien in über 65% der Berichte (46) miteinbezogen worden. In 15 Berichten (21%) sind die Kriterien nur zum Teil vorhanden und in 9 Berichten (13%) werden die Kriterien gar nicht betrachtet.

Die restlichen Fragen des Basischecks im Bereich Glaubwürdigkeit und Genauigkeit beziehen sich auf verschiedene Aspekte der Methodennutzung und -darstellung. In einer überwiegenden Anzahl (80%) sind die Methoden in den Berichten mindestens „gut-befriedigend“ (42) oder „sehr gut“ (14) dargestellt. 13 (19%) Berichte sind als eher schwach bewertet einer als schwach.<sup>33</sup>

Bezüglich der Anwendung von Methoden, die qualitative und quantitative Aspekte abdecken, wird dagegen keinmal „sehr gut“ vergeben. 24 (34%) Berichten wurden mit „gut-befriedigend“ bewertet und der überwiegenden Teil von 40 (57%) Berichten mit „eher schwach“. 6 (9%) Berichte sind mit „schwach“ bewertet.

Die Methodenvielfalt ist eine Aggregation der verschiedenen Methoden, die genutzt wurden. Insgesamt werden hier 6 (8%) Berichte mit „sehr gut“ bewertet, 41 (59%) Berichte mit „gut-befriedigend“ und 23 (33%) mit „eher schwach“. Kein Bericht wird als „schwach“ bewertet.

Abbildung 8 Glaubwürdigkeit Teil 2: Methodenvielfalt



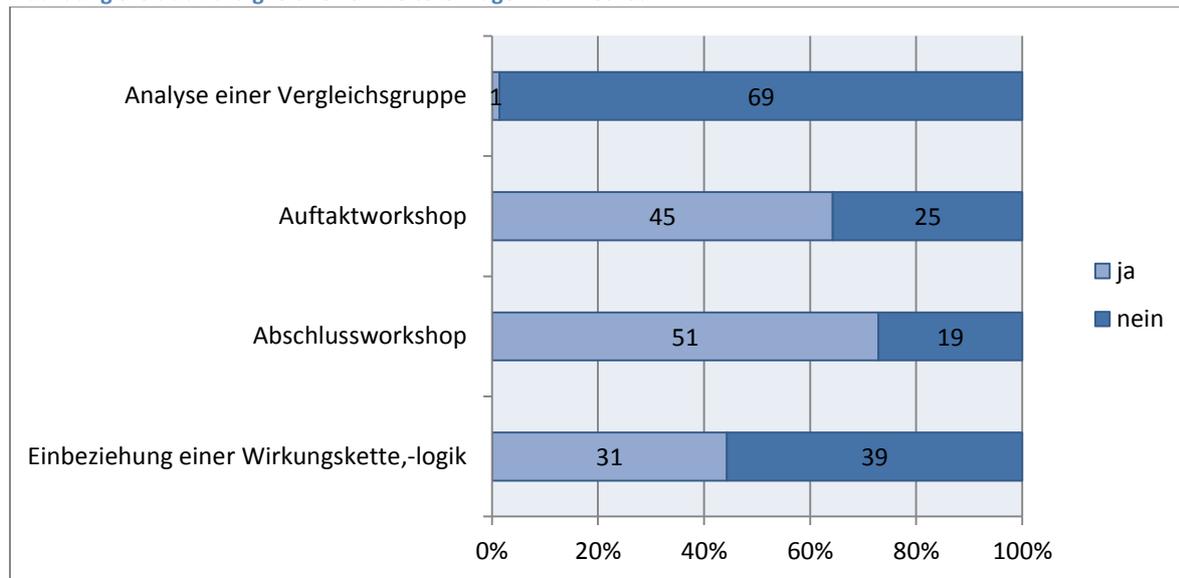
In Abbildung 8 werden die angewandten Methoden aufgeschlüsselt. In allen Evaluierungen wurden Interviews und in 49 (70%) Fokusgruppensitzungen bzw. Gruppeninterviews

<sup>33</sup> Die Unterscheidungen zwischen den Bewertungen der Methodenkategorien sind:

- Schwach: keine Aussagen zum methodischen Vorgehen
- Eher schwach: 1-3 Sätze zum Vorgehen (meist in der Einleitung);
- Gut-befriedigend: Methodenkategorien oder längerer Absatz (mehr als 3 Sätze) zum methodischen Vorgehen vorhanden, die Informationen sind aber nicht vollständig (z.B. kein Hinweis darauf, wie die Auswahl der 10 besuchten Gruppen aus der Gesamtmenge der 80 Gruppen des Projekts vorgenommen wurde)
- sehr gut: Angemessene vollständige Beschreibung der eingesetzten Methoden

durchgeführt. Zu Fragebogenerhebungen kam es in 12 Fällen (17%). Die Monitoringdaten der Projekte wurden von 41 Berichten (59%) genutzt, die Finanzdaten nur von 19 (27%). 17 Evaluierungen (24%) nutzen weitere partizipative Methoden wie eine SWOT Analyse oder Tendenzanalysen. Die häufigsten Datenerhebungsmethoden bei EQM-Evaluierungen und ELBE sind (leitfadengestützte) qualitative Interviews, gefolgt von Fokusgruppengesprächen und standardisierten Fragebogenerhebungen. Letztere kommen statistisch gesehen in jeder fünften Evaluierung vor.

Abbildung 9 Glaubwürdigkeit Teil 3: Weitere Fragen zur Methodik



In nur einem Fall wurde eine Vergleichsgruppe in der Auswahl der 70 Evaluierungen untersucht. Ein Auftaktworkshop wurde in 45 (64%) und ein Abschlussworkshop in 51 (73%) der Fälle durchgeführt. In 31 Berichten (44%) wurde eine Wirkungskette oder -logik erwähnt, diskutiert, dargestellt oder entwickelt.

Tabelle 4 Glaubwürdigkeit: ELBE vs. EQM / Regionen

	ELBE (56)	EQM (14)	Afrika (22)	Asien (10)	Lateinamerika (37)
<b>sehr gut</b>	2%(1)	7% (1)	0	0	5%(2)
<b>gut-befriedigend</b>	61%(34)	86% (12)	82% (18)	60%(6)	57% (21)
<b>eher schwach</b>	36% (20)	7% (1)	18%(4)	30% (3)	38%(14)
<b>schwach</b>	2% (1)	0	0	10%(1)	0

In Tabelle 4 werden ELBE und EQM-Evaluierungen verglichen, sowie die Evaluierungen der verschiedenen Kontinente. EQM-Evaluierungen schneiden dabei überwiegend „gut-befriedigend“ ab. Es gibt nur eine schwache aber auch nur eine sehr gute Bewertung der Glaubwürdigkeit. Bei den ELBE ist das Ergebnis weiter gestreut und insgesamt schwächer als bei den EQM-Evaluierungen. Sie teilen sich dabei hauptsächlich auf die beiden Kategorien „gut-befriedigend“ (61%, 34) und „eher schwach“ (36%, 20) auf.

Die afrikanischen Evaluierungen schneiden insgesamt methodisch besser ab, auch wenn die beiden einzigen mit sehr gut bewerteten Evaluierungen aus Lateinamerika stammen. Hierbei muss aber auch beachtet werden, dass die afrikanischen Evaluierungen durch einen größeren Anteil an EQM-Evaluierungen auch hiervon geprägt sein können. Wenn man nur die ELBE der verschiedenen Kontinente vergleicht, so ist kein deutlicher Unterschied mehr festzustellen.

### 3. Vertiefte Analysen

**Evaluierungsmethodik.** Die Darlegung und Beschreibung der Methoden und Vorgehensweise in der Evaluierung ist ein wichtiger Bestandteil um die Glaubwürdigkeit einer Evaluierung zu bewerten. Wie schon im Basischeck beschrieben ist hier die Qualität sehr unterschiedlich angefangen von einem ausführlichen Methodenkapitel bis zu einigen wenigen Sätzen. Ein besonders positives Beispiel für eine gute Beschreibung der Methoden und wichtiger Aspekte zur Datenauswertung ist die Evaluierung Nr. 301 (Mexiko, ELBE) (siehe Annex).

In den vertieften Analysen waren zusätzlich zur Abfrage der Methoden auch die Auswahl der Stichprobe, die Datenauswertung, die Evaluierungskriterien und die Schlussfolgerungen im Zentrum der Analyse.

Bezüglich der Auswahl einer Stichprobe, der Auswahlkriterien und auch der Zufallsstichprobe ist der Eindruck aus der vertieften Analyse wie auch aus Stufe 1, dass nur sehr selten Auswahlkriterien dargelegt werden und wenn nur sehr eingeschränkt. Ein Random Sample wurde in der Gesamtzahl der Studien nur bei einer EQM-Evaluierung auf den Philippinen (Nr. 16) entdeckt. Die Auswahl von schwächeren und stärkeren Elementen im Sinne einer geschichteten Stichprobe wurde keinmal dokumentiert.

Somit ist die Dokumentation der Ziehung bzw. Auswahl solcher Stichproben eher als schwach zu bewerten. Sind es überwiegend Vorschläge der Partnerorganisation, die verständlicherweise eher „starke und positive“ Stimmen vorschlagen wird? Ist es eine gemeinsame bewusste Auswahl von eher stärkeren und eher schwächeren Akteuren, die anhand von Kriterien gemeinsam zwischen EvaluatorInnen und Partnerorganisation ermittelt wurden, ggf. ergänzt durch Zufallsstichproben? Diese Schwachpunkte sind in ELBE und EQM-Evaluierungen zu finden.

Die Anzahl der InterviewpartnerInnen und DiskussionsteilnehmerInnen bei den Projektmitarbeitenden und den Verantwortlichen der Partnerorganisation sowie die Anzahl der VertreterInnen der Zielgruppen als Informationsquelle für die Evaluierung erscheint angemessen; insbesondere mit Blick auf das Projektvolumen und den für die Evaluierung zur Verfügung stehenden Mitteln.

### **Exemplarische Diskussion der Auswahl der InterviewpartnerInnen am Beispiel einer ELBE:**

In einem Projekt für Capacity-Building von Mädchen mit einem Mittelvolumen von ca. 400.000 € für 3 Jahre werden alle 4 Standorte des Projektes besucht, dabei werden 30 Schülerinnen, 7 Mitarbeitende, 3 Leitungspersonen des Projekts, 6 Personen des Vorstands, 7 ehemalige Schülerinnen sowie 5 externe Personen (LehrerInnen von anderen Schulen und Aktivisten) besucht.

**Bewertung der Auswahl.** Bei einer Evaluierung eines Capacity-Building-Projekts wäre es sinnvoll, z.B. eine semi-standardisierte Fragebogenerhebung ggf. als Online-Befragung zu initiieren. Da nach Aussage der Evaluatorin keine Datenbank der ehemaligen Teilnehmerinnen der Kurse bestand, wurden nur Interviews und Fokusgruppengespräche geführt. Dies ist eine Alternative. Deutlich aussagekräftiger wäre die Stichprobe geworden, wenn die Auswahl der Interviewpartnerinnen, insbesondere der ehemaligen Teilnehmerinnen und externen Personen dargestellt worden wäre. Waren die 7 Ehemaligen besonders positiv und intensiv mit der Einrichtung verbunden und daher erreichbar? Gab es darunter auch kritische Stimmen? Gab es Kursabbrecherinnen, die ggf. hätten kontaktiert werden können? Wie wurden die externen Personen ausgesucht? Waren es Empfehlungen des Projektpartners oder wurde pro-aktiv auf z.B. LehrerInnen in den Projektgebieten zugegangen, um ihre Einschätzung zu den Aktivitäten des Capacity Buildings einzuholen? Fünf Zeilen mehr im Bericht zu diesen Fragestellungen hätte die Auswahl transparent gemacht. Ein zusätzlicher Arbeitstag oder die Reduzierung der Besuche auf 3 anstatt 4 Einrichtungen hätte den Freiraum gegeben, gezielt externe Ressourcepersonen anzusprechen, um ihre Einschätzung der Situation in die Analyse aufzunehmen.

**Dokumentation von Veränderungen.** Über Veränderungen im Ablauf der Evaluierung wird selten berichtet. In den 20 untersuchten Berichten gab es einen Fall, in dem darauf hingewiesen wurde, dass die ursprünglich geplante Anzahl an semi-standardisierten Interviews nicht erreicht werden konnte, da die Zielgruppen durch Erntearbeiten nur eingeschränkt verfügbar waren.<sup>34</sup> In einem weiteren Beispiel wurde eine Gemeinde nicht weiter in die Evaluierung mit einbezogen, weil es zu Missverständnissen kam. Generell sind dies aber Einzelbeispiele, in denen dies dokumentiert wurde.

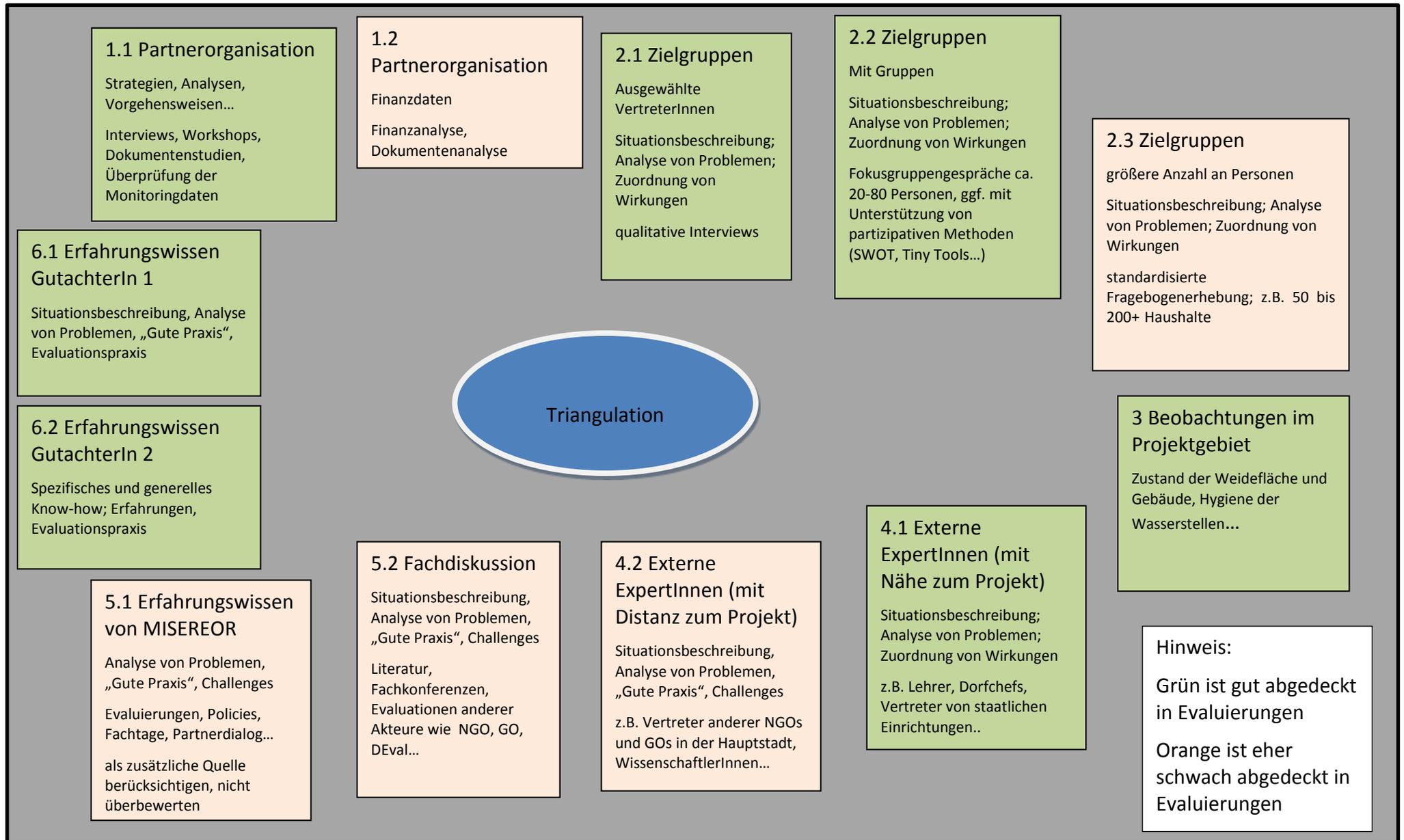
**Triangulation.** Für die Überprüfung und Einordnung der erhobenen Daten zur Bewertung der Ergebnisse des Projekts standen eine Reihe von Informationsquellen zur Verfügung, die in unterschiedlicher Weise genutzt wurden. Im Schaubild grün hinterlegte Datenquellen werden in der Regel genutzt. Die rot hinterlegten werden eher weniger eingesetzt, zumindest wird ihre Berücksichtigung nicht in den Berichten thematisiert.

---

<sup>34</sup> Siehe Nr. 182 (Malawi)

Abbildung 10 Datenquellen für Triangulation

Hinweis: In den Kästen wird zunächst die Datenquelle, dann der Grund für die Datenerhebung und ggf. auch die Erhebungsmethode angegeben



In allen untersuchten Evaluierungsberichten werden die Datenquellen *Partnerorganisation*, *Zielgruppen* und *GutachterInnen* genutzt. Die Überprüfung erfolgt vereinfacht dargestellt wie folgt: Die GutachterInnen sprechen mit der Partnerorganisation und nehmen Daten und Einschätzungen zu den Projektergebnissen auf. Die Überprüfung der Information bei den Zielgruppen erfolgt über eine Reihe von qualitativen Interviews vor Ort, die ggf. mit Gruppengesprächen ergänzt werden, um so die Anzahl der befragten Personen anzuheben und die Rückmeldungen innerhalb der Zielgruppen zu überprüfen. Alle diese Informationen werden auf der Grundlage des Erfahrungswissens und der Kenntnis der aktuellen Fachdiskussion der EvaluatorInnen überprüft.

Bei Projekten im Bereich Landwirtschaft, Viehzucht, Bildung und Gesundheit können zusätzlich Beobachtungen vor Ort die gewonnenen Informationen ergänzen: Welche Qualität haben die Weideflächen? Inwiefern konnte die Erosion aufgehalten werden? Machen die Bildungseinrichtungen einen sauberen und gepflegten Eindruck?<sup>35</sup> Bei den untersuchten Advocacy-Projekten fehlt diese Möglichkeit, da sich die Erhebungen oft auf die Gespräche mit den ZielgruppenvertreterInnen konzentrieren und oft nichts „Greifbares“ zu sehen ist.<sup>36</sup>

Bei den untersuchten EQM-Evaluierungen wurden immer eine internationale und eine nationale GutachterIn eingesetzt. Bei ELBE wurde in der Regel nur eine Person beauftragt.<sup>37</sup> Der Einsatz von mehr als einer GutachterIn ermöglicht die Triangulation der erhobenen Daten und angewandten Auswertungsmethoden vor dem Hintergrund der verschiedenen Perspektiven und Erfahrungen.

Datenquellen, die in den untersuchten EQM-Evaluierungen und ELBE praktisch nicht genutzt bzw. nicht dokumentiert wurden, sind (siehe Schaubild oben):

- a) Finanzanalysen zu den Projektaktivitäten bei der Partnerorganisation, um Einsatz und Fokussierung der Mittel zu bewerten
- b) Standardisierte Fragebogenerhebung bei der Zielgruppen, um das Feed-back zur Situationsbeschreibung, Analyse von Problemen und der Zuordnung von Wirkungen von einer größeren Anzahl von Personen zu erhalten
- c) Informationen von externen ExpertInnen (mit Distanz zum Projekt) zur Situation im fachlichen Sektor (z.B. Basis-Gesundheitsversorgung und ländliche Entwicklung sowie Zusatzinformationen zur Region); „Blick über den Tellerrand des Projektes“
- d) Fachdiskussion und „Gute Praxis“ anderer vergleichbarer Projekte für die Analyse des Projektdesigns und des Management von klassischen Herausforderungen wie z.B. Personalfuktuation im Projekt, Nehmermentalität bei der Zielgruppe, Schwierigkeiten der Gruppenbildungen bei den Zielgruppen
- e) Erfahrungswissen von MISEREOR, das sich auf die o.g. Fachdiskussion bezieht, aber noch spezifischer auf die Partnerorganisationen zugespißt ist

---

<sup>35</sup> Beispiele hierfür weisen die Berichte Nr. 217 (Burkina Faso), Nr. 228 (Madagaskar) und Nr. 185 (Palästinensische Gebiete).

<sup>36</sup> Beispiele sind die Berichte 238 (Indien) und 13 (Philippinen).

<sup>37</sup> Beispiele sind Nr. und Nr. 185 (Palästinensische Gebiete), Nr. 103 (Indien), 187 (Südafrika).

**Datenauswertung.** Auf die verschiedenen Erhebungsmethoden wird in den Berichten immer wieder Bezug genommen. Schwachpunkte sind jedoch die Einordnung der Daten: Handelt es sich bei der Aussage der zitierten Frau um eine Einzelmeinung oder wird diese auch von anderen geteilt? Wie kann die Gruppe, bei der das Instrument „Tiny Tool“ eingesetzt wurde, beschrieben werden? Besteht sie überwiegend aus Frauen oder aus Männern? Ist sie eine besonders aktive Gruppe? Wie lang arbeiten sie schon zusammen? Gehört sie eher zu den stärkeren oder schwächeren Gruppen des Projekts?

Eine Gewichtung von Ergebnissen durch „Größtenteils...“, „Eine Mehrheit...“ oder „Eine Minderheit...“ findet oftmals nicht statt. Die Begründung von Bewertungen wird zum Teil vorbildhaft aber zum Teil auch sehr wenig bis gar nicht durchgeführt. Hier ist eine große Bandbreite aufzufinden. Von den 20 Studien liegen 16 im Bereich von gut bis befriedigend. Dabei muss aber beachtet werden, dass hier einige Beispiele der guten Praxis mit untersucht wurden.<sup>38</sup> Fünf Studien haben in diesem Bereich eine eher schlechte Bewertung.

**Analyse der sozialen Differenzierung.** Eine Untergliederung der Zielgruppen nach Alter, Geschlecht, Wohlstand und anderen sozialen Kriterien ist wenig verbreitet. In der vertieften Analyse fanden wir nur in 3 von 20 Evaluierungen Analysen, die über die Unterscheidung zwischen Männern und Frauen hinausgingen. Im ELBE Nr. 16 auf den Philippinen wurden die Befragten stark untergliedert nach verschiedenen sozioökonomischen Merkmalen. In ELBE Nr. 54 in Guatemala wird nach verschiedenen landwirtschaftlichen Produzenten unterschieden. In Mexiko (ELBE Nr. 301) wird eingebracht, dass das Programm für junge und ältere Frauen unterschiedlich zu bewerten ist. Diese Untergliederungen beschränken sich alle auf die einleitenden Beschreibungen der Zielgruppen. In den Ergebnissen, Schlussfolgerungen und Empfehlungen bringt nur ein Bericht die Sichtweisen von differenzierten Zielgruppen mit ein: In ELBE Nr. 51 in El Salvador wird systematisch aus den Ergebnissen die Empfehlung abgeleitet, mehr Jugendliche durch das Projekt anzusprechen.

Der Mehrwert der differenzierten Analyse, zum Beispiel die Differenzierung zwischen Männern und Frauen besteht darin, bei Bedarf spezifische Empfehlungen für die Umsetzung der Projektaktivitäten zu entwickeln. Insgesamt besteht jedoch der Eindruck, dass die Genderperspektive wenig systematisch und wenig vertieft in den Evaluierungsberichten auftaucht. Wo Genderperspektive eingesetzt wird, führt sie in der Regel zur Erkenntnis, dass das Projekt auch Aktivitäten für Frauen anbieten müsse.<sup>39</sup> Differenziertere Schlussfolgerungen und Empfehlungen finden sich praktisch nur in Evaluierungen von „Genderprojekten“.<sup>40</sup>

**DAC-Kriterien.** Nachdem im Basischeck das bloße Vorhandensein der DAC-Kriterien bewertet wurde, prüft die vertiefte Analyse die Inhalte. Wie wurden die jeweiligen Kriterien behandelt und mit welchen Methoden? Die vertiefte Analyse zeigte: Überwiegend werden die DAC-Kriterien in den Ergebnissen behandelt. Nur in einem Fall wurden Ergebnisse dargestellt und dann in einem separaten Kapitel „Analyse der Ergebnisse“ nach DAC-

---

<sup>38</sup> Wie im Methodenkapitel ausgeführt, liefert die vertiefte Analyse der 20 Evaluierungen keine repräsentativen Daten.

<sup>39</sup> Zum Beispiel Nr. 217 (Burkina Faso)

<sup>40</sup> Zum Beispiel Nr. 182 (Malawi)

Kriterien bewertet.<sup>41</sup> Wenn Evaluierungsfragen nach DAC-Kriterien gegliedert sind, folgen in der Regel die EvaluatorInnen dieser Struktur. Zum Teil wird aber auch frei zum Thema geschrieben und es ist nicht klar, ob die Evaluierungsfragen als Grundlage genommen wurden.

Die Bereiche Relevanz, Wirkungen und Effektivität werden zum größten Teil eher breit und umfassend beantwortet. In diesen Kapiteln insbesondere bei Effektivität und Wirkungen fallen öfters Dopplungen ins Auge, da auch unter Wirkungen die langfristigen Projektziele geprüft werden. Die Kapitel Effizienz und Nachhaltigkeit werden in der Regel kurz abgehandelt.

In Abbildung 11 ist ein interessantes Beispiel für die Differenzierung der Relevanz eines Projektes nach einzelnen Aktivitäten aufgeführt.<sup>42</sup> In der linken Spalte werden die Aktivitäten benannt und rechts sind Einschätzungen des Evaluierungsteam dokumentiert. Das ermöglicht einen raschen Überblick. Im Original werden weitere 14 Aktivitäten und Bewertungen aufgelistet. Die Gefahr bei solchen kurzen Beurteilungen besteht darin, dass es Missverständnisse zwischen EvaluiererInnen und der Partnerorganisation unterstützen kann, wenn es sich um „wenig relevant“ eingestufte Aktivitäten handelt.

Abbildung 11 Gute Praxis Beispiel: Überblick zu den Ergebnissen der Relevanzbewertungen pro Aktivität

### Cohérence avec les attentes et possibilités des bénéficiaires

*Etant donné que la pertinence du projet peut différer d'une activité à une autre, nous avons préféré examiner les différentes principales activités du PIHVED par rapport à ce thème :*

Activité	Pertinence
Groupes de Recherche Action	Peu pertinent car les solutions cherchées peuvent être trouvées plus rapidement par d'autres voies
Embouche ovine, élevage, apiculture, arboriculture, savon, beurre de karité de qualité	La promotion des activités génératrices de revenus alternatives aux cultures d'hivernage est très pertinente, à condition que les activités proposées soient rentables et que leurs produits trouvent un marché rémunérateur
La promotion des pratiques DRS-CES	Très pertinente, mais il faut des stratégies de communication adéquates pour en convaincre les producteurs
L'aménagement des sites maraichers	Très pertinent si des conditions sont remplies: site approprié, route et marché pour un écoulement facile, modalités de captage et de distribution d'eau bien convenues avec les maraichers concernés
Distribution des semences maraichers et du riz	Non pertinent du tout, car les maraichers et riziculteurs concernés avaient déjà leur système d'approvisionnement en semences.
Production et distribution des arbres fruitiers et fertilisants	Pertinent pour ce qui est l'appui (raisonnable) aux pépiniéristes, pertinent quant à la sensibilisation et la formation sur l'utilité de l'agroforesterie, non pertinente quant à la constitution de groupes sans fonction particulières, et non pertinente quant à la distribution quasiment gratuite d'un nombre de plantes qui dépasse largement les capacités des membres de les soigner correctement ;
La mutualisation des moyens au sein des coopératives	Cela dépend des moyens, à voir cas par cas, dangereux de généraliser.
Mettre les coopératives en relation avec les IMF	La pertinence dépend des activités de l'OP ; ne pas limiter aux IMF, aussi voir les banques commerciales ;
Réaliser des points d'eau	Très pertinent à condition que la CFI soit raisonnable <sup>42</sup>

<sup>41</sup> Beispiel: Nr. 13 (Philippinen, ELBE)

<sup>42</sup> Evaluierung Nr. 217 (Burkina Faso, EQM)

**Effizienz.**<sup>43</sup> Die Bearbeitung des Effizienzkriteriums ist von unterschiedlicher Qualität und Ausführlichkeit. Sie orientiert sich an den von MISEREOR definierten Leitfragen zum Kriterium Effizienz. Die Herausforderung eines aussagekräftigen Vergleichs wird bei den untersuchten Evaluierungsberichten auf folgenden Ebenen bearbeitet:

- a) Vergleich der Organisationsstrukturen und des Monitoringsystems der Partnerorganisation mit effizienteren oder weniger effizienten Strukturen. Die Vergleiche erfolgten auf der Basis des Erfahrungswissens der GutachterInnen. In den Evaluierungen, bei denen dieser Maßstab eingesetzt wurde, wurde die Effizienz der Organisation jeweils als gut und angemessen und die Qualität des Monitoringsystems als eher schwach eingestuft.<sup>44</sup>
- b) Vergleich des Mittelabflusses anhand des geplanten Budgets und der Zahlen des letzten Finanzberichts. In den untersuchten Berichten wurde der Mittelabfluss als jeweils angemessen eingestuft.<sup>45</sup>
- c) Vergleich von Einheitskosten<sup>46</sup> mit internationalen Indikatoren bzw. Größenordnungen (Ausbildungskosten der Berufsbildung dürfen pro Kopf das zu erwartende zukünftige Jahreseinkommen der Person umfassen).<sup>47</sup> In einem anderen Fall wurden die Fortbildungskosten mit den von Geldgebern erwarteten Größenordnungen verglichen (Kosten für Capacity-Building-Maßnahmen sind auf dem Land doppelt so hoch wie in der Stadt).<sup>48</sup>
- d) Vergleich des Budgets für Reisekosten des Projekts mit dem einer staatlichen Organisation. Aufgrund der deutlich niedrigeren Kosten der NRO wurde dies als Hinweis für ihre Effizienz gewertet.<sup>49</sup>

Schwachpunkte im Bereich der Effizienz wurden in den untersuchten Berichten nur in wenig ausgebauten Monitoringsystemen identifiziert. Im Bereich der Effizienz fällt auf, dass nur selten Finanzdaten untersucht werden und oftmals einfach nur der Prozess eingeschätzt wird.

**Wirkungen.** Im Basischeck wurde festgestellt, dass eine Wirkungskette oder ein Wirkungsgefüge in etwa der Hälfte (31, 44%) der 70 Evaluierungen dargestellt wurde. In den anderen Evaluierungen ist es nicht erwähnt. Hier ist kein Unterschied zwischen ELBE und EQM-Evaluierungen und zwischen den Kontinenten zu finden. In der vertieften Analyse wurde die genaue Darstellung eines Wirkungsgefüges analysiert. 14 der 20 Studien haben kein Wirkungsgefüge bzw. keine Wirkungskette dargestellt, entwickelt oder diskutiert. Drei Evaluierungen nutzen ein schon vom Projekt entwickeltes Log Frame oder eine Wirkungskette.<sup>50</sup> Zwei Evaluierungen entwickeln zusammen mit den Projektmitarbeitern ein

---

<sup>43</sup> Auf Wunsch des Auftraggebers werden die Ergebnisse der beiden DAC-Kriterien „Effizienz“ und „Wirkungen“ (Outcome & Impact) im Folgenden vertieft bearbeitet.

<sup>44</sup> Beispielsweise: Nr. 185 (Palästinensische Gebiete, ELBE)

<sup>45</sup> Beispielsweise: Nr. 86 (Burkina Faso, ELBE)

<sup>46</sup> Definition der Einheitskosten: „Stückkosten oder Durchschnitts- bzw. Einheitskosten sind die auf eine Produkteinheit bezogenen Kosten“ (Arbeitskreis Entwicklungspolitik, DeGEval 2015: 8) Die angegebene Quelle stellt auch Informationen zur Berechnung und Nutzung von Einheitskosten in Projekt-Evaluierungen vor.

<sup>47</sup> Beispielsweise: Nr. 228 (Madagaskar, EQM)

<sup>48</sup> Beispielsweise: Nr. 187 (Südafrika, ELBE)

<sup>49</sup> Nr. 13 (Philippinen, ELBE)

<sup>50</sup> Siehe Nr. 12 (Philippinen, ELBE), Nr. 13 (Philippinen, ELBE), Nr. 217 (Burkina Faso, EQM)

Wirkungsgefüge oder eine Wirkungskette.<sup>51</sup> In einer Studie wird das Fehlen eines Wirkungsgefüges angemerkt und diskutiert.<sup>52</sup>

Wie in der MISEREOR-Mustergliederung für Referenzrahmen von Evaluierungen vorgesehen, wird zu den erzielten Wirkungen in den Kapiteln „Wirkungen (direkte und indirekte)“ (Outcome & Impact) sowie im Kapitel „Effektivität“ berichtet. Dabei ist festzustellen, dass es in den untersuchten Berichten in diesen Kapiteln zu Dopplungen kam oder in beiden Kapiteln nur ein Teil der Wirkungen dargestellt wurde, um sich nicht zu wiederholen. Dadurch werden Ergebnisse für Projektziele, die in der Regel auf Outcome-Ebene formuliert sind, teilweise unvollständig dargestellt.<sup>53</sup> Ein Beispiel: In den Berichten Nr. 238 (Indien, EQM) und Nr. 86 (Burkina Faso, ELBE) wird die Herausforderung so gelöst, dass unter Effektivität praktisch ohne Begründung die Erreichung der Ziele verkündet wird und die Begründung erst später im Kapitel Outcome & Impact erfolgt. Dies führte bei der Lektüre zunächst zu einer sehr schwachen Bewertung des Effektivitätskapitels. Oftmals erfolgt die Bearbeitung der DAC-Kriterien auch in der von MISEREOR gewünschten Reihenfolge, in der das Kapitel Wirkungen dem Effektivitätskapitel vorangestellt wird. Dadurch kommt es auch zu Dopplungen, da man zur Begründung der Wirkungen entlang der Wirkungskette die Wirkungen darstellt und langfristige Wirkungen mit kurzfristiger Zielerreichung, die wiederum Bestandteil des Kapitels Effektivität ist, begründet.

Um die Wirkungen plausibel auf das Projekt zurückzuführen, werden wenige Methoden angewendet, die das Kontrafaktische beachten. Bei der Analyse im Basischeck wird nur einmal eine Vergleichsgruppe hinzugezogen.<sup>54</sup> In der vertieften Analyse werden in einer Studie Familien, die nicht in der Zielgruppe sind, mit einbezogen.<sup>55</sup> Die Befragung einer größeren Anzahl von Personen aus den Zielgruppen nach Wirkungen wird eher selten vorgenommen. Ein gutes Beispiel hierfür ist die ELBE Nr. 51 aus El Salvador, in der mit einer Tendenzanalyse mit VertreterInnen der Zielgruppe und mit einer Einflussmatrix, die Wirkungen untersucht und bestimmt werden. Die Methoden *SWOT* und *Tiny Tool*, die im Basischeck bei 7 Evaluierungen identifiziert wurden, sind gleichfalls in der Lage, die Zuordnung von Wirkungen in Gruppen zu diskutieren.

---

<sup>51</sup> Siehe Nr. 51 (El Salvador, ELBE, dort im Bericht im Annex 5, Foto von erarbeiteter Wirkungskette) für gemeinsame Erstellung eines Log Frame mit Projektverantwortlichen und für Erstellung von Wirkungskette im Auftaktworkshop Nr. 302 (Bolivien, ELBE, im Auftaktworkshop wurde eine Wirkungskette diskutiert (dort S. 9 & S. 27), diese ist im Hauptteil des Berichts jedoch nicht dokumentiert, möglicherweise aber im Anhang, der fehlt).

<sup>52</sup> Siehe Nr. 211 (Kolumbien, EQM)

<sup>53</sup> Siehe z.B. Nr. 238 (Indien); bei Nr. 86 (Burkina Faso) wird die Herausforderung so gelöst, dass unter Effektivität praktisch ohne Begründung die Erreichung der Ziele verkündet wird und die Begründung erst später im Kapitel Outcome und Impact erfolgt. Dies führte bei der Lektüre zunächst zu einer sehr schwachen Bewertung des Effektivitätskapitels.

<sup>54</sup> Siehe Nr. 131 (Nigeria, EQM)

<sup>55</sup> Siehe Nr. 51 (El Salvador, ELBE)

### **Tiny Tools: Wer sagt eigentlich was?**

„Tiny Tools“ sind eine Zusammenstellung von Methoden, die den Gruppendiskussions- und Reflexionsprozess unterstützen. Die Zusammenstellung erfolgte durch das Netzwerk NGO-IDEAS<sup>56</sup> wobei insbesondere Bernward Causemann und Eberhard Gool die Initiative und Federführung hatten. Die Tools sind teilweise weiterentwickelte Methoden der sog. PRA-Toolbox (Participatory Rural Appraisal Methoden), die in den 1990er Jahren entwickelt und in der Entwicklungszusammenarbeit eingesetzt wurden.<sup>57</sup>

Die „Tiny Tools“ sind für die Arbeit von Gruppen zusammengestellt, die ihre Wirkungen und Erfolge selbst einschätzen wollen. Die Toolbox wurde mehrfach für die Wirkungsabschätzung von Projekten bei EQM-Evaluierungen genutzt. Die untersuchten Berichte machen keine Aussagen zur Nützlichkeit der eingesetzten Methoden oder zum Lerneffekt der Gruppen. Diese Dokumentation wurde auch nicht erwartet. Die Untersuchung zeigen jedoch, dass eine deutliche Steigerung der Aussagekraft für die jeweiligen Evaluierungen bei denen sie eingesetzt wurden.

Die vertiefte Analyse der Metaevaluierung zeigte, dass die Aussagekraft der Analysen, d.h. der Tools wie Trendanalyse oder die Einflussmatrix, deutlich gesteigert werden könnte, wenn im Evaluierungsbericht mehr Aussagen zur Auswahl und Zusammensetzung der Gruppen erfolgen würde: Warum wurden diese Gruppen ausgewählt? Wie können sie eingeordnet werden? Gehören sie nach Einschätzung der Partnerorganisation eher zu den „starken“, „normalen“ oder den „schwächeren“ Gruppen. Woran lässt sich das erkennen? Wie setzen sich die einzelnen Gruppen zusammen? Waren es überwiegend Frauen oder Männer, jüngere oder ältere? Waren Personen mit verschiedenem sozialem Status anwesend? Wer brachte sich insbesondere in die Diskussion ein? Welche Argumente wurden in der Diskussion ausgetauscht, die von besonderem Interesse für die Evaluierung sind? Welche Information konnten zur Triangulation wichtiger anderer Daten genutzt werden?

Diese Auswertungen und Dokumentationen der Gruppengespräche sind zeitaufwändig. Die Dokumentation von zentralen Argumenten, die belegen, warum die Gruppe z.B. die Mikrokredit- Aktivität des Projektes mit 4 von 5 Punkten als wichtig einschätzt, ist z.B. vorgesehen in MAPP-Workshops, einer umfangreicheren partizipativen Methode. In Arbeitsteilung mit einer EvaluierungskollegIn und durch die Reduzierung der Gruppengespräche, könnten diese Stärken jedoch auch für die Tiny Tools genutzt werden.

Wenn der Nachweis und die Dokumentation von Wirkungen im Bericht durch den Einsatz von Tiny Tools nicht notwendig ist, kann es auch sehr sinnvoll sein, die Anzahl der Gruppengespräche zu erhöhen. Die Rückmeldungen der Partnerorganisation auf die Evaluierung Nr. 238 in Indien zeigen, dass die eingesetzten Tools sehr geschätzt wurden. Besonders hervorgehoben wurden das Kausal-Diagramm, die Trendanalyse, das Proportionale Pilling, die Aktivitätenliste und das Kasese Tool.<sup>58</sup>

---

<sup>56</sup> Siehe <http://www.ngo-ideas.net/home/index.html>; Zugriff 31.08.2016

<sup>57</sup> Siehe (NGO-IDEAs 2012); zu den Ursprüngen und zum Kontext der Participatory Rapid Appraisal-Methoden (Chambers 1994).

<sup>58</sup> Siehe mehr Informationen zu den einzelnen Tools im Dokument (NGO-IDEAs 2012)



**Qualität der Dokumentation der Daten.** Die klarste und umfangreichste Dokumentation erfolgte in den untersuchten Berichten für die Ergebnisse von standardisierten Befragungen. In der Regel erfolgt die Darstellung in Tabellen und Schaubildern im Text, ggf. ergänzt mit Informationen im Anhang.<sup>59</sup> Darauf folgt die Dokumentation von Gruppenbefragungen, die z.B. über die Aufzählung von Argumenten im Hauptteil oder als Anlage erfolgt.<sup>60</sup> Die Darstellung von Ergebnissen von SWOT-Analysen<sup>61</sup> und Tiny-Tool-Analysen<sup>62</sup> erfolgt in Tabellen im Hauptteil des Berichts. Die Dokumentation von Interviews erfolgte in einzelnen Anekdoten (eine Frau sagte...) oder gebündelt in Serien von Zitaten, die verschiedene Sichtweisen von Personen darlegen.<sup>63</sup>

Generell wird aber wenig strukturiert z.B. in Form von Tabellen oder Grafiken. Zumeist werden die gesammelten Daten und Informationen nur im Text summarisch wiedergegeben. In vielen Evaluierungen wird durchaus eine größere Anzahl von 30 - 100 Personen befragt, so dass dies möglich wäre.

**Trennung von Beschreibung und Bewertung:** Nur in einem der 20 intensiv untersuchten Berichte wurde eine eindeutige Trennung von Beschreibung in den Kapiteln „Ergebnisse“ oder sowie der Darstellung der Bewertungen im Kapitel „Analyse der Ergebnisse“ bzw. „Schlussfolgerungen“ vorgenommen.<sup>64</sup> In den Kapiteln, in denen die Ergebnisse nach DAC-Kriterien untergliedert sind, wurden überwiegend Beschreibungen der Sachlage und getrennt davon Bewertungen vorgenommen.

Die festgestellte Logik folgt der DAC-Kriterien, die Bewertungskriterien sind. Es gibt Berichte, die die Bewertung nach DAC-Kriterien vollständig im Kapitel „Schlussfolgerungen“ vornehmen. In den untersuchten Berichten fielen dann diese DAC-Bewertungen jedoch eher knapp aus. Eine deutliche Trennung von Beschreibungen und Bewertungen konnte immer dann vorgenommen werden, wenn zunächst eine Beschreibung des Sachverhalts, möglichst mit Bezug auf die Quellen und dann eine Bewertung ggf. mit Begründung bzw. weiterer Erläuterung vorgenommen wurde. Ein gutes Beispiel ist hierfür Nr. 211 (Kolumbien, EQM). In dieser Studie wird zusätzlich das Vorgehen zur Bewertung vorab jeweils erklärt, was sich als sehr hilfreich erweist und die Unterscheidung zwischen Beschreibung und Bewertung erleichtert (vgl. Abbildung 12)

---

<sup>59</sup> Siehe Nr. 182 (Malawi) und 86 (Burkina Faso)

<sup>60</sup> Siehe Nr. 187 (Südafrika)

<sup>61</sup> Siehe Nr. 185 (Palästinensische Gebiete)

<sup>62</sup> Siehe Nr. 238 (Indien)

<sup>63</sup> Siehe Nr. 185 (Palästinensische Gebiete)

<sup>64</sup> Siehe Nr. 13 (Philippinen, ELBE)

## 1. Beschreibung des Rahmens

### VI. Eficacia de los Proyectos: Cumplimiento de los Objetivos

Las intervenciones de los dos proyectos evaluados marcan dos procesos distintos que no tienen conexión entre ellos. En este sentido no se trata de proyectos de continuación que profundizan procesos iniciados en el proyecto anterior. La eficacia es evaluada entonces proyecto por proyecto. Como el proyecto actual todavía no ha terminado la apreciación de su eficacia sólo puede ser parcial.

## 2. Limitierungen der Analysemöglichkeiten

La valoración de la eficacia de los dos proyectos se ve enfrentada con algunas dificultades e inconsistencias en su sistema de planificación.

(Auflistung von Problemen (hier nicht dargestellt))

## 3. Vorgehen zur Analyse der Effektivität

La siguiente valoración de la eficacia de los dos proyectos tiene como base los objetivos formulados en los dos contratos de cooperación. Con respecto al proyecto actual se hace también referencia a los objetivos e indicadores reformulados. Informaciones encontradas en los informes de actividades para Misereor y en informes internos o proveídas por el equipo de proyecto son complementadas por evidencias parciales de las entrevistas y visitas de campo realizadas. Con una excepción no encontramos una memoria institucional en personas que habían trabajado en el proyecto anterior y que nos hubieran podido proveer información más allá de lo escrito en los informes para Misereor. El equipo del proyecto actual no tiene conocimiento del proyecto anterior.

## 4. Analyse

### 2. Eficacia del Proyecto "Incidencia Sociopolítica ..." (No. 226-005-1109 ZG)

Objetivo 1 del Proyecto:

Las instituciones, organizaciones de base y jóvenes participantes de los barrios acompañados establecerán sinergias a través del proyecto para mejorar la articulación y la cooperación

En la preparación de la convocatoria para seleccionar a los 60 jóvenes beneficiarios el proyecto visitó 21 organizaciones juveniles y barriales, 11 universidades, 5 instancias públicas y 3 organismos de cooperación internacional para presentar el proyecto e estimular jóvenes de estas organizaciones concursar.<sup>32</sup> Durante la ejecución del proyecto este no trabajó sistemáticamente con estas organizaciones para estimular su articulación y cooperación. Algunas de ellos estaban involucradas en la elaboración de los diagnósticos participativos.

Bajo el componente "Sensibilización" el proyecto buscó la cooperación con centros educativos para realizar los eventos de sensibilización. Pero el proyecto no trabajó explícitamente con instituciones y organizaciones de base en los barrios. Como no pudimos visitar y entrevistar ninguno de los centros educativos o de las organizaciones juveniles barriales no hay evidencias que este objetivo es cumplido. Más bien es fuera del propósito del proyecto.

**Grundlagen der Schlussfolgerungen und Empfehlungen.** In den Berichten, in denen die Beschreibung und Bewertung klar getrennt werden und die Bewertungen auf den Ergebnissen basieren, sind die Schlussfolgerungen aus den Daten nachvollziehbar. In keinem der untersuchten Berichte konnte dies nur auf der Grundlage der Lektüre des Kapitels „Schlussfolgerungen“ vorgenommen werden, da dort nicht die notwendigen Verweise auf die Stellen des Hauptteils angeführt waren. Für die Prüfung der Nachvollziehbarkeit war jeweils eine komplette Lektüre des Berichts notwendig. In den Hauptteilen der untersuchten Berichte sind Querverweise Mangelware. Es gibt jedoch auch die Kategorie von Evaluierungen, deren Schlussfolgerungen plausibel sind, auch wenn diese nicht direkt mit den erhobenen Daten z.B. aus der Zielgruppen unterlegt werden. Diese Plausibilität der Schlussfolgerungen beruht dann auf dem Zusammenhang zwischen beschriebenem Kontext und den Projektaktivitäten. Ein Beispiel sind die Bündel an Empfehlungen für Frauen-Selbsthilfegruppen, die sich aus solchen Schlussfolgerungen ergaben.<sup>65</sup> Unterstützung des Zugangs von weiteren Frauengruppen zu Mikrokrediten, Unterstützung der Gruppen in der Einforderung von staatlichen Unterstützungsleistungen und Unterstützung der Frauen in der Vermarktung ihrer Produkte. Die Grenzen solcher Plausibilitätsannahmen zeigen sich z.B. in der praktisch ungeprüften Übernahme der These, dass Gruppenbildungen von benachteiligten Zielgruppen hilfreich und im Interesse der Beteiligten sind. Darauf verweist eine EQM-Evaluierung in Burkina Faso, die als Zwischenevaluierung empfiehlt, die Erfahrungen mit der Gruppenbildung von männlichen Viehzüchtern zu analysieren, bevor in einer neuen Projektphase wieder eine Gruppenbildungsphase initiiert werden sollte.<sup>66</sup>

Eine transparente und klare Ableitung im Sinne von Datenanalyse -> Schlussfolgerung -> Empfehlung lässt sich durchgängig in kaum einer der untersuchten Evaluierungen feststellen.<sup>67</sup> Für manche Empfehlungen liegt dieser Zusammenhang vor, bei anderen erscheint die Empfehlung zwar plausibel, aber nicht direkt aus den dokumentierten Schlussfolgerungen ableitbar. Teilweise wird auch die Dokumentation der Schlussfolgerung übersprungen und direkt mit den Empfehlungen begonnen. Dies geschieht nicht nur bei „schwachen“ Evaluierungen sondern durchaus auch bei als stark eingestuften ELBE.<sup>68</sup>

## Fazit

Qualitative Interviews und Fokusgruppendifkussionen prägen die Erhebungsmethoden der ELBE und EQM-Evaluierungen. Partizipative Methoden wie SWOT (Analysen von Stärken, Schwächen, Chancen und Bedrohungen) und „Tiny Tools“ (Methodensammlung auf der Grundlage von PRA-Methoden) und semi-standardisierte Fragebogenerhebungen werden weniger eingesetzt. In der Umsetzung der Datenerhebung werden, oft weit ab der städtischen Zentren, viele VertreterInnen der Zielgruppen vor Ort gesprochen und die Situation vor Ort in Augenschein genommen. In den Berichten ist das methodische Vorgehen beschrieben, Schwachpunkt ist jedoch die Dokumentation und

<sup>65</sup> Siehe zum Beispiel die Empfehlungen in Nr. 103 (Indien). Dieser Zusammenhang ist ein Hinweis, dass der Evaluator möglicherweise ein langjähriger Berater oder Projektmanager in diesem Kontext ist.

<sup>66</sup> Siehe Nr. 217

<sup>67</sup> Ein Beispiel, in dem dies weitgehend gut gelingt, ist Nr. 211 (Kolumbien, EQM)

<sup>68</sup> Ein Beispiel ist Nr. 13 (Philippinen), die seitens der Kontinentalabteilung und dem Meta-Evaluierer als gut eingestuft wurde.

Nachvollziehbarkeit der Stichproben für die Datenerhebung: Wer und wie wurde über die Auswahl der InterviewpartnerInnen, ZielgruppenvertreterInnen und Projektgebiete entschieden. Daten werden zwischen den Aussagen der Partnerorganisation und der Zielgruppen trianguliert, jedoch erfolgt dies wenig über die Projektgrenzen hinweg, z.B. bei externen ExpertInnen, der aktuellen Fachdiskussion oder „Gute-Praxis“-Beispiele und Erfahrungswissen von MISEREOR. In den Berichten wird die klare Trennung von Beschreibung und Bewertung nur teilweise eingehalten und eine Untergliederung der Dokumentation der erhobenen Daten bei den Zielgruppen nach Alter, Geschlecht, Wohlstand und anderen sozialen Kriterien ist wenig verbreitet.

## 5.4 Nützlichkeit

### 1. Hintergrund

In der MISEREOR-Mustergliederung für den Referenzrahmen von Evaluierungen, die gleichermaßen für ELBE und EQM-Evaluierungen gilt, werden im Abschnitt 6 (Bericht) Hinweise für die Berichterstattung gegeben.

- a) In einer knappen (2-3 seitigen) Zusammenfassung sollen die wichtigsten Ergebnisse dargestellt werden, damit später schneller Zugriff auf die Informationen erfolgen kann.
- b) Die Evaluierungsfragen sollen nachvollziehbar und vollständig beantwortet werden und den Schwerpunkt gemäß den TOR umsetzen.
- c) Die Empfehlungen sollen sich „logisch“ aus den Schlussfolgerungen ergeben. Sie sollen Bezug auf die Akteure nehmen und Prioritäten setzen.<sup>69</sup>

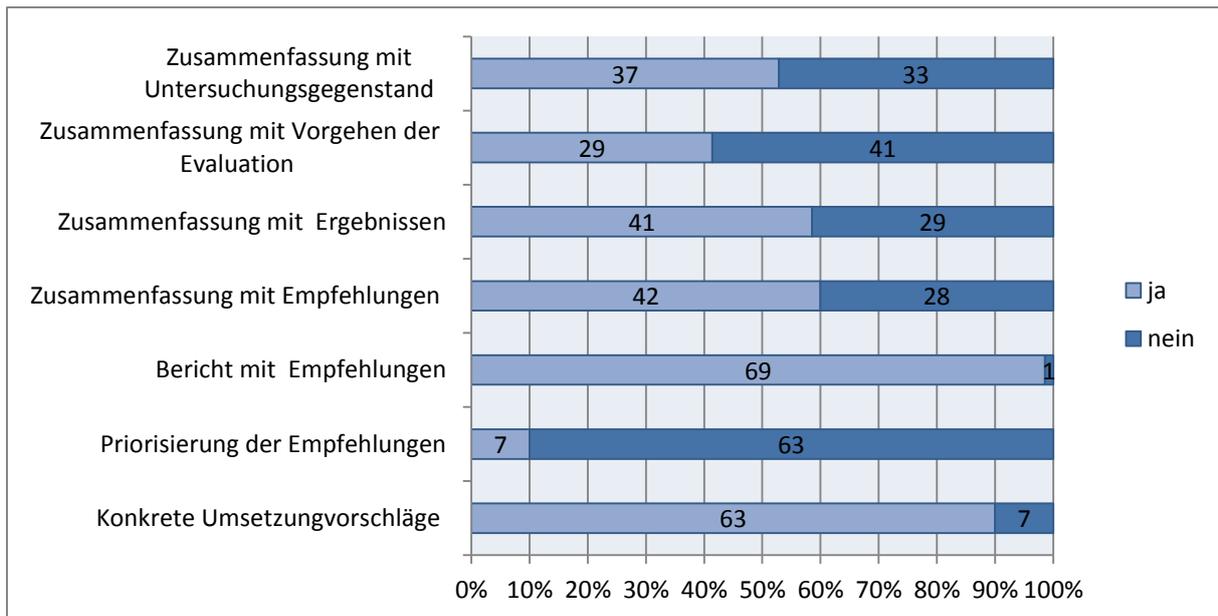
### 2. Ergebnisse des Qualitäts-Basischecks

Wie schon unter der Dokumentation beschrieben ist in 20 Evaluierungen (29%) keine Zusammenfassung im Bericht enthalten und 50 Berichte (71%) enthalten eine Zusammenfassung. In 37 Evaluierungen wird in der Zusammenfassung der Untersuchungsgegenstand vorgestellt, in 29 Berichten (41%) das Vorgehen der Evaluierung. Ergebnisse und Empfehlungen werden von 41 (59%) und 42 (60%) Berichten in der Zusammenfassung aufgegriffen (vgl. Abbildung 13).

---

<sup>69</sup> Siehe Dokumente (MISEREOR o.J.: 3 (Anlage)) und (MISEREOR, EQM 2014: 25)

Abbildung 13 Nützlichkeit



Alle Berichte bis auf einen enthalten Empfehlungen. In 63 Evaluierungen (90%) sind diese nicht priorisiert. 63 der Evaluierungen (90%) enthalten konkrete Umsetzungsvorschläge und sind pragmatisch. 7 Berichte (10%) enthalten Empfehlungen, die nur sehr allgemeine Vorschläge aussprechen und nicht direkt Ansätze zur Umsetzung mitliefern.

Tabelle 5 Nützlichkeit ELBE vs. EQM

	ELBE (56)	EQM (14)
sehr gut	21% (12)	86% (12)
gut-befriedigend	34% (19)	14% (2)
eher schwach	45% (25)	0
schwach	0	0

Im Vergleich sind EQM-Evaluierungen sehr viel höher im Hinblick auf ihre Nützlichkeit zu bewerten als die ELBE. 86% (12) der EQM-Evaluierungen sind hier sogar mit sehr gut bewertet und nur 12% (12) der 56 ELBE. Das heißt diese Evaluierungen haben sowohl bei der Zusammenfassung alle Inhalte als auch pragmatische Empfehlungen mit eingebracht. Im Bereich der ELBE sind 45% (25) mit eher schwach bewertet. Dies liegt vor allem an fehlenden Zusammenfassungen. In 20 Fällen fehlt eine Zusammenfassung, 17 Mal in Lateinamerika, zweimal in Afrika und einmal in Asien. All diese Evaluierungen sind ELBE.

### 3. Vertiefte Analysen

**Zusammenfassungen der Berichte.** In den untersuchten Berichten gibt es alle Variationen. In der Regel wurden die Zusammenfassungen als gut lesbar und informativ gewertet.<sup>70</sup> Zusammenfassungen enthalten häufig weniger kritische Elemente, als im Kapitel Schlussfolgerungen oder im Hauptteil dargestellt werden. In vielen Berichten findet sich zum Teil keine Zusammenfassung, was bei sehr langen Berichten definitiv als negativ zu bewerten ist. Wenn Berichte eher kürzer sind, wie zum Beispiel bei Nr. 109 (Indien, ELBE, 8 Seiten) oder 187 (Südafrika, ELBE, 16 Seiten) erscheint es nicht unbedingt notwendig. Zum Teil ist das Fehlen einer Zusammenfassung schon der erste Hinweis darauf, dass eine Studie insgesamt methodisch eher schwächer ist und wichtige Informationen zum Vorgehen in der Evaluierung nicht dargelegt sind.

**Beantwortung der Evaluierungsfragen.** Alle untersuchten Berichte gaben Antworten auf die Evaluierungsfragen. Unterschiedlich war jedoch die Qualität und Nachvollziehbarkeit der Ausführungen. Es gibt Beispiele, in denen die Darstellung von Ergebnissen übersprungen und immer direkt bewertet wird. Der Hauptteil Ergebnisse wird überwiegend auf der Grundlage der DAC-Kriterien strukturiert. Es gab aber auch Berichte, in denen die DAC-Kriterien erst im Kapitel Schlussfolgerungen eingebracht wurden<sup>71</sup> und einen Bericht, in dem die Kriterien Relevanz, Effektivität und Effizienz unter dem Kapitel „Ergebnisse“ dargestellt und die Kriterien Impact und Nachhaltigkeit quasi als Schlussfolgerungen dargelegt wurden.<sup>72</sup> In den untersuchten Berichten lagen die Schwerpunkte auf den Untersuchungskriterien Effektivität und wenn vorhanden, auf dem zusätzlichen Kapitel Wirkungen („Outcome und Impact“). Teilweise wird das Kriterium Relevanz intensiv behandelt, die Kriterien Effizienz und Nachhaltigkeit nehmen eher eine nachgeordnete Bedeutung ein. Dieses Vorgehen entspricht den Schwerpunktsetzungen bei den TOR.

**Empfehlungen.** In der vertieften Analyse von 20 Evaluierungsberichten führen alle, auch die seitens der MISEREOR-Kontinentalabteilungen als „schwächer“ eingestuftten Berichte, zu praxisnahen und operationalisierbaren Empfehlungen. Kein Bericht unterstützte die Nachvollziehbarkeit der Empfehlungen durch eine Dokumentation im Sinne von: Ergebnis -> Schlussfolgerung 1 -> Empfehlung 1. Die Empfehlungen richten sich überwiegend an die Partnerorganisationen. Wenn es bei ELBE eine Empfehlung an MISEREOR gibt, dann ist es die, das Projekt weiter zu fördern. Einmal wird auf die Möglichkeit eines Coaching der Partnerorganisation durch MISEREOR verwiesen, um so Capacity Building zu betreiben. Eine Priorisierung der Empfehlungen findet zumeist nicht statt. Jedoch fällt in der vertieften Analyse auf, dass zum Teil die am Schluss gegebenen Empfehlungen, eine Zusammenfassung vieler kleiner während des Berichts gegebener Empfehlungen sind.

#### Fazit

In der Regel wurden die Zusammenfassung als gut lesbar und verständlich eingestuft. Eher selten ließen sich die Empfehlungen aus dem Kapitel *Schlussfolgerungen* vollständig und logisch ableiten. Für eine solche Ableitung war die Lektüre des gesamten Kapitels *Ergebnisse* notwendig. Des Weiteren gab es Empfehlungen, die auch nach Lektüre des gesamten Berichts nicht aus dem Inhalt des Berichts abzuleiten waren.

<sup>70</sup> Beispiele für Gute Praxis sind Nr. 301 (Mexiko) und Nr. 211 (Kolumbien)

<sup>71</sup> Ein Beispiel ist Nr. 185 (Palästinensische Gebiete)

<sup>72</sup> Nr. 103 (Indien)

## 5.5 Konsistenz der TOR

### 1. Hintergrund

Für EQM-Evaluierungen und ELBE gilt die gleiche Mustergliederung. In der Einleitung des Musters wird darauf hingewiesen, dass der Referenzrahmen für jede Evaluierung individuell formuliert werden sollte, um zu gewährleisten, dass er für die jeweilige Projektsituation und die jeweiligen Entwicklungsziele geeignet sei.<sup>73</sup> Insbesondere sollen Evaluierungsfragen abgewandelt, genauer formuliert oder weggelassen werden. Relevante Querschnittsthemen sollen mit einbezogen werden.

### 2. Vertiefte Analysen

Im Qualitäts-Basischeck wurde aus Zeitgründen auf die Konsistenz der TOR in den Evaluierungsberichten nicht eingegangen. Die hier vorgenommene Analyse basiert auf der Untersuchung von 20 Evaluierungen. Die TOR waren bei allen 5 EQM-Evaluierungen sowie bei 5 von 15 untersuchten ELBE den Berichten beigelegt.

**Fragen der TOR.** Die umfangreichsten Anpassungen der TOR lassen sich bei den EQM-Evaluierungen feststellen. Hier werden in der Regel die „Standardfragestellungen“ der Mustergliederung für die DAC-Kriterien verändert, um die TOR an das Projekt anzupassen. Dies führt dazu, dass die Anzahl der Fragestellungen der Mustergliederung von 20 Fragen auf über 30 bis fast 50 Evaluierungsfragen stark erweitert werden.<sup>74</sup> Es gibt jedoch auch TOR von EQM-Evaluierungen bei denen die Fragestellungen stark reduziert wurden.<sup>75</sup> Bei den ELBE, die die TOR als Annex dokumentierten, hatten zwei deutlich mehr Fragen als die Mustergliederung, eine entsprach der Mustergliederung und bei zwei TOR war die Anzahl der Fragen reduziert.

Einzelne Fragen sind in der Regel verständlich und präzise formuliert. Wird jedoch das gesamte Bündel an Fragen mit Blick auf die DAC-Kriterien untersucht, so nimmt diese Klarheit deutlich ab und es erfolgen Dopplungen von Fragen zwischen den Kapiteln Relevanz, Wirkungen und Effektivität.

---

<sup>73</sup> Siehe die Dokumente (MISEREOR o.J.: 1 (Anlage)) und (MISEREOR, EQM 2014: 22)

<sup>74</sup> Beispiele hierfür sind Nr. 238 (Indien, EQM) mit 47 und Nr. 228 (Madagaskar, EQM) mit 36 Evaluierungsfragen.

<sup>75</sup> Ein Beispiel ist Nr. 217 (Burkina Faso, EQM) in der die TOR unter den DAC-Kriterien ausweisen, was unter diesen Kriterien zu bewerten ist. Dabei wird keine Frage gestellt.

### **Beispiel für Dopplung von Fragen in verschiedenen DAC-Kriterien:**

In einem Projekt zur der Kapazitätsbildung von Minderheiten, das insbesondere Trainings zur Erreichung ihrer Ziele durchführt, wird in den DAC-Kriterien u.a. gefragt:<sup>76</sup>

**Relevanz:** How far did the trainings enable the participants to find solutions for the problems they face in the Village/Parish, Society?

**Outcome & Impact:** What are the outcomes and impact that can be observed for the implementation of the project at Personal, Community, Village and Church level? What are the expected and unexpected outcomes of the training?

**Effektivität:**<sup>77</sup> Have the intended objectives of trainings been achieved? If so, to what extent? If not why?

In den untersuchten Berichten gab ein Evaluierungsteam die Rückmeldung zu den TOR, dass in dem vorliegenden Projekt eine Unterscheidung zwischen dem DAC-Kriterium Relevanz und Effektivität nicht gegeben sei und daher die Analyse in einem Kapitel vorgenommen würde.<sup>78</sup> Diese Unklarheit über die Abgrenzung zwischen Relevanz und Effektivität bzw. Relevanz und Wirkungen lässt sich auch in anderen Evaluierungsberichten beobachten, in denen es zu deutlichen Dopplungen in diesen Kapiteln kommt.<sup>79</sup>

**Querschnittsthemen.** Von den 5 untersuchten EQM-Evaluierungen thematisieren 3 Gender und eine Ökologie als Querschnittsthema in den TOR. Bei den untersuchten ELBE formuliert ein Referenzrahmen Gender als Querschnittsthema.

**Mengengerüst.** Bei den untersuchten Berichten gab es eine Tendenz, dass das Mengengerüst für EQM-Evaluierungen deutlich höher ist als das von ELBE. EQM-Evaluierungen sehen in den untersuchten Evaluierungen in der Regel ca. 2-3 Wochen Feldaufenthalt und insgesamt ca. 25-30 Tage pro EvaluatorIn vor. Bei ELBE stehen oft nur 1-2 Wochen Feldaufenthalt zur Verfügung mit ca. 15-20 Tagen.

Die Tabelle 6 zeigt die *Kosten der Evaluierungen*, das *Budget der aktuellen Projektlaufzeit* sowie den *Anteil der Kosten für Evaluierung am Projektbudget* sowie den *Proxy: Anteil Kosten für Evaluierung über 3 Projektphasen*. Der Proxy in der rechten Spalte wurde eingefügt, da für die Beurteilung des Anteils der Evaluierungskosten das Gesamtvolumen des untersuchten Projekts zu berücksichtigen ist. Dabei sei darauf hingewiesen, dass in der Regel MISEREOR-geförderte Projekte bis 1 Mio. Euro nur in jeder dritten Projektphase evaluiert werden. Auf internationalen Fachkonferenzen werden inzwischen Stimmen laut, dass sich der Anteil der Evaluierungskosten idealerweise auf 3-5% des Projektbudgets belaufen sollte.<sup>80</sup> So wird der seit 2009 empfohlene minimale Kostenrahmen von 1-2 %<sup>81</sup> für NRO, in

<sup>76</sup> Siehe TOR der Evaluierung Nr. 238 (Indien, EQM)

<sup>77</sup> Anmerkung zu Punkt Effektivität: Die Zielsetzungen des Projekts sind auf Vorgabe von MISEREOR auf Outcome-Ebene definiert.

<sup>78</sup> Siehe Nr. 13 (Philippinen, ELBE)

<sup>79</sup> Siehe z.B. Nr. 301 (Mexiko, ELBE)

<sup>80</sup> Austrian Development Agency (2008), WHO (2013)

<sup>81</sup> Bormann und Stockmann (2009).

dem sich auch ein Großteil der Untersuchungen von MISEREOR befindet, nicht mehr als „Best Practice“ erachtet.

**Tabelle 6 Kosten der Evaluierung und Projektkosten**

Nr.	EQM/ELBE	Evaluierung	Kosten der Evaluierung	Budget der aktuellen Projektphase	Anteil Kosten für Evaluierung an Projektbudget	Proxy: Anteil Kosten für Evaluierung über 3 Projektphasen
13	ELBE	Philippinen	854,52 €	235.000,00 €	0,36%	0,12%
128	ELBE	Brasilien	1.939,73 €	215.000,00 €	0,90%	0,30%
16	ELBE	Philippinen	2.360,49 €	54.000,00 €	4,37%	1,46%
54	ELBE	Guatemala	3.400,89 €	130.000,00 €	2,62%	0,87%
51	ELBE	El Salvador	4.190,93 €	125.000,00 €	3,35%	1,12%
12	ELBE	Philippinen	4.195,85 €	126.000,00 €	3,33%	1,11%
103	ELBE	Indien	4.731,80 €	97.000,00 €	4,88%	1,63%
185	ELBE	Palästinensische Gebiete	9.702,20 €	243.000,00 €	3,99%	1,33%
191	ELBE	Simbabwe	11.207,80 €	240.000,00 €	4,67%	1,56%
187	ELBE	Südafrika	11.199,44 €	370.000,00 €	3,03%	1,01%
182	ELBE	Malawi	13.259,17 €	180.000,00 €	7,37%	2,46%
86	ELBE	Burkina Faso	15.238,39 €	380.000,00 €	4,01%	1,34%
217	EQM	Burkina Faso	15.700,00 €	1.100.000,00 €	1,43%	<sup>82</sup>
228	EQM	Madagaskar	23.000,00 €	304.000,00 €	7,57%	2,52%
204	EQM	Indonesien	23.000,00 €	450.000,00 €	5,11%	1,70%
211	EQM	Kolumbien	23.300,00 €	382.000,00 €	6,10%	2,03%
238	EQM	Indien	25.000,00 €	109.000,00 €	22,94%	7,65%

Vier Evaluierungen, die jeweils eine internationale und eine nationale EvaluatorsIn einsetzen befinden sich im Kostenrahmen von ca. 23-25.000 €. Alle haben deutlich mehr Fragestellungen, als die Mustergliederung der TOR ausweist. Dann folgt mit gut 15.000 € eine EQM-Evaluierung in Burkina Faso, die gleichfalls von zwei GutachterInnen durchgeführt wurde. Diese hatte den oben aufgeführten reduzierten Fragenkatalog der TOR. Darauf folgt eine extern unterstützte Selbstevaluierung, die gleichfalls in Burkina Faso vorgenommen wurde. Mit gut 13.000 € schließt sich eine Evaluierung aus Malawi an, die auf den Einsatz einer umfangreichen (200 Frauen/Haushalte) semi-standardisierten Fragebogenerhebung im ländlichen Raum setzte. Schlusslichter sind die ELBE in den Philippinen, in der Universitäts-Fachpersonal eingesetzt wurde, die ELBE in Brasilien und nochmals eine ELBE aus den Philippinen.

Bei den Relationen zwischen Evaluierungskosten und Projektvolumen sticht mit ca. 7,6% (über 3 Projektlaufzeiten) die EQM-Evaluierung in Indien hervor.

<sup>82</sup> Projektvolumen von über 1 Mio. Euro werden nach den MISEREOR-Leitlinien in jeder Projektphase evaluiert.

## Fazit

Die TOR führen durch die Ausweitung des DAC-Kriteriums „Impact“ zu „Outcome & Impact“ (Wirkungen, direkte und indirekte) zu inhaltlichen Dopplungen mit dem Kapitel Effektivität. Dies wird auch durch die geänderte Reihenfolge der DAC-Kriterien weiter befördert. Das führt teilweise dazu, dass die festgestellten „Wirkungen“ auf die beiden Kapitel aufgeteilt werden, um Wiederholungen im Text zu minimieren. Zu Dopplungen kommt es auch zwischen den Kapiteln „Relevanz“ und „Wirkungen“. Querschnittsthemen werden selten in den TOR vertieft. In den untersuchten EQM-Evaluierungen war das Mengengerüst (Arbeitstage, Finanzvolumen, Umfang der Fragestellungen der TOR) teilweise deutlich höher als das der ELBE.

## 5.6 Partizipation / Beteiligung

### 1. Hintergrund

Für EQM-Evaluierungen gelten die im MISEREOR-Evaluierungskonzept festgelegten Standards zu Partizipation und Fairness. Die zentralen Punkte sind:<sup>83</sup>

- a) Partnerorganisationen sollen den Referenzrahmen mit entwickeln.
- b) Der Evaluierungsbericht soll dem Gebot der Fairness entsprechen und daher Stärken und Schwächen ausgewogen darstellen und verschiedene Perspektiven berücksichtigen.
- c) Am Ende des Evaluierungsprozesses erhalten die Partnerorganisationen die Möglichkeit der Rückmeldung zum Evaluierungsprozess.

Weitere konkrete Hinweise für GutachterInnen in puncto Partizipation werden im Anhang 1 des Evaluierungskonzepts ausgeführt. Dabei werden Mindeststandards dargelegt; die zentralen Punkte sind:

- a) Einbeziehung der Partnerorganisation bei der Konkretisierung des methodischen Vorgehens und Sicherstellung einer unparteiischen Auswahl der zu evaluierenden Projektbereiche.
- b) Nutzung der (Monitoring-) Daten der Partnerorganisation; ggf. Kommentierung der Datenqualität.
- c) Diskussion der vorläufigen Evaluierungsergebnisse am Ende des Aufenthalts (Abschlussworkshop); deutlich abweichende Einschätzungen der Partnerorganisation werden im Bericht dokumentiert.
- d) Die Perspektive der Zielgruppen soll durch Interviews und/ oder partizipative Erhebungsmethoden intensiv in die Evaluierung mit einfließen; Berücksichtigung der Heterogenität der Zielgruppen.

---

<sup>83</sup> MISEREOR, EQM 2014: 11

- e) Abwägung der Vor- und Nachteile, die durch die Unterstützung des Projektpersonals bei den Erhebungen vor Ort entstehen können (Beeinflussung versus besseren Zugang zu den Zielgruppen).

Als gute Praxis im Bereich der Partizipation wird seitens MISEREORs empfohlen, die Partnerorganisation, sofern sinnvoll, im Evaluierungsverlauf in den Erkenntnisprozess einzubeziehen.<sup>84</sup>

Für ELBE werden in der Handreichung unter dem Punkt „Methodik der Evaluierung“ deutlich weniger Hinweise gegeben:

- a) MISEREOR empfiehlt eine partizipative Methodik, d.h. die Einbeziehung der ProjektmitarbeiterInnen und der Begünstigten des Projektes in die Datenerhebung und Auswertung.
- b) Eine Selbstevaluierung kann akzeptiert werden, wenn sie durch eine unabhängige GutachterIn begleitet wird, die eine eigenständige Außensicht einbringt und diese in einem unabhängigen, ggf. ergänzenden Bericht dokumentiert.

## 2. Vertiefte Analysen

In der vertieften Analyse standen im Bereich Partizipation Fragen zur Beteiligung der Partnerorganisation oder der Zielgruppen in Planung und Durchführung des Evaluierungsprozesses im Mittelpunkt. Im Hinblick auf die Beteiligung der Partnerorganisation bei der Planung und Durchführung des Evaluierungsprozesses ist die Evaluierung oftmals partizipativ gestaltet. 11 von 20 Evaluierungen werden als partizipativ bewertet. Es finden sich in diesen Evaluierungen Hinweise auf Abstimmungsprozesse zum Vorgehen, auf Mithilfe bei Auswahl und Auffinden von Interviewpartnern und auf die gemeinsame Diskussion der Ergebnisse.

Nur wenige Evaluierungen beteiligen die Zielgruppen in Planung und Durchführung.<sup>85</sup> Hier gibt es Hinweise darauf, dass Zielgruppen zum Teil am Auftakt- oder Abschlussworkshop teilnahmen. Den Hinweis, dass Zielgruppen aktiv an der Planung und Durchführung des Evaluierungsprozesses beteiligt waren, gab es nur in einem Fall.<sup>86</sup> Dabei wurden 4 Frauen aus den 12 Gemeinden des Projekts sowie 4 Personen des Personals der Partnerorganisation in das Evaluierungsteam aufgenommen und haben gemeinsam die Instrumente entwickelt. Die drei entdeckten Fälle der Einbindung der Zielgruppen sind ELBE-Evaluierungen.

Weitere Hinweise zur Partizipation können auch aus anderen Bereichen der vertieften Analyse wie zum Beispiel aus dem Stattfinden eines Auftakt- und Abschlussworkshops, der Einbringung der Perspektiven der Zielgruppen (auch in ihrer Heterogenität) und aus der Anwendung von partizipativen Methoden abgeleitet werden. Auftakt- und Abschlussworkshops gab es bei allen untersuchten EQM-Evaluierungen. Bei ELBE sind diese Workshops nicht immer klar dokumentiert, es wurde aber in allen untersuchten ELBE Hinweise für Auftaktgespräche und Diskussionen des Berichtsentwurfs gefunden. Wer an

---

<sup>84</sup> MISEREOR, EQM 2014: 20

<sup>85</sup> In den 20 intensiv untersuchten Evaluierungen wurde dies in 3 von 20 Fällen umgesetzt: Nr. 51 (El Salvador, ELBE), Nr. 54 (Guatemala, ELBE), Nr. 301 (Mexiko, ELBE)

<sup>86</sup> Siehe Nr. 301 (Mexiko)

den Auftakt- und Abschlussworkshops, bzw. Auftakt- und Abschlussgesprächen teilnahm, ist zum Teil dokumentiert.

Eine als Selbstevaluierung deklarierte Evaluierung setzte einen externen Gutachter zur Unterstützung des Prozesses ein. Dies ist im Ablaufplan der Evaluierung ersichtlich. Welche Rollenverteilung dabei vorgenommen wurde und eine Dokumentation der „eigenständigen Außensicht“ wie von MISEREOR gefordert, wurde dabei jedoch nicht vorgenommen.

Die Einbeziehung der Sichtweisen der Zielgruppen erfolgte durch qualitative Befragungen, die teilweise durch den Einsatz von partizipativen Methoden wie *SWOT-Analysen* oder *Tiny Tools*. Im Basischeck führten insgesamt 17 (24%) Evaluierungen auch weitere partizipative Methoden durch, 7 (12%) von 56 ELBE und 10 (71%) von 14 EQM-Evaluierungen. In der vertieften Analyse wenden 9 der 20 Evaluierungen weitere partizipative Methoden an, wobei hier zu beachten ist, dass besonders partizipative Evaluierungen für die vertiefte Analyse vermehrt ausgewählt wurden.

Die Einbeziehung der Perspektiven von Partnerorganisationen ist bei 8 der 20 Evaluierungen „sehr gut“, bei 11 „gut-befriedigend“, und bei 1 „eher schwach“. Die Einbeziehung der Perspektiven von Zielgruppen ist bei 4 der 20 Evaluierungen „sehr gut“, bei 13 „gut-befriedigend“, und 3 „eher schwach“. Insgesamt werden also die Partnerorganisationen als Informationsquelle wie auch bei der Einbindung in Planung und Durchführung etwas besser eingebracht. Die meisten Evaluierungen bringen aber die Perspektiven sowohl der Partnerorganisation als auch der Zielgruppen mindestens befriedigend ein.

Die Heterogenität der Zielgruppen wird nur sehr begrenzt abgebildet, da in den meisten Evaluierungen nicht durchgängig nach verschiedenen Zielgruppen unterschieden wird. Auf spezifische Zielgruppen fokussierte Empfehlungen sind selten und beinhalten i.d.R. Empfehlungen für die Förderung von Frauengruppen oder Jugendlichen. Hier ist es also fraglich, ob im Sinne der Diversität der Zielgruppen auch alle verschiedenen Bereiche der Zielgruppen abgedeckt wurden und somit ein umfassend partizipativer Prozess durchgeführt wurde.

#### **Fazit**

Es gibt viele Anzeichen für ein Bemühen eines partizipativen Vorgehens in den Evaluierungen. Ein Auftakt- und Abschlussworkshop zur Besprechung des Vorgehens und zur Diskussion der Ergebnisse sind die Regel. Besonders die Partnerorganisation wird sehr gut mit eingebunden sowohl in Planung und Durchführung als auch als Informationsquelle. Aber auch die Zielgruppen werden von vielen Evaluierungen zumindest als Informationsquelle ausgiebig eingebunden und Perspektiven von ihr werden dargestellt. Verschiedene partizipative Methoden werden dabei von einem gewissen Teil genutzt. Eine Ausdifferenzierung der Zielgruppen wäre dabei noch wünschenswert, um eine umfassende Partizipation aller Teilgruppen von Zielgruppen zu gewährleisten.

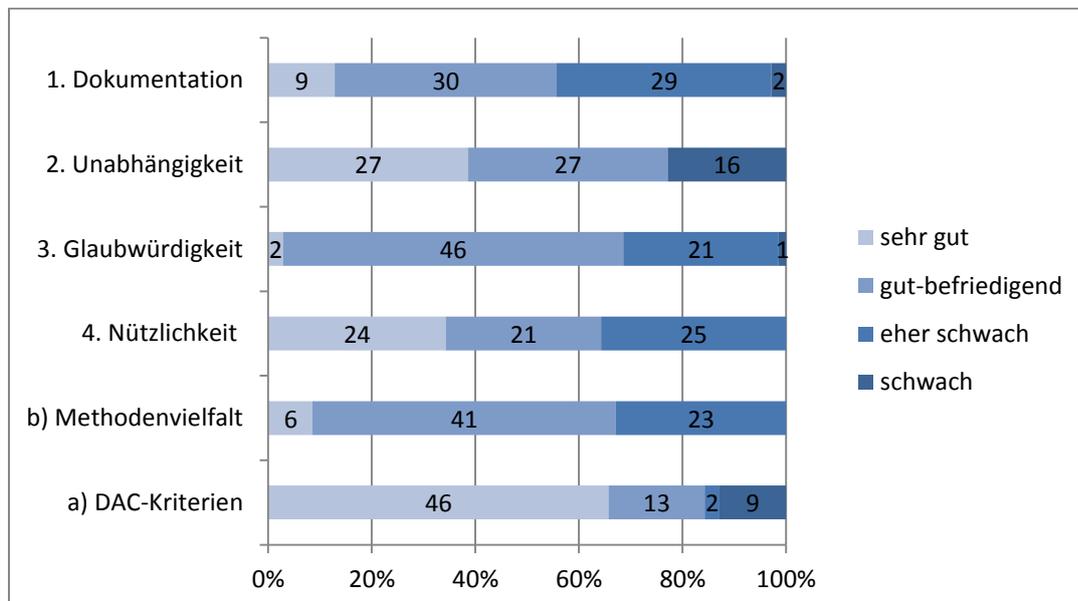
## 6 Schlussfolgerungen und Empfehlungen

### 6.1 Zusammenfassende Bewertung

Die repräsentative Auswahl von 70 ELBE und EQM-Evaluierungsberichten aus der Grundgesamtheit von 261 Berichten der Jahre 2014 und 2015 wurde in einem ersten Schritt (Basis-Qualitätscheck) auf die Kriterien *Vollständigkeit und Angemessenheit der Dokumentation, Unabhängigkeit und Überparteilichkeit, Glaubwürdigkeit und Nützlichkeit* durchgesehen. Die Prüfung erfolgte auf der Grundlage der Lektüre von Zusammenfassung, Einleitung, Methodenkapitel, Schlussfolgerungen und Empfehlungen sowie eines Durchblätterns des Berichts.

**Die Ergebnisse des Qualitäts-Basischecks:** 60-70% der Evaluierungsberichte befinden sich pro Kriterium im „grünen“ Bereich, d.h. zwischen den Bewertungen sehr gut bis befriedigend. Der andere Teil von 30-40% wurde als eher schwach bis schwach eingestuft. 23 (33%) Evaluierungen schneiden dabei in allen vier Kriterien mit sehr-gut bis befriedigend ab. Davon sind 13 (von insgesamt 14) EQM-Evaluierungen und 10 (von insgesamt 56) ELBE. Keine der Evaluierungen schneiden in allen Kriterien im roten Bereich ab d.h. zwischen eher schwach und schwach. 16 (23%) Evaluierungen sind aber in drei Kategorien mit eher schwach oder schwach bewertet. Tendenziell sind in dieser Kategorie auch vermehrt Evaluierungen aus Lateinamerika vertreten, da diese durch eine fehlende Zusammenfassung in Bereich Nützlichkeit und durch einen fehlenden Annex im Bereich Dokumentation oftmals schlecht abschneiden. Jedoch ist diese Untergruppe nicht repräsentativ für die Grundgesamtheit der Evaluierungen pro Kontinent.

Abbildung 14 Übersicht der Ergebnisse des Basischecks



In einem zweiten Schritt wurden 20 Evaluierungen einer vertieften Analyse unterzogen. Dabei wurden auch die Kriterien *Partizipation* und *Konsistenz der TOR* untersucht. Bei den Untersuchungen wurde jeweils der ganze Bericht mit Anlagen gelesen.

Insgesamt kann die Qualität der Berichte auf der Grundlage des Basis-Qualitätschecks und der vertieften Analysen als befriedigend eingestuft werden.

Die zusammenfassenden Ergebnisse, untergliedert nach den Kriterien sind:

<b>Kriterium</b>	<b>Bewertung</b>
<b>Vollständigkeit und Angemessenheit der Dokumentation</b>	<b><u>Befriedigend bis eher schwach</u></b>
<b>Hinweise in den Berichten auf die Unabhängigkeit und Überparteilichkeit der EvaluatorInnen</b>	<b><u>gut bis befriedigend</u></b>
<b>Glaubwürdigkeit der Berichte</b>	<b><u>befriedigend</u></b>
<b>Nützlichkeit der Berichte</b>	<b><u>befriedigend</u></b>
<b>Hinweise in den Berichten auf Partizipation von Partnerorganisation und Zielgruppen</b>	<b><u>gut bis befriedigend</u></b> <small><u>Partnerorganisation: gut; Zielgruppen: befriedigend</u></small>
<b>Bewertung der Konsistenz der TOR</b>	<b><u>befriedigend</u></b>

**Erläuterungen.** In der zusammenfassenden Bewertung der Qualität der Berichte werden diese als befriedigende Qualität eingestuft. Die Bewertung umfasst die Ergebnisse des Qualitäts-Basischecks und der vertieften Analyse.

**Unterschied ELBE und EQM-Evaluierungen.** EQM-Evaluierungsberichte wurden in der Regel besser eingestuft als ELBE. Die Erklärungen dafür liegen aus Sicht des Meta-Evaluierungsteams in folgenden Aspekten:

- d) Direkte Qualitätskontrolle der Berichte durch Evaluierungsfachleute von EQM
- e) Höheres methodisches Know-How/ Evaluierungs-Know-How
- f) Mehr Arbeitstage und mehr Finanzmitteleinsatz für die GutachterInnen pro EQM-Evaluierung im Vergleich zu ELBE
- g) Einhaltung des Vieraugenprinzips (i.d.R. immer 2 GutachterInnen bei EQM-Evaluierungen, bei ELBE oft nur eine GutachterIn)
- h) Mehr Distanz der GutachterInnen zum Projekt sowie die EQM-Unterstützung bei der Äußerung von Kritik am Projekt

Im Bereich Partizipation, welche nur in Stufe 2 untersucht wurde, lässt sich von der Tendenz ein vergleichbares und zum Teil besseres Abschneiden der ELBE finden. Dieses Ergebnis ist nicht repräsentativ, zeigt jedoch, dass mögliche Vorteile in der Einbeziehung der Stakeholder in die Evaluierungen liegen.

## 6.2 Einschätzung nach Untersuchungskriterien

### 1. Vollständigkeit und Angemessenheit der Dokumentation

<b>Bewertung der Vollständigkeit und Angemessenheit der Dokumentation:</b>	<b><u>Befriedigend bis eher schwach</u></b>
--	---

Die Gliederung eines Berichts ist der Schlüssel, um sich leicht in diesem zurechtzufinden und Informationen schnell zu finden. Die überwiegende Zahl der Evaluierungen wird hier mit gut-befriedigend oder sehr gut bewertet und erfüllt die Grundprinzipien einer guten Gliederung. Am häufigsten fehlen allerdings die Zusammenfassung oder das Methodenkapitel, die für die Verständlichkeit und die Einordnung der Qualität des Berichts eine wichtige Rolle spielen. Diese werden zum Teil allerdings auch nicht gemäß der TOR von der Partnerorganisation gefordert.

Die Dokumentation des Vorgehens, der Datenquellen und der Instrumente im Annex oder im Bericht ist wesentlich, um Umfang und Qualität der Evaluierung abschätzen zu können. Diese liefert ein sehr gemischtes Bild. Insbesondere der Ablaufplan und die Instrumente könnten noch besser dokumentiert werden. Die Berichte sind hier sehr unterschiedlich, etwa die Hälfte der Evaluierungen dokumentiert sehr gut bis befriedigend, die andere Hälfte eher schwach, was oftmals an einem nicht existierenden Annex liegt. Oft wird die Ergebnisdarstellung nicht in einem separaten Kapitel vorgenommen, sondern mit der Darstellung von Analysen kombiniert. In den Fällen, in denen im Text deutlich hervorgeht, was Ergebnisdarstellung und was Analyse ist, unterstützt dies die Nachvollziehbarkeit von Analysen aus den Ergebnissen.

**Diskussion.** Ist der Umfang des Evaluierungsberichts für unterschiedliche Nutzergruppen angemessen? Wie lang muss oder sollte ein Evaluierungsbericht sein?

In der vertieften Analyse der 20 Berichte ragen Extreme heraus, mit 8 Seiten Hauptteil einer Evaluierung eines Projekts zur ländlichen Entwicklung in Indien oder 96 Seiten Hauptteil bei einer Evaluierung eines Lobby- und Advocacy-Projekts auf den Philippinen. Mit Sicherheit sind die LeserInnen der verschiedenen Partnerorganisationen je nach Aufgabenbereich und persönlichem Hintergrund sehr unterschiedlich. Daher sollte den Partnerorganisationen weiterhin der Freiraum bestehen bleiben, kurze oder längere Evaluierungsberichte abzunehmen. Es besteht jedoch auch die Erfahrung, dass in sehr kurzen Berichten, die Grundanforderungen an Evaluierungen nicht eingehalten werden können und lange Evaluierungsberichte oft nur von wenigen Personen gelesen werden und teilweise hohen Aufwand in der Kommentierung durch den Auftraggeber bedeuten.

Die Mehrzahl der untersuchten Evaluierungsberichte umfasst Hauptteile, die 20-40 Seiten plus Anlagen umfassten. Diese Größenordnung wird seitens des Meta-Evaluierungsteams als praktikabel für das Informationsbedürfnis seitens der Partnerorganisation und MISEREORS angesehen.

**Schlussfolgerung aus den Ergebnissen.** Anforderungen an die Dokumentation sind nicht ausreichend klar und meist zu komprimiert dargestellt, d.h. es sind nicht ausreichend praktische Umsetzungshilfen vorhanden. Zentraler Schwachpunkt fast aller Dokumentationen ist die praktisch nicht vorhandene Darstellung der Auswahl der Stichprobe und die unvollständige Dokumentation im Annex. Das führte in der Bewertung zum Zusatz „bis eher schwach“. Gliederungen sind in der Regel angemessen, jedoch fehlen in den schwächeren Evaluierungen wichtige Kapitel wie Zusammenfassung und Methodenkapitel. Feste Vorgaben sind erforderlich und würden Abhilfe schaffen.

**Empfehlungen:**

**Empfehlung 1 (an EQM):** MISEREOR sollte die Anforderungen und die verbindlichen Inhalte für die Dokumentation in Evaluierungen klarer benennen. Grundsätzlich ist alles bereits in den TOR vermerkt, wird jedoch oft nicht umgesetzt. Neben einer verbindlicheren Darstellung von zentralen Gliederungspunkten könnten dazu Merkblätter, Checklisten und Vorlagen zur Verfügung gestellt werden. Mögliche Inhalte sind:

- a) Dokumentation der verschiedenen Datenquellen und Zuordnung zu den Kategorien wie z.B. „externe ExpertInnen“, „Gute Praxis-Beispiele/ Fachdiskussion“
- b) Dokumentation der Auswahl der Stichprobe
- c) Vorlage für die Liste der Interviewten, die untergliedert ist nach z.B. VertreterInnen der Partnerorganisation, Projektteam, VertreterInnen der Zielgruppe(n)
- d) Informationen und Erfahrungswissen zur Beauftragung von Evaluierungen (Verantwortlichkeiten, Zeitbedarf, Empfehlungen der Evaluierung können auch begründet abgelehnt werden sowie auch konkrete Hinweise wie Erfahrungswerte bei Umfang (Seitenzahlen) von Berichten etc.)

**2. Unabhängigkeit und Überparteilichkeit**

<b>Bewertung: Hinweise in den Berichten auf die Unabhängigkeit und Überparteilichkeit der EvaluatorInnen</b>	<b><u>gut bis befriedigend</u></b>
--	------------------------------------

Die Ergebnisse des Qualitäts-Basischecks zeigen, dass in ca. 3/4 der Evaluierungen (73%) unabhängige GutachterInnen explizit genannt werden und in etwa 30% der Evaluierungen weitere Informationen zum Evaluationsteam gegeben werden. Die Perspektive der Partnerorganisationen und der Zielgruppen werden größtenteils tief in die Analysen mit eingebunden. Externe Perspektiven in der Form von externen Experten oder Fachdiskussion fließen explizit eher weniger mit ein. Die meisten der Evaluierungen legen sowohl Stärken als auch Schwächen des evaluierten Projekts dar. Dies geschieht aber in unterschiedlicher Ausgewogenheit und Detailliertheit. Zum Teil wird nur sehr positiv berichtet, was aber nicht unbedingt für Parteilichkeit und für fehlende Unabhängigkeit steht.

**Schlussfolgerungen.** Die Einschätzung der Unabhängigkeit und Glaubwürdigkeit ist über die reine Analyse von Evaluationsberichten nicht abschließend überprüfbar. Im vorliegenden Fall deuten ca. 3/4 (73%) der Berichte darauf hin, dass die GutachterInnen extern waren. Es ist davon auszugehen, dass diese Zahl höher ist, da es keine spezifischen Vorgaben durch MISEREOR oder auch durch maßgebliche internationale Institutionen wie OECD-DAC<sup>87</sup> gibt, die die Dokumentation der Unabhängigkeit und Qualifikation der EvaluatorInnen im Bericht explizit einfordern. Daher wird davon ausgegangen, dass mehr externe EvaluatorInnen und GutachterInnen zum Einsatz kommen, als in den Berichten ausgewiesen wird.

#### **Diskussion: Sind ELBE unabhängig?**

In der Meta-Evaluierung der ELBE aus dem Jahr 2013/2014 werden Rückmeldungen von VertreterInnen der MISEREOR-Kontinentalabteilungen dokumentiert, die ELBE überwiegend als von den Partnerorganisationen beauftragte „Gefälligkeitsgutachten“ ansehen.

Die vertiefte Analyse von 20 Evaluierungen in der hier vorliegenden Meta-Evaluierung identifizierte keinen Evaluierungsbericht, der als „Gefälligkeitsgutachten“ eingestuft werden konnte. Wie im Methodenkapitel beschrieben, ist die Auswahl der 20 Evaluierungen nicht repräsentativ, aber sie enthielt explizit als „schwach“ von den Kontinentalabteilungen eingestufte Evaluierungen.

Eine Erklärung für diese unterschiedlichen Einstufungen könnte sein, dass es in der Tat Evaluationsberichte gibt, die über weite Teile hinweg den Eindruck vermitteln, die notwendige Distanz zur Partnerorganisation oder zum Projekt nicht einzuhalten. Dies drückt sich in der Regel im Vorwort und im Hauptteil des Berichts über sehr deutliches Lob über die Arbeit der Partnerorganisation aus, wobei gerade die Ausführungen im Vorwort oft von kulturellen Eigenheiten geprägt sind und dementsprechend nicht überinterpretiert werden sollten. Die nähere Analyse von Berichten mit dieser Auffälligkeit zeigte jedoch, dass die Kritik dieser Evaluierungen oft sehr präzise in den Empfehlungen auftauchte. Sie wurde damit zwar nur indirekt geäußert, aber sie wurde im Bericht deutlich dargelegt. In einem Fall wird auf der Grundlage des Berichts auch eher mangelnde Evaluierungskompetenz vermutet, als eine zu hohe Nähe zum Auftraggeber. Daher werden aufgrund der Berichtslage diese GutachterInnen als ausreichend unabhängig und unparteiisch eingestuft.

Jedoch unterstreicht MISEREOR im Anschreiben an die Partnerorganisationen, dass Evaluierungen „wichtige Elemente der Reflexion, der Vergewisserung und Qualifizierung der Arbeit in den Projekten“ dienen.<sup>88</sup> Daher ist es nachvollziehbar, dass Projekte und auch Partnerorganisationen, die in hohem Maße von der Förderung von MISEREOR abhängen, kaum Evaluationen beauftragen, die das Programm vom Grundsatz her in Frage stellen oder einen Stopp der Förderung durch die Geldgeber empfehlen.

#### **Empfehlungen:**

**Empfehlung 2 (an KA & EQM, Priorität):** Weitere Integration des Werkzeugs „Evaluierung“ in den Partnerdialog. Evaluierungen dienen der Rechenschaftslegung und des Lernens. Für die Stärkung des Lernens bedarf es des Vertrauens der Partnerorganisation in MISEREOR,

---

<sup>87</sup> Siehe OECD, DAC 2010

<sup>88</sup> MISEREOR o.J.: 1

dass auch gewisse Schwachpunkte nicht zum Abbruch der Förderungen führen. Anregungen für die Umsetzung:

- a) Weitere Fortbildung der MISEREOR-Mitarbeitenden im Bereich der Planung und Begleitung von Evaluierungen. Bearbeitete Fragestellungen könnten zum Beispiel sein: Wie sollte generell der Evaluierungsprozess erfolgen? Was tun bei Partnerorganisationen zu denen das Vertrauen in letzter Zeit gesunken ist und ein gewisses Misstrauen besteht? Wie sollten die Empfehlungen der Evaluierungen bearbeitet werden? Wie kann auf schlechte Berichte reagiert werden? Wie kann dem vorgebeugt werden?
- b) Partnerorganisationen verhelfen, dass sie Evaluierung für sich nutzen und sie angemessen für Rechenschaftslegung und ihr eigenes Lernen nutzen. Ansatzpunkte könnten Beratungen sein, wie diese in den vergangenen Jahren in Lateinamerika durchgeführt wurden und die weitere Thematisierung von Evaluierung im Rahmen des Partnerdialogs, z.B. bei Besuchen der Partnerorganisation durch die zuständigen MISEREOR-KollegInnen (InPro). In diesen Beratungen sollten die Erwartungen von MISEREOR an die Evaluierungen und die bestehenden Handreichungen mit Praxiserfahrungen erläutert werden.

### **Glaubwürdigkeit**

<b>Bewertung: Glaubwürdigkeit der Berichte</b>	<b><u>befriedigend</u></b>
--	----------------------------

Qualitative Interviews und Fokusgruppensitzungen prägen die Erhebungsmethoden der ELBE und EQM-Evaluierungen. Partizipative Methoden wie SWOT (Analyse von Stärken, Schwächen, Chancen und Bedrohungen), „Tiny Tools“ (Methodensammlung auf der Grundlage von PRA-Methoden) und semi-standardisierte Fragebogenerhebungen werden weniger eingesetzt. In der Umsetzung der Datenerhebung wird, in vielen Fällen weit ab der städtischen Zentren, mit VertreterInnen der Zielgruppen gesprochen und die Situation vor Ort in Augenschein genommen. In den Berichten ist zumeist das methodische Vorgehen beschrieben, Schwachpunkt ist jedoch die Dokumentation und Nachvollziehbarkeit der Stichproben für die Datenerhebung: Wer hat und wie wurde über die Auswahl der InterviewpartnerInnen, ZielgruppenvertreterInnen und Projektgebiete entschieden? Daten werden zwischen den Aussagen der Partnerorganisation und der Zielgruppen trianguliert, jedoch erfolgt dies wenig über die Projektgrenzen hinweg, z.B. mit externen ExpertInnen, der aktuellen Fachdiskussion oder Gute Praxis-Beispiele und Erfahrungswissen von MISEREOR. In den Berichten wird die klare Trennung von Beschreibung und Bewertung nur teilweise eingehalten und eine Untergliederung der Dokumentation der erhobenen Daten bei den Zielgruppen nach Alter, Geschlecht, Wohlstand und anderen sozialen Kriterien ist wenig verbreitet. In der Auswertung sind quantifizierende Aussagen im Idealfall mit Angabe der Informationsquelle weiterhin wichtig, um so Vertrauen in die Ergebnisse zu schaffen. Nur so kann transparent verdeutlicht werden, dass die Bewertung nicht auf einzelnen Meinungen beruht und dass die relevanten Zielgruppen befragt wurden bzw. eine Triangulation der Information vorgenommen wurde.

Die Wirkungen bilden einen Fokus in den meisten Evaluierungen. Viele der Evaluierungen untersuchen die Wirkungen zusammen mit den anderen DAC-Kriterien, setzen ein eigenes

Kapitel für Wirkungen an oder richten sogar die gesamte Analyse auf die Untersuchung der Wirkungen aus. Um Wirkungen richtig zu bewerten und plausibel auf das Projekt zurückzuführen, ist jedoch ein angemessenes methodisches Vorgehen unerlässlich. Viele Berichte nutzen keine Wirkungslogik zur Analyse der Wirkungen und können weder auf ex-ante Daten der Zielgruppen im Projekt zum Vorher-Nachher Vergleich zugreifen, noch erheben sie Daten einer Vergleichsgruppe. Auch die Triangulation der Daten, um mögliche gefundene Wirkungen zu bestätigen, findet im Bereich der Wirkungsbewertung eher selten statt oder wird selten dokumentiert. Messungen der Wirkungen z.B. über größere quantitative Fragebogenerhebungen sind die Ausnahme. Methodisch sehr aufwendige Verfahren können verständlicherweise mit den finanziellen Mitteln, in dem zeitlichen Rahmen und aufgrund des hohen notwendigen Know-Hows nicht verlangt werden. Jedoch kann mit einfachen Vorgehensweisen die Wirkungsanalyse glaubwürdiger gestaltet werden.

**Schlussfolgerungen.** Ein wichtiger Beitrag zur Glaubwürdigkeit einer Evaluierung ist die Darlegung der Untersuchungsmethoden und die Information zu den Befragten. Gerade die Auswahlkriterien werden fast nie dargelegt und stellen somit die Glaubwürdigkeit in Frage. Eine Darlegung der Auswahlkriterien würde zudem eine bewusstere Auswahl unterstützen, da es einen Rechtfertigungsdruck gibt.

In der Auswertung der oft umfassenden Daten gibt es noch einige Verbesserungsmöglichkeiten. Sowohl die Quantifizierung der Aussagen als auch die Differenzierung der Zielgruppen findet kaum statt, obwohl beides explizit in den Handreichungen gefordert ist.

Der Überprüfung auf Wirkungen wird sehr viel Raum in den Evaluierungen eingeräumt, jedoch ist fraglich, ob mit den angewandten Methoden sichergestellt werden kann, dass die identifizierten Wirkungen wirklich auf das Projekt zurückzuführen sind.

Weiterhin fällt auf, dass wichtige Bereiche des Projektmanagements oder der Untersuchung der konkreten inhaltlichen Ausgestaltung von Arbeitsabläufen in der Evaluierung nicht oder i.d.R. nur relativ kurz angesprochen werden. Eine Ursache dafür kann in der Fokussierung auf die Wirkungserfassung und Zielerreichung liegen, die in der MISEREOR-Mustergliederung der TOR systematisch angelegt ist.

Möglicherweise führt die Praxis, das Ergebnis-Kapitel mit den DAC-Bewertungen zu kombinieren dazu, dass in vielen Fällen die Beschreibung der Situation anhand der erhobenen Daten ganz unterbleibt und sofort die Bewertung dargelegt wird. Die Kapitel heißen zwar oft „Ergebnisse und Analysen“, die puren Ergebnisse der Datenerhebung werden dort aber selten aufgezeigt.

#### **Diskussion: Priorität von Wirkungserfassung & anderen Beurteilungskriterien**

Die Herausforderungen für eine methodisch angemessene Vorgehensweise bei der Analyse der Wirkungen von Projekten sind hoch und lassen sich bei dem hohen Bedarf an Evaluierungen im Partnerfeld von MISEREOR insbesondere bei ELBE nur langfristig durch immer qualifiziertere lokale GutachterInnen realisieren. MISEREOR steht in diesem Punkt mit ihren 2500 Partnerorganisationen vor einer Herkulesaufgabe und sollte die verfügbaren Kräfte auf mehreren Ebenen einsetzen.

Der Erfolg und die Angemessenheit von Projekten lassen sich neben der Analyse der Wirkungen auch durch die gezieltere Nutzung der DAC-Kriterien Relevanz, Effizienz und Nachhaltigkeit beurteilen: Intensiverer, detaillierterer Abgleich des Vorgehens mit anderen

Projekten (Informationsquellen dazu sind MISEREOR-Evaluierungen bei vergleichbaren Projekten, MISEREOR-Erfahrungswissen der InPros, Besuch eines Gute Praxis-Projekts im Projektgebiet), Konzentration auch auf wenige konkrete „kleine“ Verbesserungen im Arbeitsablauf der Projekte, Herstellung von mehr Transparenz bzgl. der Arbeitsabläufe und des Finanzmitteleinsatzes im Projekt; systematischere Integration von Informationen aus den Finanzaudits der Projekte und systematischere Berücksichtigung von Weichenstellungen, die für die Nachhaltigkeit der Projektarbeit wichtig sind.

### **Empfehlungen:**

**Empfehlung 3 (an EQM, Priorität):** Die Vorgaben für ELBE und EQM-Evaluierungen sollten die Wichtigkeit der der Auswahl der Stichproben und deren Dokumentation verdeutlichen und dafür klare Vorgaben entwickeln. Wichtig für die Auswahl der Stichprobe sind i.d.R. folgende Aspekte:

- a) Es muss nicht das gesamte Projektgebiet untersucht werden, aber die Auswahl, wie diese Teilmenge gewählt wird, sollte begründet werden: Welche Regionen sollen besucht werden, was sind die Auswahlkriterien? Welche Zielgruppen sollen besucht werden, was sind die Auswahlkriterien? (hohe Repräsentativität für die Zielgruppen, Abdeckung der Heterogenität der Zielgruppen, Machbarkeit)
- b) Welche Ressourcepersonen außerhalb des Projektes sollen befragt werden? Wer steht dem Projekt eher näher, wer ist eher entfernt und distanziert?

**Empfehlung 4 (an EQM):** Die erhobenen, oft umfangreichen Daten sollten noch mehr nach Zielgruppen differenziert, ausgewertet und dokumentiert werden. Das bedeutet Mehraufwand für die EvaluatorInnen und kann über eine gezieltere Datenerhebung und -auswertung ausgeglichen werden. Ein Ansatzpunkt ist z.B. die Anzahl der besuchten Gruppen oder Projektregionen zu reduzieren, um dafür die erhobene Daten differenzierter auszuwerten). Eine andere Möglichkeit ist die Verlängerung des Feldaufenthalts.

**Empfehlung 5 (an KA & EQM, Priorität):** Wirkungen nicht methodisch aufwendiger prüfen, jedoch Methoden anwenden, um diese plausibel auf das Projekt zurückzuführen. Die Wirkungen zu überprüfen ist ein wichtiger Bestandteil der Erfolgskontrolle des Projekts. Jedoch sollten, gegeben der Ressourcenausstattung der Evaluierung, die Methoden machbar sein und auch plausibel Wirkungen mit Projekt verbinden.

- a) Im gesetzten Rahmen, Mindestvorgabe: Diskussion einer Wirkungslogik, Triangulation der Informationen, selektive Befragung einer Vergleichsgemeinde/Vergleichsperson.
- b) Erhebung von Baseline-Daten zu Beginn des Projekts einführen.

**Empfehlung 6 (an EQM):** Verstärkte Nutzung von Daten und Informationen außerhalb des Projektes. Dies sollte bei Auftaktgesprächen vermittelt werden und konzeptionell in das MISEREOR Evaluierungskonzept bzw. die Musterreferenzrahmen integriert werden. Ziel ist es, die Daten der Partnerorganisation und der Zielgruppen zu triangulieren und Impulse, Ideen und Anregungen von außen einzuholen, um noch besser die Herausforderungen der Projekte umzusetzen und den Projekterfolg zu beurteilen.

**Empfehlung 7 (an KA & EQM, Priorität):** Zusätzlich zu den Wirkungen auch andere zentrale Faktoren des Projekterfolgs in den Blick der Evaluierungen nehmen. Bei der Konzeption der Evaluierung sollte stärker darauf geachtet werden, ob neben dem Schwerpunkt Wirkungserfassung verstärkt auch andere Fragestellungen wie z.B. der Arbeitsabläufe, des Projektmanagements oder der inhaltlichen Ausgestaltung des Projektes eingebracht werden. Fragestellungen für Arbeitsabläufe und Projektmanagement richten sich z.B. auf Strategie, Kooperation, Einbringen von Innovationen, Überblick über die zentralen Entwicklungsprozesse und auf die Finanzadministration. Weitere Anregungen dazu:

- a) Projektmanagement: Nutzung der Fragestellungen und Lösungsmöglichkeiten der zuständigen MISEREOR-Projektmitarbeiter, ggf. auch anderer MISEREOR-Fachkollegen; Nutzung der Grundfragestellungen der Erfolgsfaktoren von Capacity WORKS (Management-Tool der GIZ<sup>89</sup>) wie z.B. Strategie, Kooperation, Steuerungsstruktur, Prozesse sowie Lernen und Innovation.
- b) Inhaltliche Ausgestaltung von Projekten: Nutzung der MISEREOR-Fachexpertise und Nutzung der Erkenntnisse anderer Institutionen wie zum Beispiel der Welthungerhilfe oder von Brot für die Welt, die ggf. in Form von Handreichungen oder Fachexpertise zur Verfügung stehen.

**Empfehlung 8 (an KA & EQM, Priorität):** Gesamtkonzept der ELBE und EQM-Evaluierungen überprüfen und ggf. unterschiedliche Schwerpunktsetzungen und inhaltliche Ausgestaltungen für ELBE und EQM-Evaluierungen entwickeln. EQM-Evaluierungen konzentrieren sich auf komplexe Zusammenhänge und Wirkungserfassung und ggf. auf Projektkonzeptionen, die seitens MISEREOR als ineffektiv, als veraltet oder als besonders innovativ angesehen werden; ELBE konzentrieren sich auf Diskussion der Arbeitsansätze und des Abgleichs mit Guter Praxis, Erfahrungswissen und der aktuellen Fachdiskussion.

### Nützlichkeit

<b>Bewertung: Nützlichkeit der Berichte</b>	<b><u>befriedigend</u></b>
---	----------------------------

In der Regel wurden die Zusammenfassung als gut lesbar und verständlich eingestuft. In vielen ELBE fehlte jedoch auch die Zusammenfassung. Die Empfehlungen ließen sich eher selten vollständig aus dem Kapitel *Schlussfolgerungen* logisch ableiten. Für eine solche Ableitung war die Lektüre des gesamten Kapitels *Ergebnisse* notwendig. Des Weiteren gab es Empfehlungen, die auch nach Lektüre des gesamten Berichts nicht aus dem Inhalt des Berichts abzuleiten war. Die Empfehlungen wurden größtenteils als pragmatisch und konkret umsetzbar bewertet. Es fand jedoch zumeist keine explizite Priorisierung der Empfehlungen statt. Die Evaluierungsfragen wurden von den meisten Berichten abgedeckt, aber sehr unterschiedlich zufriedenstellend und detailliert beantwortet.

**Schlussfolgerungen.** Damit gerade lange Berichte auch nützlich sind, ist eine Zusammenfassung sehr wichtig. 30% der Berichte enthalten keine Zusammenfassung. Abgesehen von sehr kurzen Evaluierungen lässt dies keinen schnellen Einblick in die Evaluierung zu.

<sup>89</sup> Siehe Handreichung zu Capacity WORKS der GIZ (GIZ 2015)

Die Evaluierungsfragen werden in der Regel gut abgedeckt, jedoch wird durch die zum Teil große Anzahl und den sehr detaillierten Fragen der Überblick verloren und es fehlen Gesamteinschätzungen. Zum Teil werden so in der Beantwortung der DAC-Kriterien gewissenhaft alle Fragen zu dem Themenbereich beantwortet, ohne eine generelle Einschätzung zu geben wie relevant das Projekt einzuschätzen ist. Außerdem werden so auch nur die in den Fragen beschriebenen Themen in der Evaluierung eingebracht. Für die Vertiefung von zentralen Fragestellungen, die sich in der Feldphase ergeben können, fehlt dann oft der Raum.

Die Empfehlungen sind ein wichtiges Endprodukt einer Evaluierung, die zur Überarbeitung und Optimierung des Programms führen sollen. Wichtige Eigenschaften von Empfehlungen sind ihre Verständlichkeit und die Klarheit der möglichen Ansatzpunkte. Die Ergebnisse zeigen, dass diese Voraussetzungen weitgehend für die Empfehlungen der untersuchten Berichte gegeben sind. Die Erfahrung zeigt, dass die Nützlichkeit der Empfehlungen sehr stark verbessert werden kann, wenn für diese eine strukturierte Umsetzungsplanung vorgenommen wird. Dieser wichtige erste Schritt der Umsetzung von Empfehlungen wird jedoch bei der ELBE-Handreichung nicht berücksichtigt. Bei der Umsetzung von Evaluierungsempfehlungen ist die Erfahrung der vergangenen Jahre zu berücksichtigen, dass Evaluierungen teilweise stark die Komplexität von Abläufen und Verfahren erhöhte und damit eine gewisse hemmende „Bürokratie“ unterstützte.<sup>90</sup> Dies sollte bei der Umsetzung von neuen Verfahren wie z.B. Monitoring-Systemen überlegt und abgewogen werden: Welche Daten sind notwendig? Inwiefern führen auch einfachere Verfahren zu einem vergleichbaren Ergebnis?

In vielen Evaluierungen deutet die Lücke zwischen einer guten Kontextbeschreibung und praxisnahen Empfehlungen darauf hin, dass die Evaluierung durch GutachterInnen vorgenommen wurde, die lange Zeit inhaltlich in diesem Bereich beraten haben. Diese GutachterInnen verfügen über eine gute Fachkenntnis der Situation vor Ort, haben jedoch eher geringe methodische Kenntnisse im Bereich Evaluierung. Sie dokumentieren ihre Datenerhebungen und ihr Vorgehen zumeist nicht und es ist offen, inwieweit Datenerhebungen und Auswertungen den qualitativen Ansprüchen an Evaluationen genügen. Sie können als „Berater-EvaluatorInnen“ kategorisiert werden. Auf der anderen Seite gibt es Evaluierungen, die sehr stark auf Kenntnisse in den Bereichen Projektmanagement, Ziele, Indikatoren und Methoden der Sozialforschung hindeuten. Diese sind erkennbar an einem starken Methodenteil, guten Darstellungen von Resultaten und Empfehlungen, die sich auf Verbesserung des Projektmanagements, Monitorings, Ziel- und Indikatorenformulierung konzentrieren, aber für die „inhaltliche“ Arbeit des Projektes wenig neue Impulse einbringen. Sie können als *Management-EvaluatorInnen* kategorisiert werden.

Eine deutliche Trennung von Beschreibungen und Bewertungen konnte immer dann vorgenommen werden, wenn zunächst eine Beschreibung des Sachverhalts, möglichst mit Bezug auf die Quellen und dann eine Bewertung ggf. mit Begründung bzw. weiterer Erläuterung vorgenommen wurde. Es entspricht auch guter Praxis, dass Ergebnisse und Bewertungen in einem Kapitel zusammengefasst werden können, wenn die Inhalte deutlich voneinander getrennt sind.

---

<sup>90</sup> Siehe Vortrag von Jan-Eric Furubo (Audit Counsellor at the Swedish National Audit Office, Stockholm): After evaluation maturity: Sustainability or decay? A Swedish experience; Vortrag auf der Konferenz: The Future of Evaluation in Modern Societies, 10 Jahre CEval, 14. und 15. Juni 2012 in Saarbrücken.

## Empfehlungen:

**Empfehlung 9 (an EQM):** Die Wichtigkeit und der Nutzen der Zusammenfassung für die Arbeit von MISEREOR sollte bei ELBE den Partnerorganisationen und den von ihnen beauftragten GutachterInnen noch mehr verdeutlicht werden.

**Empfehlung 10 (an KA & EQM, Priorität):** Für die Sicherstellung der inhaltlich fachlichen Kompetenz und der Kompetenz im Bereich Evaluierung sollte auch bei ELBE der systematische Einsatz von zwei EvaluatorInnen erwogen werden.

**Empfehlung 11 (an EQM):** Es sollte darauf geachtet werden, dass der erste Schritt, der „puren“ Beschreibung nicht übersprungen wird und die anschließenden Bewertungen auch nachvollziehbar begründet werden. Die Kombination von „puren“ Beschreibungen auf der Grundlage der erhobenen Daten und die Bewertungen nach den DAC-Kriterien in einem Kapitel „Ergebnisse und Analysen“ entspricht guter Praxis. Alternativ ist es auch möglich die Bewertung nach den DAC-Kriterien in einem separaten Kapitel zu verlangen.

**Empfehlung 12 (an KA & EQM):** Die Diskussion und Erstellung von Umsetzungsplänen sollte systematisch auch für ELBE vorgesehen werden. Als hilfreich hat sich erwiesen, dass die GutachterInnen die Empfehlungen in das Umsetzungsraster einfüllen und die Ergebnisse der Diskussion in diesem Raster festhalten und mit dem Auftraggeber abstimmen. Dies muss im Vertrag inhaltlich vorgesehen und mit z.B. zwei Arbeitstagen zeitlich unterfüttert werden. Die Partnerorganisation sollte bei der Umsetzung klar analysieren, welche Empfehlungen für sie geeignet sind, was realistisch ist und was zwar modern und wünschenswert wäre, aber weniger nützlich und eher bürokratiefördernd wirkt.

## Partizipation

<b>Bewertung: Hinweise in den Berichten auf Partizipation von Partnerorganisation und Zielgruppen</b>	<b><u>gut bis befriedigend</u></b> <b>Partnerorganisation: gut; Zielgruppen: befriedigend</b>
---	--

Es gibt viele Anzeichen für ein Bemühen eines partizipativen Vorgehens in den Evaluierungen. Ein Auftakt- und Abschlussworkshop zur Besprechung des Vorgehens und zur Diskussion der Ergebnisse sind die Regel. Besonders die Partnerorganisation wird sehr gut sowohl in Planung und Durchführung als auch als Informationsquelle mit eingebunden. Aber auch die Zielgruppen werden von vielen Evaluierungen zumindest als Informationsquelle ausgiebig eingebunden und Perspektiven von ihr werden dargestellt. Verschiedene partizipative Methoden werden dabei von einem gewissen Teil genutzt. Eine Ausdifferenzierung der Zielgruppen wäre dabei noch wünschenswerter, um eine umfassende Partizipation aller Teilgruppen von Zielgruppen zu gewährleisten. In diesem Bereich ist auch kein Unterschied zwischen ELBE und EQM-Evaluierungen bzw. ein leicht positivere Tendenz bei den ELBE auszumachen, so dass es durchaus ein Vorteil der ELBE zu sein scheint, dass ein partizipatives Vorgehen einfacher gelingt. Dies kann gleichzeitig aber auch zu Einbußen im Bereich der Unabhängigkeit führen.

### **Diskussion: Spannungsfeld zwischen Partizipation und Unabhängigkeit**

Nah an den benachteiligten Zielgruppen zu sein, um sie effektiv zu unterstützen, das ist der Ansatz von MISEREOR und seinen Partnerorganisationen. Die Betroffenen selbst sollen zu Akteuren werden, vom Prozess der Evaluierung profitieren und nicht bloße DatenlieferantInnen für Evaluierungsteams sein, die die Daten abgreifen, in Berichte packen und zur Entscheidung in die Hauptquartiere der Geberorganisationen schicken. Deshalb fokussieren die Vorgaben des MISEREOR-Evaluierungskonzepts auf die möglichst partizipative Datenerhebung bei der Partnerorganisation und den Zielgruppen des Projekts. Das ist ein Unterscheidungsmerkmal zwischen der Arbeit von NRO und staatlichen Durchführungsorganisationen, die i.d.R. sich mehr auf staatliche Akteure, Wissenschaft und lokale Fach-ExpertInnen in den Zentren der Partnerländer konzentrieren.

Inwiefern besteht bei diesem partizipativen Ansatz systematisch die Gefahr, dass Evaluierungsergebnisse Daten ermitteln und Lösungsvorschläge ableiten, die jedoch langfristig die Herausforderungen nicht lösen? Wenn die Zielgruppen und die Projektmitarbeitenden zu stark an der Evaluierung beteiligt sind, muss darauf hingewiesen werden, dass dies auf Kosten der Unabhängigkeit geht. Beide Gruppen haben ein Interesse an dem Fortbestand der Förderung bspw. um über die Projektphase hinaus Transferleistungen oder Arbeitsplätzen zu erhalten. Aufgrund dieser Anreizstrukturen sollte Zurückhaltung bezüglich der Beteiligung an der direkten Erfolgskontrolle bestehen. Das soll natürlich nicht heißen, dass Zielgruppen und Projektmitarbeitende nicht partizipativ in den Evaluierungsprozess eingebunden werden sollten. Über das Bereitstellen von Daten können sie wertvolle Informationen zur Formulierung der Evaluierungsfragen und zur Diskussion und Einordnung der Evaluierungsergebnisse liefern. Um dem Unabhängigkeitsprinzip gerecht zu werden, ist es jedoch wichtig, dass weder Zielgruppen noch Partnerorganisationen an der abschließenden Datentriangulation, -analyse und -auswertung beteiligt sind.

Verzerrungen können aber auch diffuser sein. MeinungsmacherInnen der Zielgruppen können beispielsweise den Projekterfolg beschönigen, weil sie denken dass dies sozial erwünscht ist und ihnen mittelfristig persönliche Vorteile wie die Einbindung in weitere Evaluierungen durch die Partnerorganisation bringt.

Die Förderung und Entwicklung von Nehmermentalitäten ist ein Beispiel, das jedoch auch mit dem gegensätzlichen Ansatz des Top-Downs in den vergangenen Jahrzehnten oft gefördert wurde.

**Schlussfolgerungen:** Die explizit durch das MISEREOR-Evaluierungskonzept geförderte Partizipation von Partnerorganisationen und geförderten Personen steht im Spannungsfeld mit der Unabhängigkeit der Evaluierung. Im MISEREOR-Evaluierungskonzept wird diese Herausforderung nur indirekt thematisiert, indem der Anregung aus der Fachdiskussion gefolgt wird, dass Schlussfolgerungen und Empfehlungen sich aus den Ergebnissen nachvollziehbar im Evaluierungsbericht ergeben sollten.

MISEREOR und seine Partnerorganisationen haben teilweise direkten Zugang zu den Zielgruppen und können diesen noch weiter ausbauen und damit die Unterstützung für Benachteiligte noch effektiver machen.

## Empfehlungen:

**Empfehlung 13 (an EQM):** Verstärkte Nutzung von partizipativen Methoden in Evaluierungsprozessen, die die Reflexion bei Zielgruppen unterstützen und Impulse für ihre Arbeit liefern. Der Einsatz dieser Methoden benötigt Zeit für die Vorbereitung, Durchführung und ggf. Nachbereitung. Wenn Evaluation eher auf das Lernen der Gruppe fokussiert sind, dann brauchen diese Diskussionsprozesse weniger stark dokumentiert werden. Für die Rückmeldung an die Partnerorganisation, MISEREOR und für Beiträge zur laufenden Fachdiskussion ist es notwendig, deutlich mehr Kontextdaten zu den Gruppen in den Evaluierungsberichten zu dokumentieren.

**Empfehlung 14 (an EQM):** Verstärkte Analyse und Untersuchung, inwiefern die von den Projekten und Evaluierungen vorgenommene Ausdifferenzierung der Zielgruppen ausreichend ist, um kurz-, mittel- und langfristig die Situation der schwächeren Gruppen zu stärken. Inwiefern ist der zusätzliche Aufwand nützlich? Welche Unterschiede in den Empfehlungen würden dadurch vorgenommen werden? Gibt es „Fehlentwicklungen“, die bisher übersehen wurden?

## Konsistenz der TOR

<b>Bewertung: Konsistenz der TOR</b>	<b><u>befriedigend</u></b>
--------------------------------------	----------------------------

Die TOR führen durch die Ausweitung des DAC-Kriteriums „Impact“ zu „Outcome & Impact“ (Wirkungen, direkte und indirekte) zu inhaltlichen Dopplungen mit dem Kapitel Effektivität. Das führt teilweise dazu, dass die festgestellten „Wirkungen“ auf die beiden Kapitel aufgeteilt werden, um Wiederholungen im Text zu minimieren. Auch die Umstellung der Reihenfolge der DAC-Kriterien -Wirkungen vor die Effektivität ziehend- kann diese Dopplungen begünstigen, da so im Kapitel Wirkungen Begründungen aus dem Bereich der Effektivität eingebracht werden. Zu Dopplungen kommt es auch zwischen den Kapiteln „Relevanz“ und „Wirkungen“. Querschnittsthemen werden selten in den TOR vertieft. In den untersuchten Evaluierungen war das Mengengerüst (Arbeitstage, Finanzvolumen, Umfang der Fragestellungen der TOR) teilweise deutlich höher als das der ELBE.

**Schlussfolgerungen.** Die Aufteilung von Wirkungen auf der Outcome-Ebene auf die beiden Kapitel *Wirkungen* und *Effektivität* sollte aufgelöst werden, um Dopplungen vorzubeugen und die Abgrenzung der Fragestellungen zwischen den Kapiteln *Relevanz* und *Wirkungen* sollte klarer herausgearbeitet werden.

Die Erfahrung zeigt, dass die Entwicklung guter TOR Zeit und Know-how benötigt, um die Evaluierung wirklich auf die entscheidenden Fragen zu fokussieren. Daher wird es als angemessen angesehen, wenn Partnerorganisationen bei ELBE wenige zentrale Fragestellungen entwickeln und dann die Standardfragen der TOR einsetzen.

Gute TOR sind wichtig. Noch wichtiger ist für die Erstellung einer guten ELBE die Auswahl einer guten EvaluatorIn. Dazu ist es wichtig, dass die Vorgaben für das Evaluierungsbudget bzw. des Mengengerüsts (Arbeitstage) angemessen sind und die Partnerorganisation Zugang zu Netzwerken erfahrener EvaluatorInnen hat.

## Empfehlungen:

**Empfehlung 15 (an EQM, Priorität):** Darstellungen der Wirkungen des Projekts in einem Kapitel „Outcome & Effektivität“ sowie „Impact“ oder ganz klassisch nach DAC in den Kriterien „Effektivität“ und „Impact“. Klarere Abgrenzung der Fragestellungen des Kapitel Relevanz zu denen der Kapitel Wirkungen und Effektivität. Darstellung der DAC-Kriterien in internationaler Reihenfolge, da chronologisch sinnvoll. Alternativ, Bewertung der DAC-Kriterien in einem separaten Kapitel und Untergliederung der Evaluierungsergebnisse beispielsweise auf Aktivitätenebene.

**Empfehlung 16 (an KA & EQM):** Für die TOR von ELBE reichen in den meisten Fällen 4 bis 8 zentrale Fragestellungen, die die Partnerorganisation besonders interessieren, in Kombination mit den Standardfragestellungen der DAC-Kriterien aus. Details und Vertiefungen sollten dann in einem Auftaktgespräch mit den EvaluatorInnen geklärt werden.

**Empfehlung 17 (an EQM):** Systematischere Berücksichtigung von MISEREOR-Lessons learnt oder -Policies in den Fragestellungen der TOR von EQM-Evaluierungen. MISEREOR hat eine Reihe von Projekten im Bereich Bildung (Schulen, Ausbildungszentren) und Gesundheit (kirchliche Krankenhäuser) für die die Fragen nach der weiteren Strategie und ihre Einordnung in das Bildungs- und Gesundheitswesen des Landes zentral ist: Werden ggf. teure Doppelstrukturen geschaffen? Kommt es zum Crowding-out, also der Verdrängung von staatlichen Aktivitäten? Welche Perspektive besteht, das Projekt noch stärker in das staatliche System zu integrieren oder die Kohärenz (weiter) zu steigern?

**Empfehlung 18 (an KA & EQM, Priorität):** ELBE benötigen eine angemessene finanzielle Ausstattung, um qualifizierte Evaluierungsteams mit ausreichend Arbeitstagen unter Vertrag nehmen zu können. Je nach Größenordnung des Projektbudgets und Region sind mit einer Größenordnung von Kosten zwischen 8.000 € – 15.000 € zu kalkulieren.

**Empfehlung 19 (an KA & EQM, Priorität):** Unterstützung von Partnerorganisationen bei der Suche nach geeigneten EvaluatorInnen für ELBE. Der Zugang zu qualifizierten GutachterInnen ist Voraussetzung für die Umsetzung von guten TOR und die Durchführung eines guten Evaluierungsprozesses. Hier liegt einer der Schlüssel für eine deutliche Verbesserung der Evaluierungsqualität und neuer Impulse für die Partnerorganisation. Ansatzpunkt könnte zum einen das Verfahren BAZ in Lateinamerika sein. Zum anderen könnten die Identifizierung von guten nationalen EvaluatorInnen und die Beratung der InPros auch überwiegend aus der MISEREOR-Geschäftsstelle in Aachen erfolgen. Das Angebot an guten GutachterInnen nimmt zu, da ist es eher eine Frage der zur Verfügung stehenden Finanzen und der Unterscheidung zwischen eher guten und eher schwachen EvaluatorInnen, als selbst Fortbildungen für EvaluatorInnen anzubieten; ggf. könnten Gutachtertage, vergleichbar wie die in Deutschland von EQM organisierten, im Süden mit anderen NRO veranstaltet werden.

## Literatur und Quellen

- Arbeitskreis Entwicklungspolitik, DeGEval (2015):** Effizienzbewertung in Evaluationen – Handreichung für die Effizienzbewertung von Projekten und Programmen der Entwicklungszusammenarbeit, Diskussionspapier der Arbeitsgruppe Effizienz des AK Entwicklungspolitik der Gesellschaft für Evaluation.
- Austrian Development Agency (2008):** Leitfaden für Projekt- und Programmevaluierungen, Wien
- Borrmann, A. und R. Stockmann (2009):** Evaluation in der deutschen Entwicklungszusammenarbeit, Band 1 – Systemanalyse, Waxmann, Münster.
- Caspari, Alexandra (2016):** Meta-Evaluationen – die Evaluation von Evaluationen; Vortrag am Deutschen Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) am 06. April 2016 in Bonn:
- Chambers, Robert (1994):** The Origins and Practice of Participatory Rural Appraisal, in World Development, Vol. 22, No. 7, Seiten 953-969
- Furubo, Jan-Eric (2012):** After evaluation maturity: Sustainability or decay? A Swedish experience; Vortrag auf der Konferenz: The Future of Evaluation in Modern Societies, 10 Jahre CEval, 14. und 15. Juni 2012 in Saarbrücken.
- GIZ (2015):** Kooperationsmanagement in der Praxis – Gesellschaftliche Veränderungen gestalten mit Capacity WORKS, Springer Gabler, Wiesbaden.
- MISEREOR, EQM (2014):** Evaluierung in der partnerschaftlichen Entwicklungszusammenarbeit von MISEREOR, Aachen.
- MISEREOR, (o. J.):** Handreichung für MISEREOR-Partnerorganisationen zur Durchführung von lokal beauftragten externen Evaluierungen, Aachen.
- MISEREOR (2014):** Auswertung der „Externen, lokal beauftragten Evaluierungen“ (ELBEs) im Rahmen verpflichtend durchgeführter Evaluierungen von Misereor geförderten Projekten, Aachen.
- MISEREOR (2015a):** Jahresbericht 2015, Aachen
- MISEREOR (2015b):** Jahresevaluierungsbericht 2015, Aachen.
- NGO-IDEAs (2012):** Tiny Tools – Measuring Change in Communities and Groups – An Overview, IMPACT PLUS TEAM, Stuttgart.
- OECD, DAC (2008):** DAC Principles for Evaluation of Development Assistance, Development Assistance Committee, Nachdruck der Originalfassung von 1991.
- OECD, DAC (2010):** Quality Standards for Development Evaluation – DAC Guidelines and Reference Series, Paris.
- Welthungerhilfe (2015):** Sustainable Food and Nutrition Security – Orientation Framework, Bonn
- World Health Organization (2013):** Evaluation Practice Handbook, Genf

## **Anhang**



# Referenzrahmen für eine Meta-Evaluierung der Projektevaluierungen aus den Jahren 2014 – 2015

---

## I Einleitung

MISEREOR ist das katholische Hilfswerk für Entwicklungszusammenarbeit. Gemeinsam mit einheimischen Partnern unterstützt MISEREOR Menschen jeden Glaubens und jeder Kultur. Hauptziel der Arbeit mit den Partnern in Entwicklungsländern ist es, durch Förderung von vorrangig armenorientierten Projekten und Programmen zu nachhaltiger Entwicklung beizutragen. Die Projekte werden mit finanziellen Mitteln unterstützt, die Spenderinnen und Spender und das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) zur Verfügung stellen.

Mit dem BMZ wurden 2009 „Leitlinien zur ziel- und wirkungsorientierten Erfolgskontrolle“ vereinbart. Ein Element dieser Erfolgskontrolle sind Evaluierungen von Einzelvorhaben, die von MISEREOR selbst oder von seinen Partnern initiiert und verantwortet werden können. Diese sollen den Qualitätskriterien des OECD/ DAC entsprechen. Teil der Vereinbarung ist auch, dass die Qualität dieser Evaluierungen periodisch durch Metaevaluierungen überprüft wird.

Das Evaluierungssystem von MISEREOR ist ausgerichtet an den DAC Prinzipien (Unabhängigkeit und Überparteilichkeit, Glaubwürdigkeit, Nützlichkeit und Beteiligung bzw. Partnerschaftlichkeit) und an den DeGEval-Standards. Die Evaluierungen untersuchen Konzeption, Umsetzung und Wirkung von Maßnahmen und bewerten gemäß den DAC-Kriterien im Hinblick auf Relevanz, Wirkung, Effektivität, Effizienz und Nachhaltigkeit und geben konkrete Empfehlungen.

Im Verständnis von MISEREOR dienen Evaluierungen nicht nur der Erfolgskontrolle, sondern auch dem gemeinsamen Lernen. Um das Lernen aller Beteiligten sicherzustellen, ist es notwendig, dass die am Projekt Beteiligten aktiv in die Evaluierung einbezogen werden. Das heißt: Eine Evaluierung wertet nicht nur Projekte und die mit ihnen erreichten Veränderungen aus, sondern soll selbst einen Beitrag zur fortlaufenden Verbesserung zukünftiger Projektplanung leisten und das Lernen aus Erfahrungen unterstützen, sowohl bei der Partnerorganisation als auch bei MISEREOR selbst.<sup>91</sup>

Für die Evaluierung von Einzelvorhaben nutzt MISEREOR zwei verschiedene Verfahren: entweder wird der Projektträger über den Projektvertrag verpflichtet, selbst Gutachter/-innen mit der Durchführung einer externen Evaluierung zu beauftragen („externe, lokal beauftragte Evaluierung -ELBE“, ca. 80 Evaluierungen pro Jahr), oder MISEREOR initiiert und beauftragt eine Evaluierung über seinen Arbeitsbereich Evaluierung und

---

<sup>91</sup> Vgl. "Evaluierung in der partnerschaftlichen Entwicklungszusammenarbeit von MISEREOR":  
<http://www.misereor.de/fileadmin/publikationen/handreichung-evaluierung-misereor-2014.pdf>

Qualitätsmanagement - EQM (ca. 35-40 Evaluierungen pro Jahr). Solche Evaluierungen werden von zwei Gutachter(inne)n durchgeführt, je unter Vertrag genommen von MISEREOR und der Partnerorganisation. Eine erste qualitative Beurteilung der ELBE fand Ende 2013/Anfang 2014 mit einer Stichprobe von 20 Berichten statt.

## II Ziel und Gegenstand der Meta-Evaluierung

**Ziel:** MISEREOR verfügt über aussagekräftige Rückmeldungen zu der Qualität sowohl der von Partnern beauftragten Evaluierungen als auch der selbst beauftragten Evaluierungen. Diese Rückmeldungen führen bei Bedarf zu zusätzlichen Anstrengungen im Bereich der Qualitätssicherung.

**Evaluierungsgegenstand:** alle Berichte von Evaluierungsvorhaben, die in den Jahren 2014 und 2015 abgeschlossen wurden:

	2014		2015		Gesamt (davon ELBE)
	EQM	ELBE	EQM	ELBE	
Afrika/ Naher Osten	24	16	11	19	70 (35)
Asien	13	18	10	52	93 (70)
Lateinamerika	8	66	6	13	93 (79)
International	2	-	1	-	3 (0)
<b>gesamt</b>	<b>47</b>	<b>100</b>	<b>28</b>	<b>84</b>	<b>259 (184)</b>

Aus dieser Grundgesamtheit von 259 Evaluierungsberichten sollen mindestens 45 ELBE-Berichte (15 je Kontinent) und mindestens 15 Berichte von EQM-beauftragten Evaluierungen untersucht werden<sup>92</sup>. Die Berichte werden per Zufallsauswahl bestimmt.

## III Evaluierungsfragen

Die Überprüfung der Qualität der MISEREOR-Evaluierungen soll sich im Wesentlichen an den DAC-Standards und -Prinzipien orientieren. Dabei soll zwischen den von Partnern beauftragten und den von EQM beauftragten Evaluierungen differenziert werden; bei den von Partnern beauftragten Evaluierungen sollen die Ergebnisse nach Kontinent differenziert dargestellt werden.

### Vollständigkeit und Angemessenheit der Dokumentation

- Vorhandensein eines Executive Summary
- Verfügbarkeit der Terms of References - ToR in gemeinsamer Arbeitssprache oder Übersetzung im Annex
- Programm/Ablaufplan der Feldphase
- Auflistung der besuchten Orte, Organisationen, befragten Personen etc.

<sup>92</sup> Kontinentübergreifende Evaluierungen werden wegen der Besonderheiten bei ihrer Durchführung nicht in die Stichprobe einbezogen.

- Dokumentation von Fokusgruppen, Befragungen und Workshops im Annex
- Dokumentation des Auswertungsworkshops im Annex
- Verwendung von Visualisierungen, Graphiken und/oder Boxen
- Sind Form und Umfang des Evaluierungsberichts für unterschiedliche Nutzergruppen angemessen?

### **Unabhängigkeit und Überparteilichkeit**

- Gibt es Hinweise in den Evaluierungsberichten auf die Unabhängigkeit oder auch mangelnde Unabhängigkeit der Evaluierenden?
- Sind Stärken und Schwächen des Projekts im Bericht ausgewogen dargestellt?
- Mit welchen Verfahren versucht MISEREOR, die Unabhängigkeit und Überparteilichkeit einer Evaluierung sicherzustellen?

### **Glaubwürdigkeit**

- Sind die Beurteilungskriterien (zur Beantwortung der Evaluierungsfragen) klar definiert?
- Ist der Bericht klar untergliedert in einen beschreibenden und einen bewertenden Teil?
- Basieren die Schlussfolgerungen nachvollziehbar auf den erhobenen Daten? Wird in Querverweisen auf die entsprechenden Daten Bezug genommen?
- Lassen sich die Empfehlungen klar aus den Schlussfolgerungen herleiten?
- Wurden die Zielgruppen nach Alter, Geschlecht, Wohlstand und anderen sozialen Kriterien differenziert?
- Spiegelt der Bericht die verschiedenen Interessen und Perspektiven dieser Gruppen wider?
- Werden die Ergebnisse der Heterogenität der Zielgruppen entsprechend differenziert dargestellt?
- Gibt es ein Methodenkapitel und wenn ja, stellt es die verwendeten Methoden klar und hinreichend dar?
- Sind die angewandten Evaluierungsmethoden und -instrumente angemessen, um die in den ToR gestellten Evaluierungsfragen zu beantworten?
- Wie umfangreich und aussagekräftig waren die Stichproben?
- Wurden der Ablauf der Evaluierung oder die Methoden während des Evaluierungsprozesses angepasst und wenn ja, warum? Sind diese Veränderungen im Bericht kenntlich gemacht?
- Wurden die Methoden systematisch zur Beantwortung der Evaluierungsfragen eingesetzt und wird auf sie Bezug genommen?
- Sind die Ergebnisse von Fokusgruppen, Befragungen und/oder anderen Erhebungsmethoden gut nachvollziehbar dokumentiert?
- Wie werden die mit verschiedenen Instrumenten/Methoden erhobenen Daten präsentiert?

- Inwieweit wurden die Daten durch die Nutzung verschiedener Quellen und Anwendung verschiedener Methoden trianguliert?
- Mit welchen Methoden wird das Effizienzkriterium bearbeitet?
- Lassen sich beobachtete Veränderungen (Wirkungen) plausibel auf die Einwirkung des Projekts zurückführen (Attribution/Kontribution)? Wurde die kontrafaktische Situation mitgedacht?

### **Nützlichkeit**

- Beantwortet der Evaluierungsbericht die in den ToR gestellten Evaluierungsfragen vollständig?
- Enthält der Endbericht eine klar strukturierte Kurzfassung (Executive Summary), in der der Evaluierungskontext, der Untersuchungsbereich (scope), sowie die wichtigsten Ergebnisse, Schlussfolgerungen und Empfehlungen klar und gut verständlich dargestellt werden?
- Ist der Bericht sinnvoll gegliedert, z. B. nach DAC Kriterien und/oder spezifischen Evaluierungsfragen?
- Wurden alle Evaluierungskriterien ausreichend abgedeckt? Falls nicht, welche Kriterien wurden genauer betrachtet und welche entweder fallengelassen oder nur oberflächlich behandelt? Entspricht dies der Schwerpunktsetzung in den ToR?
- Stellt der Bericht die im Projekt genutzte Theorie des Wandels bzw. Wirkungslogik dar oder rekonstruiert diese?
- Sind die Schlussfolgerungen klar aus der Analyse nachvollziehbar und führen sie zu nachvollziehbaren, praxisnahen Empfehlungen?
- Wurde ein Auswertungsworkshop durchgeführt und ist er dokumentiert?
- An wen richten sich die Empfehlungen hauptsächlich? (Beteiligte, Projektpartner, MISEREOR?)
- Sind sie pragmatisch und operationalisierbar?

### **Partizipation/Beteiligung**

- Ist ersichtlich, dass die Partnerorganisationen und Zielgruppen aktiv an der Planung und Durchführung des Evaluierungsprozesses beteiligt waren?
- Wie wurden der Evaluierungsprozess und die Methodik mit den Partnerorganisationen und Zielgruppen abgestimmt? Gab es einen Auftaktworkshop?
- Welche Methoden wurden eingesetzt, um die Beteiligung der Stakeholder zu ermöglichen?
- Handelt es sich um eine durch Gutachter/-innen begleitete Selbstevaluierung?
- Sind die bei MISEREOR etablierten Verfahren geeignet, um sicherzustellen, dass partizipative Evaluierungsprozesse stattfinden?

## **Konsistenz der ToR**

- Sind die ToR und die Evaluierungsfragen ausreichend präzise formuliert?
- Sind sie im vorgegebenen Zeitraum/und mit dem vorgegebenen Mengengerüst gut umsetzbar?
- Sind sowohl DAC-Kriterien als auch zusätzliche, spezifisch auf das Projekt oder auf die zu evaluierende Partnerorganisation abgestimmte Fragen ausreichend abgedeckt?
- Wird neben der Zielerreichung auch nach unerwarteten Wirkungen gefragt?
- Werden Querschnittsthemen berücksichtigt? Wenn ja, welche (z. B. Gender, Konflikt, Frieden, Ökologie, Zivilgesellschaft)?

## **IV Organisation und Vorgehensweise**

Der Evaluierungsauftrag umfasst:

- Eine systematische Untersuchung (Deskstudie) von 60 Evaluierungen (Berichten), die in den Jahren 2014-2015 abgeschlossen wurden.
- Die Beschreibung und Analyse von Beispielen guter Praxis zu methodischen Aspekten aus verschiedenen Evaluierungen (Kontext, methodisches Vorgehen und Diskussion der Übertragbarkeit auf andere Evaluierungen.)
- Workshops mit EQM und ggf. weiteren Mitarbeitenden der Abteilung Internationale Zusammenarbeit von MISEREOR, zum Austausch über die vorläufigen Ergebnisse und Empfehlungen für das weitere Vorgehen in der Qualitätssicherung.
- Eine Abschlusspräsentation und Diskussion mit EQM und ggf. den Abteilungsleitungen der Internationalen Zusammenarbeit von MISEREOR .

In der Durchführung wird dabei wie folgt vorgegangen:

### **Phase 1: INCEPTION PHASE**

Innerhalb von drei Wochen nach Vertragsunterzeichnung erstellt das Evaluierungsteam einen Inception Report, der folgendes beinhaltet:

- Deskriptiver Teil
  - Allgemeiner Hintergrund und Informationen der zu untersuchenden Evaluierungen
  - Das Ergebnis der ersten Durchsicht der Dokumente (Sind alle erforderlichen Dokumente vorhanden? Welche Dokumente fehlen noch und müssen ggf. noch zugänglich gemacht werden?)
- Methodenteil
  - Vorschlag eines Analyserasters, das als Leitfaden für die Deskstudie verwendet werden soll. Der Leitfaden sollte eine Kombination von geschlossenen und offenen Fragen sein.
  - Vorschläge für die Auswahl und das Format der Beispiele guter Praxis

- Vorläufiger Zeitplan mit geplanten Workshops, Interviews, sodass die Deadlines der folgenden Schritte eingehalten werden können.

## Phase 2: ASSESSMENT PHASE

Innerhalb von zwei Monaten nach Abnahme des Inception Reports erstellt das Evaluierungsteam den Entwurf eines Evaluierungsberichts, der Folgendes beinhaltet:

- Die Ergebnisse der Deskstudie, die alle Fakten der geschlossenen Fragen zusammenfasst, die Synthese der Beantwortung der offenen Fragen und die wesentlichen Erkenntnisse der Desk Studie.
- Die Beispiele guter Praxis und die wichtigsten Lernerfahrungen, die daraus abgeleitet werden können.
- Die vorläufigen Schlussfolgerungen und die daraus abgeleiteten, vorläufigen Empfehlungen in Form einer PPT-Präsentation oder einer Handreichung für die geplanten Workshops.

## Phase 3: AUSWERTUNGSPHASE (WORKSHOPS)

Das Evaluierungsteam moderiert die o. g. Workshops und dokumentiert Beobachtungen, Anmerkungen und insbesondere die überarbeitete Version der Schlussfolgerungen und Empfehlungen, die das Ergebnis des Austauschs sind.

### Vorläufiger Zeitplan:

Was	Wann
Ausschreibung	18.04.16
Einreichung Angebote	06.05.16
Entscheidung Gutachter(-team)	13.05.16
Auftragsklärungsgespräch	bis Ende Mai
Inception Report	Bis Mitte Juni
Untersuchungsphase/ Berichtsentwurf	Bis Ende August
Auswertungsworkshop	Bis Mitte September
Auswertung auf Abteilungsleitungsebene	Bis Ende September
Endbericht	Bis Ende Oktober

## 7 V Bericht

Der Bericht umfasst ca. 50 Seiten (ohne Anhänge). Dem ausführlichen Bericht wird eine knappe 3-5 seitige **Zusammenfassung** vorangestellt (Executive summary).

Der Bericht untergliedert sich in einen **beschreibenden** und einen **bewertenden** Teil, aus dem sich die Schlussfolgerungen und Empfehlungen herleiten.

In einem **Methodenkapitel** wird der Evaluierungsauftrag präzise dargestellt und die Wahl der angewandten Methodik in Relation zu diesem Auftrag beschrieben und begründet.

Ausgewählte **Gute Praktiken** aus der Durchführung von Evaluierungen werden in einem separaten Teil aufgezeigt und bieten Anregungen für weitere Qualitätsverbesserungen in der ELBE- und EQM-Evaluierungspraxis.

In den **Anhängen** werden die detaillierten Ergebnisse der Analyse/Fragebogen und Workshops dokumentiert.

Aachen 18.04.2016

We

## Anlage 2: Analyseraster - Qualitäts-Basischeck (Stufe 1)

Frage	Codierung	Antwortmöglichkeiten	Gewichtung
<b>1 Vollständigkeit und Angemessenheit der Dokumentation</b>		sehr gut, gut-befriedigend, eher schwach, schwach	
<b>1.1 Gliederung ist angemessen</b>	1-4	sehr gut, gut-befriedigend, eher schwach, schwach	2
<b>1.2 Terms of References - TOR in gemeinsamer Arbeitssprache oder Übersetzung im Annex verfügbar?</b>	1,4, Annex fehlt ist nein	ja; nein	0,5
<b>1.3 Programm der Feldphase im Annex oder im Bericht?</b>	1-4, Annex fehlt ist nein	sehr gut, gut-befriedigend, eher schwach, schwach	1
<b>1.4 Besuchte Orte, Organisationen, befragten Personen etc. aufgelistet im Annex?</b>	1,4, Annex fehlt ist nein	ja; unvollständig; nein	0,5
<b>1.5 Wurden die, in der Dokumentenanalyse, genutzten Dokumente aufgelistet?</b>	1,4, Annex fehlt ist nein	ja; nein	0,5
<b>1.6 Dokumentation von Leitfäden der Fokusgruppen, Befragungen und Interviews im Annex vorhanden?</b>	1-4, Annex fehlt ist nein	sehr gut; gut-befriedigend; eher schwach; schwach	1
<b>1.7 Werden Visualisierungen, Grafiken und Boxen angemessen eingesetzt?</b>	1-4	sehr gut; gut-befriedigend; eher schwach; schwach	1
<b>2 Unabhängigkeit/Überparteilichkeit</b>			
<b>2.1 Ist es ein externes Evaluationsteam ?</b>	keine Info heißt nein, gemischt heißt nein	ja; nein; gemischt; keine Information	0,5
<b>2.2 Sind institutionelle Zugehörigkeiten und Arbeitsschwerpunkt des Evaluationsteam dargestellt?</b>		ja; nein	0,5
<b>3 Glaubwürdigkeit</b>		sehr gut, gut-befriedigend; eher schwach; schwach	
<b>DAC Kriterien</b>	Bewertung 1-4 (Durchschnitt)		1

<b>3.1 Wird auf das DAC-Kriterium Relevanz eingegangen?</b>	1,4	ja; nein	0,5
<b>3.2 Wird auf das DAC-Kriterium Wirkungen (Outcome and Impact) eingegangen?</b>	1,4	ja; nein	0,5
<b>3.3 Wird auf das DAC-Kriterium Effektivität eingegangen?</b>	1,4	ja; nein	0,5
<b>3.4 Wird auf das DAC-Kriterium Effizienz eingegangen?</b>	1,4	ja; nein	0,5
<b>3.5 Wird auf das DAC-Kriterium Nachhaltigkeit eingegangen?</b>	1,4	ja; nein	0,5
<b>3.6 Werden die angewandten Methoden dargelegt?</b>	1-4	sehr gut, gut-befriedigend; eher schwach; schwach	2
<b>3.7 Wurden verschiedene Methoden benutzt, die qualitativen und quantitativen Aspekte abdecken?</b>	1-4	sehr gut, gut-befriedigend; eher schwach; schwach	1
<b>Aggregiert Workshop (3.8, 3.9)</b>	1,3,4 (beide, einer, keiner)		1
<b>3.8 Hat ein Auftaktworkshop stattgefunden?</b>		ja; nein	
<b>3.9 Hat ein Abschlussworkshop stattgefunden?</b>		ja; nein	
<b>Methodenvielfalt</b>	1-4		2
<b>3.10 Wurden in der Evaluierung Interviews durchgeführt?</b>	1,4	ja; nein	0,5
<b>3.11 Fanden Fokusgruppensitzungen statt?</b>	1,4	ja; nein	0,5
<b>3.12 Fand eine (semi-) standardisierte Fragebogenerhebung statt?</b>	1,4	ja; nein	0,5
<b>3.13 Wurden Monitoring-Daten des Projekts analysiert?</b>	1,4	ja; nein	0,5
<b>3.14 Wurde eine Analyse der Finanzdaten durchgeführt?</b>	1,4	ja; nein	0,5
<b>3.15 Wird eine Theory of change, Wirkungskette oder Wirkungslogik dargestellt?</b>	1,4	ja; nein	1
<b>3.16 Wurden weitere partizipative Methoden genutzt?</b>	1,4	ja; nein	0,5
<b>3.17 Wenn ja welche (MAPP, Tiny Tool etc):</b>			
<b>3.18 Wurde eine Vergleichsgruppe befragt oder analysiert?</b>	nur ein Antwort "ja", keine Einbeziehung	ja;nein	

<b>4 Nützlichkeit</b>		sehr gut, gut- befriedigend, eher schwach, schwach	
<b>4.1 Executive Summary enthält Scope/ Untersuchungsbereich der Evaluation?</b>	1,4; nein wenn kein exec summ	ja; nein; kein executive summary	0,5
<b>4.2 Executive Summary enthält Vorgehen in der Evaluation?</b>	1,4 nein wenn kein exec summ	ja; nein	0,5
<b>4.3 Executive Summary enthält wichtigste Ergebnisse?</b>	1,4; nein wenn kein exec summ	ja; nein	1
<b>4.4 Executive Summary enthält Schlussfolgerungen / Empfehlungen?</b>	1,4; nein wenn kein exec summ	ja; nein	0,5
<b>4.5 Bericht enthält Empfehlungen?</b>	1,4	ja; nein	1
<b>4.6 Operationalisierbarkeit der Empfehlungen: Gibt es eine Priorisierung der Empfehlungen?</b>	1,4	ja; nein	0,5
<b>4.7 Sind die Empfehlungen pragmatisch d.h. geben Sie konkrete Ratschläge zur Umsetzung?</b>	1,4	ja; nein	0,5










## Anlage 4: Analyseraster der vertieften Analyse (Stufe 2)

Evaluierungsfrage:	Antwort-kategorie	Kategorienbeschreibung	Antwort	Kommentar für Freitext	Freitext
<b>Vollständigkeit und Angemessenheit der Dokumentation</b>					
<b>1 Ist die Dokumentation vollständig und angemessen?</b>					
<b>1.1 Gliederung ist angemessen</b>	sehr gut, gut-befriedigend, eher schwach, schwach	<b>Sehr gut:</b> Exe. Summary, Project Description, Methodology, Findings, Conclusions, Recommendations + DAC Kriterien; <b>Gut-Befriedigend:</b> Wie 1 mit ein-zwei fehlenden Unterteilungen; <b>Eher schwach:</b> weniger Unterkapitel, nicht ganz logische Struktur; <b>Schwach:</b> kein Exe. Summary, keine logische Struktur			
<b>1.2 Terms of References - TOR in gemeinsamer Arbeitssprache oder Übersetzung im Annex verfügbar?</b>	ja; nein	<b>ja:</b> in gemeinsamer Arbeitssprache im Annex; <b>nein:</b> nicht in gemeinsamer Arbeitssprache oder nicht vorhanden		Nur bei Besonderheit Freitext	
<b>1.3 Programm der Feldphase im Annex?</b>	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> detaillierte Beschreibung des Ablaufs der Evaluierung, mit Aufschlüsselung nach Datenerhebungsmethode; <b>gut-befriedigend:</b> Beschreibung des Ablaufs der Evaluierung vorhanden, nicht alle Informationen vorhanden (Beispiel 5 Tage Datenerhebung ohne Spezifizierung); <b>eher schwach:</b> Beschreibung nur unvollständig vorhanden, nicht nur im Annex; <b>schwach:</b> keine Beschreibung des Ablaufs der Evaluierung vorhanden			

1.4 Besuchte Orte, Organisationen, befragten Personen etc. aufgelistet im Annex?	ja; unvollständig; nein			Nur bei Besonderheit Freitext	
1.5 Wurden die in der Dokumentenanalyse genutzten Dokumente aufgelistet?	ja; nein				
1.6 Dokumentation von Leitfäden der Fokusgruppen, Befragungen und Interviews und der Ergebnisse im Annex vorhanden?	sehr gut; gut-befriedigend; eher schwach; schwach	<b>sehr gut:</b> umfassende und informative Dokumentation der Instrumente und der Ergebnisse vorhanden (Fragebogen, Inhalt der Fokusgruppengespräche und der Workshops); <b>gut-befriedigend:</b> Leitfaden oder Ergebnisdokumentation; <b>eher schwach:</b> Ergebnis oder Leitfadendokumentation lückenhaft; <b>schwach:</b> Keine Dokumentation der Instrumente			
1.7 Werden Visualisierungen, Grafiken und Boxen angemessen eingesetzt?	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> angemessene Anzahl, auf dem ersten Blick sinnvollen Daten, anschaulich gestaltet, leicht zu verstehen; <b>gut-befriedigend:</b> gute Anzahl, Information nicht immer leicht zu verstehen/gut aufbereitet/wichtige Information; <b>eher schwach:</b> zu viel oder zu wenig Grafiken oder Tabellen, unverständlich, unwichtige Information; <b>schwach:</b> kein Anwendung von Grafiken und Visualisierungen		Stärken und Schwächen aufzeigen	
Unabhängigkeit und Unparteilichkeit					
2 Ist der Bericht unabhängig/überparteilich?					
2.1 Ist es ein externes Evaluationsteam?	ja; nein; gemischt; keine Information			Darstellung der vorhandenen Informationen	

2.2 Sind institutionelle Zugehörigkeiten oder (Arbeits-) Schwerpunkte des Evaluationsteam dargestellt?	ja; nein			Darstellung der vorhandenen Informationen zum Evaluationsteam	
2.3 Wurden Informationsquellen so gewählt, dass eine Vielzahl von Perspektiven in der Datenerhebung, Datenauswertung und Darstellung der Ergebnisse erfasst wurde?					
a) Einbeziehung der Partnerorganisationen im gesamten Prozess	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> Die PO war eine sehr wichtige Informationsquelle und die Perspektive der PO fließt sehr gut mit ein, <b>gut-befriedigend:</b> Die PO war eine wichtige Informationsquelle und die Perspektive der PO fließt gut mit ein, <b>eher schwach:</b> Die PO diente eher nicht als Informationsquelle und die Perspektive der PO fließt wenig mit ein, <b>schwach:</b> Die PO diente nicht als Informationsquelle und die Perspektive der PO wurde nicht dargestellt.			
b) Einbeziehung der Zielgruppen (Vielzahl von Vertretern)	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> Eine Vielzahl von ZG-Vertretern war Informationsquelle und die Perspektive der ZG fließt sehr gut mit ein, <b>gut-befriedigend:</b> Eine eher hohe Zahl von ZG-Vertretern war Informationsquelle und die Perspektive der ZG fließt gut mit ein, <b>eher schwach:</b> Eine geringe Zahl von ZG-Vertretern diente als Informationsquelle und die Perspektive der ZG fließt wenig mit ein, <b>schwach:</b> Die ZG dienten nicht als Informationsquelle und die Perspektive der ZG wurde nicht dargestellt.			

<b>c) Einbeziehung von externe ExpertInnen</b>	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> Eine Vielzahl von Experten war Informationsquelle und die Perspektive der Experten fließt sehr gut mit ein, <b>gut-befriedigend:</b> Eine eher hohe Zahl von Experten war Informationsquelle, die Perspektive der Experten fließt gut mit ein, <b>eher schwach:</b> Eine geringe Zahl von Experten diente als Informationsquelle und die Perspektive der Experten fließt vereinzelt mit ein, <b>schwach:</b> Die Experten dienten nicht als Informationsquelle und die Perspektive der Experten wurde nicht dargestellt.			
<b>d) Einbeziehung der Fachdiskussion; Lessons learnt aus anderen Projekten</b>	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> Die Fachdiskussion diente deutlich als Informationsquelle und die Perspektive der Fachdiskussion fließt sehr gut mit ein, <b>gut-befriedigend:</b> Die Fachdiskussion diente zum Teil als Informationsquelle und die Perspektive der Fachdiskussion fließt gut mit ein, <b>eher schwach:</b> Die Fachdiskussion diente vereinzelt als Informationsquelle und die Perspektive der Fachdiskussion fließt vereinzelt mit ein, <b>schwach:</b> Die Fachdiskussion diente nicht als Informationsquelle, die Perspektive der Fachdiskussion wurde nicht dargestellt.			
<b>2.4 Werden Stärken und Schwächen ausgewogen dargestellt?</b>	sehr gut/gut, befriedigend, eher schwach, schwach	<b>sehr gut:</b> es wird ausgewogen und fair auf Stärken und Schwächen des Programms eingegangen; <b>gut-befriedigend:</b> Stärken und Schwächen werden dargestellt, <b>eher schwach:</b> kein ausgewogenes Verhältnis zw. Stärken und Schwächen; <b>schwach:</b> nur einseitige Darstellung		Wo erfolgt eine Wertschätzung der Arbeit? (Vorwort; Summary; Conclusions, Empf); Welche kritischen Punkte werden explizit angeführt?	
<b>Glaubwürdigkeit &amp; Genauigkeit</b>					

<b>3 Wird das Wirkungsgefüge angemessen und mit verschiedenen Stakeholdern diskutiert?</b>					
<b>3.1 Nutzt die Analyse eine Theory of Change, Wirkungskette oder Wirkungslogik? Wenn ja, wie wird Wirkungslogik mit einbezogen?</b>	ja; nein	<b>ja:</b> bereits vorhandene oder selbstaufgestellte Wirkungslogik wird dargestellt, <b>nein:</b> keine Darstellung einer Wirkungslogik		Wenn ja, wie wird Wirkungslogik mit einbezogen?	
<b>3.2 Wurden anhand der Dokumente Wirkungshypothesen bzw. ein Wirkungsgefüge entwickelt?</b>	ja; nein	<b>ja:</b> Entwicklung eines Wirkungsgefüge durch den Evaluator; <b>nein:</b> keine Entwicklung durch Evaluator			
<b>4 Sind Auswahl der Methoden und Darlegung der Methoden angemessen?</b>					
<b>4.1 Werden die angewandten Methoden dargelegt?</b>	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> Methodenkapitel mit aussagekräftiger Erklärung; <b>gut-Befriedigend:</b> Methodenkapitel oder Absatz zu Methoden, keine vollständige Information; <b>eher schwach:</b> Kaum Beschreibung der Methoden (nur 1-3 Sätze); <b>schwach:</b> keine Methodenbeschreibung			
<b>4.2 Wurde das Kontrafaktische beachtet? Wenn ja wie?</b>	ja; nein			Welche Methoden wurden verwendet, um das Kontrafaktische zu beachten? (Vergleichsgruppe, Einflussmatrix etc.)	

4.3 Wurden verschiedene Methoden benutzt, die qualitative und quantitative Aspekte abdecken?	sehr gut/gut, befriedigend, eher schwach, schwach	<b>sehr gut:</b> Es wurden qualitative und quantitative Methoden in angemessener Form benutzt; <b>gut-befriedigend:</b> Es wurden qualitative und quantitative Methoden benutzt; <b>eher schwach:</b> Es wurden nur qualitativen Methoden benutzt, aber vielfältige; <b>schwach:</b> Es wurde nur wenig qualitative Methoden benutze und kein quantitative			
4.4 Hat ein Auftaktworkshop stattgefunden?	ja; nein	ja: es wird deutlich, dass es ein Treffen gab wo nicht nur Leitungsebene dabei war;		Nur bei Besonderheit Freitext	
4.5 Hat ein Abschlussworkshop stattgefunden?	ja; nein			Nur bei Besonderheit Freitext	
4.6 Wurden in der Evaluierung Interviews durchgeführt?	ja; nein			Nur bei Besonderheit Freitext	
4.7 Fanden Fokusgruppendifkussionen statt?	ja; nein			Nur bei Besonderheit Freitext	
4.8 Fand eine (semi-) standardisierte Fragenbogenerhebung statt?	ja; nein			Nur bei Besonderheit Freitext	
4.9 Wurden Monitoring-Daten der Projektaktivitäten analysiert?	ja; nein			Nur bei Besonderheit Freitext	
4.10 Wurde eine Analyse der Finanzdaten durchgeführt?	ja; nein			Nur bei Besonderheit Freitext	
4.11 Wurden weitere partizipative Methoden genutzt?	ja; nein			Nur bei Besonderheit Freitext	

4.12 Wenn ja welche (MAPP, Tiny Tool etc):					
4.13 Wurden Informationen und Beobachtungen trianguliert? (u.a. durch Methodenmix)	sehr gut; gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> es wurden verschiedene Methoden <b>und</b> verschiedene Informationsquellen genutzt; <b>gut-befriedigend:</b> es wurden entweder verschiedene Methoden oder verschiedene Informationsquellen genutzt, <b>eher schwach:</b> geringe Informationsquellen und Methodenvielfalt; <b>schwach:</b> eine Informationsquelle oder Methode			
5 Wurden die Befragten, die Stichprobe etc. angemessen und nach klaren Kriterien ausgewählt?					
5.1 Werden die Auswahlverfahren und -kriterien im Bericht offengelegt?	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> aussagekräftige Beschreibung im Methodenteil, die Auswahlprozess dokumentiert und neben Vorschlägen der PO auch kriteriengeleitete Vorschläge und Zufallsauswahl beinhaltet; <b>gut-befriedigend:</b> Auswahlprozess ist dokumentiert und zeigt, dass nicht nur den Vorschlägen der PO gefolgt wurde; <b>eher schwach:</b> Auswahlprozess ist nicht ausreichend dokumentiert und keine präzisen Hinweise über die Kriterien, die für die Auswahl entscheidend waren; <b>schwach:</b> keine Information zum Auswahlprozess			
5.2 Wurde eine Zufallsstichprobe durchgeführt?	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> Zufallsstichprobe die stat. Ansprüchen genügt, <b>gut-befriedigend:</b> Zufallsstichprobe die erläutert wird, <b>eher schwach:</b> Zufallsstichprobe ohne nähere Angaben, <b>schwach:</b> keine Zufallsstichprobe			
5.3 Wurden bewusst stärkere, durchschnittliche und schwächere Elemente der Grundgesamtheit	ja; nein	<b>ja:</b> In Beschreibung der Auswahl wird dieses Vorgehen beschrieben, <b>nein:</b> keine Beschreibung dieses Vorgehens			

ausgewählt?					
5.4 Wie umfangreich und aussagekräftig ist die Stichprobe?				Beschreibung der Stichprobe	
6 Wurde die Datenauswertung angemessen durchgeführt?					
6.1 Wurden die Zielgruppen klar nach sozio-ökonomischen Kriterien differenziert?	sehr gut/gut, befriedigend, eher schwach, schwach	<b>sehr gut:</b> Zielgruppen sind in den Zahlenauswertungen/ Tabellen detailliert nach soz.-ök- Kriterien untergliedert; <b>gut-befriedigend:</b> es findet eine Untergliederung statt aber nach wenigen Kriterien; <b>eher schwach:</b> es wird nicht detailliert und nur nach einem Kriterium die ZG untergliedert; <b>schwach:</b> keinerlei Untergliederung der Zielgruppe			
6.2 Spiegelt der Bericht die verschiedenen Interessen und Perspektiven dieser Gruppen wieder?	sehr gut/gut, befriedigend, eher schwach, schwach	<b>sehr gut/gut:</b> Es wird in Ergebnissen, Conclusions und Recommendations nach ZG differenziert und an verschiedene Bedürfnisse angepasst; <b>befriedigend:</b> es findet zum Teil eine Anpassung an verschiedene Perspektiven innerhalb der ZG angepasst; <b>eher schwach:</b> es wird kaum auf die verschiedenen ZG eingegangen; <b>schwach:</b> es wird gar nicht auf die Interessen verschiedener ZG eingegangen			
6.3 Werden Aussagen wann immer möglich quantifiziert? (60% der Interviewpartner, oder drei von fünf Gruppen...)	sehr gut/gut, befriedigend, eher schwach, schwach	<b>sehr gut:</b> In den Ergebnisse werden wann immer möglich quantifizierte Aussagen getroffen; <b>gut-befriedigend:</b> In den Ergebnissen werden häufig quantifizierte Aussagen getroffen; <b>eher schwach:</b> in den Ergebnissen werden gelegentlich quantifizierte Aussagen getroffen; <b>schwach:</b> in den Ergebnissen werden praktisch nicht quantifizierte Aussagen getroffen			

6.4 Werden Einschätzungen und Bewertungen begründet Beantwortung der Evaluierungsfragen klar definiert? (begründete Bewertung)	sehr gut/gut, befriedigend, eher schwach, schwach	<b>sehr gut:</b> explizite Beurteilungskriterien für Bewertungen und Einschätzungen werden ; <b>gut-befriedigend:</b> Einschätzungen und Bewertungen wird häufig begründet; <b>eher schwach:</b> es finden sich selten Begründungen; <b>schwach:</b> es werden keinerlei Beurteilungskriterien deutlich			
7 Werden die Evaluierungskriterien angemessen beantwortet?					
7.1 Geht die Evaluierung auf das DAC-Kriterium Relevanz ein?	ja; nein			Nur bei Besonderheit Freitext	
7.2 Wie wird im Kapitel Relevanz folgende Frage beantwortet: Tut das Projekt das Richtige?					
7.3 Geht die Evaluierung auf das DAC-Kriterium Wirkungen (Outcome and Impact) ein?	ja; nein			Nur bei Besonderheit Freitext	
7.4 Wie wird über Wirkungen berichtet? Frage: Was verändert sich durch das Projekt tatsächlich für die Zielgruppen? (auch unerwartet und indirekte)					
7.5 Wie wird überprüft, dass sich die beobachteten Veränderungen plausibel auf das Projekt zurückführen					

lassen?					
7.6 Geht der Evaluator auf das DAC-Kriterium Effektivität ein?	ja; nein			Nur bei Besonderheit Freitext	
7.7 Wie wird im Kapitel Effektivität folgende Frage beantwortet: Erreichen wir die Ziele des Projekts? (realistische Ziele, Indikatoren realistisch)					
7.8 Geht die Evaluierung auf das DAC-Kriterium Effizienz ein?	ja; nein			Nur bei Besonderheit Freitext	
7.9 Wie wird im Kapitel Effizienz folgende Frage beantwortet: Werden die Ziele des Projekts wirtschaftlich erreicht? (Stand der Ausgaben, Qualität des Projektmanagement, Benchmarks, Qualität des M&E Systems)					
7.10 Welche Methoden werden zur Bearbeitung des Effizienzkriteriums genutzt?					
7.11 Geht die Evaluierung auf das DAC-Kriterium Nachhaltigkeit eingegangen?	ja; nein			Nur bei Besonderheit Freitext	

7.12 Wie wird im Kapitel Nachhaltigkeit folgende Frage beantwortet: Sind die positiven Wirkungen von Dauer?					
8 Sind die Schlussfolgerungen angemessen?					
8.1 Wurden beschreibende und bewertende Aussagen voneinander getrennt?	ja; nein	<b>ja:</b> in Gliederung klare Unterscheidung; <b>nein:</b> keine Untergliederung von beschreibend und bewertend			
8.2 Basieren die Schlussfolgerungen nachvollziehbar auf den Ergebnissen?	sehr gut; gut-befriedigend; eher schwach; schwach	<b>sehr gut:</b> Schlussfolgerungen lassen sich klar und vollständig aus Ergebnissen folgern; <b>gut-befriedigend:</b> Schlussfolgerungen lassen sich zum größten Teil aus Ergebnissen folgern, <b>eher schwach:</b> Schlussfolgerungen folgern eher weniger aus Ergebnissen, <b>schwach:</b> Schlussfolgerungen lassen sich nicht aus Ergebnissen folgern			
8.3 Wird dabei in Querverweisen auf die entsprechenden Daten Bezug genommen?	ja; nein				
8.4 Wie werden die Empfehlungen aus den Schlussfolgerungen hergeleitet?					
Nützlichkeit					
9 Enthält der Endbericht eine klar strukturierte, gut lesbare Zusammenfassung?					

9.1 Executive Summary enthält Scope/ Untersuchungsbereich der Evaluierung?	ja; nein			Nur bei Besonderheit Freitext	
9.2 Executive Summary enthält Vorgehen in der Evaluierung?	ja; nein			Nur bei Besonderheit Freitext	
9.3 Executive Summary enthält wichtigste Ergebnisse?	ja; nein			Nur bei Besonderheit Freitext	
9.4 Executive Summary enthält Schlussfolgerungen / Empfehlungen?	ja; nein			Nur bei Besonderheit Freitext	
9.5 Wie ist das Executive Summary bezüglich Lesbarkeit, Verständlichkeit und Informationsgehalt zu bewerten?					
10 Deckt der Evaluierungsbericht die Evaluierungsfragen der TOR vollständig ab?					
10.1 Beantwortet der Evaluierungsbericht die in den TOR gestellten Evaluierungsfragen angemessen?	sehr gut, gut-befriedigend, eher schwach, schwach	<b>sehr gut:</b> Bericht beantwortet vollständig und systematisch alle Evaluierungsfragen der TOR; <b>gut-befriedigend:</b> Bericht beantwortet die einen großen Teil Evaluierungsfragen angemessen; <b>eher schwach:</b> Bericht beantwortet nur unzureichend die Evaluierungsfragen; <b>schwach:</b> Bericht beantwortet nicht die Evaluierungsfragen			
10.2 Wird die Schwerpunktsetzung gemäß	ja; nein	Besonderes Erkenntnisinteresse			

den TOR umgesetzt?					
11 Sind die Empfehlungen des Berichts angemessen?					
11.1 Bericht enthält Empfehlungen?	ja; nein				
11.2 Operationalisierbarkeit der Empfehlungen: Gibt es eine Priorisierung der Empfehlungen? Wie wird priorisiert?	ja; nein				
11.3 Sind die Empfehlungen pragmatisch d.h. geben Sie konkrete Ratschläge zur Umsetzung? Wie werden Ratschläge zu geben? (Adressaten berücksichtigen)	ja; nein				
Partizipation					
12 Wurde die Evaluierung partizipativ durchgeführt?					
12.1 Ist ersichtlich, dass Partnerorganisationen aktiv an der Planung und Durchführung des Evaluierungsprozesses beteiligt waren? (nicht nur Interviewpartner)	ja; nein				

<b>12.2 Ist ersichtlich, dass Zielgruppen aktiv an der Planung und Durchführung des Evaluierungsprozesses beteiligt waren? (nicht nur Interviewpartner)</b>					
<b>Qualität der TOR</b>					
<b>13 Sind die TOR angemessen?</b>					
<b>13.1 Wurden die Evaluierungsfragen an den individuellen Projektkontext angepasst?</b>	ja; nein				
<b>13.2 Können die Evaluierungsfragen in dem vorgegebenen zeitlichen Rahmen angemessen bearbeitet werden?</b>	ja; nein				
<b>13.3 Werden Querschnittsthemen in den TOR berücksichtigt?</b>					
<b>a) Gender</b>	ja; nein				
<b>b) Konflikt</b>	ja; nein				
<b>c) Frieden</b>	ja; nein				
<b>d) Ökologie</b>	ja; nein				
<b>e) Zivilgesellschaft</b>	ja; nein				
<b>Sonstiges Querschnittsthema</b>					

**Anlage 5: Die 20 ausgewählten Evaluierungsberichte der vertieften Analyse (Stufe 2)**

(separate Anlage wg. Datenschutz)

Nr.	Region	Land	Vorschlag von	Titel	EQM/ELBE	Jahr



**Anlage 6: Ergebnisse der vertieften Analysen (separater Anlagenteil aus Datenschutz)**  
(separate Anlage wg. Datenschutz)

## Anlage 7: Gute Praxis Beispiel: Methodenkapitel (Nr.301, Mexiko, ELBE)

### 1. Beschreibung des (partizipativen) Vorgehens

#### IV. Metodología

La evaluación se desarrolló en diferentes fases, que se describen a continuación:

##### 3.1 Planificación y organización inicial

En la fase de planificación y organización inicial se realizaron cuatro grandes actividades:

- Se elaboró un documento de Propuesta Metodológica, que contiene la estrategia y metodología definitiva a ser utilizadas en la evaluación, mismo que fue aprobado durante la reunión inicial.
- Reunión inicial con la Coordinación del CEDIM y la coordinación del proyecto.
- Se constituyó un equipo ampliado para la evaluación. Con la participación de al menos 4 integrantes del equipo de trabajo de CEDIM y 4 mujeres de las 12 comunidades donde se implementó el proyecto para que no solo fueran solo fuente de información, además del equipo técnico que CEDIM designó. Este equipo se capacitó y ellas y ellos contribuyeron al levantamiento de la información. Este equipo quedó como valor agregado al dejar capacidades instaladas para desarrollar distintos procesos investigativos.

6

---

Septiembre 2015

---

Evaluación Externa de CEDIM, AC.  
Proyecto No. 212 – 008 – 1012 ZG.

- De igual modo los instrumentos se construyeron con la participación de este grupo seleccionado, aportando ideas desde el momento mismo de la definición del diseño de la evaluación.
- El estudio aplicó el enfoque de género y generacional, lo que permitió analizar la relación y diferencias de la participación de las mujeres jóvenes y las adultas desde su contexto social y grupo étnico al que pertenecen.
- Se realizó un taller organizativo con el equipo ampliado de la evaluación. La consultora, el equipo técnico de CEDIM, en conjunto con las mujeres seleccionadas desarrollaron un taller metodológico, donde se compartieron los aspectos básicos de la evaluación y refrescaron las bases de la implementación del proyecto. Aquí se elaboró y se aprobó el cronograma de actividades para el período de 31 días hábiles de la actividad. Se presentaron y validaron la propuesta definitiva de instrumentos para la recolección de información.

## 2. Beschreibung der Feldphase

### 3.2 Fase investigativa

Se desarrollaron dos rutas de investigación: una documental y otra de campo la que se realizó desde la perspectiva de la investigación-acción, para poder garantizar la participación de las personas como sujetos de derechos. Tuvo un enfoque formativo-pedagógico, mediante el aprendizaje de los elementos que se han constituido en lecciones aprendidas y de los cuales se identifican lecciones y pistas para proponer recomendaciones.

La investigación documental, contó con los siguientes insumos fundamentales:

- Propuestas de proyectos presentados a MISEREOR
- Documento narrativo del proyecto y sus aclaraciones
- Informes del proyecto
- Indicadores propuestos por cada uno de los resultados
- Se analizaron además como insumos: productos generados desde y para las mujeres. (Guía para promotoras Voluntarias Legales , productos comunicacionales con temas diversos, resultados de investigaciones y artículos de revistas relacionados con el desarrollo de las acciones de CEDIM en la Región de los Tuxtlas)

El trabajo de campo consistió en la aplicación de los instrumentos elaborados para esta evaluación: guías de entrevista, guía de preguntas generadoras para los grupos focales. Se realizaron 18 entrevistas de diferentes modalidades y 3 grupos focales con promotoras del proyecto en San Leopoldo, La Candelaria y El Popotal.

Las entrevistas se aplicaron de la siguiente manera:

7

---

Septiembre 2015

---

Evaluación Externa de CEDIM, AC.  
Proyecto No. 212 – 008 – 1012 ZG.

- 1 entrevista colectiva con el personal técnico del proyecto en Sihuanpan
- 1 entrevista colectiva a promotoras del Proyecto, San Andrés de Tuxtla
- 2 entrevistas colectivas a beneficiarias del Proyecto (grupos de 3 mujeres) en San Marcos y Santa Rosa
- 1 entrevista a profundidad con Servicios de Salud de Veracruz
- 2 entrevistas a profundidad con Párrocos, Catemaco y Texistepec
- 4 entrevistas a promotoras legales
- 2 entrevistas con médicas del IPN
- 3 con líderes y lideresas de comunidades
- 2 entrevistas con representante del Instituto de la Mujer Veracruzana y Representante de MOCAB.

### 3. Beschreibung der Analyse und Aufbereitung der Daten

#### 3.3 *Procesamiento y Análisis de la información*

##### a) *Procesamiento de datos*

Para el procesamiento de la información se utilizó la técnica de categorización y elaboración de matrices de salida, creadas a partir de las distintas categorías de la información, para facilitar el análisis de los datos.

##### b) *Análisis de datos*

Consistió en la generación de información a partir del cruce de variables y la información contenida en las matrices de salida.

Para el análisis de toda la información se utilizó la triangulación de los resultados obtenidos con los métodos y técnicas cualitativas aplicadas.

#### 3.4 *Elaboración de informes*

Tomando como base la triangulación del análisis de la información, se discutieron los resultados con el equipo de evaluación ampliado y se elaboró un primer informe de carácter preliminar. Para esto se realizó vía Skype una sesión para discutir los hallazgos encontrados durante el proceso.

Los insumos de esta reunión se integraron al documento preliminar, este informe fue enriquecido con aportes del personal de CEDIM a cargo de esta consultoría. Todos los aportes se incorporaron al documento, mismo que dio lugar al informe final.

## Anlage 8: Kurzprofil der GutachterInnen der Metaevaluierung

**Jens Koy** (Teamleiter und Gesamtverantwortung für die Evaluation): Mehr als 10 Jahre Berufserfahrung im Bereich Evaluierung, davon 4,5 Jahre im Bereich Konzeption, Management und Begleitung der Umsetzung von Evaluierungen als Referent für Evaluierung des Evangelischen Entwicklungsdienstes (EED). Jens Koy verfügt über Expertenkenntnisse in qualitativen und quantitativen Methoden der Datenerhebung. Kombiniert mit über 5 Jahren Arbeit als freier Evaluator und Berater verfügt er über profunde Kenntnis des Evaluierungsgeschäfts als Auftragnehmer und Auftraggeber. Seit 2013 ist Jens Koy ein Sprecher des Arbeitskreis Entwicklungspolitik der Gesellschaft für Evaluation (DeGEval).

Spezifische Evaluierungserfahrung im Bereich Metaevaluierung hat Jens Koy durch eine „Auswertung der Evaluationen aus 2011 - auf Glaubwürdigkeit der Evaluierungen und Unabhängigkeit der Gutachtenden“ für Brot für die Welt – Evangelischer Entwicklungsdienst (109 Evaluierungen) sowie zwei Qualitätschecks bzw. Metaevaluierungen von 15 bzw. 17 (dezentralen) Evaluierungen der Welthungerhilfe in 2015 und 2016 (siehe Aufstellung der Evaluierungen als Anlage des CV Jens Koy).

**Dr. Cornelia Römling** ist Senior-Projektmanagerin an der CEval GmbH. Sie ist promovierte Entwicklungsökonomin mit vier Jahren Berufserfahrung und, aufgrund mehrmonatiger Auslandsaufenthalte in Brasilien, Peru und Spanien, mit ausgezeichneten Sprachkenntnissen insbesondere im Spanischen und Portugiesischen. Frau Römling verfügt weiterhin über eine umfassende methodische Expertise sowohl hinsichtlich qualitativer als auch quantitativer Analyseverfahren. Im Rahmen der Meta-Evaluierung wird Frau Römling, in enger Abstimmung mit Herrn Koy, insbesondere für die Auswertung der spanisch- und portugiesischsprachigen Evaluierungsberichte zuständig sein.

**Dr. Susanne Johanna Väth** ist Seniorprojektmanagerin an der CEval GmbH und wird das Meta-Evaluierungsteam bei der Projektdurchführung beratend unterstützen und für die Qualitätssicherung verantwortlich sein. Sie verfügt über acht Jahre Berufserfahrung und weitreichende methodische Expertise zur Konzeption und Umsetzung von Meta-Evaluierungsvorhaben.

Abgesehen von ihrem regionalen Schwerpunkt Afrika, hat die promovierte Entwicklungsökonomin Erfahrung in der Meta-Evaluierung von Projektevaluierungen im südamerikanischen und im asiatischen Raum gesammelt. Sie arbeitete bisher für verschiedene nationale und internationale Auftraggeber wie den Common Fund for Commodities, die GIZ, die Physikalisch-Technische Bundesanstalt, den Projektträger des Deutschen Zentrums für Luft- und Raumfahrt, der Friedrich-Ebert-Stiftung, World Vision oder die Christoffel-Blindenmission und hat an Politikdialogen mit der Afrikanischen Union, der Afrikanischen Entwicklungsbank, UNECA, der FAO und der Weltbank teilgenommen.

Ihre fachlichen Schwerpunkte liegen in den Bereichen Ländliche Entwicklung, Landmärkte und Landgovernance, Integration von Kleinbauern und Capacity Development. Erfahrungen in den Bereichen Bildungsk Kooperationen, Arbeit der politischen Stiftungen, Inklusion von Menschen mit Behinderungen, Wertschöpfungsketten, Klimawandel und Natürliches Ressourcenmanagement runden ihr Profil ab.