



# EVALUIERUNGSPRAXIS IN DER DEUTSCHEN ENTWICK- LUNGSZUSAMMENARBEIT

*Umsetzungsmonitoring der letzten Systemprüfung  
und Charakterisierung wesentlicher Elemente*

2015



**DEval**

DEUTSCHES  
EVALUIERUNGSIKITUT  
DER ENTWICKLUNGS-  
ZUSAMMENARBEIT

Der vorliegende Bericht zur „Evaluierungspraxis in der deutschen Entwicklungszusammenarbeit. Umsetzungsmonitoring der letzten Systemprüfung und Charakterisierung wesentlicher Elemente“ stellt einen von zwei Untersuchungsbereichen des am DEval durchgeführten Projekts „Ausgangserhebung zum DEval und seinem Umfeld“ dar. Dieses Projekt wurde von Renate Kirsch, Dr. Kim Lücking, Heike Steckhan, Simon Freund und Simon Bettighofer konzipiert und durchgeführt. Ein Gutachterteam der Firma Rambøll Management Consulting/ Syspons GmbH unterstützte das DEval-Team in allen Arbeitsschritten.

Der andere Untersuchungsbereich wurde unter dem Titel „Erhebung zur Ausgangslage des DEval und seines Umfeldes“ (DEval, 2015a) veröffentlicht.

Zudem ist ein separater Anlagenband verfügbar, in dem das methodische Design sowie die Erhebungsinstrumente und Stichproben detailliert dargestellt werden (DEval, 2015b).

# EVALUIERUNGSPRAXIS IN DER DEUTSCHEN ENTWICK- LUNGSZUSAMMENARBEIT

*Umsetzungsmonitoring der letzten Systemprüfung  
und Charakterisierung wesentlicher Elemente*

2015

## Impressum

### Herausgeber

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

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

E-mail: [info@DEval.org](mailto:info@DEval.org)

[www.deval.org](http://www.deval.org)

### Verfasst von

Dr. Kim Lücking  
Simon Freund  
Simon Bettighofer

### Verantwortlich

Renate Kirsch

### Gestaltung

BUREAU SPÜRSINN, Aachen, Deutschland  
[www.spuersinn.biz](http://www.spuersinn.biz)

### Lektorat

Annika Mattes

### Bildnachweis

Cover: Fotolia, © vege

### Druck

network2print

Der Bericht ist auf FSC zertifiziertem Papier gedruckt.

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

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

Der vorliegende Bericht ist auch auf der DEval-Website als pdf-Download verfügbar unter:  
[www.deval.org/de/evaluierungsberichte.html](http://www.deval.org/de/evaluierungsberichte.html)

Anfragen nach einer gebundenen Ausgabe richten Sie bitte an:  
[info@DEval.org](mailto:info@DEval.org)

## Danksagung

Bereits der Titel dieser Publikation „Evaluierungspraxis in der deutschen Entwicklungszusammenarbeit. Umsetzungsmonitoring der letzten Systemprüfung und Charakterisierung wesentlicher Elemente“ weist darauf hin, dass sich die Mitarbeitenden dieses Projekts intensiv mit dem Umfeld des DEval befassen haben. Es umfasst zahlreiche Akteure: Zuvorderst das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ), das die Bundesrepublik Deutschland als Gesellschafterin vertritt, sowie die unterschiedlichen Organisationen der Entwicklungszusammenarbeit in Deutschland. Darüber hinaus gibt es weitere wichtige nationale und internationale Akteure der Entwicklungszusammenarbeit, zu denen u.a. Fachorganisationen und wissenschaftliche Institute zählen. Viele dieser Akteure waren in unserer Erhebung eingebunden und haben zu ihrem Erfolg beigetragen.

Für die Planung und Durchführung der Ausgangserhebung war eine enge Zusammenarbeit mit dem Evaluierungsreferat des BMZ wichtig, mit dessen Mitarbeiterinnen und Mitarbeitern wir viele Gespräche führten und die uns hilfreiche Informationen und Unterlagen zur Verfügung stellten. Ebenso hilfreich waren die Interviews mit Mitarbeiterinnen und Mitarbeitern verschiedener Fachreferate des BMZ zu den strategischen Evaluierungen des Evaluierungsreferates.

Davon abgesehen wäre ohne die Beteiligung der EZ-Organisationen, die wir für unsere Erhebung auswählten, die Durchführung des Projekts nicht möglich gewesen. Ein besonderer Dank gilt den Organisationen, die bereits an der letzten Systemprüfung beteiligt waren und die uns für z.T. mehrtägige Interviews in ihren Häusern zur Verfügung stellten, sowie jenen, die uns bereitwillig ihre Evaluierungsberichte zur Verfügung stellten. Nicht zuletzt möchten wir auch den Expertinnen und Experten danken, die sich für unsere Interviews Zeit genommen haben.

In einer europaweiten Ausschreibung konnte Rambøll Management Consulting/ Syspons GmbH<sup>1</sup> im November 2013 als Partner für die Durchführung der Erhebung, die bis Oktober 2014 erfolgte, gewonnen werden. Wir bedanken uns bei Dr. Christoph Emminghaus, Tobias Stern, Lennart Raetzell, Jan Oliver Scheller und Anne Kruse für die intensive, vertrauensvolle und produktive Zusammenarbeit, die maßgeblich zum Gelingen des Projektes beigetragen hat.

Des Weiteren bedanken wir uns ganz herzlich bei unserer Kollegin Heike Steckhan, die bis Mai 2014 intensiv an der Ausgangserhebung mitgearbeitet hat, um sich anschließend im Mutterschutz einem noch schöneren Projekt zu widmen, sowie bei unserer Projektadministratorin Kristen Schubert für ihre wertvolle Unterstützung.

<sup>1</sup> Im Verlauf der Ausgangserhebung wurde die Syspons GmbH von Rambøll Management Consulting als Unterauftragnehmer mit der Durchführung beauftragt.



# ZUSAMMENFASSUNG

Das Deutsche Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) nahm im Sommer 2012 seine Arbeit auf. Eines der ersten Projekte, die durchgeführt wurden, war die „Ausgangserhebung zum DEval und seinem Umfeld“:

Das Projekt umfasst zwei sich ergänzende Untersuchungsbereiche:

- Der erste Teil ist eine Erhebung zur Ausgangslage des DEval und etabliert eine Baseline zu den selbstgesteckten Zielen des Instituts.
- Der zweite Teil gründet auf einer Untersuchung des Evaluierungssystems der deutschen Entwicklungszusammenarbeit (EZ), die veranlasst durch das BMZ in den Jahren 2007/ 2008 durchgeführt wurde. Ziel dieser sogenannten „Systemprüfung“ war es, Aussagen zum Stand der Evaluation in der deutschen EZ treffen zu können und daraus Handlungsempfehlungen für deren Weiterentwicklung abzuleiten (Borrmann und Stockmann, 2009a). Die durchgeführte Erhebung umfasst ein Monitoring zur Umsetzung dieser Handlungsempfehlungen und verbindet sie mit weiteren Ergebnissen zur Evaluierungspraxis der deutschen Entwicklungszusammenarbeit (EZ).

Während die Ergebnisse zur Ausgangslage des DEval vornehmlich für das Institut von Interesse sind, ist die Zielgruppe des zweiten Erhebungsbereichs vielschichtiger. Um den Interessen der Zielgruppen, aber auch dem Umfang an Ergebnissen gerecht zu werden, wurden zwei Berichte verfasst (zum ersten Erhebungsbereich vgl. DEval, 2015a).

Dieser Bericht befasst sich mit dem Umsetzungsmonitoring, mandatiert durch das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ), und dem Stand der Evaluierungspraxis in der deutschen EZ.

## **Umsetzungsmonitoring und aktueller Stand der Evaluierungspraxis**

In den Jahren 1997/ 1998 veranlasste das BMZ erstmals eine systematische Untersuchung der Evaluierungen in der deutschen Entwicklungszusammenarbeit (Borrmann et al., 1999). Die Systemprüfung der Jahre 2007/ 2008 war demnach die zweite

Prüfung dieser Art. Auf diese wird im Folgenden als „Systemprüfung“ verwiesen.

In dieser „Systemprüfung“ wurden auch Empfehlungen zur Reform des Evaluierungssystems insgesamt gegeben. Eine zentrale Empfehlung betraf die „Gründung einer unabhängigen Evaluationsagentur, die das BMZ bei der Systementwicklung und die EZ-Organisationen bei der Instrumenten- und Qualitätsentwicklung berät und unterstützt sowie unabhängig, mit hoher Kompetenz und damit Glaubwürdigkeit Evaluationen insbesondere zur Wirksamkeit und Nachhaltigkeit der EZ länder-, sektor- und institutionenvergleichend vornimmt“ (Borrmann und Stockmann, 2009a: 165). Mit der Gründung des DEval hat das BMZ diese systemverändernde Empfehlung umgesetzt.

Die Empfehlungen auf Systemebene gründeten auf 20 Einzelfallstudien. In diesen wurde untersucht, „wie die einzelnen EZ-Organisationen Evaluation institutionalisieren, planen, durchführen und nutzen, ob sie durch Evaluation insbesondere etwas über die Wirkungen ihrer Projekte und Programme erfahren können und ob und wie sich ihre einzelnen Evaluationsysteme zu einem konsistenten Gesamtsystem für die bilaterale deutsche Entwicklungszusammenarbeit zusammenfügen“ (Borrmann und Stockmann, 2009a: 5). Ausgehend von den in den Einzelfallstudien betrachteten Organisationen wurden organisationsspezifische Handlungsempfehlungen zur Weiterentwicklung der Evaluierungspraxis ausgesprochen. Im Anschluss vereinbarte das BMZ mit den beteiligten Organisationen Umsetzungspläne zu den von den Organisationen akzeptierten Handlungsempfehlungen. Im Rahmen der „Ausgangserhebung zum DEval und seinem Umfeld“ wurde das DEval vom BMZ zur Durchführung eines „Umsetzungsmonitorings“ gebeten, um zu überprüfen, ob und in welchem Umfang die Umsetzung dieser Handlungsempfehlungen erfolgte. Eine erneute „Systemprüfung“ durchzuführen, erschien verfrüht, weshalb Organisationen nicht einzeln bewertet und auch keine Handlungsempfehlungen – wohl aber Schlussfolgerungen – ausgesprochen werden. Die Auswertungen erfolgen aggregiert auf der Ebene von Akteursgruppen (Durchführungsorganisationen, politische Stiftungen und andere zivilgesellschaftliche Organisationen).

Ergänzend folgten tiefergehende Untersuchungen zur Evaluierungspraxis verschiedener Akteure der deutschen EZ, um deren

Entwicklung und aktuellen Stand zu erfassen. Dieser Einblick trägt zu einem verbesserten Verständnis des DEval-Umfeldes bei, in dem sich das neue Institut verorten muss, um Wirkungen zu erzielen.

### Methodik

Auf der Basis eines Rapid Evidence Assessments von aktuellen Studien wurden sechs Dimensionen von Evaluierungspraxis differenziert. Die einzelnen Dimensionen sind: *Ziele der Evaluierung, Organisatorischer Kontext, Strukturen und Prozesse, Humanressourcen, Finanzielle und Technische Ressourcen.*

Bei der Datenerhebung kamen folgende Methoden zur Anwendung: Es wurde eine Online-Befragung zur Evaluierungspraxis in der deutschen EZ durchgeführt, an der sich 34 Organisationen beteiligten. Den 17 Organisationen, die an der „Systemprüfung“ beteiligt waren, wurden hierbei zusätzlich Fragen zur Umsetzung der Handlungsempfehlung der „Systemprüfung“ gestellt. Um ein besseres Verständnis für deren Umsetzungsgrad sowie weiterführende Einblicke in die Evaluierungspraxis zu erhalten, wurden in diesen Organisationen zudem ein- bis zweitägige Interviews mit der Geschäftsführung, den für Evaluierungen Zuständigen sowie den Personen des operativen Bereiches durchgeführt. Um einen Eindruck von den angewandten Designs und Methoden in der deutschen EZ zu gewinnen, wurde eine Metaevaluierung von Evaluierungsberichten der EZ-Organisationen vorgenommen. Ergänzend wurden Einschätzungen nationaler und internationaler Expertinnen und Experten der EZ erhoben.

In einem Abschlussworkshop mit den an der Erhebung teilnehmenden Organisationen wurden die gewonnenen Ergebnisse und Schlussfolgerungen abschließend diskutiert.

### Übergeordnete Ergebnisse

Die Erhebungen zeigten, dass die „Systemprüfung“ ein wichtiger Impuls für die Weiterentwicklung der Evaluierungspraxis der deutschen EZ war. Die angestoßenen Veränderungen können – aufgrund ihrer Breite und Tiefe – als systembildend beschrieben werden.

Im Umsetzungsmonitoring wurden die Handlungsempfehlungen der „Systemprüfung“ analysiert. Von den 181<sup>2</sup> Empfehlungen richteten sich 38 Prozent an die sechs politischen Stiftungen (POS), 35 Prozent an die fünf staatlichen Durchführungsorganisationen (DO) und 27 Prozent an sechs weitere zivilgesellschaftliche Organisationen (ZGO). Wird die Übereinstimmung aller Organisationen mit den Handlungsempfehlungen in den Umsetzungsplänen betrachtet, dann zeigt sich, dass die Organisationen diesen zu 59 Prozent vollständig und zu weiteren 33 Prozent zumindest teilweise zustimmten. Diese mehrheitliche Zustimmung zu den Empfehlungen spiegelt sich auch darin wider, dass sie überwiegend ganz oder zumindest teilweise umgesetzt wurden: 54 Prozent wurden laut Angabe der EZ-Organisationen vollständig umgesetzt, 29 Prozent teilweise und lediglich 12 Prozent gar nicht. Für 5 Prozent der Empfehlungen liegen keine Angaben zur Umsetzung vor.

Die Zustimmung der Leitung bzw. Geschäftsführung der betreffenden Organisation war mit Abstand der wichtigste Faktor bei der Umsetzung der Handlungsempfehlungen. Finanzielle Ressourcen hatten hingegen oft einen hemmenden Einfluss bei der Umsetzung. Dies gilt v.a. für Empfehlungen, die mit einem erheblichen finanziellen Mehraufwand verbunden waren (z.B. Schaffung einer neuen Stabsstelle, vermehrte Ex-post-Evaluierungen).

### Ergebnisse nach Dimensionen

Die Analyse der einzelnen Dimensionen von Evaluierungspraxis zeigt, dass in den Empfehlungen ein Schwerpunkt auf der Dimension „Strukturen und Prozesse“ lag. Des Weiteren betraf ein hoher Anteil der Empfehlungen die „Ziele von Evaluierungen“ sowie „Humanressourcen“. In diesen drei Kategorien wurden nahezu vier Fünftel (79 Prozent) aller Empfehlungen ausgesprochen. Auf die weiteren Kategorien „Organisatorischer Kontext“, „Finanzielle Ressourcen“ und „Technische Ressourcen“ beziehen sich deutlich weniger Empfehlungen der „Systemprüfung“.

#### Ziele von Evaluierungen

Die Empfehlungen in dieser Dimension zielten darauf ab, das Evaluierungsspektrum der EZ-Organisationen hinsichtlich ihrer Zielsetzung und damit

<sup>2</sup> In der Systemprüfung wurden 128 Handlungsempfehlungen auf Organisationsebene ausgesprochen. Da manche dieser Empfehlungen sehr umfangreich waren, wurden sie nach verschiedenen inhaltlichen Aspekten aufgeteilt.

- Art und Ziel der durchgeführten Evaluierungen (z.B. Ex-post-, Wirkungs-, Gemeinschafts- und Querschnittsevaluierungen) sowie
- Nutzung und Transparenz der Evaluierungsergebnisse zu erweitern.

Mit 47 Prozent war die vollständige Übereinstimmung zu den 49 ausgesprochenen Handlungsempfehlungen im Vergleich zu den anderen Dimensionen am geringsten. Zugleich ist mit 51 Prozent (n=25) der Anteil an Empfehlungen, die nicht bzw. nur zum Teil umgesetzt wurden, auch vergleichsweise hoch. Eine Erklärung dafür ist, dass sich viele der Empfehlungen auf eine Ausweitung der eingesetzten Evaluierungsformen bezogen. Während zwar vermehrt Querschnittsevaluierungen durchgeführt werden, bestanden bei einigen EZ-Organisationen Zweifel, was die Zweckmäßigkeit von Ex-post- und Wirkungsevaluierungen für die eigene Arbeit betrifft. Manche EZ-Organisationen verfügten nicht über die notwendigen Mittel zur Durchführung dieser Evaluierungen. Ressourcenintensivere Ex-post- und Wirkungsevaluierungen werden daher eher von den staatlichen Durchführungsorganisationen und weniger von zivilgesellschaftlichen Organisationen durchgeführt.

Evaluierungen werden den befragten Organisationen zufolge vor allem zur Verbesserung von Vorhaben bzw. zum instrumentellen Lernen genutzt. Dies geschieht v.a. durch Projektbeteiligte zur Programm-/ Projektsteuerung. Es fehlen häufig standardisierte Prozesse zur Nutzung projektübergreifender Evaluierungsergebnisse für konzeptionelles Lernen der Gesamtorganisation. Die Verpflichtung zur Rechenschaftslegung bindet Ressourcen an die Dokumentation von Projekterfolgen, die einige Organisationen gerne für konzeptionelles Lernen aus projektübergreifenden Evaluierungen nutzen würden.

Bezüglich der Nutzung und Transparenz von Evaluierungsergebnissen lässt sich feststellen, dass eine prinzipielle Veröffentlichung von vollständigen Berichten nach wie vor sehr selten ist. Die zurzeit gängige Praxis in den meisten EZ-Organisationen ist die Veröffentlichung und Verbreitung von Kurzberichten. Zum Teil wird auf Anfrage der Gesamtbericht zur Verfügung gestellt. Unter anderem halten berechnete Schutzinteressen gegenüber Partnern sowie die Sorge um das Spendenaufkommen einzelne Organisationen davon ab, Evaluierungsergebnisse

zu veröffentlichen. Darüber hinaus bestehen Befürchtungen, dass vollkommene Transparenz mit dem Ziel des Lernens aus Evaluierung in Widerspruch gerät, da kritische Ergebnisse u.U. weniger thematisiert werden, wenn vollständige Berichte unkommentiert veröffentlicht werden.

### Organisatorischer Kontext

Das Ziel der Empfehlungen in der Dimension „Organisatorischer Kontext“ war,

- den Vernetzungsgrad der EZ-Organisationen untereinander sowie mit der Wissenschaft zu evaluierungsspezifischen Themen zu stärken und
- das Verständnis bzw. die Akzeptanz von Mitarbeitenden bezüglich Evaluierung zu fördern.

Die Mehrzahl der 16 Empfehlungen in dieser Dimension wurde von den EZ-Organisationen geteilt (69 Prozent) und zu 62 Prozent (n=10) vollständig umgesetzt. Insgesamt herrschte unter den EZ-Organisationen Einvernehmen darüber, dass der Austausch in Evaluierungsnetzwerken einen Mehrwert bietet und zur Erhöhung der Evaluierungsqualität beitragen kann. Der Vernetzungsgrad der deutschen EZ-Organisationen im Bereich Evaluierung untereinander hat in den letzten Jahren deutlich zugenommen. Dies gilt auch für die Zusammenarbeit mit der Wissenschaft, wenn auch in geringerem Ausmaß. Mitarbeitende der EZ-Organisationen sind Evaluierungen gegenüber aufgeschlossen.

### Strukturen und Prozesse

Das Ziel der Empfehlungen in der Dimension „Strukturen und Prozesse“ war,

- die Stellung der Evaluierungseinheit in der Institution,
- institutionalisierte Evaluierungsgrundlagen,
- festgelegte Evaluierungsprozesse und
- Einbindung von Partnern in den Evaluierungsprozess zu stärken.

Werden die absoluten Werte betrachtet, so sind in dieser Dimension die meisten Empfehlungen (n=56) ausgesprochen und umgesetzt worden (31 vollständig, 19 zumindest teilweise). Relativ gesehen ist die Anzahl der vollständig umgesetzten

Empfehlungen mit 55 Prozent als durchschnittlich zu bewerten; geteilt wurden die Empfehlungen in dieser Dimension zu 64 Prozent.

Innerhalb der vier inhaltlichen Cluster der Empfehlungen zeigen sich große Unterschiede in deren Umsetzung: Während eine hohe Bereitschaft bestand, klarere Strukturen und Abläufe von Evaluierungen zu etablieren, bezieht sich der vergleichsweise hohe Anteil an nur teilweise umgesetzten Empfehlungen auf die institutionelle Stellung der Evaluierungseinheit sowie die stärkere Einbindung von Partnern.

Evaluierungseinheiten und -stellen wurden seit der „Systemprüfung“ vielfach neu geschaffen oder gestärkt. Die Evaluierungsverantwortlichen verstehen sich dabei überwiegend als Dienstleistende für die Organisation und weniger als unabhängige Prüfungsinstanz. Sie warfen hierzu auch die Frage auf, inwieweit eine interne Einheit einer Organisation als unabhängig bezeichnet werden kann bzw. Unabhängigkeit erreichen kann. Zudem befürchten einige Evaluierungsverantwortliche, dass eine stärkere Unabhängigkeit ihrer Arbeit zu Lasten der Relevanz und Akzeptanz der Evaluierungsergebnisse geschehen würde.

Die Berücksichtigung der OECD DAC-Kriterien ist in den Evaluierungsrichtlinien der Organisationen der deutschen EZ verankert. Ihre Umsetzung in den Evaluierungsberichten scheint jedoch nicht konsequent.

Hinsichtlich institutionalisierter Evaluierungsgrundlagen lässt sich feststellen, dass die Organisationen der deutschen EZ ihre Evaluierungspraxis formalisiert und systematisiert haben; Evaluierungen folgen dabei überwiegend Evaluierungsrichtlinien und -plänen. Der Formalisierungsgrad variiert je nach Prozessphase: Die Umsetzung bzw. Nutzung der Evaluierungsergebnisse ist dabei am geringfügigsten formalisiert, wenngleich einige Organisationen standardisierte Prozesse zur Bewertung und Umsetzung von Evaluierungsergebnissen haben.

Hinsichtlich der Einbindung von Partnerorganisationen der Empfängerländer in den Evaluierungsprozess lässt sich feststellen, dass es stark von der Art der Kooperation abhängt, ob und in welchem Maße Partnerorganisationen eingebunden werden.

Generell kann konstatiert werden, dass eine systematische und gleichberechtigte Einbeziehung von Partnern im Evaluierungsprozess nicht die Regel ist.

### *Humanressourcen*

Das Ziel der Empfehlungen in der Dimension „Humanressourcen“ war die Weiterentwicklung der

- Anzahl der Mitarbeitenden in der Evaluierungseinheit,
- Monitoring- und Evaluierungskompetenz der Mitarbeitenden,
- Monitoring- und Evaluierungskompetenz der Partner,
- Einsätze von Gutachterinnen und Gutachtern und der
- eingesetzten Methoden.

Die 38 Empfehlungen wurden zu 66 Prozent geteilt und zu 61 Prozent vollständig umgesetzt. Seit der „Systemprüfung“ werden mehr Personalressourcen für Evaluierung eingesetzt, wobei diese Ausbauprozesse zum Zeitpunkt der Erhebung in einigen Organisationen noch nicht abgeschlossen waren. Zugleich haben sich oftmals auch die methodischen Ansprüche an Evaluierungen erhöht, wodurch sie arbeits- und betreuungsintensiver sind.

Die fachliche Kompetenz in den Evaluierungseinheiten wurde erhöht. Die Qualität von Evaluierungen hängt aber nicht allein von den Evaluierungseinheiten ab, sondern auch von der Kompetenz der beauftragten Gutachterinnen und Gutachter sowie jener der operationellen Bereiche und Partner, sofern diese Evaluierungen managen. Es besteht nicht überall die Möglichkeit (v.a. aufgrund der Vielzahl von Evaluierungen in einigen Organisationen), eine systematische Qualitätssicherung durch die Evaluierungseinheit vorzunehmen.

Bei dem Einsatz externer Gutachterinnen und Gutachter lässt sich feststellen, dass nahezu alle untersuchten Organisationen diese vermehrt beauftragen und ihnen auch die volle Berichtsverantwortung überlassen. Nach Ansicht einzelner Organisationen ist das methodische Repertoire von Gutachterinnen und Gutachtern ausbaufähig. Länder- und organisationsspezifische Kenntnisse werden bei der Auswahl oftmals stärker berücksichtigt.

Den Aufbau von Evaluierungskapazitäten in Partnerorganisationen sehen viele deutsche EZ-Organisationen nicht als ihren Verantwortungsbereich, wenngleich sich einige Organisationen sehr stark in diesem Bereich engagieren.

Evaluierungsdesigns basieren häufig nicht auf überprüfbareren Wirkungshypothesen. Das verwendete Methodenspektrum bzw. die Nutzung eines angemessenen Methodenmixes ist ausbaufähig.

### **Finanzielle Ressourcen**

Die in der Dimension „Finanzielle Ressourcen“ ausgesprochenen Empfehlungen hatten zum Ziel, die Budgets für Evaluierungen und Evaluierungssysteme zu erhöhen.

12 Empfehlungen an die EZ-Organisationen griffen Themen der finanziellen Rahmenbedingungen auf. Insgesamt wurden die Empfehlungen zu 58 Prozent geteilt und zu 42 Prozent vollständig umgesetzt. Die vollständig umgesetzten Handlungsempfehlungen bezogen sich unter anderem auf die hinreichende Budgetierung von Schlussevaluierungen oder anspruchsvollere Erhebungsdesigns. Die beiden Handlungsempfehlungen, die nicht umgesetzt wurden, bezogen sich auf die generelle Erhöhung von finanziellen Mitteln und eine verbindliche Festlegung eines Evaluierungsbudgets.

Die Evaluierungspraxis ist in vielerlei Hinsicht abhängig von den finanziellen Ressourcen, die den Evaluierungsverantwortlichen in deren Organisationen zur Verfügung gestellt werden. Seit der „Systemprüfung“ sind die Mittel, die in deutschen EZ-Organisationen für Evaluierungen eingesetzt werden, gestiegen. Die zur Verfügung gestellten finanziellen Ressourcen werden von den Befragten überwiegend als ausreichend empfunden. Diese Aussage gilt allerdings nicht für die Durchführung von methodisch anspruchsvollen oder projektübergreifenden Evaluierungen.

### **Technische Ressourcen**

Die in der Dimension „Technische Ressourcen“ ausgesprochenen Empfehlungen zielten darauf ab, die Evaluierungsarbeit zu erleichtern bzw. zu verbessern, indem z.B. der Zugang zu Statistik- oder Befragungssoftware, zu Evaluierungsliteratur oder zu einem internen Wissensmanagementsystem ermöglicht wird.

Die Mehrheit der 10 Empfehlungen wurde geteilt und vollständig umgesetzt (jeweils 60 Prozent; n=6). Mitarbeitende in deutschen EZ-Organisationen haben einen guten Zugang zu Fachliteratur und Leitfäden zum Themenbereich Evaluierung. Zudem verfügt die Mehrheit der befragten Organisationen über Wissensmanagementsysteme (WMS), die vornehmlich zum Abruf von Informationen genutzt werden und weniger für institutionelles Lernen. Vertreterinnen und Vertreter der EZ-Organisationen wiesen darauf hin, dass bestimmte technische Ressourcen nicht dem Auftraggebenden zur Verfügung stehen müssten, sondern den beauftragten Gutachterinnen und Gutachtern.

### **Fazit**

Das Umsetzungsmonitoring der „Systemprüfung“ hat gezeigt, dass es bemerkenswerte Weiterentwicklungen in der Evaluierungspraxis der deutschen Entwicklungszusammenarbeit gibt. Die ausgesprochenen Empfehlungen trugen bei den untersuchten Organisationen beispielsweise zur Einrichtung von Evaluierungseinheiten bei oder zur Aufstockung der für Evaluierung zuständigen Personalstellen und des Budgets für Evaluierung. Zudem wurden vermehrt Evaluierungsrichtlinien eingeführt. Daneben zeigen sich – ebenso wie in den bisherigen Systemprüfungen – aber auch Themenfelder, in denen Ansprüche der „Systemprüfung“ sowie der OECD DAC-Prinzipien und die Interessen der untersuchten EZ-Organisationen nicht vollständig deckungsgleich sind. Dies betrifft beispielsweise die in den Systemprüfungen geforderten Ex-post-Evaluierungen zur Überprüfung von Wirksamkeit und Nachhaltigkeit der Vorhaben sowie einen transparenten Umgang mit Evaluierungsergebnissen.

Insgesamt zeigen die Ergebnisse, dass die „Systemprüfung“ ein wichtiger Impuls für die Evaluierungspraxis der deutschen EZ war und viele Entwicklungen gefördert oder angestoßen hat.

# INHALT

<b>Danksagung</b>	<b>v</b>
<b>Zusammenfassung</b>	<b>vi</b>
<b>Abkürzungen und Akronyme</b>	<b>1</b>

## 1. Einleitung **2**

---

## 2. Methodisches Vorgehen **5**

---

2.1	Erhebungsphasen und -instrumente	7
2.1.1	<i>Explorative Phase</i>	7
2.1.2	<i>Datenerhebungsphase</i>	9
2.2	Methodische Herausforderungen	10
2.2.1	<i>Adäquate Berücksichtigung unterschiedlicher Stakeholder</i>	10
2.2.2	<i>Grenzen von Erhebungsinstrumenten</i>	11
2.2.3	<i>Entwicklung eines geeigneten Analyserasters</i>	11
2.2.4	<i>Überprüfung von Selbsteinschätzungen</i>	12
2.2.5	<i>Transparenz und Vertraulichkeit</i>	12

## 3. Umsetzungsmonitoring und Evaluierungspraxis **13**

---

3.1	Ziele von Evaluierungen	19
3.1.1	<i>Art und Ziele der durchgeführten Evaluierung</i>	20
3.1.2	<i>Nutzung und Transparenz der Evaluierungsergebnisse</i>	23
3.2	Organisatorischer Kontext	26
3.2.1	<i>Austausch zwischen EZ-Organisationen und Netzwerkbindung</i>	27
3.2.2	<i>Akzeptanz von Evaluierungen seitens der Mitarbeitenden</i>	28

3.3	Strukturen und Prozesse	29
3.3.1	<i>Stellung der Evaluierungseinheit innerhalb der Organisation</i>	32
3.3.2	<i>Institutionalisierte Evaluierungsgrundlagen</i>	33
3.3.3	<i>Festgelegte Prozesse zur Evaluierungspraxis</i>	34
3.3.4	<i>Einbindung von Partnern in den Evaluierungsprozess</i>	35
3.4	Humanressourcen	37
3.4.1	<i>Anzahl der Mitarbeitenden in der Evaluierungseinheit</i>	39
3.4.2	<i>Monitoring- und Evaluierungskompetenz der Mitarbeitenden und Partner</i>	40
3.4.3	<i>Einsatz von Gutachterinnen und Gutachtern</i>	42
3.4.4	<i>Eingesetzte Designs und Methoden</i>	43
3.5	Finanzielle Ressourcen	46
3.6	Technische Ressourcen	50

## 4. Fazit und Schlussfolgerungen **53**

---

<b>Literatur</b>	<b>58</b>
------------------	-----------

## Abbildungen

Abbildung 1:	Methodisches Konzept der Ausgangserhebung	6	Abbildung 20:	Umsetzungsgrad nach inhaltlichen Clustern (Strukturen und Prozesse)	31
Abbildung 2:	Datenquellen des Umsetzungsmonitorings und der Evaluierungspraxis	14	Abbildung 21:	Umsetzungsgrad nach Akteursgruppe (Stellung der Evaluierungseinheit innerhalb der Organisation)	33
Abbildung 3:	Umsetzungsgrad der Empfehlungen (gesamt)	15	Abbildung 22:	Umsetzungsgrad nach Akteursgruppe (Institutionalisierte Evaluierungsgrundlagen)	34
Abbildung 4:	Umsetzungsgrad der Empfehlungen nach Dimensionen	16	Abbildung 23:	Umsetzungsgrad nach Akteursgruppe (Festgelegte Prozesse zur Evaluierungspraxis)	35
Abbildung 5:	Umsetzungsgrad der Empfehlungen (gesamt; nach Akteursgruppe)	16	Abbildung 24:	Umsetzungsgrad nach Akteursgruppe (Einbindung von Partnern in den Evaluierungsprozess)	36
Abbildung 6:	Einflussfaktoren für die Umsetzung der Empfehlungen (gesamt)	17	Abbildung 25:	Umsetzungsgrad der Empfehlungen (Humanressourcen)	38
Abbildung 7:	Verfügbarkeit finanzieller Ressourcen als Einflussfaktor für die Umsetzung der Empfehlungen (nach Akteursgruppe)	18	Abbildung 26:	Umsetzungsgrad nach inhaltlichen Clustern (Humanressourcen)	38
Abbildung 8:	Umsetzungsgrad der Empfehlungen (Ziele von Evaluierungen)	20	Abbildung 27:	Umsetzungsgrad nach Akteursgruppe (Anzahl der Mitarbeitenden in der Evaluierungseinheit)	39
Abbildung 9:	Umsetzungsgrad nach inhaltlichen Clustern (Ziele von Evaluierungen)	20	Abbildung 28:	Vollzeit-Stellen Evaluierungsverantwortliche	40
Abbildung 10:	Umsetzungsgrad nach Akteursgruppe (Art und Ziele der durchgeführten Evaluierungen)	21	Abbildung 29:	Umsetzungsgrad nach Akteursgruppe (Monitoring- und Evaluierungskompetenz der Mitarbeitenden)	41
Abbildung 11:	Ziele von Evaluierungen	22	Abbildung 30:	Umsetzungsgrad nach Akteursgruppe (Monitoring- und Evaluierungskompetenz der Partner)	41
Abbildung 12:	Umsetzungsgrad nach Akteursgruppe (Nutzung und Transparenz der Evaluierungsergebnisse)	23	Abbildung 31:	Qualifizierungsmaßnahmen im Bereich Evaluierung	42
Abbildung 13:	Verbreitungsmedien von Evaluierungsergebnissen	25	Abbildung 32:	Umsetzungsgrad nach Akteursgruppe (Einsatz von Gutachterinnen und Gutachtern)	43
Abbildung 14:	Umsetzungsgrad nach inhaltlichen Clustern (Organisatorischer Kontext)	26	Abbildung 33:	Umsetzungsgrad nach Akteursgruppe (Eingesetzte Designs und Methoden)	44
Abbildung 15:	Umsetzungsgrad der Empfehlungen (Organisatorischer Kontext)	27	Abbildung 34:	Einsatz verschiedener Datenerhebungsmethoden	45
Abbildung 16:	Umsetzungsgrad nach Akteursgruppe (Austausch zwischen EZ-Organisationen und Netzwerkbindung)	27	Abbildung 35:	Umsetzungsgrad der Empfehlungen (Finanzielle Ressourcen)	47
Abbildung 17:	Netzwerke für fachlichen Austausch	28	Abbildung 36:	Umsetzungsgrad nach Akteursgruppe (Finanzielle Ressourcen)	47
Abbildung 18:	Umsetzungsgrad nach Akteursgruppe (Akzeptanz von Evaluierungen seitens der Mitarbeitenden)	29			
Abbildung 19:	Umsetzungsgrad der Empfehlungen (Strukturen und Prozesse)	31			

Abbildung 37:	Folgen von unzureichendem Evaluierungsbudget	49
Abbildung 38:	Umsetzungsgrad der Empfehlungen (Technische Ressourcen)	50
Abbildung 39:	Umsetzungsgrad nach Akteursgruppe (Technische Ressourcen)	51
Abbildung 40:	Verfügbarkeit von Evaluierungsmaterialien, Software und Wissensmanagementsystemen	51

#### **Tabellen**

Tabelle 1:	Dimensionen und inhaltliche Cluster von Evaluierungspraxis	8
Tabelle 2:	Anteil der evaluierten, überwiegend durch das BMZ finanzierten Projekte/Programme	24
Tabelle 3:	Anteil der evaluierten Projekte/Programme (gesamt)	24
Tabelle 4:	Anteil der Programmmittel, die für unabhängige Evaluierungen aufgewendet werden	43
Tabelle 5:	Evaluierungsbudget als Anteil am Gesamtbudget	48

# ABKÜRZUNGEN UND AKRONYME

**AK Epol**

Arbeitskreis Entwicklungspolitik  
der DeGEval

**BfdW**

Brot für die Welt

**BHO**

Bundeshaushaltsordnung

**BGR**

Bundesanstalt für Geowissen-  
schaften und Rohstoffe

**BMZ**

Bundesministerium für wirt-  
schaftliche Zusammenarbeit und  
Entwicklung

**BMZ-E**

BMZ-Evaluierungsreferat

**CEval**

Centrum für Evaluation

**Ci**

Caritas International

**DAC**

Development Assistance  
Committee der OECD

**DED**

Deutscher Entwicklungsdienst

**DEG**

Deutsche Investitions- und  
Entwicklungsgesellschaft mbH

**DEval**

Deutsches Evaluierungsinstitut der  
Entwicklungszusammenarbeit

**DeGEval**

Gesellschaft für Evaluation

**DIE**

Deutsches Institut für  
Entwicklungspolitik

**DO**

Durchführungsorganisation

**dvv international**

Institut für internationale  
Zusammenarbeit des Deutschen  
Volkshochschul-Verbandes

**DWHH**

Deutsche Welthungerhilfe

**EED**

Evangelischer Entwicklungsdienst

**EWDE**

Evangelisches Werk für Diakonie  
und Entwicklung

**EZ**

Entwicklungszusammenarbeit

**EZ-O**

EZ-Organisation(en)

**FEEZ**

Fortbildungsprogramm  
Evaluation in der Entwick-  
lungszusammenarbeit

**FES**

Friedrich-Ebert-Stiftung

**FNF**

Friedrich-Naumann-Stiftung  
für die Freiheit

**FZ**

Finanzielle Zusammenarbeit

**GIZ**

Deutsche Gesellschaft für  
Internationale Zusammenarbeit

**GTZ**

Deutsche Gesellschaft für  
Technische Zusammenarbeit

**hbs**

Heinrich-Böll-Stiftung

**HSS**

Hanns-Seidel-Stiftung

**ICAI**

Independent Commission  
for Aid Impact

**InWEnt**

Internationale Weiterbildung  
und Entwicklung

**KAS**

Konrad-Adenauer-Stiftung

**KfW**

Kreditanstalt für Wiederaufbau

**KZM**

Kompetenzzentrum Methoden

**MISEREOR**

Bischöfliches Hilfswerk MISEREOR

**NRO**

Nichtregierungsorganisation

**OECD**

Organisation for Economic  
Cooperation and Development

**PTB**

Physikalisch-Technische  
Bundesanstalt

**POS**

Politische Stiftungen

**RMC**

Rambøll Management Consulting

**REA**

Rapid Evidence Assessment

**RLS**

Rosa-Luxemburg-Stiftung

**SEK**

Sozial- und Entwicklungshilfe  
des Kolpingwerkes

**SEVAL**

Schweizerische  
Evaluationsgesellschaft

**SLE**

Seminar für ländliche Entwicklung

**SST**

Sozialstrukturträger

**ToR**

Terms of Reference

**TZ**

Technische Zusammenarbeit

**VENRO**

Verband Entwicklungspolitik  
deutscher Nichtregierungsorga-  
nisationen e.V.

**WMS**

Wissensmanagementsystem

**ZGO**

Zivilgesellschaftliche  
Organisationen

1.

EINLEITUNG

Das Deutsche Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) nahm im Sommer 2012 seine Arbeit auf. Übergeordnetes Ziel des DEval ist es, einen Beitrag zur Verbesserung der Wirksamkeit der deutschen Entwicklungszusammenarbeit (EZ) zu leisten. Dazu führt das DEval insbesondere unabhängige Evaluierungen durch, die auch Empfehlungen für die Verbesserung der analysierten Maßnahmen enthalten. Mandat, Ziel und Aufgaben des DEval wurden in ihren wesentlichen Zügen im Gesellschaftsvertrag durch die Gesellschafterin, die Bundesrepublik Deutschland, vertreten durch das BMZ, vorgezeichnet. Diese können in die folgenden Leistungsbereiche untergliedert werden:

- Durchführung strategischer Evaluierungen: Es ist die Aufgabe des DEval, strategische Evaluierungen durchzuführen, die insbesondere für politische Entscheidungsträger (BMZ, AwZ), aber auch für die deutschen EZ-Organisationen Steuerungsrelevanz besitzen.
- Methodenforschung: Das DEval ist beauftragt, Methodenforschung zu betreiben sowie Evaluierungsdesigns, -methoden und -standards weiterzuentwickeln.

Das Aufgabenspektrum umfasst weitere Leistungsbereiche, die jedoch zu Beginn der Ausgangserhebung noch nicht operativ waren, wie die Förderung von Evaluierungskapazitäten in Kooperationsländern der EZ. Zur Vermittlung der Ergebnisse der Leistungsbereiche nutzt das DEval Netzwerke, Kooperationen, Publikationen sowie Qualifizierungsmaßnahmen. Auf diese Weise wird zugleich Lernen aus Evaluierungen ermöglicht.

Das DEval ist ein vergleichsweise neuer Akteur in der Evaluierungslandschaft der deutschen Entwicklungszusammenarbeit (EZ). Die Erstellung einer „Ausgangserhebung zum DEval und seinem Umfeld“ als eines der ersten und umfangreich angelegten Projekte des Instituts verfolgt demnach die folgenden Ziele: DEval möchte einerseits seine eigene Arbeit und die Wirkung seiner Tätigkeit überprüfen können und andererseits einen tieferen Einblick in die Evaluierungspraxis größerer und kleinerer Akteure in der deutschen EZ erhalten.

Die „Ausgangserhebung zum DEval und seinem Umfeld“ umfasst zwei separate Erhebungsbereiche: Zum einen die Durchführung einer „Baseline-Studie“ bzw. „Erhebung zur Ausgangslage“ für das Institut an sich, als Standortbestimmung eines neuen EZ-Akteurs und Referenzpunkt für zukünftige Leistungsüberprüfungen (vgl. DEval, 2015a). Zum anderen wurde das DEval vom BMZ mandatiert, ein *Umsetzungsmonitoring der letzten Systemprüfung* durchzuführen. Als „Systemprüfung“ gilt hierbei die im Jahr 2009 von Axel Borrmann und Reinhard Stockmann veröffentlichte Studie „Evaluation in der deutschen Entwicklungszusammenarbeit“ (vgl. Borrmann und Stockmann, 2009a, 2009b), die im Auftrag des BMZ die Evaluierungspraxis deutscher EZ-Organisationen untersuchte, und in der Handlungsempfehlungen zu deren Verbesserung ausgesprochen wurden. Im Rahmen des Umsetzungsmonitorings wurde die Umsetzung dieser Handlungsempfehlungen überprüft. Zudem wurde in diesem Zuge der *aktuelle Stand der Evaluierungspraxis* in der deutschen Entwicklungszusammenarbeit erfasst. Die Ergebnisse hierzu werden in dem vorliegenden Bericht dokumentiert.

### Umsetzungsmonitoring und aktueller Stand der Evaluierungspraxis

In der „Systemprüfung“ wurde untersucht, „wie die einzelnen EZ-Organisationen Evaluation institutionalisieren, planen, durchführen und nutzen, ob sie durch Evaluation insbesondere etwas über die Wirkungen ihrer Projekte und Programme erfahren können und ob und wie sich ihre einzelnen Evaluationssysteme zu einem konsistenten Gesamtsystem für die bilaterale deutsche Entwicklungszusammenarbeit zusammenfügen“ (Borrmann und Stockmann, 2009a: 5). Es handelt sich bei dieser Studie um die zweite Systemprüfung<sup>3</sup> und die bislang umfassendste und detaillierteste Darstellung des Evaluierungssystems in der deutschen Entwicklungszusammenarbeit. Anhand von Einzelfallstudien wurden das BMZ sowie 19 deutsche EZ-Organisationen untersucht (für eine Auflistung vgl. Kapitel 2.1.2). Zudem wurde eine Fallstudie zum Evaluierungssystem der Nichtregierungsorganisationen (NRO) der deutschen EZ durchgeführt.<sup>4</sup>

<sup>3</sup> In den Jahren 1997/98 veranlasste das BMZ erstmals eine systematische Untersuchung der Evaluierungen in der deutschen Entwicklungszusammenarbeit (Borrmann et al., 1999).

<sup>4</sup> In dieser Fallstudie wurden 49 Mitgliedsorganisationen des Verbands Entwicklungspolitik deutscher Nichtregierungsorganisationen e.V. (VENRO) einbezogen sowie vier Zuwendungsempfänger des BMZ, die nicht Mitglied bei VENRO sind.

Auf Grundlage der Einzelfallstudien wurden auf Ebene der untersuchten EZ-Organisationen organisationspezifische Handlungsempfehlungen zur Weiterentwicklung der Evaluierungspraxis entwickelt. Für das BMZ wurden Handlungsempfehlungen festgehalten<sup>5</sup>, die sich auf die gesamte Systementwicklung bezogen (Borrmann und Stockmann, 2009b: 7ff). Darüber hinaus wurden Handlungsempfehlungen allgemeiner Natur ausgesprochen, die aus der Synthese der Ergebnisse aus den Einzelfallstudien sowie einem Fazit anhand der OECD DAC-Kriterien abgeleitet wurden (Borrmann und Stockmann, 2009a: 165ff). Eine zentrale Empfehlung betraf dabei die „Gründung einer unabhängigen Evaluationsagentur, die das BMZ bei der Systementwicklung und die EZ-Organisationen bei der Instrumenten- und Qualitätsentwicklung berät und unterstützt, sowie unabhängig, mit hoher Kompetenz und damit Glaubwürdigkeit Evaluationen insbesondere zur Wirksamkeit und Nachhaltigkeit der EZ länder-, sektor- und institutionenvergleichend vornimmt“ (Borrmann und Stockmann, 2009a: 165). Mit der Gründung des DEval hat das BMZ diese systemverändernde Empfehlung umgesetzt und die Reform des Evaluierungssystems eingeleitet.

Im Nachgang der „Systemprüfung“ vereinbarte das BMZ mit den 19 beteiligten EZ-Organisationen zwischen November 2008 und Dezember 2009 Umsetzungspläne für die Handlungsempfehlungen. Dabei wurde dokumentiert, ob die Empfehlungen von der betreffenden EZ-Organisation geteilt wurden. Anschließend wurden in den Umsetzungsplänen einzelne Teilschritte oder Indikatoren sowie Terminierungen und Verantwortlichkeiten für die Umsetzung der akzeptierten Empfehlungen festgelegt.

Der Fokus des Umsetzungsmonitorings liegt auf diesen organisationspezifischen Handlungsempfehlungen, die sich an die 19 EZ-Organisationen der Einzelfallstudien richteten. Um darüber hinaus Aussagen über die Weiterentwicklung der Evaluierungslandschaft der deutschen Entwicklungszusammenarbeit

auf Systemebene treffen zu können, wurde untersucht, wie sich der gegenwärtige Status quo der Evaluierungspraxis deutscher EZ-Organisationen charakterisieren lässt. Dieser Einblick ist hilfreich für die zukünftige Arbeit des DEval, da er dem Institut ein besseres Verständnis seines Umfeldes ermöglicht.

Das angewandte Erhebungsdesign basierte wesentlich auf Selbsteinschätzungen und Eigenangaben, die im Rahmen von Befragungen gewonnen wurden. Diese Angaben wurden um eine Metaevaluierung ergänzt, bei der bereits abgeschlossene Evaluierungen nach zuvor festgelegten Kriterien untersucht wurden. Es handelt sich insofern um keine erneute „Systemprüfung“, da keine Bewertung einzelner Organisationen vorgenommen wurde. Stattdessen erfolgte die Auswertung auf aggregierter Ebene, wobei nach Akteursgruppen differenziert wurde. Der Bericht stellt somit einen zusammenfassenden, deskriptiven Überblick zu der Umsetzung der Handlungsempfehlungen sowie zu darüber hinausgehenden Entwicklungen der Evaluierungspraxis dar.

Der vorliegende Bericht zu den Ergebnissen des Umsetzungsmonitorings sowie zum Stand der Evaluierungspraxis in der deutschen EZ ist folgendermaßen aufgebaut:

- Kapitel 1 führt in die Erhebung zur „Evaluierungspraxis in der deutschen Entwicklungszusammenarbeit“ ein.
- Kapitel 2 beschreibt das methodische Vorgehen der Erhebung.
- Kapitel 3 präsentiert die Ergebnisse zu den Erhebungen des Umsetzungsmonitorings und der Evaluierungspraxis entlang verschiedener Dimensionen.
- Kapitel 4 fasst die Ergebnisse in einem Fazit und in Schlussfolgerungen zusammen.

Die Anlagen können in dem separat veröffentlichten Anlagenband eingesehen werden (vgl. DEval, 2015b).

<sup>5</sup> Für die Handlungsempfehlungen, die sich an das Ministerium richteten, wurde damals kein Umsetzungsplan entwickelt; dies wurde vom BMZ aber retrospektiv nachgeholt. DEval ist infolge der Empfehlungen der „Systemprüfung“ gegründet worden. Damit sind die ans Ministerium gerichteten Empfehlungen entweder umgesetzt oder obsolet. Bei den Empfehlungen, die nicht vom BMZ geteilt werden bzw. obsolet sind, handelt es sich überwiegend um Alternativen zur Gründung eines unabhängigen Evaluierungsinstitutes wie z.B. die organisatorische Aufhängung des Evaluierungsreferates als Stabsstelle (vgl. BMZ, 2014).

2.

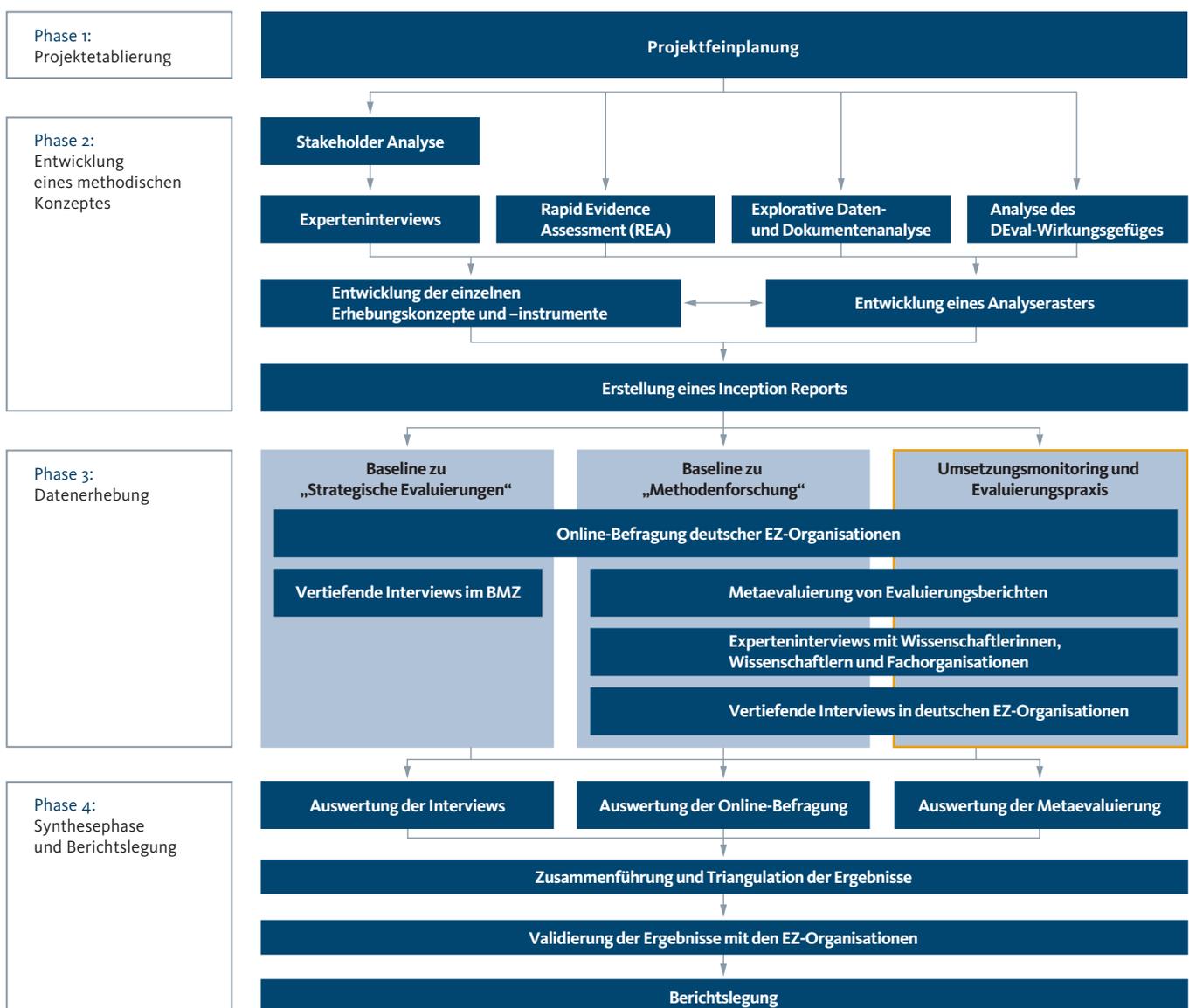
## METHODISCHES VORGEHEN

Das Projektdesign der „Ausgangserhebung zum DEval und seinem Umfeld“ wurde als Gesamtkonzept entwickelt, d.h. es umfasst sowohl den Untersuchungsbereich der *Erhebung zur Ausgangslage* des DEval (vgl. Kapitel 1) sowie den Untersuchungsbereich *Umsetzungsmonitoring und Evaluierungspraxis*. Das Projektdesign gliedert sich in vier Phasen: Projektetablierung, Entwicklung

eines methodischen Konzepts, Datenerhebung sowie Synthese und Berichtslegung (vgl. Abbildung 1).

Nach Festlegung des Projektdesigns wurden (explorative) Vorarbeiten und -untersuchungen Anfang April 2014 abgeschlossen (Phasen 1 und 2). Es folgte bis August 2014 eine umfangreiche Erhebungsphase, bei der das DEval durch die

Abbildung 1: Methodisches Konzept der Ausgangserhebung



Teilnahme der einbezogenen Akteure und Organisationen an den Erhebungen und die Bereitstellung von verschiedensten Unterlagen unterstützt wurde (Phase 3). Die Auswertung und Synthese der erhobenen Daten fand in den Monaten August und September 2014 statt. In einem Abschlussworkshop mit den beteiligten EZ-Organisationen wurden Ende Oktober 2014 die Ergebnisse gemeinsam diskutiert und Ergänzungen für die anschließende Berichtslegung aufgenommen (Phase 4).

Im Verlauf der Datenerhebung wurden unterschiedliche Erhebungsinstrumente für einzelne Untersuchungsbereiche gezielt eingesetzt.<sup>6</sup> Die Erhebungen zu *Umsetzungsmonitoring und Evaluierungspraxis* sind in der Abbildung durch eine farbliche Umrandung hervorgehoben. Sie werden im Folgenden näher betrachtet und die damit einhergehenden, methodischen Herausforderungen beschrieben.

## 2.1 Erhebungsphasen und -instrumente

Für die Erhebung wurden sowohl qualitative als auch quantitative Methoden verwendet. Im Folgenden werden zunächst jene Instrumente beschrieben, die in der explorativen Phase der methodischen Konzeptentwicklung Anwendung fanden (Phase 2). Im Anschluss werden die Instrumente der Datenerhebungsphase beschrieben (Phase 3).

### 2.1.1 Explorative Phase

Das Ziel der explorativen Phase war es, den Kontext der Untersuchung zu erfassen, einzugrenzen und ein methodisches Konzept für die „Ausgangserhebung zum DEval und seinem Umfeld“ zu entwickeln.

- **Stakeholder-Analyse**

Zu Beginn wurde eine Stakeholder-Analyse durchgeführt, die der Identifikation relevanter Akteure für die „Ausgangserhebung zum DEval und seinem Umfeld“ diente. Die Analyse unterstützte die Entscheidungen,

- a. mit welchen Stakeholdern explorative Experteninterviews geführt wurden (siehe unten) und
- b. welche Stakeholder mit welchen Instrumenten in der Datenerhebung untersucht wurden.

Zudem konnte bereits ein Teil der identifizierten Stakeholder im Dezember 2013 zu einer Auftaktveranstaltung der Ausgangserhebung ins DEval eingeladen werden, um das methodische Konzept der Ausgangserhebung zu diskutieren.

- **Explorative Experteninterviews**

Im Zeitraum von Ende Januar bis Ende Februar 2014 wurden zehn der durch die Stakeholder-Analyse identifizierten Akteure staatlicher EZ-Organisationen, Organisationen der Zivilgesellschaft, EZ-Verbänden sowie der Wissenschaft interviewt. Sie wurden zu deren Sichtweise des DEval, zum Evaluierungssystem der deutschen Entwicklungszusammenarbeit sowie zu Anregungen und Anmerkungen bezüglich der Ausgangserhebung befragt. Diese Informationen waren grundlegend für die weitere Gestaltung des methodischen Konzepts.

- **Rapid Evidence Assessment (REA) und Entwicklung eines Analyserasters**

Für die Charakterisierung der Evaluierungspraxis bedurfte es einer Systematisierung, um eine analytische Beschreibung des Evaluierungssystems zu ermöglichen. Zur Erfassung des aktuellen Forschungsstandes zur Evaluierungspraxis wurde in der Vorbereitungsphase daher ein Rapid Evidence Assessment (REA) durchgeführt.

<sup>6</sup> Für die Baseline-Studie wurden hierzu die Bereiche „Strategische Evaluierungen“ und „Methodenforschung“ voneinander abgegrenzt. Auf die Hintergründe dieser Differenzierung wird an dieser Stelle jedoch nicht näher eingegangen. Ebenfalls nicht näher betrachtet werden die im Projektdesign aufgeführte „Explorative Daten- und Dokumentenanalyse“ sowie die „Analyse des Wirkungsgefüges“ des DEval, die sich auf den Untersuchungsbereich der Baseline bezogen. Ausführliche Erläuterungen zu diesen Aspekten finden sich im dazugehörigen Bericht (vgl. DEval, 2015a).

Tabelle 1: Dimensionen und inhaltliche Cluster von Evaluierungspraxis			
Dimension	Untersuchungsfragen	Inhaltliche Cluster	Kategorie Systemprüfung
Ziele von Evaluierungen	Mit welchem Ziel wird evaluiert?	Art und Ziele der durchgeführten Evaluierungen	Nutzung, Formen und Umfang
	Welcher Nutzen wird aus Evaluierungen generiert?	Nutzung und Transparenz der Evaluationsergebnisse	
	Wie werden Evaluationsergebnisse in Organisationen genutzt?		
Organisatorischer Kontext	Welche Faktoren des organisatorischen Kontextes beeinflussen die Evaluierungspraxis?	Akzeptanz von Evaluierungen seitens der Mitarbeitenden	Nutzung, Partner, Vernetzung
		Austausch zwischen EZ-O und Netzwerkbindung	
Strukturen und Prozesse	Wie ist die Evaluierungspraxis strukturiert und in Prozessen festgelegt?	Stellung der Evaluationseinheit in der Institution	Planung und Ablauf, Grundlagen und Ziele, Institutionalisierung
		Festgelegte Prozesse zur Evaluierungspraxis	
		Institutionalisierte Evaluierungsgrundlagen	
		Einbindung von Partnern	
Humanressourcen	Welche für Evaluierungspraxis relevanten Humanressourcen sind in den Organisationen vorhanden?	Anzahl der Mitarbeitenden in der Evaluationseinheit	Personal
		Monitoring- und Evaluierungskompetenz der Mitarbeitenden	
		Monitoring- und Evaluierungskompetenz der Partner	
		Einsatz von Gutachterinnen und Gutachtern	
		Eingesetzte Designs und Methoden	
Finanzielle Ressourcen	Wie hoch sind die finanziellen Mittel, die für die Evaluierungspraxis zur Verfügung stehen?	Budget für Evaluierungen und Evaluationssysteme	Formen und Umfang
Technische Ressourcen	Welche für die Evaluierungspraxis relevanten technischen Ressourcen sind in der Organisation vorhanden?	Zugang zu Wissensmanagementsystem und Evaluationsliteratur	Qualität

Dabei wurden insbesondere Ansätze, Methoden und Konzepte identifiziert, um ein theoriegeleitetes Analysesystem für Evaluierungspraxis zu entwickeln.<sup>7</sup> Auf Grundlage der Ergebnisse des REA wurden Dimensionen für die Erfassung von Evaluierungspraxis entwickelt und diesen Dimensionen untersuchungsleitende Fragestellungen zugeordnet, die

eine Erfassung des Evaluierungssystems der deutschen EZ-Landschaft ermöglichen (vgl. Tabelle 1). Anschließend erfolgte die Zuordnung der in der „Systemprüfung“ auf Ebene einzelner Organisationen ausgesprochenen Empfehlungen zu diesen Dimensionen. In Tabelle 1 sind die Dimensionen der Evaluierungspraxis sowie die ihnen zugeordneten Kategorien

<sup>7</sup> In den durchsuchten Datenbanken (JSTOR, Emerald insight, Sage, Wiley Online Library, Science direct, Springer Link, OECD iLibrary, Weltbank eLibrary, Google Scholar, Academia.edu) wurden unter Verwendung von „Evaluierungspraxis“ abgeleiteter Synonyme (deutsch und englisch) 72 Studien identifiziert. Nach einer ersten Überprüfung des Forschungsthemas in den Titeln und Kurzbeschreibungen (Abstracts) sowie des Erscheinungsjahrs wurden 40 Studien, welche eine hohe Relevanz für die Forschungsfrage besaßen, ausgewählt. Berücksichtigt wurde sowohl Peer-Review- als auch nicht Peer-Review-Literatur, die zwischen 2005 und 2013 veröffentlicht wurde, nicht aber Graue-Literatur, da diese in der Regel keinem unabhängigen Qualitätssicherungsprozess unterliegt. In einem weiteren Selektionsprozess wurde die Qualität der Studien nach methodischer Qualität bewertet und 14 Studien für die weitere Analyse des REA ausgewählt.

der „Systemprüfung“ dargestellt. Basierend auf einer Analyse der Kriterien, die in der „Systemprüfung“ impliziert waren und entlang der im REA entwickelten Kategorien wurden in einem Analyseraster die zentralen Fragen und Aspekte für die Ausgangserhebung erfasst. Ferner wurden jedem Untersuchungsaspekt Indikatoren zugeordnet, anhand derer der entsprechende Aspekt im Laufe der Erhebung identifiziert und analysiert werden konnte (vgl. Anlage 1.1 bis 1.4).

### 2.1.2 Datenerhebungsphase

In der Datenerhebungsphase wurden für den Untersuchungsbereich Umsetzungsmonitoring und Evaluierungspraxis folgende Erhebungsinstrumente eingesetzt:

- **Online-Befragung**

Die im Analyseraster identifizierten Fragestellungen zur Evaluierungspraxis der deutschen EZ wurden in das Design einer Online-Befragung von EZ-Organisationen integriert. Als Stichprobe deutscher EZ-Organisationen wurden für die Befragung alle EZ-O angeschrieben, die als Mittelempfänger des BMZ in den Jahren 2012 und 2013 jeweils mindestens eine Million Euro erhielten<sup>8</sup> (dies schloss auch alle an der „Systemprüfung“ beteiligten EZ-Organisationen ein). Nach der Prüfung anhand eines Pretests wurde die Befragung von Mitte April bis Anfang Juni 2014 durchgeführt. Von den 52 angeschriebenen Organisationen beteiligten sich 34 (Rücklaufquote: 65,4 Prozent; eine Auflistung der angeschriebenen Organisationen findet sich in Anlage 2.1.).

Neben (überwiegend quantitativen) Angaben zur Evaluierungspraxis wurden in der Online-Befragung bei den an der „Systemprüfung“ beteiligten Organisationen Angaben zur Umsetzung der ausgesprochenen Handlungsempfehlungen erhoben. Dazu sollten die Evaluierungseinheiten bzw. Evaluierungsbeauftragten der betreffenden Organisationen für jede Handlungsempfehlung beurteilen, ob diese vollständig, teilweise oder gar nicht umgesetzt wurde. Darüber hinaus wurde für jede Handlungsempfehlung anhand vorgegebener Kategorien abgefragt, welche Faktoren förderlich oder hinderlich für die Umsetzung waren (Grad der Übereinstimmung mit der Handlungsempfehlung, Verfügbarkeit von finanziellen Ressourcen, Haltung der Geschäftsführung).

- **Vertiefende Interviews in deutschen EZ-O**

Um die Daten aus der Online-Befragung zu der Umsetzung von Handlungsempfehlungen zu validieren und ggfs. um qualitative Informationen zu ergänzen, wurden halbstandardisierte Interviews mit den 17 Organisationen durchgeführt, die an der „Systemprüfung“ teilnahmen (vgl. Leitfaden in Anlage 6). Die Interviews wurden ferner genutzt, um einen ausführlicheren Einblick in die Evaluierungspraxis der einbezogenen EZ-Organisationen zu erhalten. Im Rahmen von ein- bis zweitägigen Besuchen wurden Gespräche mit der Geschäftsführung, der Evaluierungseinheit, der operativen Leitung und Mitarbeitenden des operativen Bereichs geführt. Im Einzelnen waren dies (in alphabetischer Reihenfolge): Bischöfliches Hilfswerk MISEREOR (MISEREOR), Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), Caritas International (Ci), Deutsche Investitions- und Entwicklungsgesellschaft mbH (DEG), Deutsche Welthungerhilfe (DWHH), Evangelisches Werk für Diakonie und Entwicklung (EWDE), Friedrich-Ebert-Stiftung (FES), Friedrich-Naumann-Stiftung für die Freiheit (FNF), Gesellschaft für internationale Zusammenarbeit (GIZ), Hanns-Seidel-Stiftung (HSS), Heinrich-Böll-Stiftung (hbs), Institut für internationale Zusammenarbeit des Deutschen Volkshochschul-Verbandes (dvv international), Konrad-Adenauer-Stiftung (KAS), Kreditanstalt für Wiederaufbau (KfW), Physikalisch-Technische Bundesanstalt (PTB), Rosa-Luxemburg-Stiftung (RLS), Sozial- und Entwicklungshilfe des Kolpingwerkes (SEK).

- **Metaevaluierung**

Die Zielsetzung der Metaevaluierung war, Erkenntnisse zu den angewandten Designs und Methoden in strategischen oder projektübergreifenden Evaluierungen in der deutschen EZ zu gewinnen. Hierzu wurden die an der Online-Befragung teilnehmenden Organisationen um Folgendes gebeten: Sie sollten die in den Jahren von 2008 bis 2012 von ihnen beauftragten durchgeführten strategischen und/ oder projektübergreifenden Evaluierungen angeben und gemäß ihres jeweiligen Evaluierungstyps spezifizieren (Sektor-, Länderprogramm-, Portfolio- oder Instrumentenevaluierung). Von den hierbei 213 genannten Evaluierungen wurde anschließend in Abhängigkeit von Akteursgruppe

<sup>8</sup> Direkt oder über Engagement Global vermittelt.

und Evaluierungstyp eine Stichprobe von 45 Evaluierungsberichten<sup>9</sup> für die Metaevaluierung gezogen. Daraufhin wurden die entsprechenden Evaluierungsberichte (inklusive Anhänge und Terms of Reference) bei den Organisationen erbeten und anhand eines im Vorfeld entwickelten Analyserasters untersucht (vgl. Anlage 1.2). Zur Bestimmung relevanter Analyseaspekte – wie der Systematik der Berichtslegung, der Art des Evaluierungsdesigns oder der Darstellung des methodischen Vorgehens – wurden etablierte Qualitätsstandards und darauf basierende Kriterienkataloge berücksichtigt (vgl. BMZ, 2006; Caspari, 2010; DeGEval, 2008; European Commission, 2006; OECD, 2010; Scriven, 2010; Widmer, 1996).

- **Interviews mit Wissenschaftlerinnen, Wissenschaftlern und Fachorganisationen**

Um die in deutschen EZ-Evaluierungen angewandten Designs und Methoden in den internationalen Kontext einordnen zu können, wurden im Zeitraum von Juli bis August 2014 insgesamt 9 Experteninterviews durchgeführt. Die Auswahl der Expertinnen und Experten erfolgte im Rahmen der Stakeholder-Analyse (vgl. Kapitel 2.1.1). Personen, die bereits in der explorativen Phase befragt wurden, blieben hiervon unberücksichtigt (vgl. Kapitel 2.1). Bei diesen Interviews wurden Expertinnen und Experten folgender Organisationen befragt: DAC Evaluation Resource Center, Deutsches Institut für Entwicklungspolitik (DIE), Independent Commission for Aid Impact (ICAI), Netherlands Ministry of Foreign Affairs, Schweizerische Evaluationsgesellschaft (SEVAL), Seminar für ländliche Entwicklung (SLE), Swedish Expert Group on Aid, United Nations Evaluation Group, Weltbank (der bei den Gesprächen verwendete Leitfaden kann in Anlage 4 eingesehen werden).

Die einzelnen Erhebungsinstrumente finden sich in den Anlagen 3 bis 7. Die Stichproben der einzelnen Erhebungsinstrumente finden sich in Anlage 2.

## 2.2

### Methodische Herausforderungen

Als methodische Herausforderungen können insbesondere die adäquate Berücksichtigung unterschiedlicher Stakeholder, die Entwicklung eines geeigneten Analyserasters sowie die Berücksichtigung von Transparenz und Vertraulichkeit bezeichnet werden.

#### 2.2.1 Adäquate Berücksichtigung unterschiedlicher Stakeholder

Eine wesentliche methodische Herausforderung stellte die Einbeziehung und adäquate Berücksichtigung zahlreicher unterschiedlicher Akteure (EZ-Organisationen, Fachorganisationen und -referate, nationale und internationale Expertinnen und Experten) im Erhebungsdesign dar.

Zunächst wurde im Rahmen der Vorbereitungsphase festgestellt, dass ein Teil der Organisationen sich nur in geringem Maße von Aktivitäten des DEval betroffen sah und daher der Mehrwert der Ausgangserhebung für die eigene Organisation als gering eingestuft wurde. Diese Einschätzung barg die Gefahr einer niedrigen Bereitschaft für die Teilnahme an den mitunter aufwendigen Datenerhebungen. Diesem Risiko wurde mit einer proaktiven Ansprache der Akteure und dem Werben um Beteiligung entgegengewirkt. Im Zuge der Studie wurde erfreulicherweise eine hohe Bereitschaft zur Teilnahme festgestellt (so hatte bspw. die Online-Befragung eine sehr hohe Rücklaufquote von 65,4 Prozent; vgl. Kapitel 2.1.2).

Es musste zudem dem Umstand Rechnung getragen werden, dass in der „Systemprüfung“ noch 19 EZ-Organisationen untersucht wurden, bei denen zwischenzeitlich jedoch Fusionen erfolgt waren: So waren an der „Systemprüfung“ die Organisationen Deutscher Entwicklungsdienst (DED), Gesellschaft für technische Zusammenarbeit (GTZ) sowie die Internationale Weiterbildung und Entwicklung gGmbH (InWEnt) beteiligt, die inzwischen in der Deutschen Gesellschaft für Internationale Zusammenarbeit (GIZ) vereinigt sind. Zudem wurde im Rahmen der „Systemprüfung“ der Evangelische Entwicklungsdienst (EED) untersucht, der sich zwischenzeitlich mit Brot für

<sup>9</sup> Ursprünglich war geplant, eine repräsentative Zufallsstichprobe von 52 Berichten zu ziehen (Konfidenzniveau 90%). Es zeigte sich jedoch, dass einzelne Akteursgruppen und Evaluierungstypen stark überrepräsentiert waren. Um einen möglichst breiten Überblick zu gewährleisten, wurde daher eine gleichverteilte, geschichtete Ziehung nach Akteursgruppe (DO, POS, kirchliche ZGO, andere ZGO) und Evaluierungstyp (Instrumenten-, Länder-, Portfolio-, Sektorevaluierung) vorgenommen. Da nicht alle Felder der dabei entstandenen Kombinationen besetzt waren, verringerte sich die Stichprobengröße auf letztlich 45 Evaluierungsberichte, die in die Metaevaluierung eingingen.

die Welt (BfdW) und Diakonie Deutschland – Evangelischer Bundesverband zum Evangelischen Werk für Diakonie und Entwicklung (EWDE) zusammengeschlossen hat. Für das Umsetzungsmonitoring bedeuteten diese organisatorischen Veränderungen, dass die Handlungsempfehlungen an Organisationen gerichtet waren, die in ihrer damaligen Form nicht mehr existierten. Im Zuge der Online-Befragung wurde daher in beiden Fällen (GIZ und EWDE) zunächst die Umsetzung der Handlungsempfehlungen an die Vorgängerorganisation abgefragt. Anschließend wurde in den vertiefenden Interviews erörtert, inwiefern die Empfehlungen an die Vorgängerorganisationen noch relevant sind und welche Veränderungen oder Weiterentwicklungen stattgefunden haben.

### 2.2.2 Grenzen von Erhebungsinstrumenten

Eine weitere Herausforderung bestand darin, die Erhebungsinstrumente derart zu gestalten, dass sie den Eigenheiten der verschiedenen Organisationen und Akteursgruppen (DO, POS und andere ZGO) gerecht wurden. So konnten beispielsweise bei der Gestaltung der Online-Befragung – deren Stichprobe 52 Organisationen umfasste – nicht die Merkmale und Besonderheiten aller teilnehmenden Organisationen berücksichtigt und in Antwortoptionen widerspiegelt werden. Die Gefahr, Merkmale und Besonderheiten der teilnehmenden Organisationen nicht im Einzelnen berücksichtigen zu können, ist zudem direkt mit der Auswertungsform verbunden, in der das Umsetzungsmonitoring erfolgte: Sie erfolgte ausschließlich auf aggregierter Ebene, da sie eine Momentaufnahme auf Systemebene darstellt und bewusst auf eine Bewertung von Sachverhalten auf Organisationsebene verzichtet. Waren jedoch spezifische Besonderheiten der Akteursgruppen wichtig für die Interpretation der Ergebnisse, so wurden sie – soweit möglich – auch abgefragt und berücksichtigt. Die beteiligten Organisationen hatten zudem die Möglichkeit, sich in der Befragungsphase an eine Ansprechpartnerin zu wenden, die aufkommende Fragen klärte.

In der Online-Befragung wurde mit Blick auf die Metaevaluierung die Frage nach durchgeführten strategischen und/oder projektübergreifenden Evaluierungen gestellt. Die Fokussierung auf diese Evaluierungen diente dazu, die Grundgesamtheit aller Evaluierungen des Untersuchungszeitraums

zu reduzieren und zugleich deren Vergleichbarkeit mit strategischen Evaluierungen des DEval im Zuge der Baseline-Studie herzustellen. Diese Kategorisierung der Evaluierungen als strategisch (im Sinne übergeordneter Steuerungsrelevanz) stellte sich jedoch bezüglich eines einheitlichen Begriffsverständnisses als problematisch heraus, wie die Heterogenität der Stichprobe zeigte. So enthielt bspw. über die Hälfte der untersuchten Evaluierungen keine projektübergreifenden Empfehlungen. Da auch jene Evaluierungen ohne übergeordnete Steuerungsrelevanz einbezogen wurden, kann die untersuchte Stichprobe als breite Streuung unterschiedlicher Evaluierungstypen gelten. Falls die Untersuchung bestimmter Analyseaspekte für einzelne Evaluierungstypen als unpassend beurteilt wurde, wurden diese ausgeschlossen.

### 2.2.3 Entwicklung eines geeigneten Analyserasters

Bei der Zuordnung der Handlungsempfehlungen zu den Dimensionen von Evaluierungspraxis (vgl. Tabelle 1) bestand eine methodische Herausforderung darin, dass einzelne Handlungsempfehlungen in den Umsetzungsplänen mehrere thematische Aspekte zugleich aufgriffen. Falls keine eindeutige thematische Zuordnung dieser Handlungsempfehlungen möglich war, wurden sie derart in einzelne Aspekte aufgeteilt, bis eine Zuordnung möglich war.<sup>10</sup> Für die beteiligten Organisationen hatten diese Anpassungen zur Folge, dass die Handlungsempfehlungen nicht durchgängig in ihrer ursprünglichen Formulierung abgefragt wurden. Ausdifferenzierungen oder Kürzungen der Handlungsempfehlungen erfolgten so nah an der ursprünglichen Formulierung wie möglich; zudem wurde jeweils auf die ursprüngliche Form der Empfehlung verwiesen.

Hierbei ist zu beachten, dass die Handlungsempfehlungen teilweise deskriptive Passagen enthielten, die keinen Empfehlungscharakter aufwiesen. Sofern sich daraus keine Verständnisschwierigkeiten ergaben, blieben diese Passagen bei der Abfrage unberücksichtigt. Eine weitere Herausforderung bestand darin, dass in einigen Handlungsempfehlungen Defizite beschrieben wurden, ohne dass daraus eine abgeleitete, normative Vorgabe formuliert wurde. In einigen Fällen konnte die Festlegung, ob die Formulierung auch eine implizite Empfehlung enthielt, letztlich nur anhand der Einschätzung durch das Evaluatorenteam erfolgen.

<sup>10</sup> Auf Nachfrage kann DEval eine Übersicht zu der vorgenommenen Zuordnung zur Verfügung stellen.

Falls Empfehlungen an einzelne EZ-Organisationen gerichtet wurden, sie zugleich aber organisationsübergreifenden Charakter hatten – z.B. als übergreifende Leitlinie für die politischen Stiftungen –, so wurde der Umsetzungsstand der entsprechenden Empfehlung zwar erfasst, aber nicht in der organisationsspezifischen Auswertung des Anteils umgesetzter Handlungsempfehlungen berücksichtigt.

#### **2.2.4 Überprüfung von Selbsteinschätzungen**

Die Angaben zur Umsetzung der Handlungsempfehlungen wurden zunächst im Rahmen der Online-Befragung erhoben und basieren insofern auf Selbsteinschätzungen der EZ-Organisationen. Um diese Eigenangaben zu überprüfen, wurden sie in den vertiefenden Interviews trianguliert, indem die Umsetzung jeder Handlungsempfehlung erneut aufgegriffen wurde. Für vollständig oder teilweise umgesetzte Empfehlungen wurde erhoben, wie sich die Umsetzung gestaltete bzw. welche Teile der Empfehlung umgesetzt werden konnten. Für nicht umgesetzte Empfehlungen wurde in Erfahrung gebracht, welche Gründe einer Umsetzung entgegenstanden. Sofern

möglich, wurde die Umsetzung zudem belegt, bspw. durch ein Organigramm der Organisation, das aufzeigte, dass die Unabhängigkeit der Evaluierungseinheit durch die Schaffung einer Stabsstelle gefördert wurde, oder indem Evaluierungsrichtlinien zur Verfügung gestellt wurden.

#### **2.2.5 Transparenz und Vertraulichkeit**

Bei einigen EZ-Organisationen wurde für die Durchführung der Metaevaluierung die Zusendung von Evaluierungsberichten erbeten (vgl. Kapitel 2.1.2). Bereits im Auftaktworkshop wurden hierzu Bedenken geäußert, da der Schutz von Partnern durch die Weitergabe von Berichten gefährdet werden könnte. Einige EZ-O holten daher die schriftliche Zusicherung einer vertraulichen Behandlung der Berichte ein, bevor sie diese zur Verfügung stellten. Der damit verbundene Klärungs- und Bearbeitungsprozess zog zeitliche Verschiebungen in der Projektplanung nach sich, die durch inhaltliche Verschiebungen im Projektablauf ausgeglichen werden konnten (vgl. Anlage 10).

# 3.

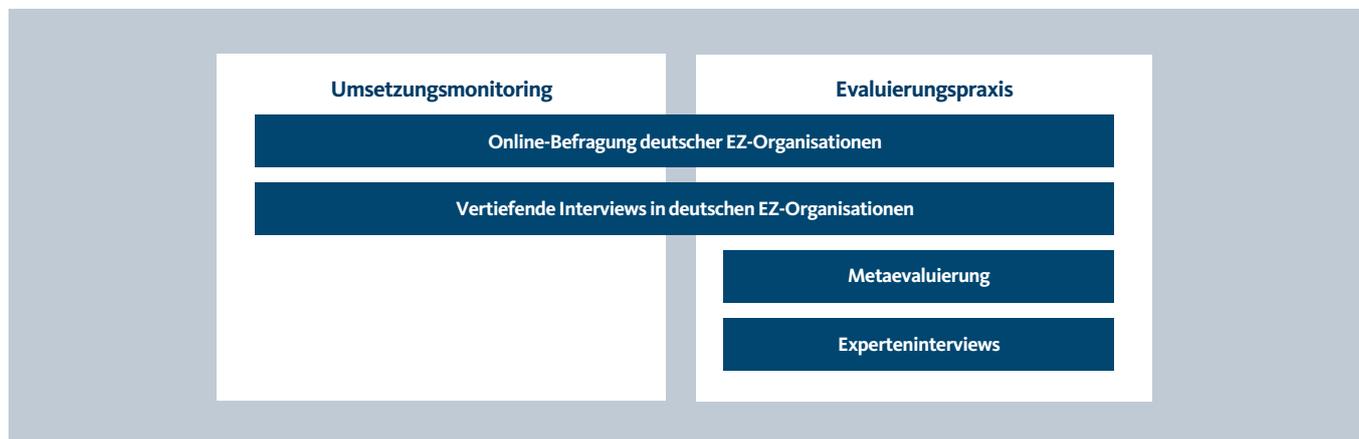
## UMSETZUNGS- MONITORING UND EVALUIERUNGS- PRAXIS

In diesem Kapitel wird aufgezeigt, inwiefern die Empfehlungen der „Systemprüfung“ umgesetzt wurden und wie sich der aktuelle Stand der Evaluierungspraxis darstellt. Zunächst erfolgt ein Überblick, bei dem die Umsetzung der organisationsspezifischen Handlungsempfehlungen zusammengefasst wird: Dabei wird der Umsetzungsgrad insgesamt sowie differenziert nach Dimensionen von Evaluierungspraxis und Akteursgruppen betrachtet. Zudem wird auf förderliche und hemmende Faktoren bei der Umsetzung eingegangen. Anschließend folgt die Gliederung den Dimensionen von Evaluierung: *Ziele von Evaluierungen, Organisatorischer Kontext, Strukturen und Prozesse, Humanressourcen, Finanzielle und Technische Ressourcen*. Dabei werden die Ergebnisse des Umsetzungsmonitorings entlang dieser Dimensionen näher betrachtet und mit übergeordneten Ergebnissen zum Stand

der Evaluierungspraxis zusammengeführt (vgl. Kapitel 3.1 bis 3.6).

Dabei gilt es zu beachten, dass den Aussagen zu Umsetzungsmonitoring und Evaluierungspraxis unterschiedliche Datenquellen zu Grunde liegen: Die Ergebnisse des Umsetzungsmonitorings basieren ausschließlich auf den Aussagen der 17 Organisationen, die an der „Systemprüfung“ beteiligt waren. Diese Informationen wurden im Rahmen vertiefender Interviews sowie mit einem gesonderten Frageblock bei der Online-Befragung erhoben. Demgegenüber sind die Ergebnisse zur Evaluierungspraxis um weitere Informationen ergänzt worden, die im Rahmen der Online-Befragung von 34 EZ-O sowie in den Experteninterviews und der Metaevaluierung gewonnen wurden (vgl. auch Kapitel 2.1.2).

**Abbildung 2: Datenquellen des Umsetzungsmonitorings und der Evaluierungspraxis**

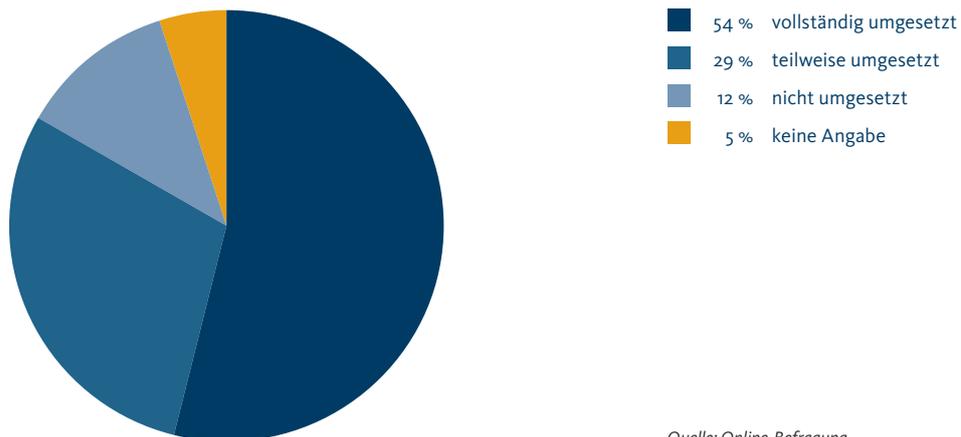


### Umsetzung der Handlungsempfehlungen der „Systemprüfung“ – ein Überblick

Von den 181 Handlungsempfehlungen an die EZ-Organisationen gingen 68 (38 Prozent)<sup>11</sup> an die sechs politischen Stiftungen (POS), 64 (35 Prozent) an die fünf staatlichen Durchführungsorganisationen (DO) und 49 (27 Prozent) an

sechs weitere zivilgesellschaftliche Organisationen (andere ZGO). Die EZ-O stimmten den Handlungsempfehlungen in den Umsetzungsplänen zu 59 Prozent vollständig und zu weiteren 33 Prozent zumindest teilweise zu.

<sup>11</sup> Prozentzahlen sind im Text auf ganze Zahlen gerundet und in Tabellen auf eine Stelle hinter dem Komma.

**Abbildung 3: Umsetzungsgrad der Empfehlungen (gesamt)**

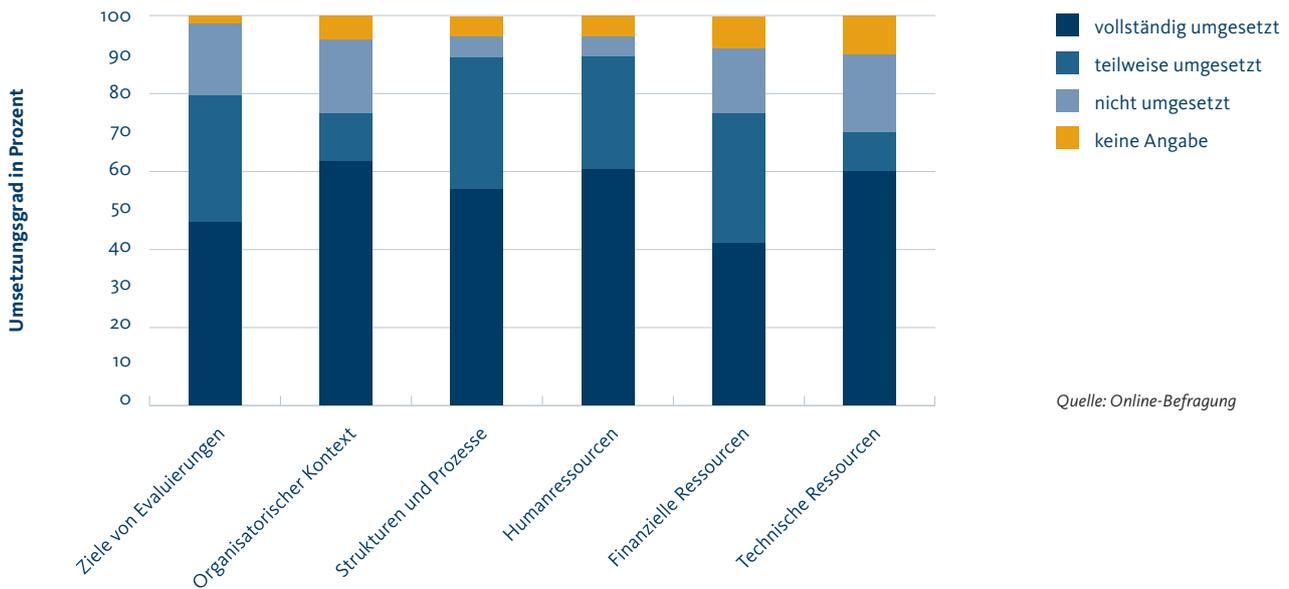
Quelle: Online-Befragung

Diese mehrheitliche Zustimmung spiegelt sich auch darin wider, dass sie überwiegend ganz oder zumindest teilweise umgesetzt wurden: 98 (54 Prozent) wurden laut Angaben der EZ-Organisationen vollständig umgesetzt, 53 (29 Prozent) teilweise und lediglich 21 (12 Prozent) gar nicht (vgl. Abbildung 3). Im Abschlussworkshop sowie in den vertiefenden Interviews begründeten einige Organisationen die Nichtumsetzung von Handlungsempfehlungen mit der Erklärung, dass sie die Empfehlung entweder schon 2008 nicht geteilt hätten oder aber die Relevanz der Empfehlung aufgrund von Entwicklungen innerhalb der letzten Jahre nicht mehr gegeben sei.

Bei Betrachtung der einzelnen Dimensionen (vgl. Abbildung 4) wird in der „Systemprüfung“ mit 31 Prozent (n=56) der Handlungsempfehlungen ein Schwerpunkt auf der Dimension „Strukturen und Prozesse“ deutlich, wovon 55 Prozent vollständig umgesetzt wurden. Des Weiteren betraf mit 27 Prozent (n=49) ein hoher Anteil der Empfehlungen die „Ziele von Evaluierungen“, von denen 47 Prozent vollständig umgesetzt wurden. Insgesamt 21 Prozent (n=38) der Empfehlungen bezogen sich auf die „Humanressourcen“, die mit 60 Prozent

vollständiger Umsetzung einen vergleichsweise hohen Umsetzungsgrad aufweisen. Insgesamt wurden in diesen drei Kategorien nahezu vier Fünftel (79 Prozent) aller Empfehlungen ausgesprochen. Auf die weiteren Kategorien „Organisatorischer Kontext“ (9 Prozent, n=16), „Finanzielle Ressourcen“ (7 Prozent, n=12) und „Technische Ressourcen“ (6 Prozent, n=10) entfielen hingegen deutlich weniger Handlungsempfehlungen.

Abbildung 4: Umsetzungsgrad der Empfehlungen nach Dimensionen

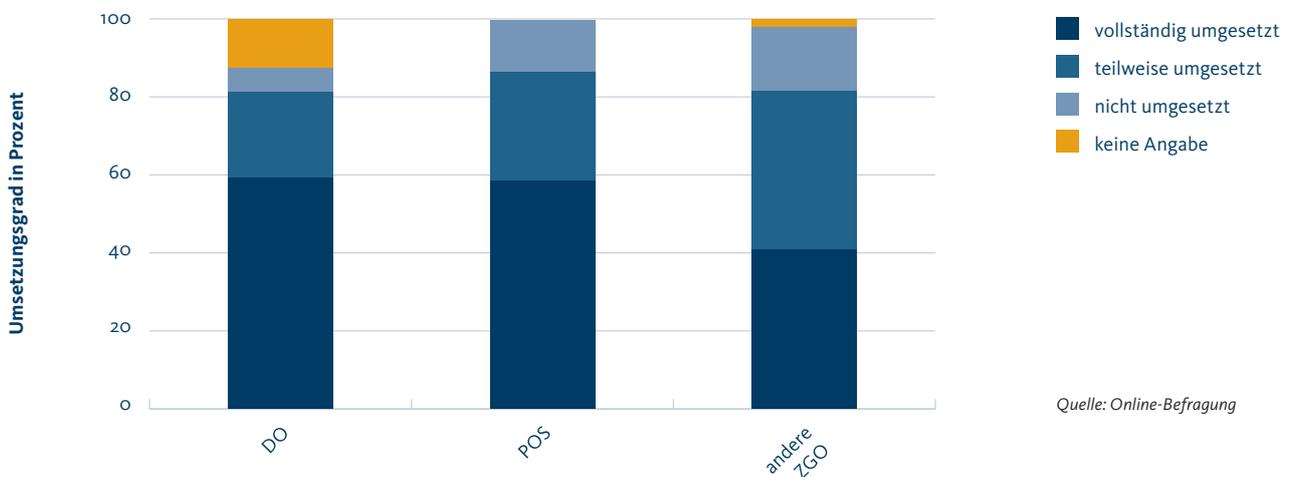


Quelle: Online-Befragung

Den höchsten Umsetzungsgrad weisen die staatlichen Durchführungsorganisationen und die politischen Stiftungen auf. Gemäß deren Angaben wurden jeweils 59 Prozent der

Empfehlungen vollständig umgesetzt. Die verbleibenden zivilgesellschaftlichen Organisationen weisen hierbei einen Wert von 41 Prozent auf (vgl. Abbildung 5).

Abbildung 5: Umsetzungsgrad der Empfehlungen (gesamt; nach Akteursgruppe)

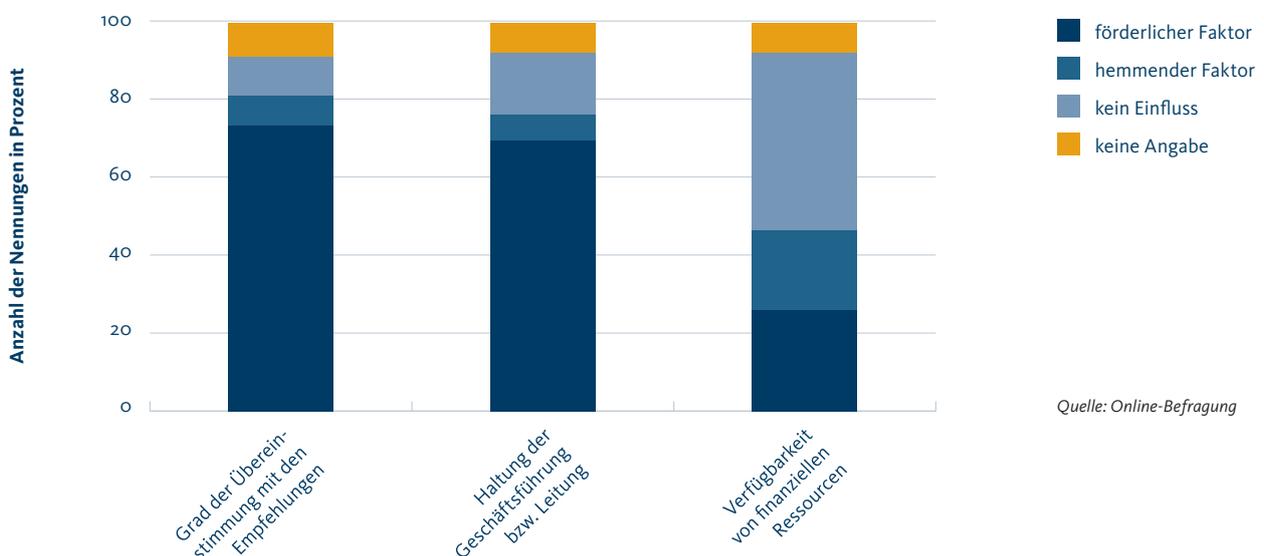


Quelle: Online-Befragung

In der Online-Befragung konnten die Organisationen für jede Handlungsempfehlung angeben, inwiefern eine Übereinstimmung mit der jeweiligen Empfehlung, die Haltung der Geschäftsführung bzw. Leitung sowie die Verfügbarkeit von finanziellen Ressourcen einen förderlichen, hemmenden oder keinen Einfluss auf die Umsetzung hatten (vgl. Abbildung 6). Zudem konnten die befragten Organisationen weitere Gründe für die Umsetzung oder Nichtumsetzung der Handlungsempfehlungen angeben. Der Grad der Übereinstimmung wurde bei rund drei Vierteln (n=133) der Empfehlungen als förderlicher Faktor genannt (Korrelation<sup>12</sup> Übereinstimmung und Umsetzungsgrad:  $r_s=0,413^{***}$ ; n=164). Die Haltung der Geschäftsführung bzw. Leitung wurde bei ähnlich vielen Empfehlungen (n=126) als förderlich eingestuft (Korrelation Haltung Geschäftsführung bzw. Leitung und Umsetzungsgrad:

$r_s=0,517^{***}$ ; n=166). Hingegen wurde als hemmender Faktor für die Umsetzung am häufigsten (n=37) die Verfügbarkeit finanzieller Ressourcen genannt (Korrelation Verfügbarkeit finanzieller Ressourcen und Umsetzungsgrad:  $r_s=0,407^{***}$ ; n=166). Allerdings wurde bei den finanziellen Ressourcen auch mit Abstand am häufigsten angegeben, dass diese keinen Einfluss gehabt hätten. Dieser vermeintliche Widerspruch lässt sich damit erklären, dass die Umsetzung mancher Empfehlungen erhebliche finanzielle Mehrkosten für die EZ-O bedeuteten (z.B. Schaffung neuer Stabsstellen, vermehrte Durchführung von Ex-post- und Wirkungsevaluierungen), während andere Empfehlungen ohne großen finanziellen Mehraufwand umsetzbar waren (z.B. Teilnahme an Netzwerktreffen oder Veröffentlichung von Evaluierungsberichten).

**Abbildung 6: Einflussfaktoren für die Umsetzung der Empfehlungen (gesamt)**



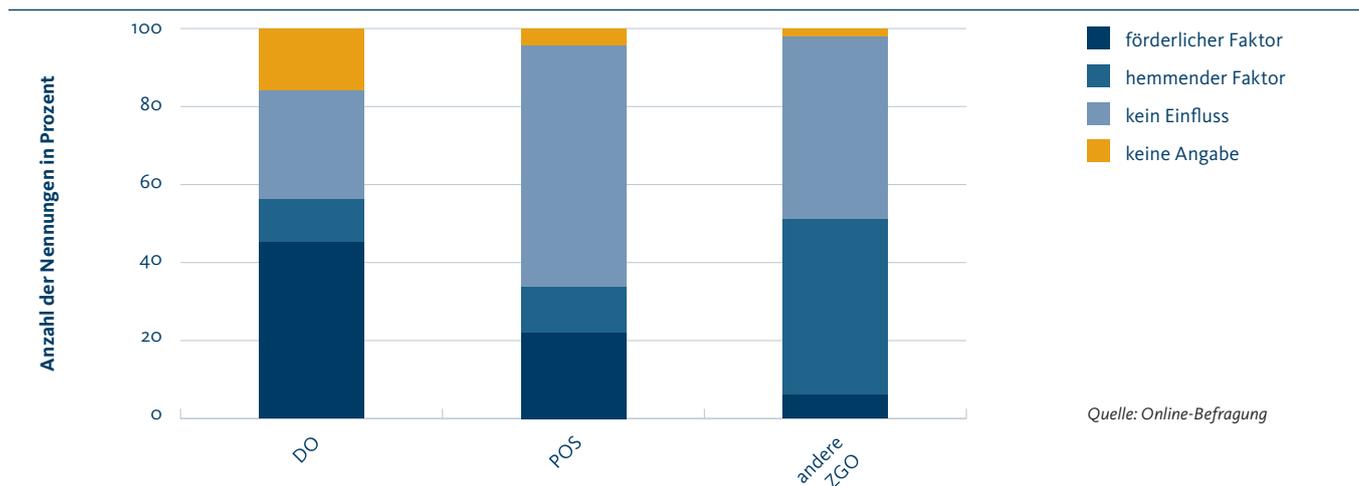
Auffällig ist, dass die Verfügbarkeit von finanziellen Ressourcen (vgl. Abbildung 7) von den ZGO mit Abstand häufiger als hemmender Faktor bei der Umsetzung der Handlungsempfehlungen angeführt wurde (45 Prozent; n=22) als bei den DO (13 Prozent; n=7) und den POS (9 Prozent; n=8). Zugleich ist die Einschätzung von Übereinstimmung und Haltung der

Geschäftsführung bzw. Leitung als förderlicher Faktor bei den ZGO ähnlich hoch wie bei den DO und POS.<sup>13</sup> Eine mögliche Erklärung hierfür stellen die unterschiedlichen Finanzierungsmodalitäten und der finanzielle Spielraum der verschiedenen Akteursgruppen dar (vgl. Kapitel 3.5).

<sup>12</sup> Um den Zusammenhang zwischen der Umsetzung der Handlungsempfehlungen und förderlichen bzw. hemmenden Faktoren (Übereinstimmung mit Empfehlungen, Haltung der Geschäftsführung bzw. Leitung und zur Verfügung stehende Ressourcen) zu messen, wurde Spearmans Rangkorrelationskoeffizient  $r_s$  berechnet: Ab einem Wert von 0,5 kann von einem mittleren positiven Zusammenhang gesprochen werden. Mit  $p < 0,001$  (\*\*\*) sind die Angaben hochsignifikant. Aufgrund des paarweisen Fallausschlusses ist es möglich, dass die in den Korrelationen ausgewiesenen Fallzahlen nicht mit jenen der Handlungsempfehlungen insgesamt übereinstimmen.

<sup>13</sup> Förderlicher Grad der Übereinstimmung mit Handlungsempfehlungen: DO: 69% (n=44); POS: 79% (n=54); andere ZGO: 71% (n=35); Förderliche Haltung der Geschäftsführung bzw. Leitung: DO: 72% (n=46); POS: 65% (n=44); andere ZGO: 74% (n=36).

**Abbildung 7: Verfügbarkeit finanzieller Ressourcen als Einflussfaktor für die Umsetzung der Empfehlungen (nach Akteursgruppe)**



### Exkurs: Förderrichtlinien und Leitlinien

EZ-Organisationen evaluieren, weil sie es müssen und weil sie es wollen. Für ein besseres Verständnis der Evaluierungspflichten der EZ-Organisationen sind die Förderrichtlinien und Leitlinien zur ziel- und wirkungsorientierten Erfolgskontrolle des BMZ für die jeweiligen Akteursgruppen daher von zentraler Bedeutung. In ihnen ist die Häufigkeit der Evaluierung von Programmen und Projekten festgeschrieben: So muss für jede Einzelmaßnahme im Zuwendungsverfahren mindestens eine Zielerreichungskontrolle durchgeführt werden. Die Zielerreichungskontrolle sieht also einen Deckungsgrad von 100 Prozent vor. Die Zielerreichungskontrolle beinhaltet eine Prüfung der Zielerreichung, Wirksamkeit und Wirtschaftlichkeit von finanzwirksamen Maßnahmen z.B. in Berichtsform, aber nicht notwendigerweise eine Evaluierung (vgl. BMZ, 2009a). Die *staatlichen Durchführungsorganisationen* sind darüber hinaus verpflichtet, für eine aussagekräftige Stichprobe von abgeschlossenen und gegebenenfalls laufenden Entwicklungsmaßnahmen eigenständig Evaluierungen in mit dem BMZ abgestimmten Verfahren sowie unter Berücksichtigung der OECD DAC-Kriterien und Standards durchzuführen. Die Evaluierungsberichte von Schluss- und Ex-post-Evaluierungen sind dem BMZ, dem Kooperationspartner sowie der allgemeinen Öffentlichkeit zur Verfügung zu stellen (vgl. BMZ, 2008). Im Förderbereichsverfahren

für die *Kirchen* wird die Zielerreichungskontrolle auf der Ebene der Einzelvorhaben nach Maßgabe BMZ-spezifischer Anforderungen durch Evaluierungen ergänzt. Sie erfolgen auf der Grundlage eines Stichprobenverfahrens mit „angemessenem Deckungsgrad“. Die Zusammenstellung der jährlichen Stichprobe erfolgt unter qualitativen sowie quantitativen Gesichtspunkten, wobei Sektor, Region und Fördervolumen berücksichtigt werden. Bezüglich der *politischen Stiftungen* (POS) und der *Sozialstrukturträger* (SST) wird über die o.g. Zielerreichungskontrolle hinaus in der gemeinsamen Förderrichtlinie von POS und SST aufgeführt, dass die Projekte in „angemessenen Zeitabständen“ zu evaluieren und Berichte und Stellungnahme zu den Empfehlungen einzureichen sind (vgl. BMZ, 2002). In den neuen Förderrichtlinien für POS von 2013 wird spezifiziert: „Innerhalb von drei (dreijährigen) Förderphasen werden alle Vorhaben von einer Evaluierung erfasst. [...]“. Die politischen Stiftungen berichten dem BMZ jährlich über durchgeführte Evaluierungen mittels aussagekräftiger Zusammenfassungen der Evaluierungsberichte (Hintergrund der Evaluierung, wesentliche Feststellungen, Schlussfolgerungen und Empfehlungen); auf Wunsch erhält das BMZ die vollständigen Evaluierungsberichte (vgl. BMZ, 2013). Bei privaten Trägern kann bei mehrjährigen komplexen Projekten oder solchen mit Pilotcharakter eine Evaluierung durch unabhängige Gutachterinnen und Gutachter mitfinanziert werden. Die Ausgaben dafür dürfen 15.000,- € nicht übersteigen.

Es folgt die Darstellung von Umsetzungsmonitoring und Evaluierungspraxis gegliedert nach den sechs Dimensionen von Evaluierungspraxis. Diese Unterkapitel sind wie folgt aufgebaut:

1. Einleitend werden – der Dimension von Evaluierung entsprechend – die übergeordneten, auf Systemebene ausgesprochenen Handlungsempfehlungen<sup>14</sup> dargestellt. Diese veranschaulichen und resümieren die Inhalte der organisationsspezifischen Handlungsempfehlungen. Zudem erfolgt eine zusammenfassende Darstellung der Ergebnisse zum Umsetzungsmonitoring in der jeweiligen Dimension.
2. Anschließend werden die einzelnen inhaltlichen Cluster behandelt (vgl. Kapitel 2.1.2). Dabei werden zunächst die Ergebnisse des Umsetzungsmonitorings dargestellt, die zur deutlichen Abgrenzung kursiv gehalten sind. Im Anschluss werden die Ergebnisse zum Stand der Evaluierungspraxis zusammengefasst.
3. Am Ende der jeweiligen Unterkapitel befindet sich eine Box mit einer Kurzzusammenfassung der Ergebnisse.

### 3.1 Ziele von Evaluierungen

Viele Handlungsempfehlungen aus der „Systemprüfung“ lassen sich schwerpunktmäßig der Dimension „Ziele von Evaluierungen“ zuordnen. Um ein besseres Verständnis dafür zu erhalten, worauf die Handlungsempfehlungen in dieser Dimension abzielen, werden die vier übergeordneten Empfehlungen der „Systemprüfung“ zusammenfassend dargestellt (Borrmann und Stockmann, 2009a: 165ff):

- *Vermeehrt projektübergreifende und Ex-post-Evaluationen durchführen, um das projektübergreifende, konzeptionelle Lernen zu fördern und nach außen glaubwürdig Rechenschaft über Wirksamkeit und Nachhaltigkeit ablegen zu können.*

- *Vermehrte Wirkungsevaluierungen, um der konzeptionellen Ausrichtung der EZ auf Wirkung gerecht zu werden.*
- *Mehr vergleichende Evaluierungen von Programmen einer EZ-Organisation innerhalb eines Sektors oder Landes sowie interinstitutionelle Vergleiche zwischen den Organisationen für konzeptionelles Lernen.*
- *Erhöhte Transparenz zur Förderung interinstitutionellen Lernens. Hierfür reicht es nicht aus, nur Kurzzusammenfassungen öffentlich zugänglich zu machen, sondern (bis auf eng begrenzte schützenswerte Daten) den gesamten Evaluationsbericht.*

Die Empfehlungen in dieser Dimension zielen darauf ab, das Evaluierungsspektrum der EZ-Organisationen hinsichtlich ihrer Zielsetzung und damit

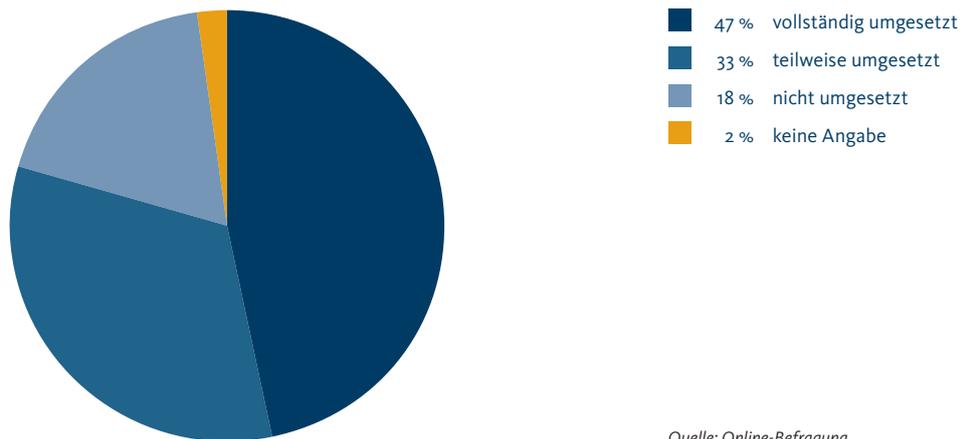
- Evaluierungsformen (z.B. Ex-post-, Wirkungs-, Querschnittsevaluierungen)
- Nutzungsmöglichkeiten (z.B. institutionelles Lernen, Rechenschaftslegung)
- und ihrer Verbreitung an Dritte (Transparenz) zu erweitern.

Die Ergebnisse werden dementsprechend anhand der beiden inhaltlichen Cluster „Art und Ziele der durchgeführten Evaluierungen“ sowie „Nutzung und Transparenz der Evaluierungsergebnisse“ (vgl. Tabelle 1) dargestellt.

Gut ein Viertel (27 Prozent, n=49) aller organisationsspezifischen Handlungsempfehlungen bezog sich auf die Art der durchgeführten Evaluierungen, deren Zielsetzungen sowie die Nutzung und Transparenz der Ergebnisse (vgl. Abbildung 8). Von den 49 Empfehlungen in dieser Dimension wurden 23 und damit knapp die Hälfte vollständig umgesetzt. Weitere 16 wurden teilweise (33 Prozent) sowie 9 (18 Prozent) überhaupt nicht umgesetzt (vgl. Abbildung 8 und Abbildung 9). Der Umsetzungsgrad fällt damit geringer aus als in den meisten anderen Dimensionen.

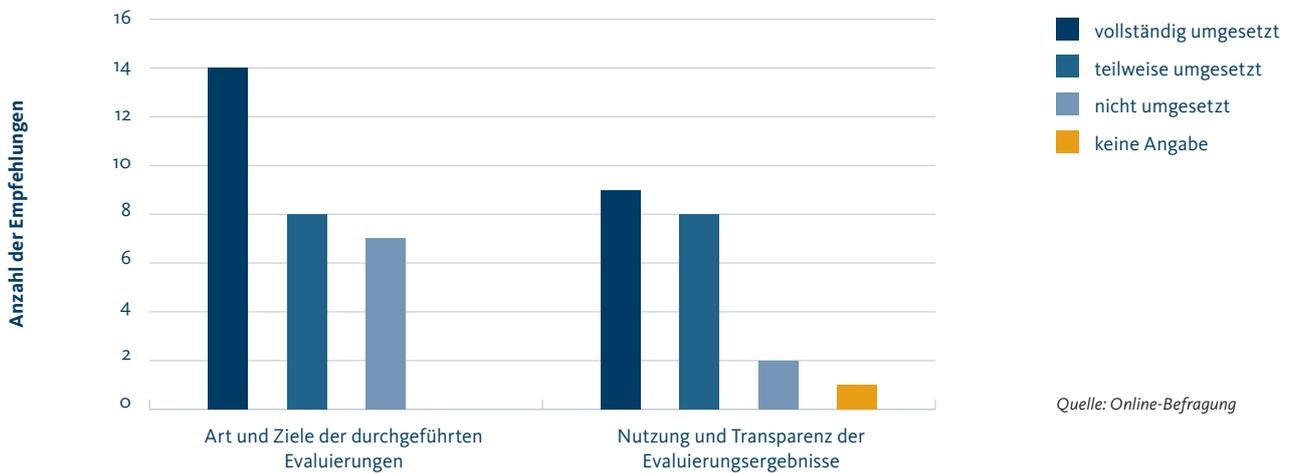
<sup>14</sup> vgl. Borrmann und Stockmann, 2009a: 165ff.

Abbildung 8: Umsetzungsgrad der Empfehlungen (Ziele von Evaluierungen)



Quelle: Online-Befragung

Abbildung 9: Umsetzungsgrad nach inhaltlichen Clustern (Ziele von Evaluierungen)



Quelle: Online-Befragung

Viele der Empfehlungen bezogen sich auf eine Ausweitung der eingesetzten Evaluierungsformen, insbesondere auf eine vermehrte Durchführung von Querschnitts-, Wirkungs- und Ex-post-Evaluierungen. Bei einigen EZ-Organisationen bestanden Zweifel, was die Zweckmäßigkeit dieser Schwerpunktsetzung für die eigene Arbeit betrifft. Dies erklärt einerseits, warum die Handlungsempfehlungen nicht immer befürwortet wurden: Mit 45 Prozent war die vollständige Zustimmung zu den ausgesprochenen Handlungsempfehlungen in den

Umsetzungsplänen im Vergleich zu den anderen Dimensionen am geringsten.<sup>15</sup> Andererseits lässt sich dadurch auch der vergleichsweise hohe Anteil jener Empfehlungen erklären, die nicht bzw. nur zum Teil umgesetzt wurden.

### 3.1.1 Art und Ziele der durchgeführten Evaluierung

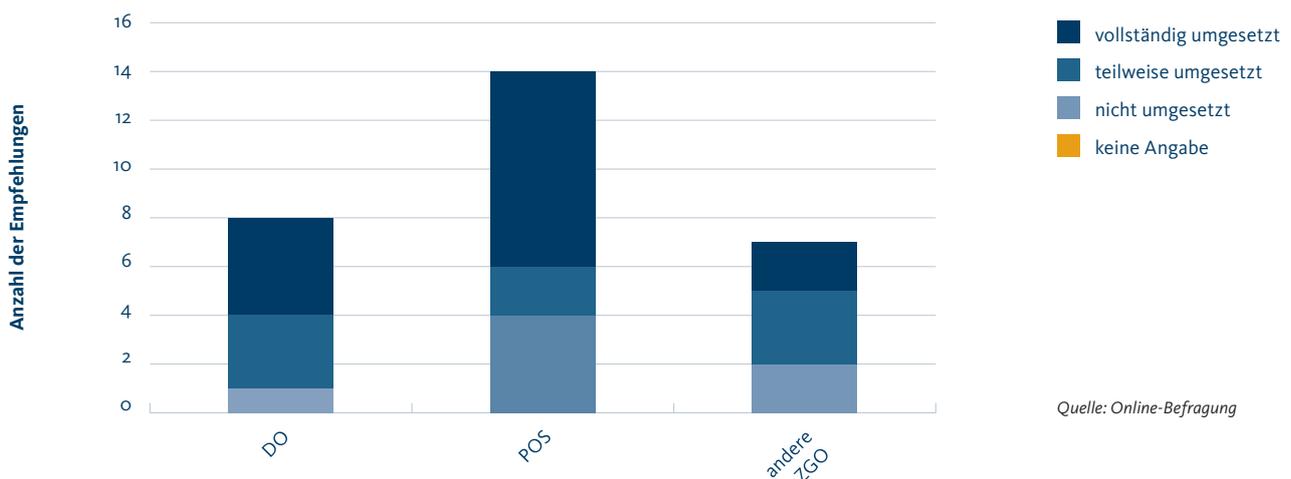
Insgesamt bezogen sich 29 Empfehlungen auf die Diversifizierung der eingesetzten Evaluierungsformen. Die Handlungsempfehlungen zielten insbesondere darauf ab, die Evaluierungspraxis durch

<sup>15</sup> DO: 57 Prozent, POS: 22 Prozent, andere ZGO: 75 Prozent.

Wirkungs- und Ex-post-Evaluierungen<sup>16</sup> zu erweitern sowie mehr Querschnittsevaluierungen durchzuführen. Knapp die Hälfte der Empfehlungen im Cluster „Art und Ziele der durchgeführten Evaluierungen“ richtete sich an die politischen Stiftungen (48 Prozent, n=14; vgl. Abbildung 10). Von diesen wurden 8 (57 Prozent) Handlungsempfehlungen vollständig, 2 (14 Prozent) teilweise und nur 4 (29 Prozent) gar nicht umgesetzt. Die nicht umgesetzten Empfehlungen bezogen sich insbesondere auf die Durchführung von Ex-post-Evaluierungen. In den Interviews begründeten die

Vertreterinnen und Vertreter der politischen Stiftungen die nicht erfolgte Umsetzung der Empfehlungen bzgl. Ex-post-Evaluierung damit, dass die politische Arbeit fortlaufend sei und es in diesem Sinne kein Projektende gäbe und somit keine Bestimmung eines Ex-post-Zeitpunktes möglich sei. Umgesetzt wurden dagegen die als sinnvoller erachteten Empfehlungen zur verstärkten Durchführung von Querschnittsevaluierungen sowie zu Evaluierungen nach Beendigung einzelner Teilabschnitte der Zusammenarbeit.

**Abbildung 10: Umsetzungsgrad nach Akteursgruppe (Art und Ziele der durchgeführten Evaluierungen)**



Weitere 8 (28 Prozent) Empfehlungen, v.a. zum verstärkten Einsatz von Ex-post- und Querschnittsevaluierungen, waren an die Durchführungsorganisationen gerichtet. Davon wurde eine gar nicht, drei weitere teilweise und vier vollständig umgesetzt, da sie laut Angaben in der Online-Befragung nur begrenzt geteilt wurden und aus Sicht der Durchführungsorganisationen nur zum Teil in das bestehende Evaluierungssystem integriert werden konnten. Den zivilgesellschaftlichen Organisationen wurden insgesamt sieben (24 Prozent) Empfehlungen zur Erweiterung von Evaluierungsarten um mehr Ex-post-, Wirkungs- und projektübergreifende bzw. Querschnittsevaluierungen ausgesprochen. Alle sieben Empfehlungen wurden geteilt, aber nur zwei wurden vollständig und drei teilweise umgesetzt. Als hemmende Faktoren für die Umsetzung wurden in der Online-Befragung und in Interviews die Verfügbarkeit von finanziellen Ressourcen (Finanzierung von Wirkungs- und Ex-post-Evaluierungen aus Projektbudgets,

Verwaltungskosten und zweckungebundenen Spenden) sowie begrenzte personelle Kapazitäten genannt. Zur Durchführung von Ex-post-Evaluierungen gab es hinsichtlich deren Komplexität und der Übertragbarkeit ihrer Ergebnisse auch Bedenken.

Querschnittsevaluierungen werden, wie auch in den vertiefenden Interviews betont wurde, in zunehmendem Maße eingesetzt, insbesondere weil sie konzeptionelles Lernen ermöglichen. Allerdings gehen sie mit einem finanziellen Mehraufwand einher, der nicht ohne Weiteres aus Projektbudgets zur Verfügung steht. Projektübergreifende Finanzierungsmodalitäten wären hier eine Lösungsoption, sie existieren aber in der Praxis bisher eher selten.

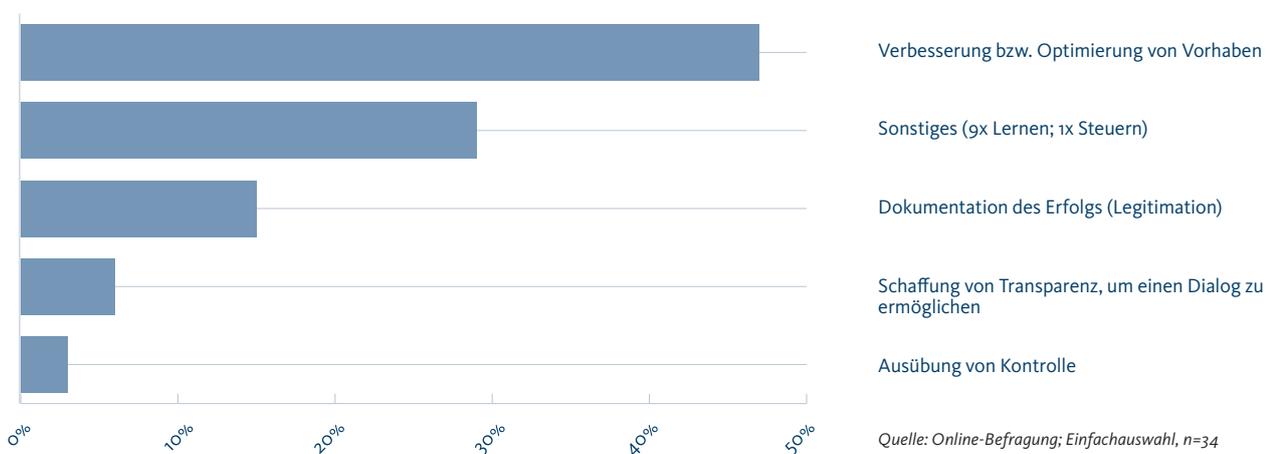
Als primäres Ziel von Evaluierungen nennen EZ-Organisationen in der Online-Befragung die Verbesserung bzw. Optimierung von Vorhaben (47 Prozent). Lernen als Evaluierungsziel

<sup>16</sup> Bzw. „Quasi-ex-post-Evaluierungen“ nach abgrenzbaren Programmphasen bei den POS.

wurde mehrfach explizit in der Kategorie „Sonstiges“ genannt (29 Prozent). Im Vergleich hierzu werden die Dokumentation des Erfolgs für die Legitimation von Entwicklungsmaßnahmen

gegenüber Geldgebern oder der Öffentlichkeit, die Schaffung von Transparenz für Dialog sowie die Ausübung von Kontrolle nicht als vorrangige Ziele gewertet (vgl. Abbildung 11).

**Abbildung 11: Ziele von Evaluierungen**



In den vertiefenden Analysen zeigte sich, dass unterschiedliche Funktionsgruppen in Organisationen unterschiedliche Erwartungen an „Lernen“ aus Evaluierung haben: Operative Einheiten möchten für ihre spezifischen Projekte instrumentelle Lernerfahrungen sammeln, Abteilungsleitungen wünschen sich konzeptionelles Lernen für ihren Zuständigkeitsbereich und Geschäftsführungen möchten ihre strategischen Lerninteressen sowie die Rechenschaftslegung gegenüber dem Geldgeber bedient sehen. Dabei ist die Dokumentation von Erfolgen durchaus von zentraler Bedeutung: Der Zeitpunkt, zu dem eine Evaluierung stattfindet, wird in der Regel durch die Projektförderung bestimmt. Die Ergebnisse dienen der Programm- bzw. Projektsteuerung und werden bei Entscheidungen über die Verlängerung von Maßnahmen bzw. deren Abschluss berücksichtigt. Eine Kontrolle über die Effizienz des Mitteleinsatzes wird von den befragten EZ-Organisationen eher – sofern vorhanden – bei Controlling, Revision oder vergleichbaren Einheiten verortet.

Daneben wurde in den vertiefenden Interviews deutlich, dass Mitarbeitende der EZ-Organisationen Evaluierungen

gegenüber grundsätzlich aufgeschlossen sind. Zudem besteht ein starkes Interesse an mehr übergreifenden Evaluierungen (v.a. Querschnittsevaluierungen), um konzeptionelles Lernen der Gesamtorganisation zu ermöglichen. Allerdings sind Evaluierungssysteme häufig nicht systematisch auf institutionelles Lernen ausgerichtet. Lernerfahrungen aus evaluierten Projekten werden oftmals allein mit den Stakeholdern der betreffenden Projekte/ Programme geteilt. Im Abschlussworkshop wurde angemerkt, dass es sich hierbei durchaus auch um institutionelles Lernen handele.

Darüber hinaus wird der Nutzen von Evaluierungen im Zusammenspiel mit anderen Instrumenten zur Verbesserung von Maßnahmen und Bestimmung der Wirksamkeit gesehen. Im Widerspruch zum Ziel einer erhöhten Wirksamkeit durch Lernprozesse aus Evaluierungen steht, dass selten Wirkungs- und Ex-post-Evaluierungen durchgeführt werden<sup>17</sup> und auch Ergebnisse projektübergreifender Evaluierungen oftmals nicht systematisch in der Organisation verbreitet werden.

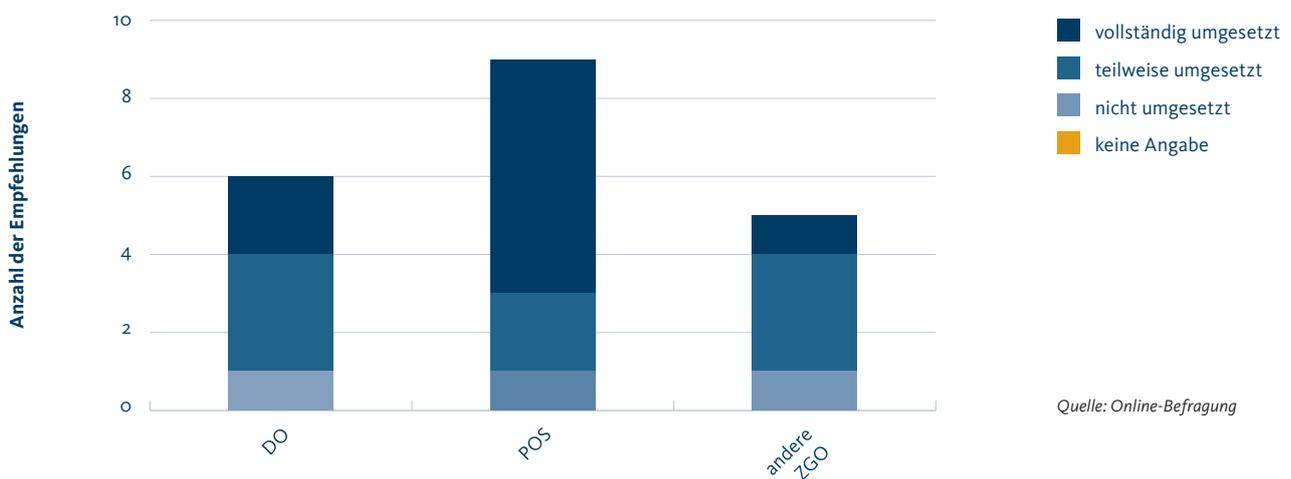
<sup>17</sup> In der Online-Befragung wurde erhoben, wie viele Evaluierungen die befragten EZ-O insgesamt durchführen und wie viele davon z.B. Ex-post- oder Wirkungsevaluierungen waren. Aufgrund der Tatsache, dass zahlreiche Organisationen unter Ex-post-Evaluierungen auch Abschlussevaluierungen eingetragen haben, sind die Zahlen nicht aussagekräftig genug, um hier wiedergegeben zu werden. Die Angaben beziehen sich daher auf Aussagen in den vertiefenden Interviews.

### 3.1.2 Nutzung und Transparenz der Evaluierungsergebnisse

Neben Empfehlungen zu Evaluierungsformen wurden den EZ-Organisationen Hinweise zur verbesserten Nutzung von Evaluierungsergebnissen gegeben, insbesondere im Zusammenhang mit ihrer systematischen Erfassung und Abrufbarkeit zur Förderung des institutionellen Lernens (vgl. Kapitel 3.1.1). Darüber hinaus wurde bei vielen Empfehlungen eine transparentere

Veröffentlichungspolitik der Evaluierungen angeregt. Einerseits, um transparent gegenüber der Öffentlichkeit zu sein, andererseits, um durch das Teilen von Evaluierungsergebnissen auch Lernprozesse über den Kreis der direkten Projektbeteiligten hinaus zu ermöglichen. Insgesamt wurden 9 (45 Prozent) der Handlungsempfehlungen vollständig umgesetzt, 8 (40 Prozent) teilweise und 2 (10 Prozent) gar nicht (vgl. Abbildung 12).<sup>18</sup>

**Abbildung 12: Umsetzungsgrad nach Akteursgruppe (Nutzung und Transparenz der Evaluierungsergebnisse)**



Eine verstärkte Veröffentlichung von Evaluierungsergebnissen und -berichten wird von diversen EZ-Organisationen als Herausforderung angesehen (vertiefende Interviews). Diesbezügliche Empfehlungen an staatliche EZ-Organisationen wurden teilweise erfüllt, wenn z.B. Partnern zumindest Kurzberichte zur Verfügung gestellt wurden, jedoch nicht die Langfassung der Evaluierungsberichte. Langfassungen werden im Allgemeinen äußerst selten veröffentlicht. Politischen Stiftungen wurde insbesondere die Empfehlung ausgesprochen, Evaluierungsergebnisse in der Landessprache der Partnerländer bereitzustellen. Laut der Angaben in den vertiefenden Interviews wurde diese Handlungsempfehlung aufgrund politisch sensibler und brisanter Arbeitskontexte von einer Stiftung gar nicht und von zwei weiteren lediglich teilweise umgesetzt, die weiteren Handlungsempfehlungen wurden hingegen vollständig umgesetzt.

Welche Akteure bzw. Personen Evaluierungsergebnisse nutzen, hängt in hohem Maße vom Typ der durchgeführten

Evaluierungen ab: Angesichts des hohen Anteils von Projektevaluierungen im Portfolio der untersuchten EZ-Organisationen erfolgt die Nutzung der Evaluierungsergebnisse im Wesentlichen durch die direkt beteiligten Projektverantwortlichen in den evaluierten Projekten, meist unter Einbeziehung der Evaluierungsabteilungen bzw. -beauftragten. Vor diesem Hintergrund wird auch die Nutzbarkeit dieser Ergebnisse für Akteure außerhalb der eigenen Organisation und sogar außerhalb des jeweiligen Projektes, kritisch bewertet. Nur 59 Prozent der an der Online-Befragung beteiligten Organisationen geben den vollständigen Evaluierungsbericht an den Partner weiter.

Der Geldgeber (BMZ) ist kein direkter Nutzer der Evaluierungsergebnisse: Während die staatlichen Durchführungsorganisationen dem BMZ vollständige Evaluierungsberichte übermitteln, stellen die POS Evaluierungsergebnisse ausschließlich als Kurzberichte zur Verfügung. Bei Bedarf kann das BMZ vollständige Berichte jedoch anfragen (vgl. Exkurs,

<sup>18</sup> Zu einer Empfehlung (5 Prozent) wurden in der Online-Befragung keine Angaben zur Umsetzung gemacht.

Kapitel 3). Die weiteren befragten zivilgesellschaftlichen Organisationen stellen dem BMZ zu 78 Prozent (n=18) vollständige Berichte zur Verfügung, ansonsten Kurzberichte. In einem Fall erfolgte ein mündliches Debriefing bzw. eine Präsentation. Eine Nutzung der Evaluierungsergebnisse für

Steuerungsentscheidungen durch das BMZ findet im Allgemeinen nicht statt. Tabelle 2 bildet den Deckungsgrad, also den Anteil der evaluierten Projekte und Programme ab, die überwiegend vom BMZ finanziert wurden; Tabelle 3 den Anteil aller evaluierten Projekte und Programme.

**Tabelle 2: Anteil der evaluierten, überwiegend durch das BMZ finanzierten Projekte/ Programme**

	2008	2012
DO	84,0 %	90,0 %
POS	6,1 %	7,7 %
andere ZGO	23,5 %	35,6 %

Quelle: Online-Befragung, arithmetisches Mittel in Prozent für die Jahre 2008, n=27 und 2012, n=29

**Tabelle 3: Anteil der evaluierten Projekte/ Programme (gesamt)**

	2008	2012
DO	80,0 %	88,0 %
POS	6,1 %	5,4 %
andere ZGO	18,9 %	18,8 %

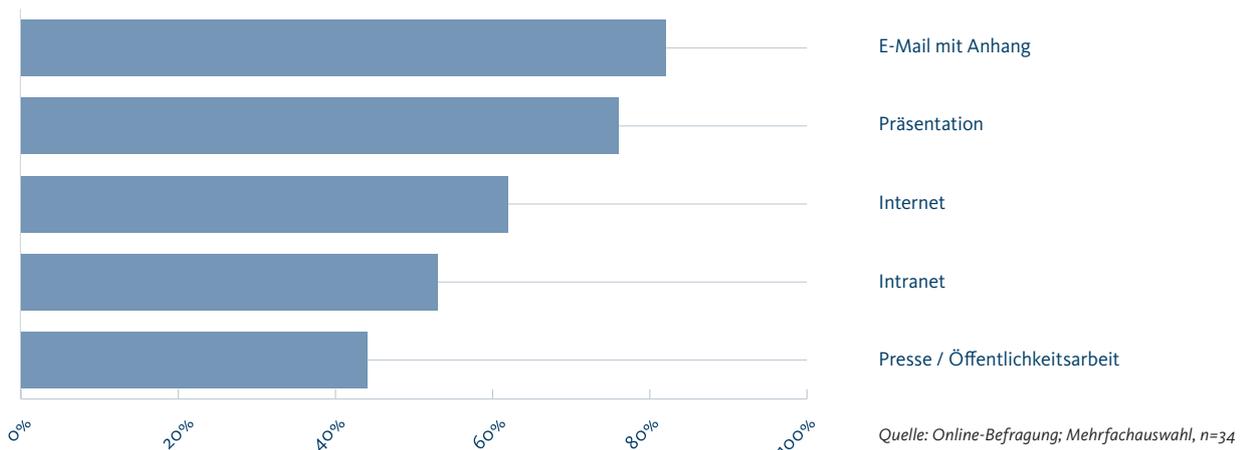
Quelle: Online-Befragung, arithmetisches Mittel in Prozent für die Jahre 2008, n=22 und 2012, n=24

Die allgemeine Öffentlichkeit ist kein Adressat der Evaluierungsergebnisse: Nur drei der an der Online-Befragung teilnehmenden EZ-Organisationen veröffentlichen vollständige Evaluierungsberichte. Häufig werden Kurzzusammenfassungen von Evaluierungen über das Internet bereit gestellt oder es wird auf die Möglichkeit verwiesen, den Gesamtbericht auf Anfrage zu erhalten. Eine Verwendung von Evaluierungsergebnissen für die Presse und Öffentlichkeitsarbeit findet insgesamt nur in Ausnahmefällen statt, beispielsweise, um für ein bestimmtes Förderinstrument Lobbying zu betreiben. Es wird befürchtet, dass vollkommene Transparenz mit dem Ziel des Lernens aus Evaluierung in Widerspruch gerät, da kritische Ergebnisse evtl. weniger thematisiert werden, wenn vollständige Berichte unkommentiert veröffentlicht werden.

Die Verbreitungswege sind ebenso wie die Nutzer selbst vom Typ der Evaluierung abhängig: Während Projektevaluierungen in der Regel nicht, bzw. nicht systematisch, verbreitet werden, bestehen bei projektübergreifenden Evaluierungen oft Verfahren, um die Ergebnisse innerhalb der Organisation zu teilen (z.B. Präsentationen oder Lern-Cafés). Evaluierungsergebnisse werden am häufigsten per E-Mail (82 Prozent) kommuniziert (vgl. Abbildung 13).

Die Erkenntnis, dass die Ergebnisse der Evaluierung in Form eines Debriefings vorgestellt wurden, geht aus knapp zwei Dritteln (64 Prozent) der in der Metaevaluierung untersuchten Berichte hervor (n=45). Etwas seltener (in 58 Prozent der Fälle) wird deutlich, dass beteiligten Akteuren explizit die Möglichkeit gegeben wurde, zu den Ergebnissen der Evaluierung Stellung zu nehmen.

Abbildung 13: Verbreitungsmedien von Evaluierungsergebnissen



### Zentrale Ergebnisse zu der Dimension „Ziele von Evaluierungen“

Das Ziel der Empfehlungen war, das Evaluierungsspektrum der EZ-Organisationen hinsichtlich ihrer Zielsetzung und damit

- Evaluierungsformen (z.B. Ex-post-, Wirkungs-, Querschnittsevaluierungen)
- Nutzungsmöglichkeiten (z.B. institutionelles Lernen, Rechenschaftslegung)
- und ihre Verbreitung an Dritte (Transparenz)

zu erweitern. Die 49 Empfehlungen wurden zu 45 Prozent geteilt und zu 47 Prozent vollständig umgesetzt.

In der heutigen Evaluierungspraxis zeigt sich:

- Bei der Diversifizierung von Evaluierungsformen lässt sich feststellen, dass vermehrt Querschnittsevaluierungen durchgeführt werden. Bei einigen EZ-Organisationen bestehen Zweifel hinsichtlich der Zweckmäßigkeit von Ex-post- und Wirkungsevaluierungen für die eigene Arbeit oder sie verfügen nicht über die nötigen Mittel zu deren Durchführung. Ressourcenintensivere Ex-post-, Wirkungs- und Gemeinschaftsevaluierungen werden daher eher

von den staatlichen Durchführungsorganisationen und weniger von zivilgesellschaftlichen Organisationen durchgeführt.

- Evaluierungen werden den befragten Organisationen zufolge vor allem zur Verbesserung von Vorhaben bzw. zum Lernen genutzt. Dies geschieht v.a. durch Projektbeteiligte zur Programm- oder Projektsteuerung. Es fehlen häufig standardisierte Prozesse zur Nutzung projektübergreifender Evaluierungsergebnisse für institutionelles Lernen. Die Verpflichtung zur Rechenschaftslegung bindet Ressourcen für die Dokumentation von Projekterfolgen, die einige Organisationen gerne für institutionelles Lernen aus projektübergreifenden Evaluierungen nutzen würden.
- Eine grundsätzliche Veröffentlichung von vollständigen Evaluierungsberichten ist nach wie vor sehr selten. Es werden vor allem Kurzberichte veröffentlicht und zum Teil auf Anfrage der Gesamtbericht zur Verfügung gestellt. Unter anderem halten berechnete Schutzinteressen gegenüber Partnern und Sorge um Spendenaufkommen Organisationen davon ab, Evaluierungsergebnisse zu veröffentlichen. Darüber hinaus wird befürchtet, dass vollkommene Transparenz mit dem Ziel des Lernens aus Evaluierung in Widerspruch gerät, da kritische Ergebnisse evtl. weniger thematisiert werden, wenn vollständige Berichte unkommentiert veröffentlicht werden.

### 3.2 Organisatorischer Kontext

In der „Systemprüfung“ wurden Empfehlungen ausgesprochen, die sich auf den organisatorischen Kontext von Evaluierungen und den Austausch zwischen Evaluierungseinheiten verschiedener EZ-Organisationen bezogen. Auf übergeordneter Ebene wurde in dieser Dimension eine systemweite Handlungsempfehlung formuliert (Borrmann und Stockmann, 2009a: 165ff):

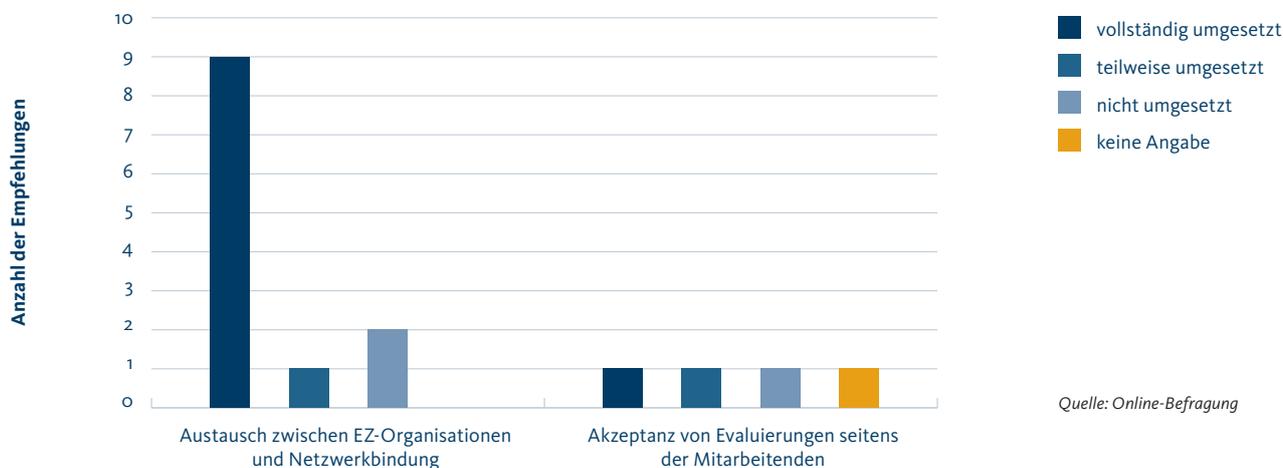
- *Der Vernetzungsgrad der deutschen EZ-Organisationen im Bereich Evaluation untereinander sowie mit internationalen Organisationen und der Wissenschaft hat in den letzten Jahren*

*deutlich zugenommen, ist jedoch noch verbesserungswürdig. Vor allem das Erkenntnispotenzial der Wissenschaft sollte von den EZ-Organisationen besser genutzt werden.*

Das Ziel der Empfehlung in der Dimension organisatorischer Kontext war es, den

- Austausch und Vernetzungsgrad zwischen EZ-Organisationen sowie zwischen diesen und der Wissenschaft zu stärken sowie
- Verständnis bzw. Akzeptanz von Mitarbeitenden bezüglich Evaluierung zu fördern (vgl. Abbildung 14).

Abbildung 14: Umsetzungsgrad nach inhaltlichen Clustern (Organisatorischer Kontext)

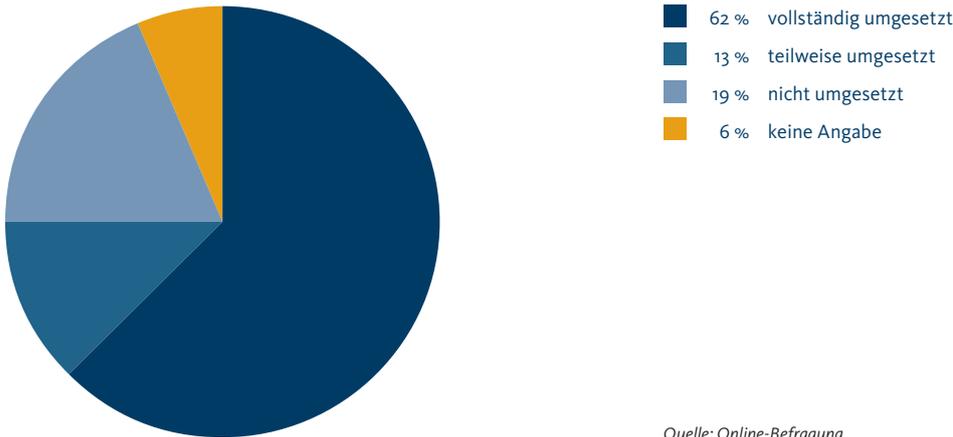


Bei 16 aller 181 organisationsspezifischen Empfehlungen (9 Prozent) wurden den EZ-Organisationen Hinweise zur besseren Einbindung von Mitarbeitenden in den Evaluierungsprozess, zum verstärkten Austausch mit anderen EZ-Organisationen sowie der Wissenschaft gegeben. Die Mehrheit dieser Empfehlungen (69 Prozent) wurde von den EZ-Organisationen

geteilt<sup>19</sup> (vgl. Abbildung 15). Von 16 Empfehlungen wurden 10 (62 Prozent) vollständig umgesetzt. Verglichen mit den anderen Dimensionen, ist dies der höchste Anteil vollständig umgesetzter Empfehlungen. Weitere 2 (13 Prozent) Empfehlungen wurden teilweise und 3 (19 Prozent) gar nicht umgesetzt. Zu einer weiteren Empfehlung liegt keine Angabe vor.

<sup>19</sup> DO: 75 Prozent, POS: 50 Prozent, andere ZGO: 75 Prozent.

Abbildung 15: Umsetzungsgrad der Empfehlungen (Organisatorischer Kontext)



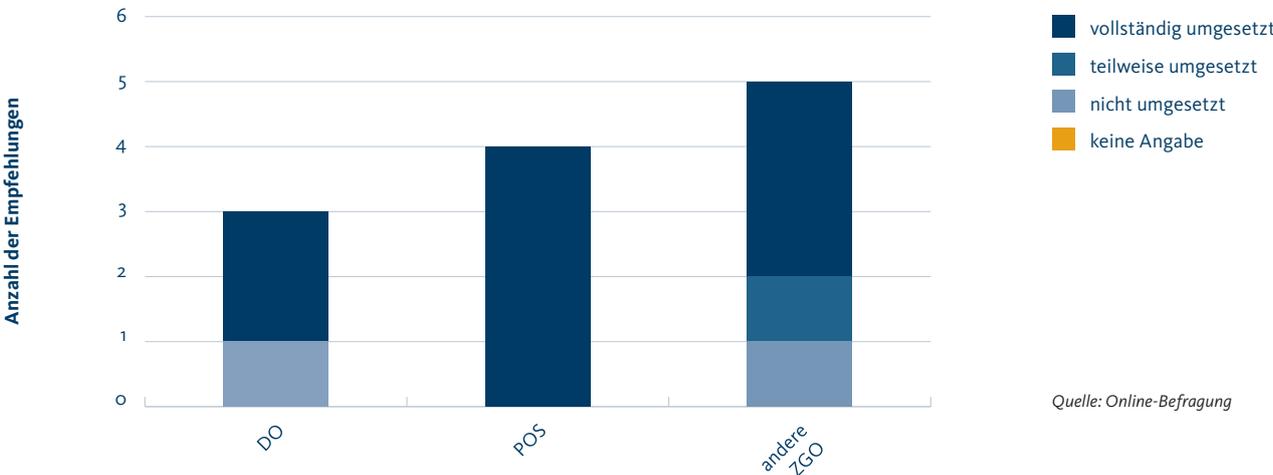
Quelle: Online-Befragung

**3.2.1 Austausch zwischen EZ-Organisationen und Netzwerkbindung**

Insbesondere die Empfehlungen, die der Vernetzung und dem Austausch mit anderen EZ-Organisationen dienen sollen, weisen mit 75 Prozent (n=9) vollständiger und 8 Prozent (n=1) teilweiser Umsetzung einen hohen Erfüllungsgrad auf. Lediglich 2 (17 Prozent)

Empfehlungen wurden gar nicht umgesetzt (vgl. Abbildung 16). Insgesamt herrschte unter den EZ-Organisationen Einvernehmen darüber, dass der Austausch in Evaluierungsnetzwerken einen Mehrwert bietet und zur Erhöhung der Evaluierungsqualität beitragen kann.

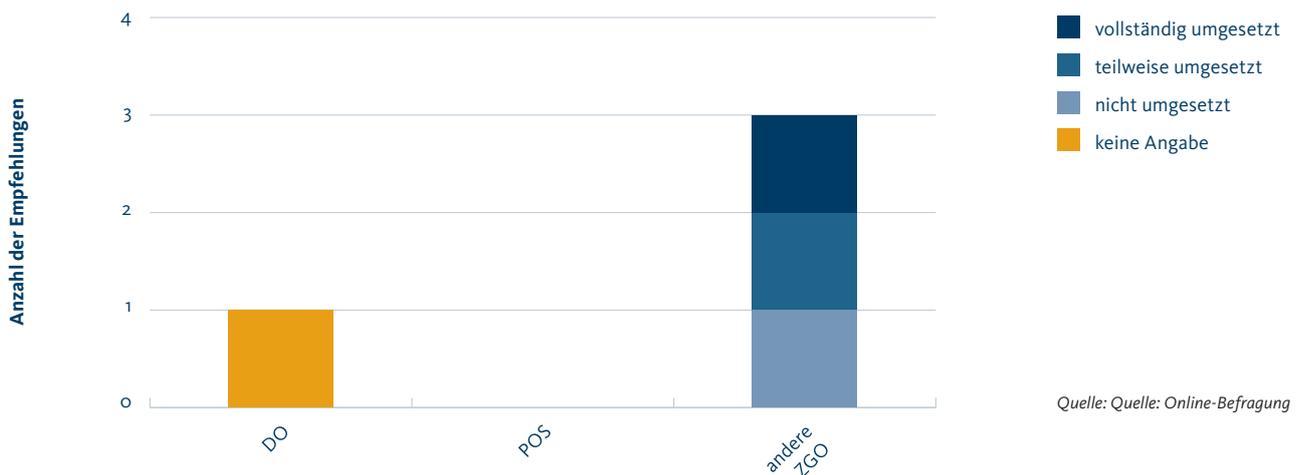
Abbildung 16: Umsetzungsgrad nach Akteursgruppe (Austausch zwischen EZ-Organisationen und Netzwerkbindung)



Quelle: Online-Befragung



Abbildung 18: Umsetzungsgrad nach Akteursgruppe (Akzeptanz von Evaluierungen seitens der Mitarbeitenden)



Grundsätzlich wurde durch die vertiefenden Interviews ersichtlich, dass unter den Mitarbeitenden der EZ-Organisationen eine hohe Nachfrage und Aufgeschlossenheit gegenüber Evaluierungen besteht. Die Einbindung von Mitarbeitenden in den Evaluierungsprozess ist zunehmend systematisiert, wobei viele Organisationen erhebliche Anstrengungen unternehmen, um Mitarbeitende in M&E zu schulen (vgl. Kapitel 3.4).

### Zentrale Ergebnisse zu der Dimension „Organisatorischer Kontext“

Das Ziel der Empfehlungen in dieser Dimension war,

- den Austausch und Vernetzungsgrad von EZ-Organisationen untereinander sowie zwischen ihnen und der Wissenschaft zu stärken sowie
- das Verständnis bzw. die Akzeptanz von Mitarbeitenden bezüglich Evaluierung zu fördern.

Die 16 Empfehlungen wurden zu 69 Prozent geteilt und zu 62 Prozent vollständig umgesetzt.

Der Vernetzungsgrad der deutschen EZ-Organisationen im Bereich Evaluierung hat in den letzten Jahren zugenommen. Mitarbeitende sind Evaluierungen gegenüber aufgeschlossen.

## 3.3 Strukturen und Prozesse

Viele übergeordnete Empfehlungen, die in der „Systemprüfung“ ausgesprochen wurden, beziehen sich auf die Dimension „Strukturen und Prozesse“ (Borrmann und Stockmann, 2009a: 165ff; hier gekürzt und umformuliert zusammengestellt):

- *Unabhängige Evaluationseinheiten oder -beauftragte sollten Evaluierungsfunktionen wahrnehmen und wesentlich besser ausgestattete Budgets erhalten, über die sie eigenständig verfügen.*
- *Zur Stärkung der Unabhängigkeit der Evaluationseinheiten und -beauftragten sollte die oberste Leitungsebene auf ihr Weisungsrecht verzichten.*
- *Die Evaluationssysteme sollten bestmöglich mit dem Qualitäts- und Wissensmanagement verbunden werden.*
- *Evaluationseinheiten oder Evaluationsbeauftragte sollten für die Qualitätssicherung aller Evaluationen zuständig sein und diese effektiv vornehmen.*
- *Die EZ-Organisationen sollten bestehende Lücken in ihren programmatischen, institutionellen, organisatorischen und methodischen Grundlagen ihrer Evaluationssysteme schließen und an (inter-)nationale Standards und Vorgaben anpassen.*
- *Die OECD DAC-Prinzipien sollten weiter handlungsleitend für die Ausgestaltung von Evaluationssystemen sein, ebenso wie die*

OECD DAC-Qualitätsstandards für Evaluationssysteme (DAC 2006). Das BMZ sollte diesen Prozess durch einen regelmäßigen Dialog und die Vereinbarung und periodische Überprüfung von Vorgaben aktiv fördern.

- In der Planungsphase neuer EZ Maßnahmen sollten spätere Evaluationen obligatorisch und explizit vorgesehen und budgetiert sein.
- Evaluationsaufträge sollten möglichst öffentlich anhand transparenter Kriterien ausgeschrieben werden, um die Abhängigkeit von wenigen Gutachtern zu vermeiden und den Qualitätswettbewerb zu stärken.
- Die Unabhängigkeit der Gutachter ist systematisch zu überprüfen.
- Unabhängige Gutachter erhalten die uneingeschränkte Berichtsverantwortung.
- Ein Inception-Report ist vorzusehen, inklusive einer Ausführung der methodischen Konzeption.
- Eine Management Response seitens des Auftraggebers und ein möglichst unabhängiges Umsetzungsmonitoring sollten durchgeführt werden und Evaluierungsergebnisse in die Neukonzeption von Vorhaben einfließen.
- Um der Paris Declaration und anderen internationalen Abkommen zu entsprechen, sollten gemeinsame Evaluationen der deutschen EZ-Organisationen untereinander sowie mit anderen internationalen Gebern vermehrt durchgeführt werden. Hierzu sollte eine klare institutionelle Regelung ausgearbeitet werden.
- Partnerorganisationen sollten in weitaus größerem Umfang als bisher in die Evaluationsplanung, -durchführung und -auswertung mit einbezogen werden.
- Aufklärung und Sensibilisierung über den Nutzen von Evaluation durch die aktive Einbindung der Partnerorganisation in den Evaluationsprozess sowie die Beteiligung lokaler Gutachter.
- EZ-Organisationen, die die ownership ihrer Partner besonders betonen, sollten ihr eigenes Interesse an Evaluation und ihre Verantwortung hierfür nicht zurückstellen. Im Gegenteil, sie sollten unabhängige Evaluationen aktiv initiieren und gemeinsam mit ihren Partnern planen und durchführen sowie ihre Ergebnisse nutzen.

Die Empfehlungen in der Dimension „Strukturen und Prozesse“ zeigen Verbesserungspotentiale zu folgenden Punkten auf:

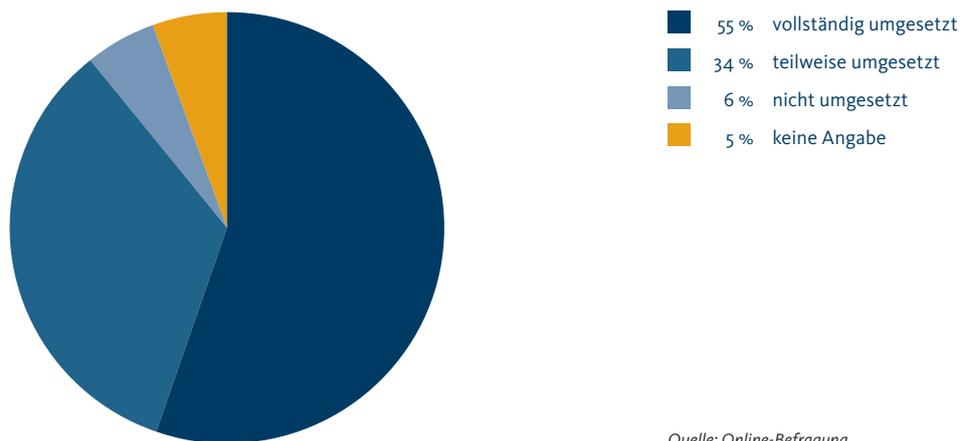
- Stellung der Evaluierungseinheit innerhalb der Institution
- institutionalisierte Evaluierungsgrundlagen
- festgelegte Prozesse zur Evaluierungspraxis und
- Einbindung von Partnern in den Evaluierungsprozess

Einen Schwerpunkt der Handlungsempfehlungen aus der „Systemprüfung“ bildet mit 56 Empfehlungen (31 Prozent) die Kategorie „Strukturen und Prozesse“. Die absoluten Werte zeigen, dass mit 31 vollständig sowie weiteren 19 zum Teil umgesetzten Empfehlungen, die meisten Empfehlungen in dieser Dimension umgesetzt wurden (vgl. Abbildung 19). Von den entsprechenden EZ-O wurden die Empfehlungen in dieser Dimension zu 64 Prozent geteilt.<sup>21</sup> Gemessen an der Anzahl der ausgesprochenen Empfehlungen, ist der Umsetzungsgrad mit 55 Prozent durchschnittlich.<sup>22</sup> Innerhalb der verschiedenen Cluster von Empfehlungen lassen sich große Unterschiede in der Umsetzung der Empfehlungen ausmachen: Während eine hohe Bereitschaft dazu bestand, klarere Strukturen und Abläufe von Evaluierungen zu etablieren, wurden Empfehlungen, die sich auf die stärkere Einbindung von Partnern und die Stellung der Evaluierungseinheit bezogen, häufiger nur teilweise umgesetzt (vgl. Abbildung 20).

<sup>21</sup> DO: 68 Prozent, POS: 56 Prozent, andere ZGO: 69 Prozent.

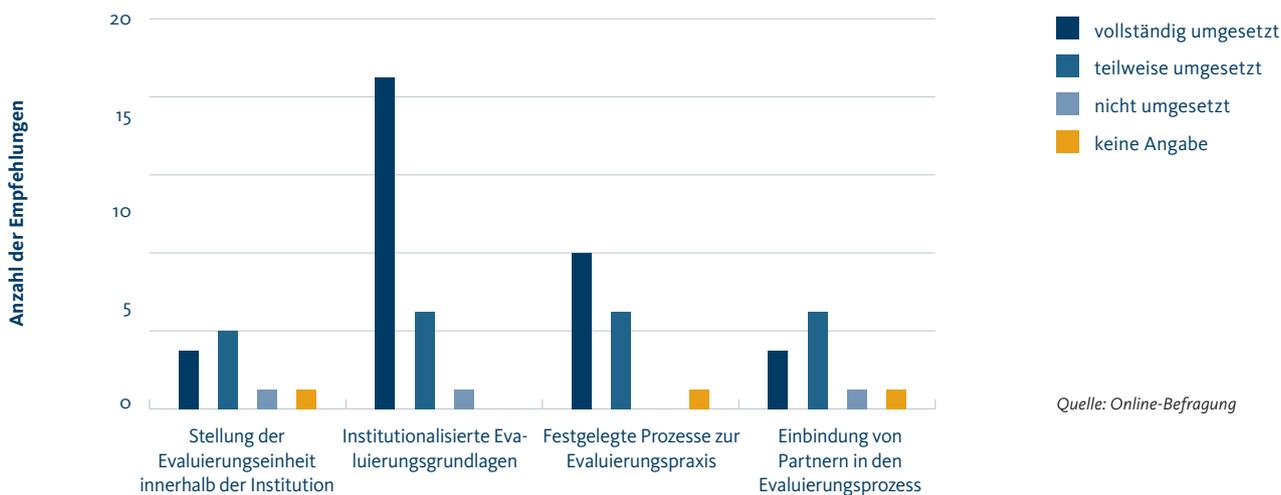
<sup>22</sup> Zu der Umsetzung von drei Handlungsempfehlungen wurden keine Angaben gemacht.

Abbildung 19: Umsetzungsgrad der Empfehlungen (Strukturen und Prozesse)



Quelle: Online-Befragung

Abbildung 20: Umsetzungsgrad nach inhaltlichen Clustern (Strukturen und Prozesse)



Quelle: Online-Befragung

### 3.3.1 Stellung der Evaluierungseinheit innerhalb der Organisation

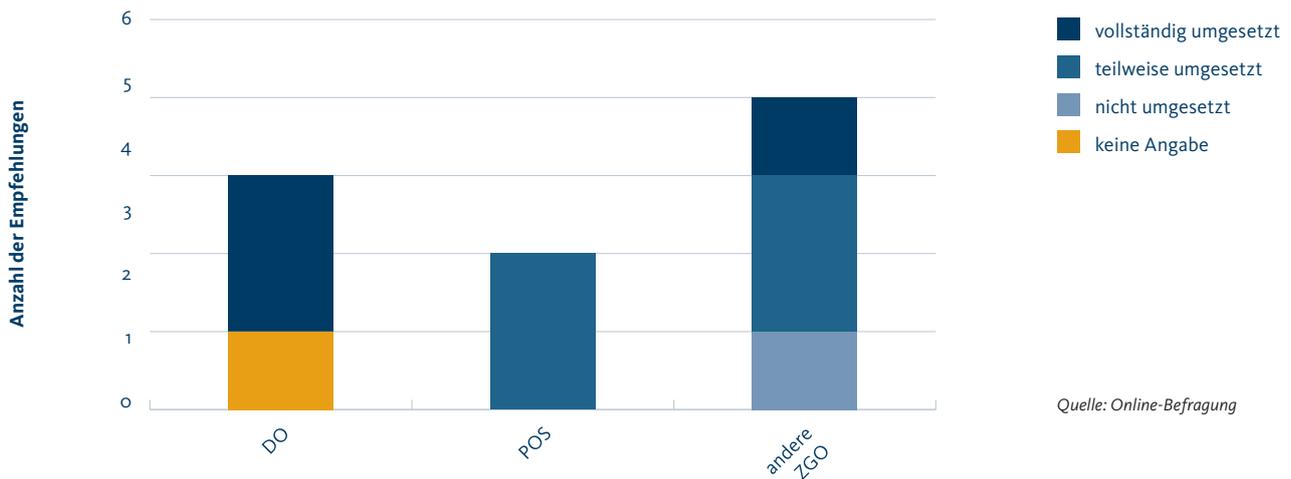
*In Bezug auf das thematische Cluster zur „Stellung der Evaluierungseinheit innerhalb der EZ-Organisationen“ wurden neun Empfehlungen in der „Systemprüfung“ ausgesprochen. Diese verteilten sich auf alle Akteursgruppen zu ähnlichen Anteilen. 3 Empfehlungen (33 Prozent) wurden vollständig, 4 (44 Prozent) teilweise umgesetzt. Eine weitere Handlungsempfehlung wurde nicht umgesetzt (vgl. Abbildung 21).<sup>23</sup> Als hemmende Faktoren wurden in der Online-Befragung mangelnde finanzielle Ressourcen sowie die Haltung der Geschäftsführung bzw. Leitung zu der Empfehlung genannt. In den vertiefenden Interviews wurde insgesamt bezweifelt, inwieweit eine interne Einheit einer Organisation als unabhängig gelten kann bzw. Unabhängigkeit erreichen kann. Zudem stellt sich einigen Evaluierungsverantwortlichen die Frage, ob zusätzliche Unabhängigkeit ihrer Arbeit nicht zu Lasten der Relevanz und Nutzung der Evaluierungsergebnisse geschehen würde.*

Die institutionelle Verortung der Evaluierungseinheiten in den Organisationen fällt sehr unterschiedlich aus. Das Spektrum reicht von eigenständigen Referaten und Stabsstellen bis hin zu definierten Stellenanteilen von Mitarbeitenden in operativen Bereichen. Die Qualitätssicherung von Evaluierungen ist daher in vielen Organisationen nicht in einer zentralen Evaluierungseinheit verortet<sup>24</sup>, sondern liegt in der dezentralen Evaluierungsverantwortung von operativen Bereichen oder Partnern:

Während beispielsweise in manchen Organisationen mit hunderten von Einzelprojekten die Partner die Evaluierungen verantworten, sind bei den großen Regionalevaluierungen der POS die Evaluierungseinheiten dafür zuständig. Die Mitarbeitenden in allen Organisationen betonten in den vertiefenden Interviews die Unabhängigkeit der Evaluierungen, die sie durchführen bzw. in Auftrag geben. Unabhängigkeit bezieht sich hierbei primär auf die wissenschaftliche Unabhängigkeit bezüglich der Ergebnisse. Es besteht keine völlige Unabhängigkeit bei der Bestimmung des Evaluierungsgegenstands durch die Evaluierungseinheit, sondern vielmehr ein Recht auf Mitentscheidung. Auf diese Weise soll sichergestellt werden, dass die gewählten Themen eine größtmögliche Relevanz für die Organisation aufweisen. Zudem dienen die Ergebnisse laut den Mitarbeitenden der Evaluierungseinheiten und der operativen Bereiche eher dem Ziel des instrumentellen Lernens (vgl. Kapitel 3.1) und der besseren Projekt- oder Programmsteuerung und weniger der Kontrolle oder Rechenschaftslegung. Die Rolle der Evaluierungseinheiten geht dabei über das Management der eigenen Evaluierungen hinaus und umfasst oft auch die Qualitätssicherung dezentraler (von operativen Bereichen oder auch Partnern beauftragten oder durchgeführten) Evaluierungen sowie deren Unterstützung im Gesamtprozess. Dies gilt z.B. bei der Auswahl von Gutachterinnen und Gutachtern, der Erstellung von ToRs oder der Berücksichtigung von Standards und Kriterien.

<sup>23</sup> Zu einer Empfehlung wurden in der Online-Befragung keine Angaben zur Umsetzung gemacht.

<sup>24</sup> Von den befragten Organisationen (n=34) geben 17 Organisationen (4 DO, 1 POS und 12 andere ZGO) an, keine eigenständige Einheit zu haben, die sich ausschließlich mit Evaluierung, Monitoring und Qualitätssicherung beschäftigt.

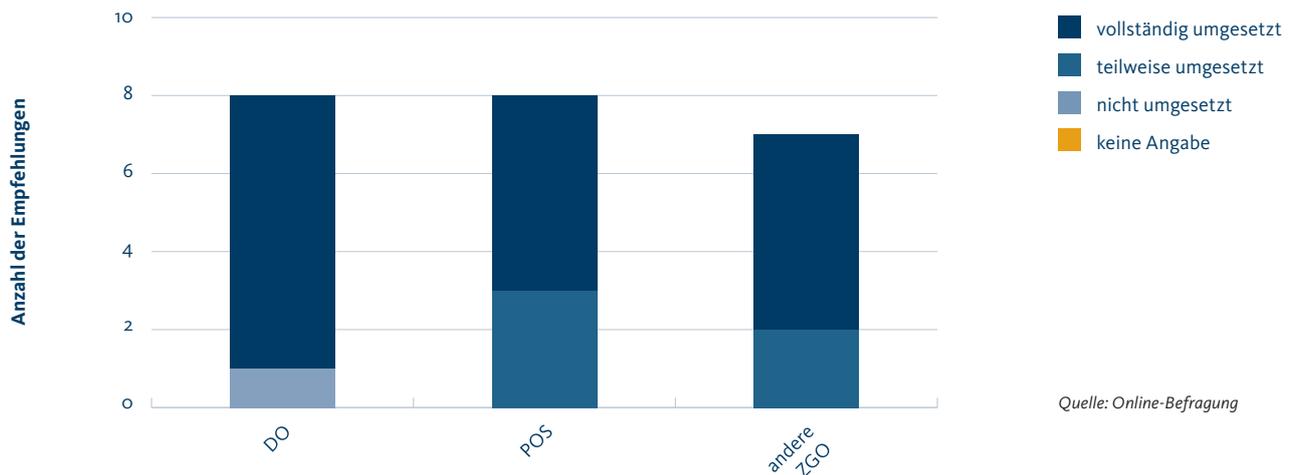
**Abbildung 21: Umsetzungsgrad nach Akteursgruppe (Stellung der Evaluierungseinheit innerhalb der Organisation)**

Dementsprechend gibt es auch große Unterschiede, in welchem Maße operative Einheiten in die Durchführung von Evaluierungen eingebunden sind. Evaluierungseinheiten und -beauftragte verstehen sich nicht als Kontrollinstanz, sondern als bedarfsorientierte Dienstleistende. Daher werden nur äußerst selten Evaluierungen ausschließlich durch die Evaluierungseinheiten alleine durchgeführt.

### 3.3.2 Institutionalisierte Evaluierungsgrundlagen

Die OECD DAC-Prinzipien und Qualitätsstandards sind Leitlinien, die EZ-Organisationen zur Institutionalisierung von Evaluierungsgrundlagen zur Verfügung stehen. Damit wird zum einen die Vergleichbarkeit von Evaluierungen erhöht und zum anderen sichergestellt, dass zentrale (Qualitäts-)Aspekte in Evaluierungen systematisch berücksichtigt werden. Grundsätzlich konnte in der Auswertung festgestellt werden, dass Empfehlungen in diesem

Bereich mehrheitlich vollständig umgesetzt wurden (74 Prozent;  $n=17$ ; vgl. Abbildung 22). Die Empfehlungen richteten sich an alle Akteursgruppen und bezogen sich inhaltlich vornehmlich darauf, die Evaluierungspraxis stärker an die OECD DAC-Prinzipien heranzuführen sowie die Erarbeitung einer Evaluierungsrichtlinie im Sinne der OECD DAC-Prinzipien weiter zu fördern (vgl. Kapitel 3.3.3). Als hemmende Faktoren bei der Umsetzung wurden in der Online-Befragung personelle Engpässe genannt sowie geringe Möglichkeiten zur Einflussnahme auf Partner, Gutachterinnen und Gutachter, um der Empfehlung nachzugehen und die Evaluierungsqualität zu erhöhen.

**Abbildung 22: Umsetzungsgrad nach Akteursgruppe (Institutionalisierte Evaluierungsgrundlagen)**

In den vertiefenden Interviews wurde deutlich, dass EZ-Organisationen Evaluierungen als Teil ihres professionellen Selbstverständnisses verstehen. Dabei orientieren sie sich zunehmend an etablierten nationalen und internationalen Standards (DeGEval oder OECD DAC) und verankern deren Berücksichtigung in organisationsinternen Richtlinien. Manche Akteursgruppen wie die politischen Stiftungen weichen bewusst von internationalen Standards ab. So werden zum Beispiel Schutzinteressen von Partnern über einen transparenten Umgang mit Evaluationsergebnissen gestellt (vgl. Kapitel 3.1). Institutionalisierte Evaluierungsrichtlinien der EZ-O orientieren sich immer auch an den Förderrichtlinien des BMZ, wobei nicht in allen Organisationen Evaluierungsrichtlinien eine Pflicht zu evaluieren beinhalten oder einen Zyklus festlegen. Zudem hat das BMZ viele Rahmenbedingungen der EZ-Evaluierungspraxis weiterentwickelt: Dazu zählen z.B. neue Leitlinien zur ziel- und wirkungsorientierten Erfolgskontrolle, die BMZ-Handreichungen zur Erfolgskontrolle, die BMZ-Grundsätze der Evaluierung bei Zuwendungen zur Projektförderung<sup>25</sup> sowie die Arbeit in der AG „Evaluierung aus einem Guss“ und andere Austauschformate mit den EZ Evaluierungseinheiten. Darüber hinaus ist die OECD DAC-Terminologie (Wirkungslogik) in der Wirkungsmatrix bei Programmvorschlägen und Berichten ebenso obligatorisch wie die Bearbeitung der OECD DAC-Kriterien (ergänzt durch das für die Durchführungsorganisationen verpflichtende BMZ-Evaluierungskriterium "Kohärenz,

Komplementarität und Koordination" sowie die Bearbeitung von Querschnittsthemen) in Evaluierungen.

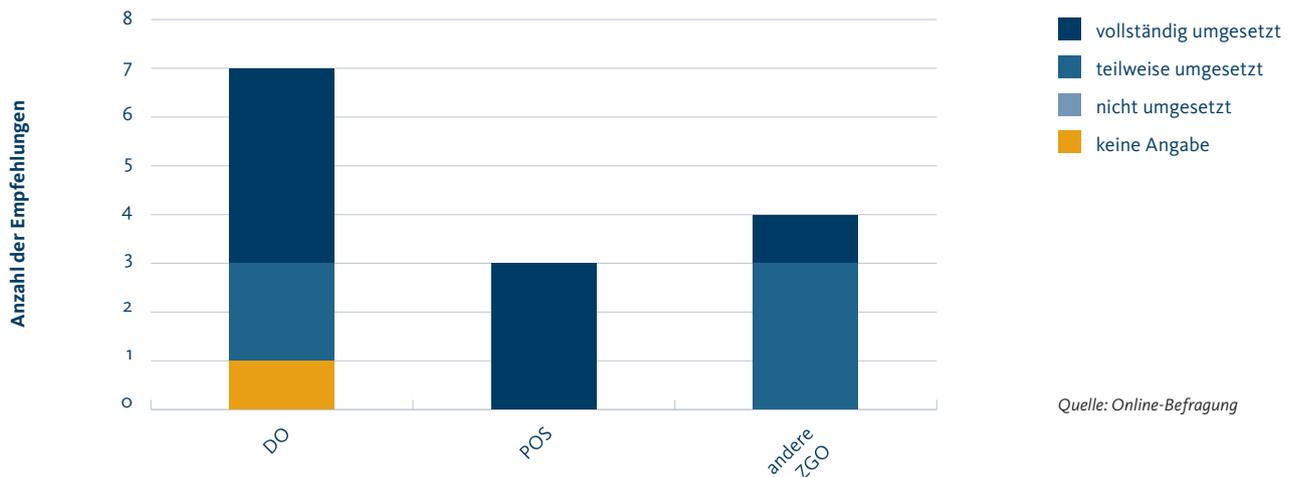
Im Rahmen der Metaevaluierung wurde festgestellt, dass gut ein Drittel der insgesamt 45 Berichte eine vollständige Darstellung der Wirkungslogik der untersuchten Entwicklungsmaßnahmen (nach Inputs, Outputs, Outcomes und Impacts) enthielt. Die Mehrheit der betrachteten Evaluierungen berücksichtigte die OECD DAC-Kriterien (in 59 Prozent der Fälle). Das zusätzliche Kriterium „Kohärenz, Komplementarität und Koordination“ des BMZ wurde hingegen nur sehr selten (bei 8 Prozent) behandelt, wobei es für die bilaterale staatliche EZ, aber nicht für ZGO verpflichtend ist. Zudem wurden bei zwei Fünftel der Evaluierungen (40 Prozent) Querschnittsthemen aufgegriffen (insbesondere Gender).

### 3.3.3 Festgelegte Prozesse zur Evaluierungspraxis

Die Arbeitsbelastung von Mitarbeitenden sowie eingeschränkte finanzielle Ressourcen spielten bei der Umsetzung von Empfehlungen in Bezug auf geregelte Prozesse eine hemmende Rolle. Inhaltlich reichten die Empfehlungen von Themen wie Qualitätssicherung von Partnerevaluierungen über Systematisierung von Evaluierungen bis hin zu einer noch engeren Heranführung der Evaluierungspraxis an OECD DAC-Prinzipien. Von diesen Empfehlungen wurden 8 (57 Prozent) vollständig und 5 (36 Prozent) teilweise umgesetzt (vgl. Abbildung 23).<sup>26</sup>

<sup>25</sup> vgl. BMZ, 2009a; BMZ, 2008; BMZ, 2002 und BMZ, 2013.

<sup>26</sup> Zu einer Empfehlung wurde in der Online-Befragung keine Angabe zur Umsetzung gemacht.

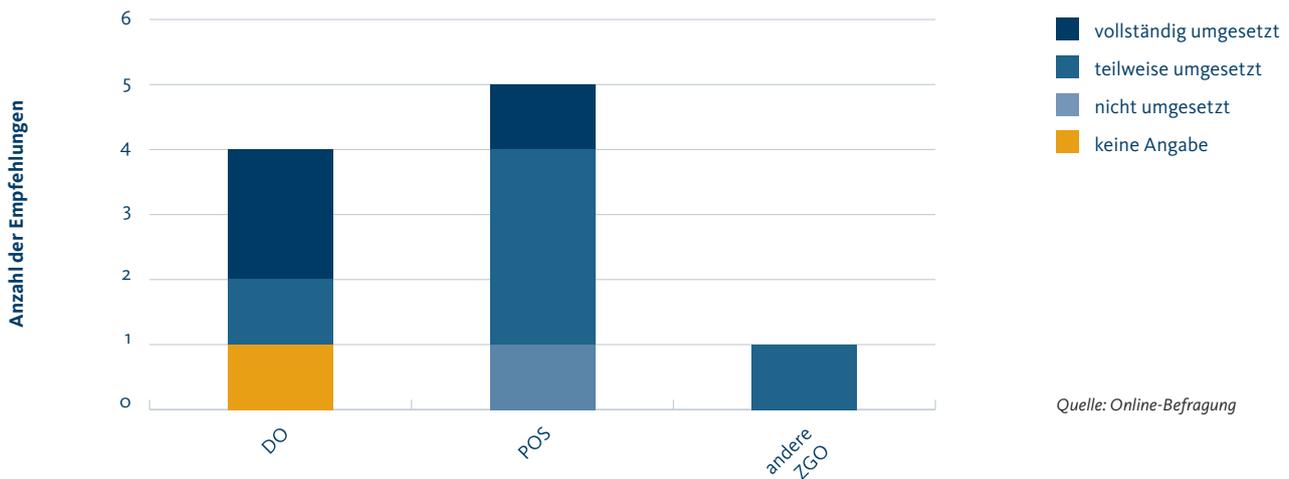
**Abbildung 23: Umsetzungsgrad nach Akteursgruppe (Festgelegte Prozesse zur Evaluierungspraxis)**

Bei der großen Mehrzahl der EZ-Organisationen gibt es feste Regeln, zu welchem Zeitpunkt, auf welche Art und unter Beteiligung welcher Akteure evaluiert werden sollte (vertiefende Interviews). Es existieren jedoch keine spezifischen Anreizsysteme in den Organisationen, um Evaluierungen durchzuführen. Diese werden durch Planungszyklen sowie aufgrund eigener Motivation, die Projekte und Programme zu verbessern, initiiert. Die Leitung ist häufig bei strategischen Themen involviert oder initiiert diese. Die Evaluierungen werden in allen Organisationen, die an der „Systemprüfung“ beteiligt waren, in (Mehr-)Jahresplänen für die Gesamtorganisation zusammengeführt. Dabei wird oftmals zwischen verpflichtend durchzuführenden Evaluierungen und optionalen unterschieden. Mit letzteren wird auf operationelle und manchmal auf strategische Bedarfe reagiert.

Die Organisationen unterscheiden sich in der Qualität der Prozesse und dem Standardisierungsgrad unterschiedlicher Prozessteile: Die Formalisierung der Konzeptions- und Vorbereitungsphase ist häufig stark ausgeprägt, während die Durchführungsphase weniger formalisiert ist. Die Unterschiede sind bei der Standardisierung der Umsetzungsphase von Handlungsempfehlungen am größten. So fehlen in vielen Organisationen standardisierte Prozesse, um die Umsetzung von Empfehlungen aus Evaluierungen nachzuhalten.

### 3.3.4 Einbindung von Partnern in den Evaluierungsprozess

Auch die Einbindung von Partnern stellt einen wichtigen Bestandteil von Evaluierungsprozessen dar. Die Auswertung der Umsetzung der insgesamt 10 ausgesprochenen Handlungsempfehlungen zeigt hierzu jedoch, dass der Einbeziehungsgrad stark von der Art der Kooperation sowie vom Interventionsbereich der zu evaluierenden Projekte abhängt: Politische Stiftungen, die die Hälfte der Empfehlungen zum Thema verbesserte Einbindung von Partnern erhielten (50 Prozent,  $n=5$ , vgl. Abbildung 24), teilten diese z.B. nur teilweise. Als Gründe wurden in den vertiefenden Interviews die teils politisch sehr brisanten Kontexte genannt und angeführt, dass der Partner oftmals Gegenstand der Evaluierung sei und daher nicht stärker in den Evaluierungsprozess eingebunden werden könne. Darüber hinaus nannten staatliche EZ-Organisationen in der Online-Befragung finanzielle Ressourcen als hemmende Faktoren, um der Empfehlung einer stärkeren Partnereinbindung nachgehen zu können.

**Abbildung 24: Umsetzungsgrad nach Akteursgruppe (Einbindung von Partnern in den Evaluierungsprozess)**

Die an den Evaluierungen beteiligten Akteure sowie deren Aufgaben und Rollen unterscheiden sich je nach Organisation, aber auch nach Art der Evaluierung: In den vertiefenden Interviews zeigte sich, dass verpflichtend durchzuführende Evaluierungen oftmals in der Verantwortung der Partner oder des Landesbüros liegen und nicht immer Teil der zentralen Qualitätssicherung sind. Demgegenüber verantworten bei strategischen Evaluierungen meistens Leitung und Evaluierungsverantwortliche den gesamten Prozess. Die Einbeziehung von Partnern in den Evaluierungsprozess ist sehr unterschiedlich und ist in vielen Organisationen durch Richtlinien o.Ä. geregelt. Nur ein Drittel der online-befragten Organisationen gab an, Partner partizipativ in den Evaluierungsprozess einzubinden.

#### Zentrale Ergebnisse zu der Dimension „Strukturen und Prozesse“

Das Ziel der Empfehlungen in dieser Dimension war,

- die Stellung der Evaluierungseinheit in der Institution
- institutionalisierte Evaluierungsgrundlagen
- festgelegte Evaluierungsprozesse und
- die Einbindung von Partnern in den Evaluierungsprozess

zu stärken.

Die 56 Empfehlungen wurden zu 64 Prozent geteilt und zu 55 Prozent vollständig umgesetzt.

- Evaluierungseinheiten bzw. -stellen wurden seit der letzten "Systemprüfung" vielfach neu geschaffen bzw. gestärkt. Evaluierungsverantwortliche verstehen sich überwiegend als Dienstleistende innerhalb der Organisationen. Sie sind bei der Festlegung von Evaluierungsgegenständen beteiligt, bestimmen diese jedoch nicht allein. Die Einbindung in die Durchführung und Qualitätssicherung von Evaluierungen ist unterschiedlich stark. Ein vergleichsweise hoher Anteil von Evaluierungen wird durch die operativen Bereiche beauftragt und gemanagt (bzw. durch die Partner).
- Die Berücksichtigung der OECD DAC-Kriterien ist in den Evaluierungsrichtlinien der Organisationen der deutschen EZ verankert. Die Umsetzung in den Evaluierungsberichten ist jedoch nicht konsequent.
- Die Organisationen der deutschen EZ haben ihre Evaluierungspraxis systematisiert; Evaluierungen folgen dabei überwiegend Evaluierungsrichtlinien und -plänen. Der

Formalisierungsgrad variiert je nach Prozessphase: Die Umsetzung bzw. Nutzung der Evaluierungsergebnisse ist dabei am geringfügigsten formalisiert, wenngleich einige Organisationen Prozesse zur Bewertung und Umsetzung von Evaluierungsergebnissen vorweisen.

- Bei der Einbindung von Partnern lässt sich feststellen, dass diese stark von der Art der Zusammenarbeit abhängt. Eine systematische und gleichberechtigte Einbeziehung von Partnern im Evaluierungsprozess ist nicht die Regel.

*Ursachenzuschreibung möglichst zuverlässig erlauben, sind für den Regelbetrieb zu entwickeln und standardmäßig zu nutzen.*

Das Ziel der Empfehlungen in der Dimension „Humanressourcen“ war die Weiterentwicklung der

- Anzahl der Mitarbeitenden in der Evaluierungseinheit
- Monitoring- und Evaluierungskompetenz der Mitarbeitenden
- Monitoring- und Evaluierungskompetenz der Partner
- Einsätze von Gutachterinnen und Gutachtern und
- eingesetzten Methoden.

### 3.4 Humanressourcen

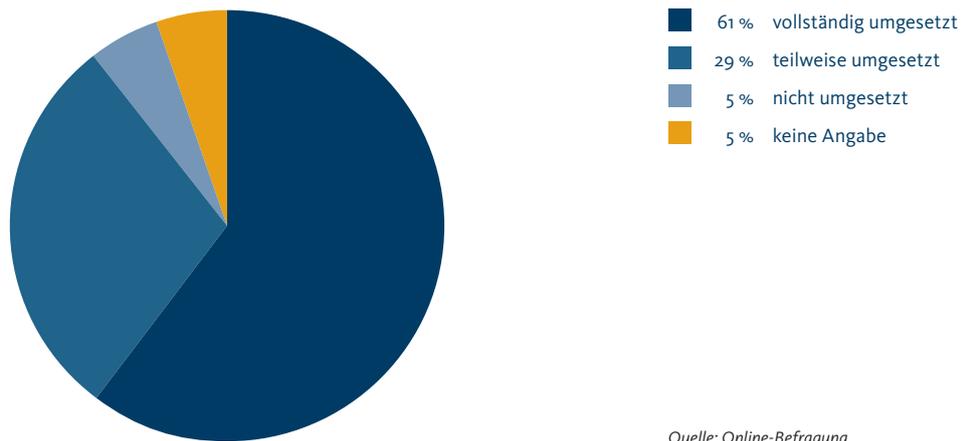
Folgende Handlungsempfehlungen auf Systemebene können der Dimension „Humanressourcen“ zugeordnet werden (Borrmann und Stockmann, 2009a: 165ff):

- *EZ-Organisationen sollten für ihre Evaluationseinheiten ausgebildete und erfahrene Evaluatoren von außen rekrutieren, und/ oder das interne Personal entsprechend sorgfältig vorbereiten und laufend weiterbilden.*
- *Methodenkompetenz von Gutachterinnen und Gutachtern sollte gestärkt werden.*
- *Systematische Aus- und Weiterbildung der Partner im Bereich der Evaluation.*
- *Entwicklung, Erprobung und Anwendung von angepassten Evaluationsdesigns, Erhebungs- und Auswertungsmethoden sollten gezielt (in Bezug auf zunehmende Gemeinschaftsevaluierungen und projektübergreifende Evaluierungen auf immer höherem Aggregationsniveau und von neuen EZ-Instrumenten) gefördert werden.*
- *Robuste Methoden, die die Erfassung von Wirkungen und eine auf Wirkungshypothesen beruhende (probabilistische)*

21 Prozent (n=38) aller organisationsspezifischen Handlungsempfehlungen beziehen sich auf Humanressourcen wie die Größe der Evaluierungseinheit sowie die Evaluierungskompetenzen der Mitarbeitenden und Partner. Die Empfehlungen wurden zu 66 Prozent von den EZ-Organisationen als relevant eingestuft.<sup>27</sup> Der Umsetzungsgrad in dieser Kategorie ist mit 61 Prozent (n=23) vollständiger Umsetzung vergleichsweise hoch. Weitere 29 Prozent (n=11) der Empfehlungen wurden teilweise umgesetzt, lediglich 5 Prozent (n=2) wurden nicht umgesetzt (vgl. Abbildung 25).

<sup>27</sup> DO: 79 Prozent, POS: 44 Prozent, andere ZGO: 88 Prozent.

Abbildung 25: Umsetzungsgrad der Empfehlungen (Humanressourcen)

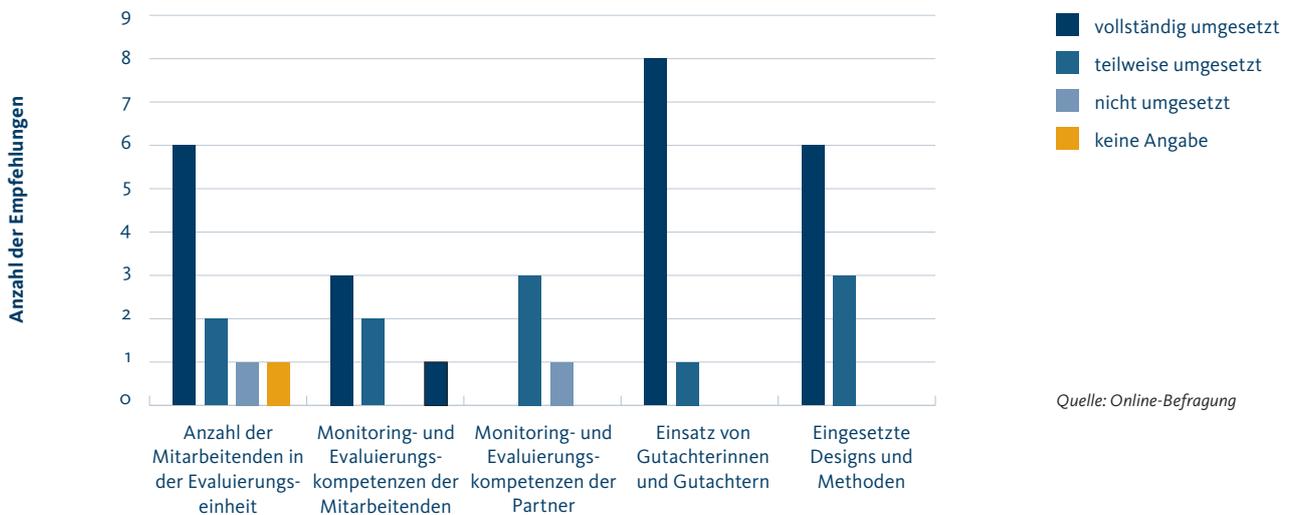


Quelle: Online-Befragung

Ähnlich viele Empfehlungen wurden zu Stellenbesetzungen in Evaluierungseinheiten (26 Prozent; n=10) sowie dem Methoden- und Gutachtereinsatz (jeweils 24 Prozent; n=9) ausgesprochen. Seltener wurden Hinweise zur Monitoring- und

Evaluierungskompetenz von Mitarbeitenden (16 Prozent; n=6) und von Partnern (10 Prozent; n=4) an EZ-Organisationen gegeben (vgl. Abbildung 26).

Abbildung 26: Umsetzungsgrad nach inhaltlichen Clustern (Humanressourcen)



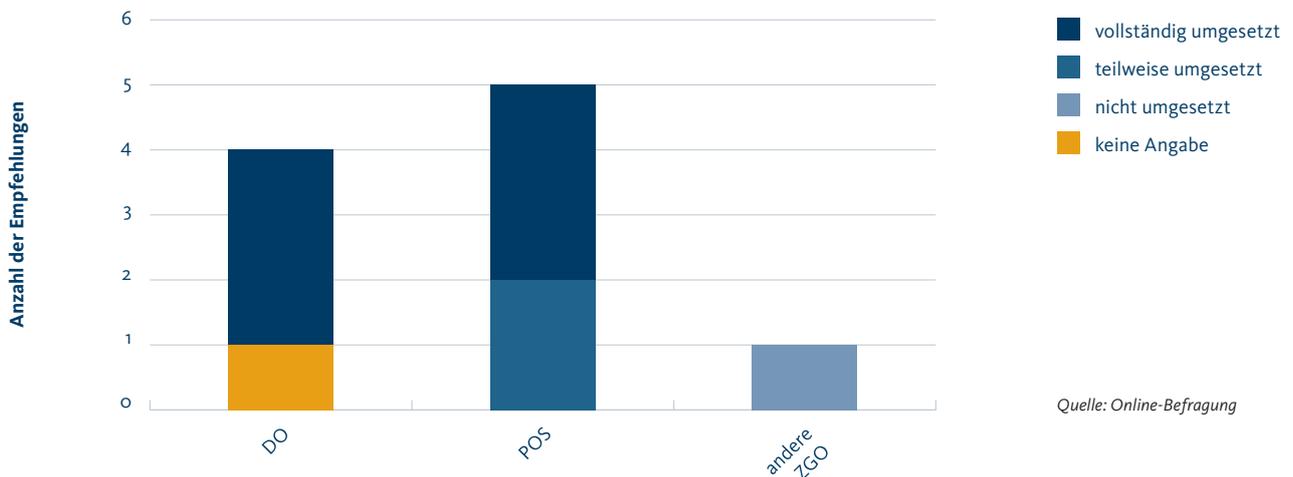
Quelle: Online-Befragung

### 3.4.1 Anzahl der Mitarbeitenden in der Evaluierungseinheit

Zur Verbesserung der Evaluierungspraxis wurde sowohl politischen Stiftungen als auch staatlichen EZ-Organisationen im Rahmen der „Systemprüfung“ empfohlen, die Anzahl der Mitarbeitenden in der Evaluierungseinheit zu erhöhen. Dies wurde von insgesamt zehn Empfehlungen sechs Mal vollständig

umgesetzt sowie von zwei politischen Stiftungen zum Zeitpunkt des Umsetzungsmonitorings teilweise (vgl. Abbildung 27). In den vertiefenden Interviews wurde angemerkt, dass der Einstellungsprozess noch nicht abgeschlossen sei. Eine ZGO setzte eine entsprechende Empfehlung nicht um, da diese aufgrund anderweitiger Reorganisation nicht prioritär war.

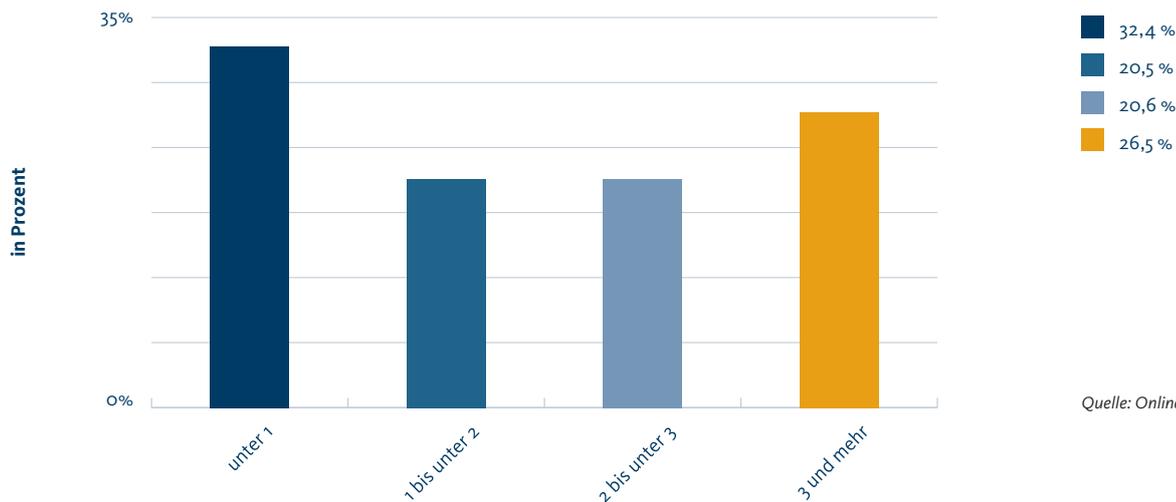
Abbildung 27: Umsetzungsgrad nach Akteursgruppe (Anzahl der Mitarbeitenden in der Evaluierungseinheit)



Seit der „Systemprüfung“ hat sich die Anzahl der Mitarbeitenden in den jeweiligen Evaluierungseinheiten erhöht oder diese sind momentan im Prozess, die Mitarbeiterzahl zu erhöhen (vertiefende Interviews): Der Anteil der Evaluierungsverantwortlichen am Gesamtpersonal einer Organisation liegt im Mittel bei 1,7 Prozent (Median). Bei rund der Hälfte der Organisationen (52,9 Prozent) beschäftigen sich weniger als zwei Personen hauptsächlich mit Evaluierungen, die andere Hälfte an Organisationen verfügt hierfür über zwei oder mehr Personen (vgl. Abbildung 28). Lediglich die großen DO

unterhalten große Evaluierungseinheiten. Ein knappes Drittel der EZ-Organisationen aus der „Systemprüfung“ und die Hälfte aller online-befragten Organisationen unterhalten hingegen keine eigenen Evaluierungseinheiten. In diesen Organisationen sind Personen aus den operativen Bereichen mit weniger als einem Vollzeitäquivalent für Evaluierungen zuständig. Evaluierungseinheiten bzw. -beauftragte haben oftmals noch weitere Aufgaben zu erfüllen, wie z.B. Verantwortlichkeit für Qualitätsmanagement, Monitoring, Schulungen und Beratung bei Angebotserstellung.

Abbildung 28: Vollzeit-Stellen Evaluierungsverantwortliche



Quelle: Online-Befragung; n=34

### 3.4.2 Monitoring- und Evaluierungskompetenz der Mitarbeitenden und Partner

Monitoring- und Evaluierungskompetenzen der Mitarbeitenden und der Partner waren ebenfalls Gegenstand der Empfehlungen: Über Fort- und Weiterbildungsmaßnahmen für Mitarbeitende im In- und Ausland sowie über Capacity Building Maßnahmen bei Partnern zu Evaluierungsmethoden sollten deren Kompetenzen weiterentwickelt und geschärft werden. Während Weiterbildungsmaßnahmen für Mitarbeitende zur Erweiterung ihrer Methoden- und Evaluierungskompetenzen überwiegend durchgeführt wurden (3 (50 Prozent) Empfehlungen wurden vollständig, 2 (33 Prozent) teilweise<sup>28</sup> umgesetzt; vgl. Abbildung 29), stellte dies in Bezug auf Partner größere Herausforderungen dar. Von den vier

ausgesprochenen Empfehlungen zur Förderung der Monitoring- und Evaluierungskompetenz der Partner wurden drei teilweise und eine Empfehlung nicht umgesetzt (vgl. Abbildung 30, zur Begründung vgl. nachstehende Absätze).

Jüngere Mitarbeitende verfügen häufig über einen evaluierungsspezifischen Ausbildungshintergrund, ältere Mitarbeitende bringen oft viel Erfahrung aus dem operativen Bereich mit und haben fachspezifische Fortbildungen besucht (z.B. FFEZ). Aufgrund relativ geringer Fluktuation ist daher in vielen Evaluierungseinheiten großes Erfahrungswissen und Kompetenz in der Beurteilung von Designs und Methoden vorhanden, selbst wenn meistens keine eigene Evaluierungserfahrung vorliegt.

<sup>28</sup> Zu einer Handlungsempfehlung wurde in der Online-Befragung keine Angaben zur Umsetzung gemacht.

Abbildung 29: Umsetzungsgrad nach Akteursgruppe (Monitoring- und Evaluierungskompetenz der Mitarbeitenden)

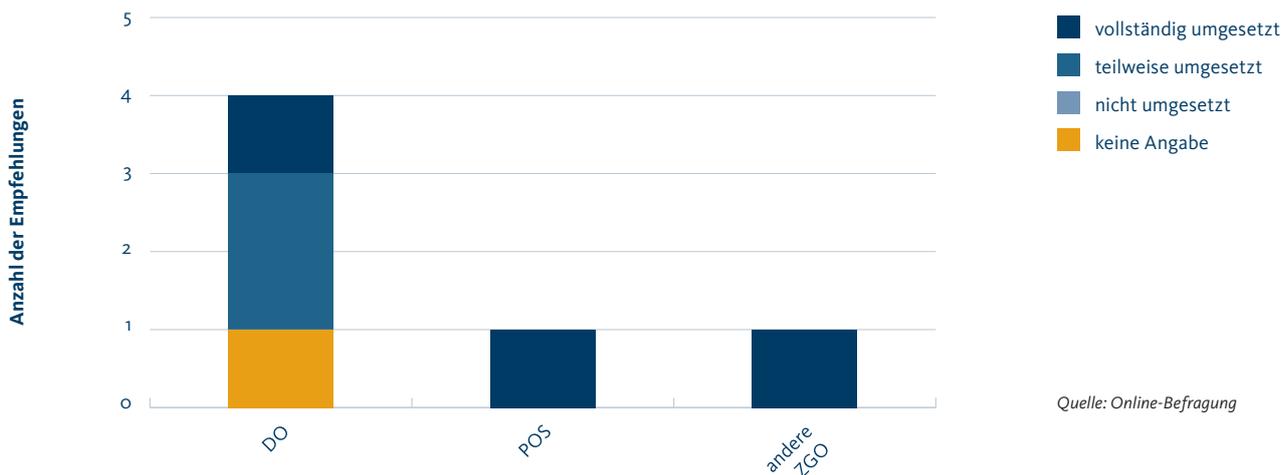
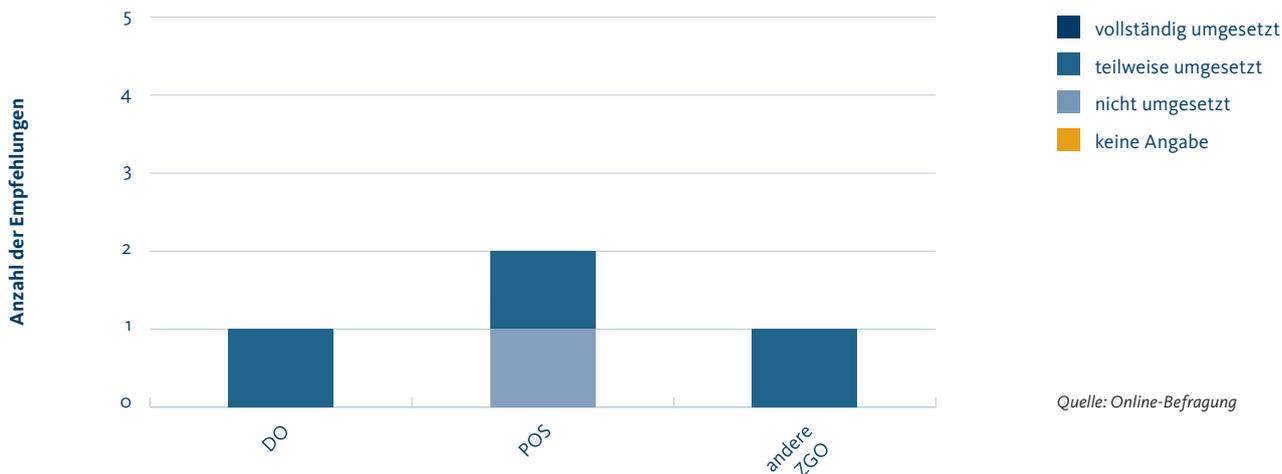


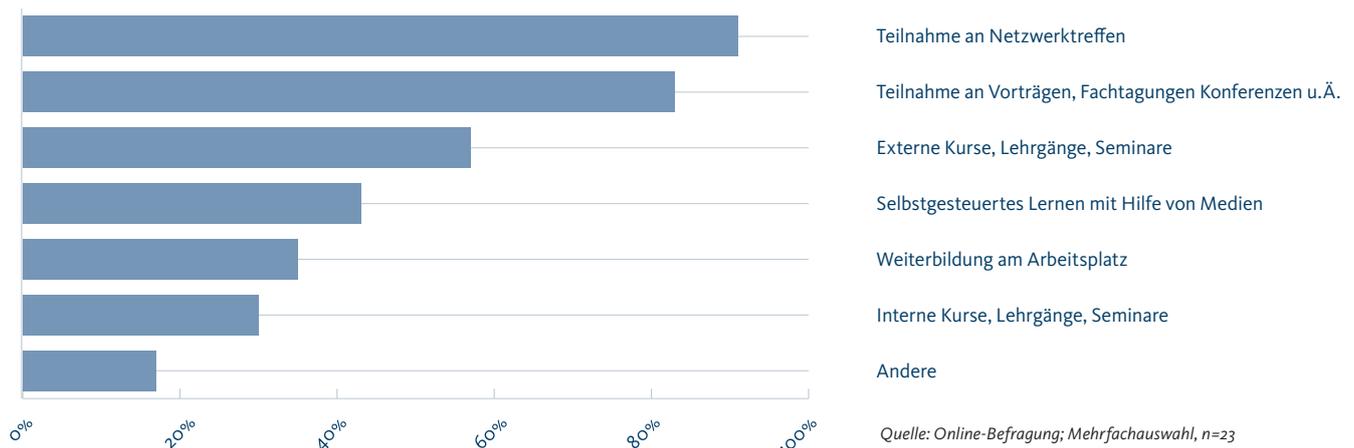
Abbildung 30: Umsetzungsgrad nach Akteursgruppe (Monitoring- und Evaluierungskompetenz der Partner)



Mehrheitlich nahmen Evaluierungsverantwortliche in 2013 an Qualifizierungen im Bereich Evaluierung, v.a. Netzwerktreffen (91,3 Prozent), und an Vorträgen und Fachtagungen (82,6

Prozent) teil (vgl. Ergebnisse der Online-Befragung in Abbildung 31). Fortbildungen sind möglich und werden unterstützt (vgl. Kapitel 3.2.2).

Abbildung 31: Qualifizierungsmaßnahmen im Bereich Evaluierung



Ein methodischer Austausch erfolgt v.a. über fachspezifische Netzwerke (DeGEval AK Epol/ Methoden, Jahrestreffen der Evaluierungseinheiten, VENRO Wirkungsbeobachtung). Als Innovatoren oder Verbreiter gelten vor allem CEval, DeGEval, GIZ und das BMZ (vgl. Kapitel 3.2.2).

Wie bereits einleitend erwähnt, wurden die Handlungsempfehlungen zur Qualifizierung von Partnern nur teilweise umgesetzt. Ein Fortbildungsbedarf für Partner wird auch von den betroffenen EZ-O oft erkannt (vertiefende Interviews). Allerdings fühlen sich die Organisationen in sehr unterschiedlichem Maße dafür verantwortlich, Partnern Fortbildungsangebote zu unterbreiten bzw. deren Qualifizierung nachzuhalten. Während manche ZGO und hier v.a. die kirchlichen EZ-Organisationen die Qualifizierung ihrer Partner als eine ihrer zentralen Aufgaben sehen, stellten einige Organisationen in den vertiefenden Interviews klar, dass sie das empfohlene Capacity Building beim Partner im Bereich Evaluierung nicht zu ihrem Aufgabenbereich zählen. Als weitere Gründe für eine fehlende Umsetzung der Empfehlungen wurden u.a.

mangelndes Interesse oder mangelnder Bedarf der Partner genannt sowie mangelnde Fortbildungsangebote vor Ort.

### 3.4.3 Einsatz von Gutachterinnen und Gutachtern

Die Stärkung der Methodenkompetenzen der Gutachterinnen und Gutachter wurde in insgesamt neun Empfehlungen berücksichtigt, die an alle Akteursgruppen gerichtet wurden. Daneben wurde auch empfohlen, die Methodenkompetenzen als zentrales Auswahlkriterium bei der Gutachterausswahl stärker zu beachten. Zur Verbesserung der Evaluierungspraxis erhielten die EZ-Organisationen zudem den Hinweis, mehr externe Gutachterinnen und Gutachter einzusetzen. Bei diesen Aspekten bestand unter den Organisationen eine hohe Akzeptanz und von neun Empfehlungen wurden nahezu alle vollständig (eine teilweise) umgesetzt (vgl. Abbildung 32).

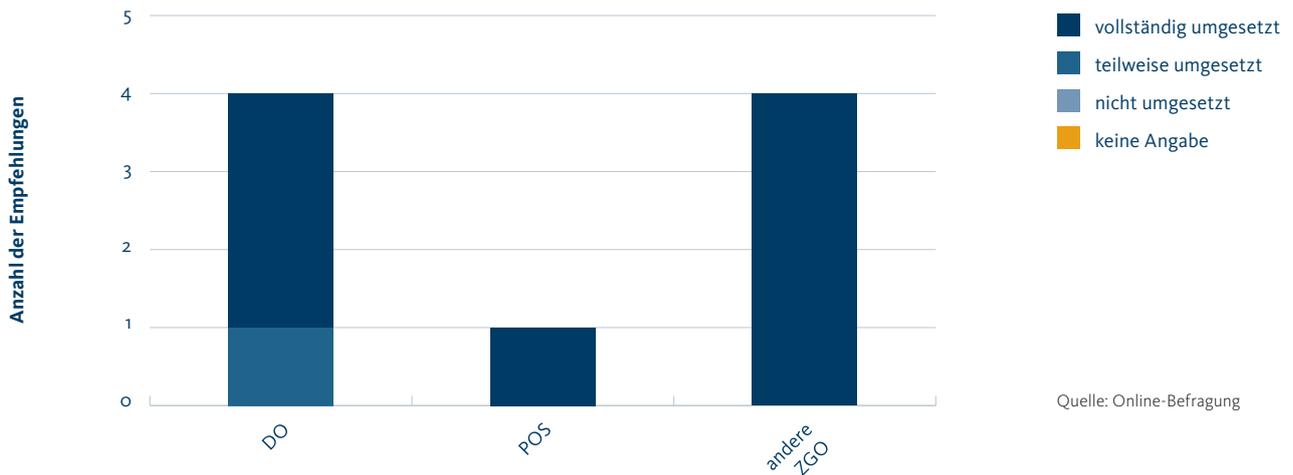
**Abbildung 32: Umsetzungsgrad nach Akteursgruppe (Einsatz von Gutachterinnen und Gutachtern)**

Tabelle 4 zeigt, dass Evaluierungen zunehmend auch durch unabhängige Gutachtende bzw. extern durchgeführt werden. Die in der Metaevaluierung betrachteten Evaluierungen wurden nahezu ausschließlich (98 Prozent) von externen Gutachterinnen und Gutachtern sowie nur in einem Fall (2 Prozent) von eigenen Mitarbeitenden durchgeführt. Als externe Gutachtende waren in 41 Fällen (91 Prozent) privatwirtschaftliche Gutachterinnen und Gutachter, Agenturen oder Institute für die Durchführung verantwortlich, in 3 Fällen (7 Prozent)

wurde sie von Universitäten übernommen. In den vertiefenden Interviews wurde deutlich, dass ein Teil der befragten Organisationen das methodische Repertoire der Gutachtenden für ausbaufähig hält, jedoch werden diese auch kaum von Organisationen geschult. Länder- und organisationspezifische Kenntnisse wurden bei der Frage nach Auswahlkriterien für Gutachtende meist vor methodischen Kenntnissen genannt (vertiefende Interviews).

**Tabelle 4: Anteil der Programmmittel, die für unabhängige Evaluierungen aufgewendet werden**

	2008	2012
DO	3,8 %	7,3 %
POS	0,3 %	0,3 %
andere ZGO	0,2 %	0,5 %

Quelle: Online-Befragung, arithmetisches Mittel in Prozent für die Jahre 2008, n=22 und 2012, n=25

### 3.4.4 Eingesetzte Designs und Methoden

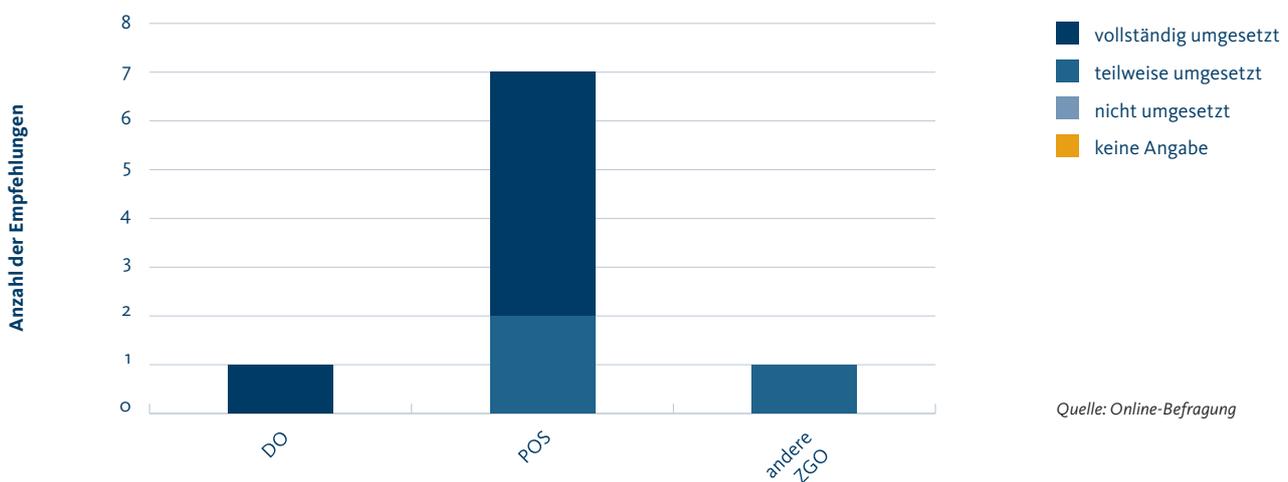
Neben einer Aufstockung des Personals wurde im Zuge der „Systemprüfung“ zur Verbesserung der Evaluierungspraxis deutscher EZ-Organisationen ebenfalls empfohlen, das Methodenspektrum der Evaluierungseinheiten zu erweitern. Dadurch soll sichergestellt werden, dass robustere Daten erhoben werden

können und über eine Methodentriangulation Erkenntnisse validiert werden. Die Empfehlungen stießen weitestgehend auf Zustimmung und wurden zu 67 Prozent (n=6) vollständig umgesetzt. Die Mehrheit der Empfehlungen (78 Prozent; n=7) richtete sich an politische Stiftungen, denen ein Mix aus qualitativen und quantitativen Methoden empfohlen wurde sowie die Anwendung

anspruchsvollerer Evaluationsdesigns und -methoden. Den Empfehlungen wurde weitestgehend nachgegangen (67 Prozent (n=6) wurden vollständig sowie 33 Prozent (n=2) teilweise umgesetzt; die POS setzten fünf Empfehlungen vollständig und zwei teilweise um, vgl. Abbildung 33). In den vertiefenden Interviews merkten Vertreterinnen und Vertreter der politischen Stiftungen jedoch an, dass anspruchsvolle Designs sowie der Einsatz quantitativer

Methoden Herausforderungen darstellten: Zum einen ließen finanzielle Limitierungen nicht immer anspruchsvolle Designs zu. Zum anderen würden Gutachterinnen und Gutachter nicht allein nach ihrer Methodenkenntnis ausgewählt, sondern auch aufgrund weiterer Kriterien wie z.B. Regional-/ Landeskompentzen (vgl. Kapitel 3.4.3). Außerdem sei der ausschließliche Einsatz qualitativer Methoden in bestimmten Fällen durchaus zweckmäßig.

**Abbildung 33: Umsetzungsgrad nach Akteursgruppe (Eingesetzte Designs und Methoden)**



In den Interviews mit internationalen Expertinnen und Experten wurde angeführt, dass in der deutschen EZ bis zur Jahrtausendwende ein gewisser Vorbehalt gegenüber quantitativen Methoden vorherrschte und nach wie vor rigorose Methoden weitaus kritischer gesehen werden als etwa in anglophonen Ländern. Es verbreite sich aber zunehmend die Sichtweise, dass ein der Evaluierungsfragestellung angemessener Methodenmix Anwendung finden sollte.<sup>29</sup>

Die Metaevaluierung von Evaluierungsberichten führte bezüglich der angewandten Designs und Methoden zu folgenden Ergebnissen: Der Evaluierungsgegenstand wird größtenteils (in 87 Prozent der 45 Fälle) ausführlich beschrieben, insofern wesentliche Aspekte wie Akteure, Strukturen und Prozesse des zu evaluierenden Gegenstandes erläutert werden. Zudem werden in der Regel (93 Prozent) auch die Ziele der Evaluierung deutlich, indem dokumentiert wird, warum und für wen die Evaluierung durchgeführt wurde. Die Arbeitsschritte

des methodischen Vorgehens werden bei einem Großteil der Evaluierungen (84 Prozent) zumindest in Grundzügen beschrieben. Eine detaillierte Darstellung der Erhebungs- und Auswertungsmethoden – im Sinne einer Beschreibung einzelner methodischer Schritte, bspw. Auswahlkriterien bei Interviewpartnerinnen und -partnern – wird hingegen weit seltener gegeben (in 38 Prozent der Fälle). Eine Begründung der Methodenauswahl bzw. eine Thematisierung der Vor- und Nachteile angewandter Methoden wird ebenfalls selten vorgenommen (in 16 Prozent der Fälle).

Bei knapp einem Drittel der untersuchten Evaluierungen werden explizit Wirkungshypothesen formuliert oder Indikatoren verwendet, die SMART-Kriterien entsprechen (in jeweils 32 Prozent der Fälle). Dementsprechend werden die Datenerhebungs- und Auswertungsmethoden auch nur bei einem Fünftel der Evaluierungen (20 Prozent) systematisch von Wirkungshypothesen oder Indikatoren abgeleitet.

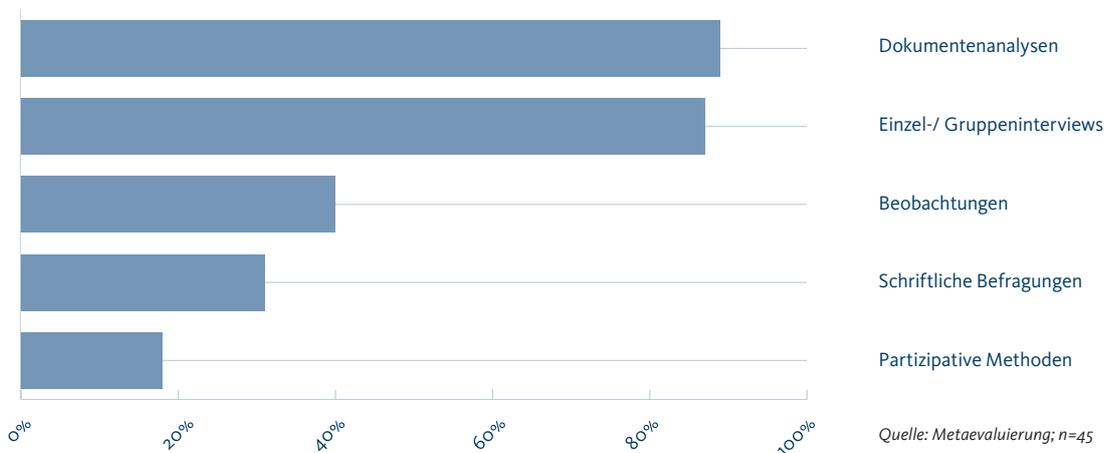
<sup>29</sup> vgl. hierzu auch Neubert, Mack und Roxin (2014) sowie BMZ (2009).

Knapp zwei Drittel der Evaluierungen (62 Prozent) bedienen sich eines *single-method designs*, insofern ausschließlich qualitative Datenerhebungsmethoden verwendet werden.<sup>30</sup> Ein *mixed-method design* wird dementsprechend bei gut einem Drittel der Fälle (38 Prozent) verwendet. Die Einbeziehung von Kontroll- oder Vergleichsgruppen in das Untersuchungsdesign wurde selten (bei 7 Prozent) vorgenommen.

Ein Großteil der Evaluierungen geht multimethodisch vor, indem mindestens zwei unterschiedliche Datenerhebungsmethoden eingesetzt werden (82 Prozent). Sehr häufig werden Dokumentenanalysen (89 Prozent) und Interviews<sup>31</sup> eingesetzt

(87 Prozent), wobei sich rund die Hälfte der Berichte (51 Prozent) ausschließlich auf eine oder die Kombination dieser beiden Methoden stützt. Auf Beobachtungen – bspw. im Zuge von *field visits* oder der Teilnahme an Veranstaltungen – wird in 40 Prozent der Evaluierungen verwiesen, schriftliche Fragebögen werden bei knapp einem Drittel verwendet (31 Prozent). Bei einem Fünftel der untersuchten Evaluierungen (20 Prozent) kamen alle diese Erhebungsmethoden zum Einsatz (Dokumentenanalyse, Interviews, Fragebögen und Beobachtung). Zudem wurden im Zuge der Datenerhebung bei knapp einem Fünftel der Evaluierungen (18 Prozent) partizipative Methoden<sup>32</sup> eingesetzt (vgl. Abbildung 34).

**Abbildung 34: Einsatz verschiedener Datenerhebungsmethoden**



Eine Datentriangulation wird ausnahmslos vorgenommen (zumeist in Form der Durchführung mehrerer Interviews). Eine nachvollziehbare Methodentriangulation konnte in 38 Prozent der Fälle festgestellt werden. Hinsichtlich der Forschertriangulation werden gut die Hälfte der Evaluierungen (53 Prozent) nur von einer Person durchgeführt, bei gut einem Fünftel (22 Prozent) waren Durchführende aus der Projektregion an der Evaluierung beteiligt.

In Bezug auf die inhaltliche Systematik der Evaluierungsberichte findet bei zwei Dritteln (67 Prozent) eine klare Trennung

zwischen Beschreibung, Analyse, Bewertung und ggf. Handlungsempfehlungen statt. Bei gut der Hälfte (53 Prozent) ist jedoch nicht durchgehend verständlich, was die methodischen Quellen der Informationen sind. Selbst wenn für die Analyse der Evaluierungen auch deren ToR mit schriftlich dokumentierten Fragestellungen vorlagen, war bei knapp zwei Drittel (65 Prozent) nicht ersichtlich, anhand welcher Datenerhebung einzelne Untersuchungsfragen beantwortet wurden. Zudem wurden bei gut einem Viertel (27 Prozent) nicht alle Fragen der ToR beantwortet oder deren Beantwortung war nicht klar ersichtlich.

<sup>30</sup> Bei keiner Evaluierung der Stichprobe wurden ausschließlich quantitative Datenerhebungsmethoden verwendet.

<sup>31</sup> Als Interviews wurden hierbei (mündliche) Einzelgespräche mit Gruppendiskussionen zusammengefasst. Eine ausführliche Unterscheidung sozialwissenschaftlicher Datenerhebungsmethoden und ihrer Nutzung bei Evaluierungen findet sich bspw. bei Meyer (2007: 224ff).

<sup>32</sup> Als partizipativ wurden jene Methoden bewertet, die eine aktive Einbeziehung von Projektbeteiligten oder Zielgruppen in den Prozess der Informationsgewinnung oder Entscheidungsfindung gewährleisten, beispielsweise die Durchführung einer SWOT-Analyse im Rahmen eines Workshops (vgl. BMZ, 2009b: 19; Caspari, 2006: 369; Weltbank, 2004: 16). Partizipative Methoden sind insofern nicht als eigenständige Erhebungsmethode zu verstehen, sondern stellen eine besondere Form der Durchführung klassischer sozialwissenschaftlicher Datenerhebungsmethoden dar.

Im Zusammenhang mit den Ergebnissen der Evaluierungen werden in nahezu allen Berichten (96 Prozent) auch konkrete Handlungsempfehlungen ausgesprochen. Bei 67 Prozent dieser Evaluierungen werden Ziel oder Zweck der Empfehlungen eindeutig benannt. In ebenso vielen Fällen ist deren Umsetzung ebenfalls gut messbar und deren Realisierbarkeit inhaltlich abgesichert. Projektübergreifende Empfehlungen werden in knapp der Hälfte der Berichte (47 Prozent) ausgesprochen.

### Zentrale Ergebnisse zu der Dimension „Humanressourcen“

Das Ziel der Empfehlungen war die Weiterentwicklung von

- der Anzahl der Mitarbeitenden in der Evaluierungseinheit
- der Monitoring- und Evaluierungskompetenz der Mitarbeitenden
- der Monitoring- und Evaluierungskompetenz der Partner
- Einsätzen von Gutachterinnen und Gutachtern und
- eingesetzten Methoden.

Die 38 Empfehlungen wurden zu 66 Prozent geteilt und zu 61 Prozent vollständig umgesetzt.

- Seit der „Systemprüfung“ hat eine Erhöhung der Personalressourcen stattgefunden. Zugleich haben sich oftmals auch die methodischen Ansprüche an die Evaluierungsverantwortlichen erhöht.
- Die fachliche Kompetenz in den Evaluierungseinheiten wurde gestärkt. Die Qualität von Evaluierungen hängt aber nicht allein von den Evaluierungseinheiten ab, sondern auch von der Kompetenz der beauftragten Gutachterinnen und Gutachter sowie der operationellen Bereiche und Partner, sofern diese Evaluierungen managen. Es besteht nicht überall die Möglichkeit (v.a. aufgrund der Vielzahl von Evaluierungen in einigen Organisationen), eine systematische Qualitätssicherung durch die Evaluierungseinheit vorzunehmen.

- Die Qualifizierung von Partnern für Evaluierung sehen viele Organisationen nicht in ihrem Verantwortungsbereich, wenngleich einige ZGO sich sehr stark in diesem Bereich engagieren.
- Evaluierungsdesigns basieren häufig nicht auf überprüfbareren Wirkungshypothesen. Das genutzte Methodenspektrum bzw. die Nutzung eines angemessenen Methodenmixes ist ausbaufähig.

## 3.5 Finanzielle Ressourcen

Der Dimension „finanzielle Ressourcen“ kann eine Handlungsempfehlung auf Systemebene zugeordnet werden (vgl. Borrmann und Stockmann, 2009a):

- *Die Evaluationssysteme der meisten EZ-Organisationen sind deutlich unterfinanziert. Den größten Nachholbedarf hat das BMZ. Eine Aufstockung ist umso mehr erforderlich, wenn sich Evaluation den neuen Herausforderungen einer veränderten EZ sowie der Ausrichtung auf Wirkungen stellen soll. Für unabhängige Evaluationen sollte mindestens 1 Prozent des Portfolios zur Verfügung stehen. Da zivilgesellschaftliche Organisationen Schwierigkeiten haben, projektübergreifend und Ex-post zu evaluieren, sollte das BMZ einen externen Evaluationsfonds einrichten.*

Die in dieser Dimension ausgesprochene Empfehlung hatte zum Ziel, die Budgets für Evaluierungen und Evaluationssysteme zu erhöhen.<sup>33</sup>

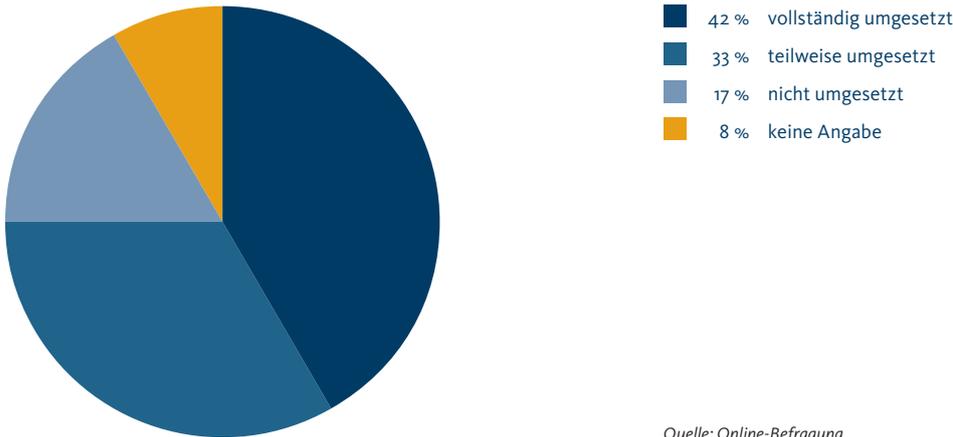
Die Evaluierungspraxis hängt in vielerlei Hinsicht von finanziellen Ressourcen ab, die den Evaluierungsverantwortlichen in deren Organisationen zur Verfügung gestellt werden: 12 (7 Prozent) der Empfehlungen an die EZ-Organisationen griffen Themen der finanziellen Rahmenbedingungen auf, wovon mit 5 (42 Prozent) etwas weniger als die Hälfte der Empfehlungen vollständig umgesetzt wurden. 4 Empfehlungen (33 Prozent) wurden teilweise umgesetzt und 2 (17 Prozent) gar nicht (vgl. Abbildung 35).<sup>34</sup> Geteilt wurden die Empfehlungen von den betroffenen Organisationen zu 58 Prozent.<sup>35</sup>

<sup>33</sup> Die Empfehlungen, die sich an das BMZ richteten, sind durch die Gründung des DEval umgesetzt (vgl. Kapitel 1).

<sup>34</sup> Zu einer Handlungsempfehlung wurde in der Online-Befragung keine Angabe zur Umsetzung gemacht.

<sup>35</sup> DO: 100 Prozent, POS: 20 Prozent, andere ZGO: 50 Prozent.

Abbildung 35 : Umsetzungsgrad der Empfehlungen (Finanzielle Ressourcen)

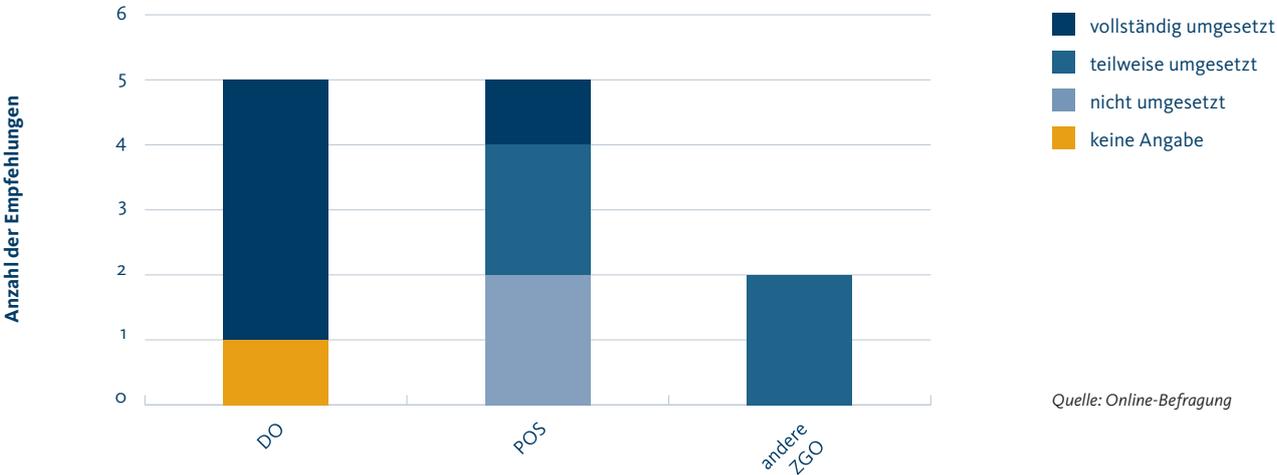


Quelle: Online-Befragung

Die vollständig umgesetzten Handlungsempfehlungen bezogen sich thematisch unter anderem auf die hinreichende Budgetierung von Schlussevaluierungen oder anspruchsvollere Erhebungsdesigns und wurden von vier staatlichen EZ-Organisationen sowie einer politischen Stiftung umgesetzt (vgl. Abbildung 36). Die vier teilweise umgesetzten Handlungsempfehlungen richteten sich an Nichtregierungsorganisationen sowie politische Stiftungen

und empfahlen die Erhöhung des Evaluierungsbudgets im Wert von einem Prozent der Gesamtprogrammmittel. Die beiden nicht umgesetzten Handlungsempfehlungen bezogen sich inhaltlich ebenfalls auf die Erhöhung von finanziellen Mitteln und eine verbindliche Festlegung eines Evaluierungsbudgets, wurden jedoch bereits im Anschluss an die „Systemprüfung“ von den Organisationen nicht geteilt und somit nicht umgesetzt.

Abbildung 36: Umsetzungsgrad nach Akteursgruppe (Finanzielle Ressourcen)



Quelle: Online-Befragung

In der Online-Befragung wurde deutlich, dass etwa die Hälfte der zehn EZ-Organisationen, die entsprechende Empfehlungen erhielten, mit ihren finanziellen Ressourcen für Evaluierungen zufrieden war. Einige Organisationen erläuterten in den vertieften Interviews, dass keine Notwendigkeit für eine Erhöhung der Mittel für Evaluierungen gesehen wurde, da die Anzahl und der Umfang an Evaluierungen als ausreichend empfunden werde.

Seit 2008 ist ein Anstieg der finanziellen Ressourcen für Evaluierungen zu verzeichnen (vgl. Tabelle 5). In der Online-Befragung wurden die Teilnehmenden nach der Höhe des zur Verfügung stehenden Evaluierungsbudgets gefragt. Zum

Budget zählt das von der Evaluierungseinheit verwaltete Budget für Evaluierungsstudien sowie das Budget für Stellen der Evaluierungseinheit bzw. des Evaluierungsbeauftragten. 2008 verfügten 17 Prozent (n=6) der Organisationen über ein Evaluierungsbudget von 1 Prozent oder mehr des Gesamtbudgets, während 2012 bereits 38 Prozent (n=13) ein entsprechendes Budget besaßen (Online-Befragung; vgl. Tabelle 5). Jedoch führt die zunehmende Professionalisierung auch zu höheren Ansprüchen (z.B. Finanzierung von Wirkungs- oder Ex-post-Evaluierungen), für die der Anstieg im Evaluierungsbudget bei vielen Organisationen nicht ausreichend ist.

**Tabelle 5: Evaluierungsbudget als Anteil am Gesamtbudget**

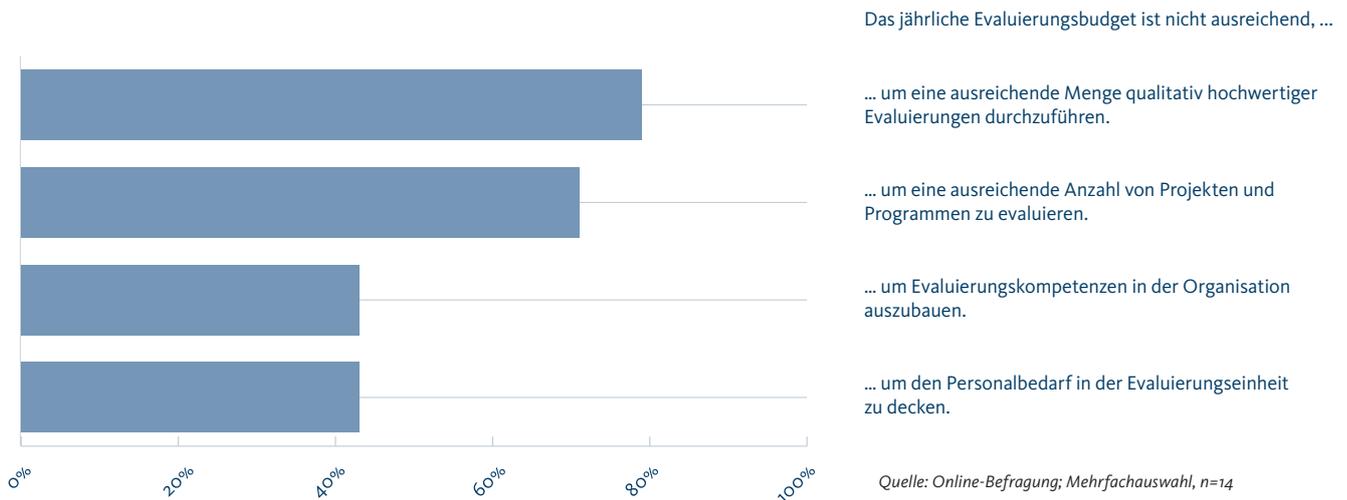
	2008	2012
DO	4,3 %	4,1 %
POS	0,6 %	0,7 %
andere ZGO	0,6 %	1,5 %

Quelle: Online-Befragung, arithmetisches Mittel in Prozent für die Jahre 2008, n=23 und 2012, n=26

Manche Evaluierungseinheiten verfügen über kein eigenes Budget, die Finanzierung der Evaluierungen erfolgt hier aus Projektmitteln. Das bereitgestellte Budget für Evaluierungen wird von 18 (53 Prozent) EZ-Organisationen als ausreichend bewertet. Weitere 14 Organisationen<sup>36</sup> (41 Prozent) bewerten

ihr Evaluierungsbudget als zu niedrig. Auf die Frage, in welchem Bereich das Budget nicht ausreichend sei, verwies die Mehrheit dieser Organisationen auf Einschränkungen bezüglich Qualität und Quantität der durchgeführten Evaluierungen (vgl. Abbildung 37).

<sup>36</sup> 2 DO, 1 POS, 11 andere ZGO.

**Abbildung 37: Folgen von unzureichendem Evaluierungsbudget**

In den vertiefenden Interviews fiel die Bewertung der Angemessenheit des Budgets durch die Leitung deutlich positiver aus als durch die Evaluierungseinheiten. Für über die Pflicht hinausgehende Evaluierungen (z.B. projektübergreifende Evaluierungen) ist die Finanzierung schwieriger bzw. nicht ausreichend. Insbesondere für zivilgesellschaftliche Organisationen ist die Finanzierung von projektübergreifenden oder methodisch anspruchsvollen Evaluierungen aus Projektbudgets (mit 25 Prozent Eigenbeitrag), zweckungebundenen Spenden und 4 Prozent Verwaltungskostenzuschuss schwierig.<sup>37</sup>

Wenig überraschend ist, dass in Organisationen, in denen Evaluierungen ein hoher Nutzen seitens der Leitung zugesprochen wird, auch eine größere Bereitschaft zur Bereitstellung von Ressourcen besteht (vertiefende Interviews).

### Zentrale Ergebnisse zu der Dimension „Finanzielle Ressourcen“

Die ausgesprochenen Empfehlungen hatten zum Ziel, die Budgets für Evaluierungen und Evaluierungssysteme zu erhöhen.

Die 12 Empfehlungen wurden zu 58 Prozent geteilt und zu 42 Prozent umgesetzt.

Seit der „Systemprüfung“ sind die finanziellen Ressourcen für Evaluierungen gestiegen. Diese werden überwiegend als ausreichend betrachtet. Für einen nicht zu vernachlässigenden Teil von Organisationen sind sie jedoch weiterhin nicht ausreichend, insbesondere für methodisch anspruchsvolle, projektübergreifende Evaluierungen.

<sup>37</sup> Entsprechend den jeweiligen Förderrichtlinien erhalten POS und SST Vollfinanzierung und bis zu 14 Prozent Verwaltungskostenzuschuss. Bei privaten Trägern/ sonstigen NRO liegt der Verwaltungskostenzuschuss bei 4 Prozent und 25 Prozent der Projekt-/ Programmkosten müssen als Eigenbeitrag erbracht werden (die kirchlichen Organisationen erhalten keine Verwaltungskostenzuschüsse). Bei mehrjährigen komplexen Projekten oder Projekten mit Pilotcharakter kann eine Evaluierung durch unabhängige Gutachterinnen und Gutachter beantragt und mitfinanziert werden. Die Ausgaben dafür dürfen 15.000,- € nicht übersteigen (vgl. BMZ, 2002 und BMZ, 2007).

## 3.6

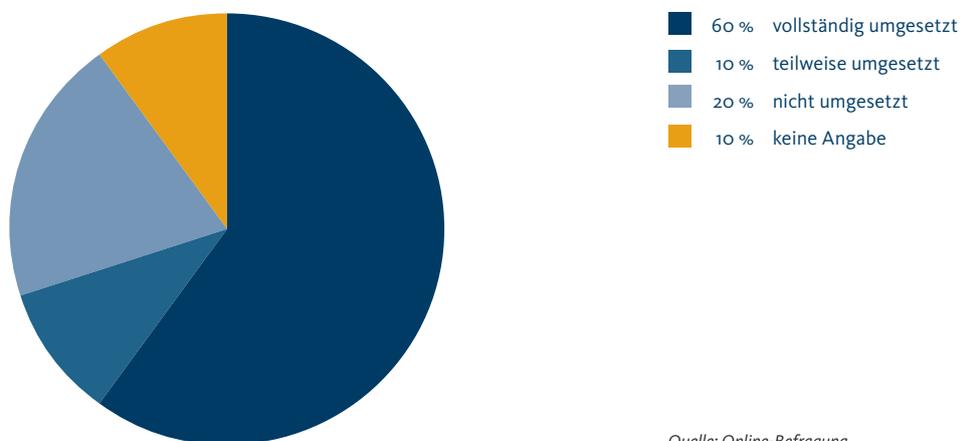
### Technische Ressourcen

Die ausgesprochenen Empfehlungen, die sich der Dimension „Technische Ressourcen“ zuordnen lassen, bezogen sich auf die Verbesserung des Zugangs zu Wissensmanagementsystemen und Evaluierungsliteratur. Dieser Dimension von Evaluierungspraxis konnte keine übergeordnete Empfehlung zugeordnet werden.

Neben finanziellen sind auch technische Ressourcen wichtig für die Verbesserung der Evaluierungspraxis. Technische Ressourcen sind jene, die Evaluierungsarbeit ermöglichen,

erleichtern oder verbessern. Beispiele sind der Zugang zu einer Statistik- oder Befragungssoftware, zu Evaluierungsliteratur oder zu einem internen Wissensmanagementsystem. Insgesamt bezogen sich 10 (6 Prozent) Handlungsempfehlungen auf diese Themen. Damit sind in dieser Dimension von Evaluierungspraxis die wenigsten Handlungsempfehlungen ausgesprochen worden. Die Hälfte dieser Empfehlungen richtete sich an die staatlichen Durchführungsorganisationen. Die Mehrheit der Empfehlungen (60 Prozent)<sup>38</sup> wurde geteilt und vollständig umgesetzt (jeweils 60 Prozent; n=6). Eine Empfehlung wurde teilweise, zwei gar nicht umgesetzt (vgl. Abbildung 38 und Abbildung 39).<sup>39</sup>

**Abbildung 38: Umsetzungsgrad der Empfehlungen (Technische Ressourcen)**



Quelle: Online-Befragung

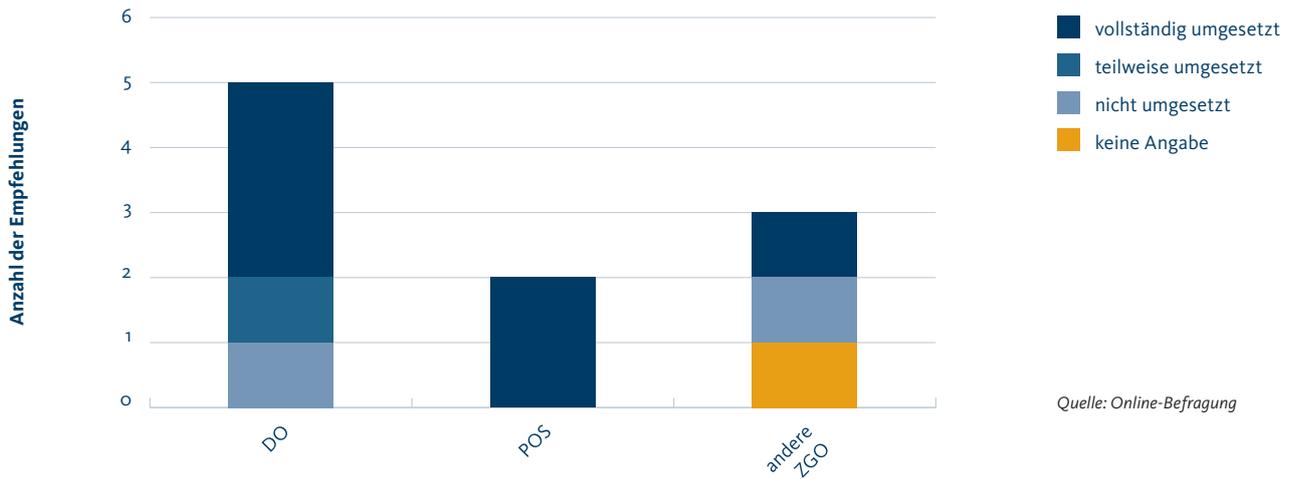
Die vollständig umgesetzten Handlungsempfehlungen bezogen sich inhaltlich u.a. auf den Abbau von technischen Zugangsbarrieren, um Evaluierungserkenntnisse – beispielsweise im Intranet der Organisation – verfügbar zu machen. Als Gründe für die zwei nicht umgesetzten Handlungsempfehlungen wurden von den beiden Organisationen im Freitextfeld der Online-Befragung die Koordination mit der IT-Abteilung und die Akquise neuer Software als hemmende Faktoren angegeben. Weiter wurde eine

mangelnde Übereinstimmung mit der Handlungsempfehlung als Grund genannt, was als nicht zielführend erachtet wurde. Im Abschlussworkshop äußerten sich Vertreterinnen und Vertreter der EZ-Organisationen dahingehend, dass bestimmte technische Ressourcen nicht dem Auftraggebenden zur Verfügung stehen müssten, sondern den beauftragten Gutachterinnen und Gutachtern.

<sup>38</sup> DO: 60 Prozent, POS: 100 Prozent, andere ZGO: 33 Prozent.

<sup>39</sup> Zu einer Handlungsempfehlung wurde in der Online-Befragung keine Angabe zur Umsetzung gemacht.

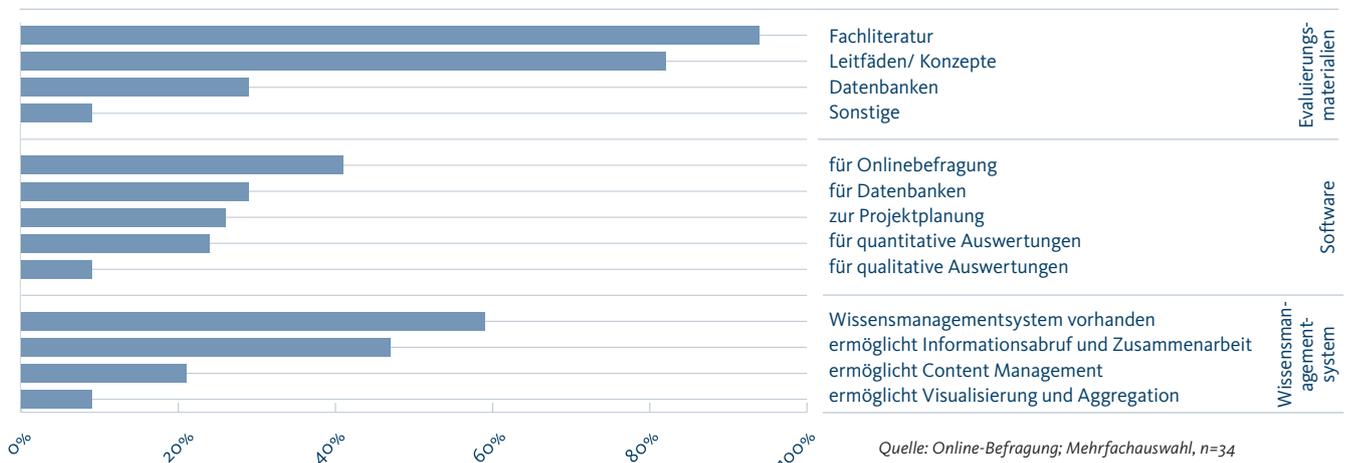
Abbildung 39: Umsetzungsgrad nach Akteursgruppe (Technische Ressourcen)



Technische Ressourcen werden überwiegend ausreichend bereitgestellt. Da in der Regel Evaluierungen nicht selbst durchgeführt werden, bestehen auch kaum spezielle Bedarfe. So verfügen beispielsweise einige Organisationen über Software für Online-Befragungen, die allerdings bislang kaum bis gar nicht genutzt werden. Generell ist der Zugang zu unterstützender Software, mit Ausnahme von Software zu Online-Befragungen, eher gering. Werden Evaluierungsressourcen

betrachtet, dann wird ersichtlich, dass fast alle Organisationen Zugang zu Fachliteratur und Evaluierungsleitfäden haben. Datenbanken werden von weniger als einem Drittel der Organisationen genutzt. 20 Organisationen (59 Prozent) verfügen zudem über ein Wissensmanagementsystem, das ihnen vornehmlich den Abruf von Informationen ermöglicht (vgl. Abbildung 40).

Abbildung 40: Verfügbarkeit von Evaluierungsmaterialien, Software und Wissensmanagementsystemen



### Zentrale Ergebnisse zu der Dimension „Technische Ressourcen“

Ziel der ausgesprochenen Empfehlungen war

- die Verbesserung des Zugangs zu Wissensmanagementsystemen und Evaluierungsliteratur.

Die 10 Empfehlungen wurden zu 60 Prozent geteilt und zu 60 Prozent vollständig umgesetzt.

Es besteht ein guter Zugang zu Fachliteratur und Leitfäden für Evaluierung. Zudem verfügt die Mehrheit der befragten Organisationen über Wissensmanagementsysteme (WMS), die überwiegend zum Abruf von Informationen genutzt werden und weniger für institutionelles Lernen.

# 4.

## FAZIT UND SCHLUSS- FOLGERUNGEN

**D**ie in diesem Bericht aufgezeigten Ergebnisse geben Aufschluss darüber, welche Entwicklungen die Durchführung der „Systemprüfung“ von 2009 und die in ihrem Zuge ausgesprochenen Empfehlungen im Evaluierungssystem der deutschen EZ gefördert oder angestoßen haben. Darüber hinaus wird deutlich, wie der aktuelle Stand der Evaluierungspraxis beschaffen ist.

Wird die Verteilung der 181 Handlungsempfehlungen auf einzelne Dimensionen von Evaluierungspraxis betrachtet, wird deutlich, dass sich mit 31 Prozent (n=56) die meisten Empfehlungen auf den Bereich „Strukturen und Prozesse“ bezogen. Mit 27 Prozent (n=49) thematisierte ein ebenfalls hoher Anteil der Empfehlungen die „Ziele von Evaluierungen“. Weitere 21 Prozent (n=38) der Empfehlungen bezogen sich auf „Humanressourcen“. Folglich wurden in diesen drei Kategorien nahezu vier Fünftel (79 Prozent) aller Empfehlungen ausgesprochen. Auf die weiteren Bereiche „Organisatorischer Kontext“ (9 Prozent, n=16), „Finanzielle Ressourcen“ (7 Prozent, n=12) und „Technische Ressourcen“ (6 Prozent, n=10) entfielen dagegen weniger Empfehlungen der „Systemprüfung“.

Das Umsetzungsmonitoring dieser 181 Handlungsempfehlungen ergab, dass diese nach Angaben der EZ-Organisationen überwiegend umgesetzt wurden: 54 Prozent – und somit mehr als die Hälfte aller Empfehlungen – wurden vollständig umgesetzt, 29 Prozent zumindest teilweise und lediglich 12 Prozent gar nicht. Zu 5 Prozent der Empfehlungen konnte der Umsetzungsgrad nicht ermittelt werden.

Diese mehrheitliche Umsetzung der Empfehlungen erklärt sich dadurch, dass ihnen überwiegend zugestimmt wurde: Wird die Zustimmung mit den Handlungsempfehlungen in den Umsetzungsplänen betrachtet, zeigt sich, dass die EZ-Organisationen diesen zu 59 Prozent vollständig und zu weiteren 33 Prozent zumindest teilweise zustimmten. Als Gründe für die nicht erfolgte Umsetzung von Handlungsempfehlungen wurden in den vertiefenden Interviews sowie im Abschlussworkshop des Projekts von einzelnen Organisationen daher auch darauf verwiesen, dass sie die Empfehlung von Anfang an nicht geteilt hätten oder aber deren Relevanz durch (insbesondere organisationsinterne) Entwicklungen der letzten Jahren nicht mehr gegeben sei.

Die Übereinstimmung der betreffenden Organisation mit den Empfehlungen begünstigte daher am ehesten deren Umsetzung. Als hemmender Faktor wurde am häufigsten die Verfügbarkeit finanzieller Mittel genannt. Dies gilt v.a. für Empfehlungen, die mit einem erheblichen finanziellen Mehraufwand verbunden waren (z.B. Schaffung einer neuen Stabsstelle, vermehrte Ex-post-Evaluierungen). Bei Empfehlungen, die mit geringen finanziellen Mitteln umsetzbar waren (z.B. Teilnahme an Netzwerktreffen oder Veröffentlichung von Evaluierungsberichten), hatten finanzielle Ressourcen hingegen keinen Einfluss.

Zusammenfassend hat das Umsetzungsmonitoring der „Systemprüfung“ gezeigt, dass es bemerkenswerte Weiterentwicklungen in der Evaluierungspraxis der deutschen Entwicklungszusammenarbeit gibt: Die ausgesprochenen Empfehlungen bewirkten oder förderten zahlreiche konkrete Maßnahmen und Prozesse in den jeweils untersuchten Organisationen. So kam es beispielsweise zu einer Einrichtung oder Verstärkung der neuen Evaluierungseinheiten, zu einer Erhöhung der verfügbaren Budgets für Evaluierungen sowie einer vermehrten Einführung von Evaluierungsrichtlinien. Zudem bezieht sich inzwischen die überwiegende Zahl der staatlichen wie der nicht-staatlichen EZ-Organisationen explizit auf internationale und nationale Standards, wie die des OECD DAC oder der DeGEval. Daneben zeigen sich – ebenso wie in den beiden bisherigen Systemprüfungen – auch Themenfelder, in denen Ansprüche der „Systemprüfung“ sowie der OECD DAC-Prinzipien und die Interessen der untersuchten EZ-Organisationen nicht vollständig deckungsgleich sind. Dies betrifft beispielsweise die in den Systemprüfungen geforderten Ex-post-Evaluierungen zur Überprüfung von Wirksamkeit und Nachhaltigkeit der Vorhaben sowie einen transparenten Umgang mit Evaluationsergebnissen. Durch die Einbettung von Evaluierungen in den Projektzyklus und einzelne Förderphasen wird gegenüber dem Geldgeber Rechenschaft abgelegt. Es mangelt dabei zum einen an Evaluierungsdesigns, die auf robusten Methoden und nachvollziehbaren Wirkungshypothesen basieren sowie an Wirkungs- und Ex-post-Evaluierungen, um die Wirksamkeit und Nachhaltigkeit der Maßnahmen zu belegen. Hierbei verweisen jedoch viele Organisationen darauf, dass sie nicht über die nötigen Mittel (Personal, Finanzierung oder Qualifikation) verfügen. Zum anderen fehlt es, z.T. aufgrund von

Schutzinteressen, an einer transparenten Veröffentlichungspraxis, die es ermöglicht, auch der Öffentlichkeit gegenüber Rechenschaft abzulegen. Gleichzeitig sollte Transparenz mit dem Ziel des Lernens aus Evaluierung nicht in Widerspruch geraten, indem beispielsweise bei einer vollständigen und unkommentierten Veröffentlichung von Berichten kritische Ergebnisse weniger thematisiert werden. Um den Prozess des Lernens aus Evaluierungen zu fördern, werden zudem stärker standardisierte organisationsinterne Prozesse benötigt.

Die zentralen Ergebnisse bzw. Schlussfolgerungen dieses Berichts zum aktuellen Stand der Evaluierungspraxis in der deutschen EZ sind nachfolgend zusammengefasst und gegliedert nach den betrachteten Dimensionen dargestellt:

### Ziele

1. Bei der Diversifizierung von Evaluierungsformen lässt sich feststellen, dass vermehrt Querschnittsevaluierungen durchgeführt werden. Bei einigen EZ-Organisationen bestehen Zweifel hinsichtlich der Zweckmäßigkeit von Ex-post- und Wirkungsevaluierungen für die eigene Arbeit oder sie verfügen nicht über die nötigen Mittel zu deren Durchführung. Ressourcenintensivere Ex-post-, Wirkungs- und Gemeinschaftsevaluierungen werden daher eher von den staatlichen Durchführungsorganisationen und weniger von zivilgesellschaftlichen Organisationen durchgeführt.
2. Evaluierungen werden den befragten Organisationen zufolge vor allem zur Verbesserung von Vorhaben bzw. zum Lernen genutzt. Dies geschieht v.a. durch Projektbeteiligte zur Programm-/ Projektsteuerung. Es fehlen häufig standardisierte Prozesse zur Nutzung projektübergreifender Evaluierungsergebnisse für institutionelles Lernen. Die Verpflichtung zur Rechenschaftslegung bindet Ressourcen für die Dokumentation von Projekterfolgen, die einige Organisationen gerne für konzeptionelles Lernen aus projektübergreifenden Evaluierungen nutzen würden.
3. Eine grundsätzliche Veröffentlichung von vollständigen Evaluierungsberichten ist nach wie vor sehr selten. Die zurzeit gängige Praxis in den meisten EZ-Organisationen ist die Veröffentlichung und Verbreitung von Kurzberichten. Zum Teil wird auf Anfrage der Gesamtbericht zur Verfügung gestellt. Unter anderem halten berechnigte Schutzinteressen gegenüber Partnern und die Sorge um Spendenaufkommen

Organisationen davon ab, Evaluierungsergebnisse zu veröffentlichen. Darüber hinaus wird befürchtet, dass vollkommene Transparenz mit dem Ziel des Lernens aus Evaluierung in Widerspruch gerät, da kritische Ergebnisse evtl. weniger thematisiert werden, wenn vollständige Berichte unkommentiert veröffentlicht werden.

### Organisatorischer Kontext

4. Die Aufgeschlossenheit von Mitarbeitenden hinsichtlich Evaluierungen ist vorhanden. Der Vernetzungsgrad der deutschen EZ-Organisationen im Bereich Evaluierung untereinander sowie mit internationalen Organisationen hat in den letzten Jahren deutlich zugenommen. Dies gilt in geringerem Ausmaß auch für die Zusammenarbeit mit der Wissenschaft.

### Strukturen und Prozesse

5. Evaluierungseinheiten/-stellen wurden seit der „Systemprüfung“ vielfach neu geschaffen bzw. gestärkt. Die Mitarbeitenden im Bereich Evaluierung verstehen sich überwiegend als Dienstleistende für und innerhalb der Organisationen und weniger in der Rolle einer Prüfungsinstanz.
6. Die Berücksichtigung der OECD DAC-Kriterien ist in den Evaluierungsrichtlinien der Organisationen der deutschen EZ verankert. Die Umsetzung in den Evaluierungsberichten ist jedoch nicht konsequent.
7. Die Organisationen der deutschen EZ haben ihre Evaluierungspraxis formalisiert und systematisiert; Evaluierungen folgen dabei überwiegend Evaluierungsrichtlinien und -plänen. Der Formalisierungsgrad variiert je nach Prozessphase: Die Umsetzung bzw. Nutzung der Evaluierungsergebnisse ist am geringfügigsten formalisiert, wenngleich einige Organisationen Prozesse zur Bewertung und Umsetzung von Evaluierungsergebnissen haben.
8. Hinsichtlich der Einbindung von Partnerorganisationen der Empfängerländer in den Evaluierungsprozess lässt sich feststellen, dass es stark von der Art der Kooperation abhängt, ob und in welchem Maße Partnerorganisationen eingebunden werden. Generell kann konstatiert werden, dass eine systematische und gleichberechtigte Einbeziehung von Partnern im Evaluierungsprozess nicht die Regel ist.

### Humanressourcen

9. Seit der „Systemprüfung“ werden mehr Personalressourcen für Evaluierung eingesetzt. Zugleich haben sich oftmals auch die an Evaluierungen gestellten methodischen Ansprüche erhöht, was sie arbeits- und betreuungsintensiver werden lässt.
10. Die fachliche Kompetenz in den Evaluierungseinheiten wurde erhöht. Die Qualität von Evaluierungen hängt aber nicht allein von den Evaluierungseinheiten ab, sondern auch von der Kompetenz der beauftragten Gutachterinnen und Gutachter und der operationellen Bereiche und Partner, sofern diese Evaluierungen managen. Es besteht nicht überall die Möglichkeit (v.a. aufgrund der Vielzahl von Evaluierungen in einigen Organisationen), eine systematische Qualitätssicherung durch die Evaluierungseinheit vorzunehmen.
11. Bei dem Einsatz externer Gutachterinnen und Gutachter, ist festzuhalten, dass fast alle untersuchten EZ-O diese vermehrt beauftragen und diesen auch die volle Berichtsverantwortung überlassen. Nach Ansicht einzelner Organisationen ist das methodische Repertoire von Gutachterinnen und Gutachtern ausbaufähig. Länder- und organisationsspezifische Kenntnisse werden bei der Auswahl oftmals stärker berücksichtigt.
12. Den Aufbau von Evaluierungskapazität in Partnerorganisationen sehen viele deutsche EZ-Organisationen nicht in ihrem Verantwortungsbereich, obgleich sich einige dieser Organisationen sehr stark in diesem Bereich engagieren.
13. Evaluierungsdesigns basieren häufig nicht auf überprüfbareren Wirkungshypothesen. Das verwendete Methodenspektrum bzw. die Nutzung eines angemessenen Methodenmixes ist ausbaufähig.

### Finanzielle Ressourcen

14. Seit der „Systemprüfung“ sind die Mittel, die in deutschen EZ-Organisationen für Evaluierungen eingesetzt werden, gestiegen. Die zur Verfügung gestellten finanziellen Ressourcen werden von den Befragten überwiegend als ausreichend eingestuft. Diese Aussage gilt allerdings nicht mehr, wenn methodisch anspruchsvolle oder projektübergreifende Evaluierungen durchgeführt werden sollen.

### Technische Ressourcen

15. Mitarbeitende in deutschen EZ-Organisationen haben einen guten Zugang zu Fachliteratur und Leitfäden zum Themenbereich Evaluierung. Zudem verfügt die Mehrheit der befragten Organisationen über Wissensmanagementsysteme, die zum Abruf von Informationen genutzt werden und weniger für institutionelles Lernen.

Zusammenfassend lässt sich feststellen, dass die Weiterentwicklung der Evaluierungspraxis in der deutschen EZ in den letzten Jahren beachtliche Fortschritte gemacht hat. Auch das DEval hat sich der Weiterentwicklung der Evaluierung von Vorhaben der deutschen EZ verschrieben. In der „Ausgangserhebung zum DEval und seinem Umfeld“ wurde deutlich, dass viele EZ-Organisationen dieses Ziel teilen und Evaluierung nicht als Selbstzweck verstehen, sondern als ein wichtiges Instrument der Weiterentwicklung erachten, das zielführend einzusetzen ist.



# LITERATUR

Alle Internetquellen wurden zuletzt am 12.03.2015 abgerufen.

**BMZ (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung) (2002)**, *Richtlinien für die Förderung von Maßnahmen der Gesellschafts- und Sozialstrukturpolitik*, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn.

**BMZ (2006)**, *Evaluierungskriterien für die deutsche bilaterale Entwicklungszusammenarbeit. Eine Orientierung für Evaluierungen des BMZ und der Durchführungsorganisationen*, Evaluierungsreferat des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn. [http://www.bmz.de/de/zentrales\\_downloadarchiv/erfolg\\_und\\_kontrolle/evaluierungskriterien.pdf](http://www.bmz.de/de/zentrales_downloadarchiv/erfolg_und_kontrolle/evaluierungskriterien.pdf)

**BMZ (2007)**, *Richtlinien für die Förderung entwicklungswichtiger Vorhaben privater deutscher Träger in Entwicklungsländern aus Kapitel 2302 Titel 687 06*, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn. [http://bengo.engagement-global.de/aktuelle-mitteilung/items/fazilitaet-emissionsminderung-anpassung-an-den-klimawandel-klimaschutz-wald-und-biodiversitaetserhalt.html?file=tl\\_files/microsite/\\_media/content/bengo/PDF-Dokumente/BMZ-Foerderrichtlinien\\_private-Traeger.pdf](http://bengo.engagement-global.de/aktuelle-mitteilung/items/fazilitaet-emissionsminderung-anpassung-an-den-klimawandel-klimaschutz-wald-und-biodiversitaetserhalt.html?file=tl_files/microsite/_media/content/bengo/PDF-Dokumente/BMZ-Foerderrichtlinien_private-Traeger.pdf)

**BMZ (2008)**, *Leitlinien für die bilaterale Finanzielle und Technische Zusammenarbeit mit Kooperationspartnern der deutschen Entwicklungszusammenarbeit*, BMZ Konzepte 165, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn. [http://www.bmz.de/de/mediathek/publikationen/themen/entwicklungspolitik\\_allgemein/konzept165.pdf](http://www.bmz.de/de/mediathek/publikationen/themen/entwicklungspolitik_allgemein/konzept165.pdf)

**BMZ (2009a)**, *Erfolgskontrolle bei Zuwendungen. Handreichung für Mitarbeiterinnen und Mitarbeiter des BMZ*, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn.

**BMZ (2009b)**: *Wirkungsevaluierungen. Zum Stand der internationalen Diskussion und dessen Relevanz für die Evaluierung der deutschen Entwicklungszusammenarbeit*. Evaluierungsreferat des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung: Evaluation Working Papers. Verfügbar unter: [http://www.bmz.de/de/zentrales\\_downloadarchiv/erfolg\\_und\\_kontrolle/Evaluation\\_Working\\_Papers/BMZ\\_WP\\_Wirkungsevaluierung\\_2009.pdf](http://www.bmz.de/de/zentrales_downloadarchiv/erfolg_und_kontrolle/Evaluation_Working_Papers/BMZ_WP_Wirkungsevaluierung_2009.pdf)

**BMZ (2013)**, *Leitlinien zur ziel- und wirkungsorientierten Erfolgskontrolle im Bereich der Förderung entwicklungswichtiger Vorhaben der politischen Stiftungen*, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn.

**BMZ (2014)**, *Umsetzungsplan Systemprüfung BMZ-Stand 03.06.2014*, nicht veröffentlicht.

**Borrmann, A., K. Fasbender, M. Holthus, A. von Gleich, B. Reichl und R. Shams (1999)**, *Erfolgskontrolle in der deutschen Entwicklungszusammenarbeit. Studie im Auftrag des Bundesministeriums für Wirtschaftliche Zusammenarbeit und Entwicklung*, NOMOS-Verlags-gesellschaft, Baden-Baden.

**Borrmann, A. und R. Stockmann (2009a)**, *Evaluation in der deutschen Entwicklungszusammenarbeit, Band 1 – Systemanalyse*, Waxmann, Münster.

**Borrmann, A. und R. Stockmann (2009b)**, *Evaluation in der deutschen Entwicklungszusammenarbeit, Band 2 – Fallstudien*, Waxmann, Münster.

**Caspari, A. (2006)**, „Partizipative Evaluationsmethoden – zur Entmystifizierung eines Begriffs der Entwicklungszusammenarbeit“, in U. Flick (Hrsg.), *Qualitative Evaluationsforschung. Konzepte – Methoden – Umsetzungen*, Rowohlt, Reinbek bei Hamburg, S. 365-384.

**Caspari, A. (2010)**, *Lernen aus Evaluierungen. Meta-Evaluation & Evaluationssynthese von InWEnt-Abschlussequalulierungen 2009*, InWEnt, Bonn.

**Caspari, A. und R. Barbu (2008)**, *Wirkungsevaluierung: Zum Stand der internationalen Diskussion und dessen Relevanz für Evaluierungen der deutschen Entwicklungszusammenarbeit*, Evaluation Working Papers, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, Bonn. [http://www.bmz.de/de/zentrales\\_downloadarchiv/erfolg\\_und\\_kontrolle/Evaluation\\_Working\\_Papers/BMZ\\_WP\\_Wirkungsevaluierung\\_2009.pdf](http://www.bmz.de/de/zentrales_downloadarchiv/erfolg_und_kontrolle/Evaluation_Working_Papers/BMZ_WP_Wirkungsevaluierung_2009.pdf)

**Court, J., I. Hovland und J. Young (Hrsg.) (2005)**, *Bridging Research and Policy in Development: Evidence and the Change Process*, Overseas Development Institute/ ITDG Publishing, United Kingdom.

**DeGEval (Gesellschaft für Evaluation e.V.) (Hrsg.) (2008)**, *Standards für Evaluation*, DeGEval, Mainz.

**DEval (2015a)**, *Erhebung zur Ausgangslage des DEval und seines Umfeldes*, DEval, Bonn.

**DEval (2015b)**, *Anlagenband zur „Ausgangserhebung zum DEval und seinem Umfeld“*, DEval, Bonn.

**European Commission (2006)**, *Evaluation Methods for the European Union's External Assistance. Guidelines for Project and Programme Evaluation*, Office for Official Publications of the European Communities, Luxemburg. [http://ec.europa.eu/europeaid/evaluation/methodology/examples/guide3\\_en.pdf](http://ec.europa.eu/europeaid/evaluation/methodology/examples/guide3_en.pdf)

**Meyer, W. (2007)**, *Datenerhebung: Befragungen – Beobachtungen – Nicht-reaktive Verfahren*, in R. Stockmann (Hrsg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*. Waxmann, Münster, S. 223-277.

**Neubert, S., D. Mack und H. Roxin (2014)**, *Vom Methodenstreit zum Methodenmix – Chronologie und Stand der Entwicklungsevaluierung in Deutschland*, in W. Böttcher et al. (Hrsg.), *Evaluation in Deutschland und Österreich. Stand und Entwicklungsperspektiven in den Arbeitsfeldern der DeGEval – Gesellschaft für Evaluation*, Waxmann, Münster, S.61-72.

**OECD (Organisation for Economic Cooperation and Development) (2010)**, *Quality Standards for Development Evaluation*, DAC Guidelines and Reference Series OECD Publishing, Paris. <http://www.oecd.org/dac/evaluation/qualitystandards.pdf>

**Scriven, M. (2010)**, *Evaluating Evaluations: A Meta-Evaluation Checklist*. <http://dme4peace.org/sites/default/files/EvaluatingEvals-Checklist.pdf>

**Weltbank (2004)**, *Monitoring & Evaluation: Some Tools, Methods & Approaches*, Weltbank, Washington D.C.

**Widmer, T. (1996)**, *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*, Haupt, Bern.





DEval  
Deutsches Evaluierungsinstitut der  
Entwicklungszusammenarbeit

Fritz-Schäffer-Straße 26  
D-53113 Bonn

Tel: +49 (0)228 33 69 07-0  
Fax: +49 (0)228 33 69 07-904

Mail: [info@DEval.org](mailto:info@DEval.org)  
[www.DEval.org](http://www.DEval.org)



**DEval**

DEUTSCHES  
EVALUIERUNGsinstitut  
DER ENTWICKLUNGS-  
ZUSAMMENARBEIT

---