



WISSEN WAS WIRKT

WISSEN WAS WIRKT
WISSEN
WISSEN WAS WIRKT

Querschnittsauswertung (QSA) von Projektevaluierungen (PEV) 2015 – Meta- Evaluierung

Bericht

Teil I: Methodische Qualität (Genauigkeit)

Teil II: Prozessqualität (Durchführbarkeit
und Fairness)

Teil III: Nützlichkeit

Als Bundesunternehmen unterstützt die GIZ die deutsche Bundesregierung bei der Erreichung ihrer Ziele in der Internationalen Zusammenarbeit für nachhaltige Entwicklung.

Herausgeber

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH
T +49 228 44 60-0 (Bonn)
T +49 61 96 79-0 (Eschborn)

Sitz der Gesellschaft
Bonn und Eschborn

Friedrich-Ebert-Allee 40
53113 Bonn
T +49 228 44 60-1877
F +49 228 44 60-2877

Dag-Hammarskjöld-Weg 1-5
65760 Eschborn
T +49 61 96 79-1408
F +49 61 96 79-80 1408

E evaluierung@giz.de
I www.giz.de/monitoring

Verantwortlich

Dr. Ricardo Gomez (GIZ, Leiter Stabsstelle Evaluierung)

Projektleitung

Dr. Annette Backhaus (GIZ, Gruppenleiterin Konzeption, Steuerung und Qualitätssicherung, Stabsstelle Evaluierung)

Konzeption und Koordination

Dr. Vera Hundt (GIZ, Fachkonzeptionistin, Stabsstelle Evaluierung)
Dr. Alexander Erich (GIZ, Senior-Fachkonzeptionist, Stabsstelle Evaluierung)

Autoren/Autorinnen

Dieses Gutachten wurde von unabhängigen externen Sachverständigen erstellt.
Es gibt ausschließlich deren Meinung und Wertung wieder.
Isabelle Freimann (Syspons GmbH)
Matías Krämer (Syspons GmbH)

Evaluierungsteam

Isabelle Freimann (Syspons GmbH)
Matías Krämer (Syspons GmbH)
Lennart Raetzell (Syspons GmbH)
Lukas Verfürden (Syspons GmbH)
Olga Almqvist (Syspons GmbH)

Erscheinungsort und -jahr

Bonn und Eschborn 2016

Executive Summary

Die Stabsstelle Evaluierung (STS 08) der Deutschen Gesellschaft für Internationale Zusammenarbeit (GIZ) führt seit einigen Jahren Meta-Evaluierungen von Evaluierungsberichten durch. Die Meta-Evaluierungen dienen dem organisationalen Lernen der GIZ. Bis 2014 wurden in die Meta-Evaluierungen die Berichte der Unabhängigen Evaluierungen (UE), der Projektfortschrittskontrollen (PFK) sowie weiterer dezentral gesteuerter Evaluierungen einbezogen. Im Januar 2014 wurden die UE und PFK vom dezentralen Instrument der Projektevaluierung (PEV) abgelöst. Mit der Einführung der PEV wurden Änderungen an dem bisherigen Verfahren der PFK vorgenommen. Während PFK allein den Anspruch eines internen Beratungsinstruments hatten, wird an PEV auch der Anspruch der Rechenschaftslegung und Transparenz gestellt. Gleichzeitig wurden der methodische und evaluatorische Anspruch an PEV im Vergleich zu den PFK erhöht. PEV wurden gemeinsam mit einer Handreichung und verschiedenen Unterstützungsmaterialien eingeführt und durch Informationsveranstaltungen und Fortbildungen begleitet. Nach der flächendeckenden Einführung der PEV im April 2014 in allen Vorhaben, die im BMZ-Auftrag durchgeführt werden und ein Mindestvolumen von 1 Mio. Euro aufweisen, wurde nun die Qualität der PEV durch eine Meta-Evaluierung überprüft.

Zielsetzung und methodische Grundlagen der Metaevaluierung

Das **Ziel der Meta-Evaluierung** war es, die Qualität der PEV anhand eines umfassenden Qualitätsverständnisses zu untersuchen, welches sowohl die methodische Qualität als auch die Prozessqualität und die Nützlichkeit und den Nutzen des Instruments erhebt, und Ansatzpunkte für die Qualitätsverbesserung des PEV-Instruments zu identifizieren. Dazu wurde die PEV-Qualität auf der Grundlage anerkannter (inter-)nationaler Evaluierungsstandards geprüft (insbesondere der Evaluierungsstandards der Deutschen Gesellschaft für Evaluation – DeGEval).

Darüber hinaus dienen die Ergebnisse dieser Meta-Evaluierung als Baseline für die zukünftige Weiterentwicklung des Instruments. Demnach werden die nachfolgend dargestellten Ergebnisse als Referenzpunkt für zukünftige Meta-Evaluierungen der PEV dienen. Damit entsteht ein Monitoringsystem, mit dem die Qualitätsentwicklung des PEV-Instruments erfasst und die Wirkungen von Veränderungen im Instrument auf dessen Qualität nachvollzogen werden kann.

Die Meta-Evaluierung basiert methodisch auf der **Textanalyse** von PEV-Berichten und einer **Online-Befragung von Auftragsverantwortlichen (AV)**. Die Textanalyse basierte auf einer repräsentativen Stichprobe von PEV-Berichten, die zwischen Oktober 2014 und September 2015 erstellt wurden. Die Online-Befragung der AV ermöglichte die Prüfung von Qualitätsaspekten, die sich einer Textanalyse entziehen (v.a. Aspekte der Prozessqualität und der Nützlichkeit). Auch in der Online-Befragung wurden alle AV, die zwischen Oktober 2014 und September 2015 eine PEV beauftragt haben, um ihre Teilnahme gebeten. Der hohe Rücklauf ermöglicht auch hier repräsentative Aussagen.

Dem verwendeten Bewertungssystem zur Prüfung der PEV-Qualität liegt der **Anspruch einer vollständigen Zielerreichung** zugrunde. Dies bedeutet, dass es Anspruch der GIZ ist, alle (inter-)nationalen Qualitätsstandards für Evaluierungen vollständig zu erfüllen. Um darauf aufbauend die Bewertungen einheitlich und nachvollziehbar zu gestalten, wird die **Erfüllung von Standards wie folgt angegeben**: Erfüllung zu 90-100 % (größtenteils bis vollständig erfüllt); Erfüllung zu 80-89 % (zumeist erfüllt); Erfüllung zu 60-79 % (eher erfüllt); Erfüllung geringer als 60 % (bedingt erfüllt).

Zusammenfassung der wesentlichen Ergebnisse und Handlungsempfehlungen

Die Meta-Evaluierung zeigt, dass das PEV-Instrument bislang die (inter-)nationalen Qualitätsstandards für Evaluierungen eher erfüllt. Über alle Evaluierungsstandards hinweg erfüllt die PEV die geprüften

Qualitätsstandards für Evaluierungen zu 68 %. Die Qualität der PEV variiert jedoch stark im Hinblick auf die methodische Qualität (Evaluierungsstandard Genauigkeit), die Prozessqualität (Evaluierungsstandards Durchführbarkeit und Fairness) und die Nützlichkeit (Evaluierungsstandard Nützlichkeit). Die deutlichsten Schwächen weist die PEV hinsichtlich der methodischen Qualität auf, während die Nützlichkeit positiv einzuschätzen ist.

Ergebnisse und Empfehlungen zur methodischen Qualität (Standard Genauigkeit)

Die Evaluierungsstandards zur methodischen Qualität sollen sicherstellen, „*dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt*“ (DeGEval 2008). Dafür wurden in der Meta-Evaluierung die DeGEval-Kriterien zur Genauigkeit überprüft, deren Einhaltung die methodische Qualität der PEV gewährleisten soll. Das zentrale Ergebnis ist, dass die methodische Qualität der PEV in der Breite schwach ist (Erfüllungsgrad 59 %) und angesichts des GIZ-Qualitätsanspruchs, die Standards vollständig zu erfüllen, ein wesentliches Handlungsfeld für die Weiterentwicklung der PEV darstellt.

Zu den zentralen methodischen Schwächen zählen der **geringe Reflexionsgrad** zur methodischen Vorgehensweise und die eingeschränkte Methodenvielfalt. Wie die Meta-Evaluierung zeigt, stellen PEV nur in Ausnahmen ihre methodische Vorgehensweise mit Hinweis auf die Beantwortung von Evaluierungsfragen oder die Besonderheiten des Vorhabens begründet dar. Dies führt dazu, dass zum einen die methodische Bearbeitung der OECD-DAC-Kriterien schlecht bewertet wird (u.a. wird die Zielerreichung der Vorhaben teils auf Grundlage nicht geeigneter Indikatoren gemessen) und zum anderen eine **eingeschränkte Auswahl an Evaluierungsmethoden** eingesetzt wird. Zumeist beschränken sich PEV auf qualitative Interviews sowie Daten- und Dokumentenanalysen. Diese Einschränkung wirkt sich jedoch negativ auf die methodische Gesamtqualität der PEV aus. Hierbei zeigen sich aus Sicht des Meta-Evaluierungsteams auch die Auswirkungen einer fehlenden Inception-Phase im PEV-Prozess, in der das Evaluierungsdesign abwägend entwickelt wird.

Empfehlung 1: Die Handreichung und die Unterstützungsmaterialien der PEV greifen bereits eine Vielzahl der Vorgaben der Evaluierungsstandards auf. Allerdings sollten die Unterstützungsmaterialien der PEV auch die Auseinandersetzung und Reflexion mit methodischen Entscheidungen der PEV als Analysefragen und zu bewertende Aspekte benennen. Andernfalls bleibt die Nachvollziehbarkeit der Analysen der PEV eingeschränkt.

Empfehlung 2: Die Unterstützungsmaterialien der PEV sollten stärker auf die Bedeutung der methodischen Vielfalt der PEV für die methodische Qualität einer PEV hinweisen.

Empfehlung 3: Die in der Meta-Evaluierung identifizierten Einflussfaktoren für die methodische Qualität sollten in der Auftragsklärung einer PEV thematisiert werden.

Insgesamt sollte kritisch reflektiert werden, dass die Ergebnisse zur methodischen Qualität einer PEV, die durchschnittlich auf eine bedingte Erfüllung hinweisen, keinen Einfluss auf die Einschätzungen zur Nützlichkeit durch die AV haben. Zudem liegt eine Diskrepanz zwischen der bedingten Erfüllung des Genauigkeitsstandards und der Einschätzung der Methodenkompetenz der PEV-Teams durch die AV vor. Dies könnte gegebenenfalls damit zusammenhängen, dass die AV die methodische Kompetenz von PEV-Teams nicht adäquat einschätzen oder der methodischen Qualität nicht ausreichend Aufmerksamkeit schenken. Dies wiederum führt dazu, dass die AV die PEV-Ergebnisse unabhängig von ihrer methodischen Qualität nutzen (zum Beispiel für Folgemaßnahmen). Zum anderen wird auch deutlich, dass der Anspruch, GIZ-Vorhaben auf der Basis eines evidenz-informierten Handelns umzusetzen, unzureichend greift.

Empfehlung 4: Die methodische Qualität könnte durch den Aufbau eines Pools von Expert/innen befördert werden, die die AV bei der Wahl geeigneter externer Evaluator/innen unterstützen. Bei der Auswahl der Expert/innen könnte bewusst auf Profile gesetzt werden, die ein breites Spektrum sowohl qualitativer als auch quantitativer Methoden abdecken, um die Methodenvielfalt zu fördern. Die AV

könnten gleichzeitig durch Profilbeschreibungen unterstützt werden, die aufzeigen, welche Kenntnisse Evaluations-Expert/innen mitbringen sollten.

Empfehlung 5: Zugleich sollten die TOR deutlicher machen, welche Kompetenzprofile gesucht werden, und darauf aufbauend auch die Verantwortung der Evaluierungs-Expert/innen im PEV-Prozess stärken.

Empfehlung 6: Da Mitarbeiter/innen des Fach- und Methodenbereichs (FMB) in der Regel Teil der PEV-Teams sind, könnten Investitionen in deren Training und Weiterbildung eine Möglichkeit sein, um eine Sensibilisierung für grundlegende methodische Aspekte der PEV zu erreichen, ohne zugleich die Verantwortung der Evaluierungs-Expert/innen im PEV-Prozess zu unterminieren.

Eine weitere wichtige Erkenntnis aus der Meta-Evaluierung ist, dass die **Anzahl der veranschlagten Tage für PEV-Gutachter/innen** einen Einfluss auf die methodische Qualität hat. PEV unter 55 Gutachtertage fallen methodisch schwächer aus als PEV mit mehr Gutachtertage. Dies gilt unabhängig von dem Auftragsvolumen eines Vorhabens. Zugleich ist methodische Qualität von PEV mit zunehmenden Gutachtertage bis zu einem Mengengerüst von circa 86 Tagen tendenziell besser wird. Ab diesem Wert ist der Effekt jedoch nicht mehr vorhanden bzw. sogar gegenläufig. Diesbezüglich gilt zu berücksichtigen, dass die Streuung der untersuchten PEV ab diesem Wert verhältnismäßig hoch und die Fallzahlen verhältnismäßig klein sind. Ein ähnlicher Trend lässt sich für die **Größe des PEV-Teams** erkennen. Die methodische Qualität einer PEV ist höher, wenn mehr als zwei Gutachter/innen an der Evaluierung mitwirken. Dieses zentrale Ergebnis zeigt auf, dass es für die Steigerung der methodischen Qualität der PEV von Nutzen sein könnte, den AV Orientierungspunkte für den Umfang des Mengengerüsts einer PEV und der Größe des PEV-Teams zu geben.

Empfehlung 7: Es wird empfohlen, Orientierungspunkte für den Ressourceneinsatz einer PEV zu formulieren. Dies betrifft sowohl die Anzahl der Gutachter/innen (mehr als zwei Gutachter/innen) als auch die Anzahl der Gutachtertage (mehr als rund 55 Tage). Hierbei sollte berücksichtigt werden, dass diese Orientierungspunkte größtenteils von PEV mit Folgevorhaben abgeleitet wurden.

Weitere relevante methodische Schwächen zeigen die PEV in der Analyse und Bewertung einzelner OECD-DAC-Kriterien. Die Meta-Evaluierung verdeutlicht, dass PEV in ihren **Impact**-Untersuchungen kaum Bezug zu den relevanten EZ-Programmen nehmen und diesen somit eine geringe Bedeutung zuschreiben. Zugleich schneidet die Untersuchung der **Effizienz** als schwächstes OECD-DAC-Kriterium ab. Es besteht hierbei scheinbar teils kein klares grundlegendes Verständnis der PEV-Teams darüber, was „Effizienz“ bedeutet und wie sie gemessen werden kann. Beide Kriterien sind zugleich jedoch von unternehmensstrategischer Bedeutung. Das BMZ ist verstärkt daran interessiert ist, in Zukunft die Beiträge der Vorhaben zu EZ-Programmzielen zu messen. Zudem bestehen seit geraumer Zeit Diskussionen zum Nachweis von „Value for Money“ in der EZ, die verstärkt im BMZ aufgegriffen werden. Schließlich ist bei den Effektivitätsanalysen auffällig, dass zwar die Qualität der Indikatoren der Vorhaben (nach SMART) zumeist von den PEV-Teams kritisiert wird, doch wird daraufhin nur sehr selten auch die geforderte Anpassung der Indikatoren vorgenommen.

Empfehlung 8: Es wird empfohlen, in den Unterstützungsmaterialien und in den Trainings für PEV-Teams die Bedeutung der Einbettung der Vorhaben in EZ-Programme hervorzuheben.

Empfehlung 9: Es sollte verstärkt in die Qualität der Effizienzuntersuchungen investiert werden. Demnach sollten in den Unterstützungsmaterialien zu den OECD-DAC-Kriterien die Erwartungen an eine Effizienzanalyse geschärft werden, u.a. um darzulegen, welche Effizienzuntersuchung (Implementierungs-, Produktions- oder Allokationseffizienz) mit welcher Datengrundlage möglich ist.

Empfehlung 10: Es sollte im Rahmen einer PEV eine Analyse der Allokationseffizienz durchgeführt werden, um zu verdeutlichen, wie in der Effizienzanalyse eines Vorhabens der GIZ Ressourcen und Wirkungen zueinander in Bezug gesetzt werden können.

Empfehlung 11: Es sollte eine Diskussion mit Gutachter/innen des FMB, die zuletzt PEV durchgeführt haben, darüber erfolgen, warum als nicht-SMART identifizierte Indikatoren oftmals nicht umformuliert

werden. Zugleich könnten Unterstützungsmaterialien der PEV Beispiele aufzeigen, wie Indikatoren umformuliert werden, ohne deren Inhalte und damit die Ansprüche der Vorhaben zu verändern.

Aus Sicht des Meta-Evaluierungsteams ist ein weiteres zentrales Ergebnis der Meta-Evaluierung, dass **nicht-bilaterale Vorhaben** methodisch signifikant schlechter abschneiden als bilaterale Vorhaben. Dies konnte u.a. darauf zurückgeführt werden, dass das Instrument der PEV eher auf bilaterale Vorhaben zugeschnitten ist. Derzeit sind die PEV vom spezifischen Wirkungsverständnis und Verständnis des Kooperationsystems von bilateralen Vorhaben geprägt. Die PEV berücksichtigt aus Sicht des Meta-Evaluierungsteams nicht ausreichend, dass Sektorvorhaben nicht auf der Outcome-Ebene gesteuert werden können, u. a. weil der Partner zugleich der Auftraggeber ist.

Empfehlung 12: Die Handreichung und die Unterstützungsmaterialien zur annotierten Berichtsgliederung sollten durch spezifische Hinweise auf die Wirkungsanforderungen von Sektorvorhaben ergänzt werden.

Ergebnisse und Empfehlungen zur Prozessqualität (Standards Durchführbarkeit und Fairness)

Die Überprüfung der Prozessqualität der PEV wurde entlang von zwei Evaluierungsstandards durchgeführt. Als Erstes wurde die Qualität der Durchführbarkeit untersucht. *„Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird“* (DeGEval 2008). Die hierunter gefassten Kriterien befassen sich mit der Qualität der Verfahren und der Effizienz der PEV. Im Ergebnis werden diese Kriterien durch das PEV-Instrument eher erfüllt (63 %), sodass die Herausforderungen dominieren.

Als Zweites wurde die Prozessqualität anhand des Evaluierungsstandards der Fairness geprüft. *„Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird“* (DeGEval 2008). Dazu zählen unter anderem Kriterien wie die Offenlegung von Ergebnissen und der Schutz individueller Rechte der Beteiligten und Betroffenen. Hier zeigt das PEV-Instrument mit einem Erfüllungsgrad von 76 % seine Stärken.

Mit Blick auf die Prozessqualität in der Durchführbarkeit ist die zentrale Erkenntnis, dass die **Aufwände und Kosten** einer PEV von den AV kontrovers und insgesamt durchwachsen eingeschätzt werden. Die Angemessenheit der Verfahren wird insbesondere in Bezug auf die zeitlichen Aufwände für die Mitarbeiter/innen der evaluierten Vorhaben hinterfragt. Zugleich wird der finanzielle Aufwand einer PEV kontrovers seitens der AV diskutiert. Insgesamt zeigt die Meta-Evaluierung, dass die AV bei ihrer Einschätzung zur Effizienz der PEV-Verfahren primär die verursachten Kosten und Aufwände diskutieren und diese selten in Relation zum Nutzen setzen, gleichwohl dieser Nutzen von den AV im Rahmen der Meta-Evaluierung positiv eingeschätzt wird. Aus Sicht des Meta-Evaluierungsteams muss diese starke Aufmerksamkeit der AV auf die Aufwände, auch vor dem Hintergrund der Ergebnisse zur methodischen Qualität, kritisch reflektiert werden. Diese zeigten, dass eine zu geringe Investition in ein ausreichend großes Mengengerüst und Team sich negativ auf die methodische Qualität der PEV auswirkt.

Empfehlung 13: In Abstimmung mit dem BMZ sollte eine Diskussion über die Aufwände für M&E (Monitoring & Evaluierung) im Rahmen einer eigenen Budgetlinie erfolgen. Durch eine eigene Budgetlinie könnte zum einen sichergestellt werden, dass genug Ressourcen für die PEV zur Verfügung stehen und zugleich der zum Teil negativen Wahrnehmung zu den Kosten und Aufwänden entgegengewirkt werden.

Empfehlung 14: Die unternehmensstrategische Bedeutung von M&E könnte durch verbindliche und international akzeptierte Orientierungswerte gesteigert werden, die von anderen Gebern bereits genannt werden (i.d.R. 3 bis 5 % des Projektbudgets).

Ein weiteres zentrales Ergebnis der Meta-Evaluierung ist, dass die **Handreichung und die Unterstützungsleistungen der Stabsstelle (STS) 08** nicht in der Breite den gewünschten Effekt auf den PEV-Prozess entfalten konnten. Die Meta-Evaluierung untersuchte, inwieweit sich die den PEV-Teams bereitgestellte Handreichung und die Unterstützungsmaterialien positiv auf die Durchführbarkeit auswirken. Im Ergebnis weist die Meta-Evaluierung auf einen Überarbeitungsbedarf hin. Zwar wurde festgestellt, dass die Inhalte der Handreichung und der Unterstützungsmaterialien die meisten

methodischen Anforderungen enthalten und auch teils als hilfreich für die Durchführung angesehen werden. Allerdings werden die Handreichung und Unterstützungsmaterialien von den AV gleichzeitig oftmals als zu „überladen“ und aufgrund der Vielzahl der Materialien als unübersichtlich wahrgenommen.

Empfehlung 15: Es wird empfohlen, die Nutzerfreundlichkeit der Unterstützungsmaterialien zu erhöhen. Hierzu könnte die zentrale Handreichung weiterentwickelt werden (Starter-Kit PEV, FAQ PEV o. Ä.), dabei entlang des PEV-Umsetzungsprozesses strukturiert werden und die Unterstützungsmaterialien an geeigneten Stellen verlinkt.

Auch hinsichtlich der pilothaft eingeführten **Qualitäts-Checks** der PEV-Berichte durch die StS 08 zeigen sich Schwächen. Bislang sind diese Qualitäts-Checks als eine „end-of-pipe“-Qualitätssicherung der PEV-Berichte angelegt, mit deren Hilfe die AV und die PEV-Teams eine qualifizierte Rückmeldung zur Qualität der Berichte erhalten. Damit hängt die Wirksamkeit der Qualitäts-Checks stark von effektiven und zeitnahen Überarbeitungsprozessen ab. Zugleich können Überarbeitungsschleifen den Prozess in die Länge ziehen und eine zeitnahe Rückmeldung durch die STS 08 erschweren. Zugleich können die PEV-Berichte durch die „end-of-pipe“-Konstruktion der Qualitäts-Checks nur rückwirkend Rückmeldungen zur methodischen Qualität der PEV-Berichte geben, haben aber keinen Einfluss auf die Auswahl und Umsetzung der Vorgehensweise und der Methoden während der Durchführung.

Empfehlung 16: Es wird eine Diskussion zur Weiterführung der PEV-Checks empfohlen. Sollte sich dafür entschieden werden, müsste eine zeitnahe Prozessgestaltung der Qualitäts-Checks garantiert werden, um den Evaluierungsprozess nicht in die Länge zu ziehen.

Bei der Durchführung der Meta-Evaluierung fiel zudem auf, dass **die Grundlagendokumente der PEV nicht alle Qualitätskriterien für Evaluierungen aufgreifen**. In Bezug auf die methodische Qualität werden bislang u. a. die Darstellung des Auftragswerts der Vorhaben und die Ableitung der Ziele und Zwecke einer Evaluierung nicht gefordert. Zugleich ist bislang keine Reflexion der methodischen Vorgehensweise hinsichtlich der Zuordnungs- und/ oder Beitragsanalyse sowie bezüglich einiger Indikatoren der OECD-DAC-Kriterien gefordert. Auch werden in den Grundlagendokumenten bisher noch keine Orientierungspunkte für die Planung von zum Beispiel Aufwänden einer PEV gegeben.

Empfehlung 17: Die genannten Lücken in den Grundlagendokumenten sollten geschlossen werden, indem die fehlenden Standards aufgenommen oder begründet ausgeschlossen werden.

Zu den identifizierten Stärken in der PEV-Durchführung gehört die **ethisch korrekte Umsetzung** des PEV-Prozesses. Diese umfasst eine kultur- und gendersensible Vorgehensweise sowie eine unparteiische Umsetzung. Die Ergebnisse der Meta-Evaluierung weisen darauf hin, dass die PEV zumeist den Grundsätzen von Integrität, Menschenrechten und Unparteilichkeit entspricht. Darüber hinaus scheint eine ausreichend hohe Anzahl an Gutachtertagen einen signifikant positiven Einfluss auf eine faire Evaluierung zu haben (vergleiche hierzu Empfehlung 5).

Ergebnisse und Empfehlungen zur Nützlichkeit

Bei der Überprüfung der Nützlichkeit des PEV-Instruments stellte die Meta-Evaluierung fest, dass die darunter gefassten Qualitätskriterien eher erfüllt werden (Erfüllung zu 74 %). „Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an den geklärten Evaluierungszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet“ (DeGEval 2008). Aus Sicht des Meta-Evaluierungsteams schneidet die Nützlichkeit vergleichsweise gut ab. Es überwiegen hier die Stärken.

Zu den Stärken der PEV gehört die **Einbeziehung relevanter Stakeholder**. Die Analysen hierzu zeigen, dass der Austausch mit den Stakeholdern im Evaluierungsprozess einen positiven Einfluss auf die Nützlichkeit der PEV nimmt und somit eine wichtige Stellschraube für die Einhaltung der Nützlichkeit darstellt. Dabei sollte positiv hervorgehoben werden, dass die Evaluierungszwecke oftmals gemeinsam diskutiert werden. Dies schafft die Grundlage dafür, dass die PEV auch für die beteiligten und betroffenen Stakeholder nützlich sind.

Empfehlung 18: Es wird empfohlen, auf diesen Stärken aufzubauen und spezifische Kommunikationsformate für einzelne Stakeholdergruppen zu entwickeln, die bislang noch nicht ausreichend adressiert wurden (z. B. Zielgruppen), um die zielgruppengerechte Aufbereitung der Evaluierungsergebnisse weiter zu verbessern.

Empfehlung 19: Um die Einbindung der Stakeholder weiter zu stärken, sollte die Sensibilisierung für partizipative Ansätze stärker als bisher in die Handreichung und die Unterstützungsmaterialien der PEV aufgenommen werden.

Eine weitere Stärke der PEV liegt aus Sicht des Meta-Evaluierungsteams in ihrer hohen **Nützlichkeit für die GIZ-interne Projekt- und Programmebene**. Damit erfüllt ein Großteil der PEV eine der zentralen Nutzererwartungen, die an das Instrument gestellt werden. Hierzu zeigt die Meta-Evaluierung, dass die Ergebnisse der PEV vor allem der Planung des Angebots einer Folgemaßnahme und dem Austausch im Projektteam dienen. Die Ergebnisse werden ebenfalls häufig unter Steuerungsgesichtspunkten genutzt, wie zum Beispiel für die Steuerung des Folgevorhabens und des laufenden Vorhabens. Bei der Nutzung der PEV-Ergebnisse **auf der GIZ-internen Sektor-/ Regionalebene** besteht hingegen Verbesserungsbedarf. Hierzu zeigen die Ergebnisse der Meta-Evaluierung, dass die Ergebnisse der PEV bislang kaum für die Fachverbandsarbeit im Sektor und nur teils für die Auseinandersetzung mit fachlichen und regionalen Ansätzen sowie das unternehmensweite Wissensmanagement in Wert gesetzt werden.

Empfehlung 20: Es sollte mehr Aufmerksamkeit auf die Nutzung der Evaluierungsergebnisse auf der sektoralen und regionalen Ebene gelegt werden. Hierbei sollte geprüft werden, inwieweit sich bereits vorhandene Formate (z. B. Fachverbands- und Lernveranstaltungen) eignen oder inwieweit neue Formate des regionalen und sektoralen Austauschs zu Evaluierungsergebnissen sinnvoll sind.

Die Interessen der **Partnerorganisationen** sollten konsequent in Vorbereitung, Durchführung und Nachbereitung verankert werden. Die Meta-Evaluierung zeigt, dass Interessen von Partnern in der Konzeption berücksichtigt und die relevanten Stakeholder weitgehend in die Vorbereitung einbezogen werden. Zugleich setzen sich die PEV-Berichte bei den Rahmenbedingungen mit den Kontextfaktoren im Partnerland auseinander. Diese Aspekte werden zugleich noch nicht ausreichend bei den Analysen der OECD-DAC Kriterien in Wert gesetzt. Trotz dessen werden die Ergebnisse der PEV wiederum mit den Partnern geteilt.

Empfehlung 21: Es wird empfohlen, die vorhandenen Stärken (Partnerdialog) weiter in Wert zu setzen und, um die Relevanz der Ergebnisse für Partnerländer und Partnerorganisationen zu erhöhen, die Kontextfaktoren in den Analysen stärker zu berücksichtigen. Es sollte keinen Bruch in der Kette der Partnerorientierung zwischen Vorbereitung, Durchführung und Nachbereitung einer PEV geben.

Inhalt

1 Einleitung	1
2 Methodik	2
2.1 Grundlagen und Struktur des Analyserasters	2
2.2 Kurzdarstellung der methodischen Vorgehensweise	4
3 Evaluierungsgegenstand und Datengrundlage	7
3.1 Evaluierungsgegenstand	7
3.2 Beschreibung der Datengrundlage	8
4 Teil I: Ergebnisse zur methodischen Qualität (Genauigkeit)	12
4.1 Beschreibung und Analyse des Evaluierungsstandards Genauigkeit	12
4.2 Einflussfaktoren auf die methodische Qualität	20
4.3 Bewertung und Schlussfolgerungen zur methodischen Qualität	23
5 Teil II: Ergebnisse Prozessqualität (Durchführbarkeit und Fairness)	26
5.1 Beschreibung und Analyse des Evaluierungsstandards Durchführbarkeit	26
5.2 Beschreibung und Analyse des Evaluierungsstandards Fairness	29
5.3 Einflussfaktoren auf die Prozessqualität	31
5.4 Bewertung und Schlussfolgerungen zur Prozessqualität	32
6 Teil III: Ergebnisse zur Nützlichkeit	34
6.1 Beschreibung und Analyse des Evaluierungsstandards Nützlichkeit	34
6.2 Einflussfaktoren auf die Nützlichkeit	39
6.3 Bewertung und Schlussfolgerungen zur Nützlichkeit	40
7 Schlussfolgerungen und Handlungsempfehlungen	42
7.1 Handlungsfeld 1: Methodische Qualität	42
7.2 Handlungsfeld 2: Prozessqualität	44
7.3 Handlungsfeld 3: Nützlichkeit	46
8 Literaturverzeichnis	Fehler! Textmarke nicht definiert.
9 Anlagen	49
9.1 Überblick über die Ergebnisse in den DeGEval-Evaluierungsstandards	50
9.2 Ergebnisse der Online-Befragung	69
9.3 Fragebogen der Online-Befragung	76
9.4 Analyseraster	84

Abbildungsverzeichnis

Abbildung 1: Anforderungen an die Bewertungsinstrumente von Meta-Evaluierungen	2
Abbildung 2: Schematische Darstellung des Analyserasters	3
Abbildung 3: Methodische Vorgehensweise der Meta-Evaluierung	4
Abbildung 4: Bewertungssystem	5
Abbildung 5: Regionalbereiche und fachliche Schwerpunkte der evaluierten Vorhaben	9
Abbildung 6: Kreuzung des Auftragswerts und der Laufzeit der untersuchten Vorhaben	9
Abbildung 7: Kreuzung des Auftragswerts der Vorhaben mit der Anzahl der Gutachtertage	10
Abbildung 8: Anzahl der Gutachter/innen	11
Abbildung 9: Erfüllungsgrad der Bewertungskriterien im Evaluierungsstandard Genauigkeit	12
Abbildung 10: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien G 1–G 3	13
Abbildung 11: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien G 4–G 6	14
Abbildung 12: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien G 7 (A-F)–G 8	16
Abbildung 13: Einflussfaktoren Regionalbereich, Umfang der Methoden, Anzahl der Gutachter/innen	20
Abbildung 14: Kreuzung der Genauigkeit und der Anzahl der Gutachtertage	21
Abbildung 15: Kreuzung zwischen Regionalbereich, Anzahl der Gutachter/innen und Methodenvielfalt	22
Abbildung 16: Erfüllungsgrad der Bewertungskriterien im Evaluierungsstandard Durchführbarkeit	26
Abbildung 17: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien D 1–D 3	27
Abbildung 18: Erfüllungsgrad der Bewertungskriterien im Evaluierungsstandard Fairness	29
Abbildung 19: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien F 1–F 5	30
Abbildung 20: Erfüllungsgrad der Bewertungskriterien im Evaluierungsstandard Nützlichkeit	34
Abbildung 21: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien N 1–N 2	35
Abbildung 22: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien N 3–N 7	36
Abbildung 23: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien N 8 (A)–N 8 (B)	37

Abkürzungsverzeichnis

AV	Auftragsverantwortliche/r
B 1	Regionalbereich Afrika
B 2	Regionalbereich Asien/Pazifik, Lateinamerika/Karibik
B 3	Regionalbereich Mittelmeer, Europa und Zentralasien
B 4, B 8	Sektorvorhaben, Globalvorhaben, überregionale Vorhaben
BMZ	Bundesministerium für Wirtschaftliche Zusammenarbeit und Entwicklung
D	Durchführbarkeit (Evaluierungsstandard)
DAC	Development Assistance Committee
DeGEVal	Deutsche Gesellschaft für Evaluation e.V.
EZ	Entwicklungszusammenarbeit
F	Fairness (Evaluierungsstandard)
FMB	Fach- und Methodenbereich
FZ	Finanzielle Zusammenarbeit
G	Genauigkeit (Evaluierungsstandard)
GIZ	Gesellschaft für Internationale Zusammenarbeit GmbH
InWEnt	Internationale Weiterbildung und Entwicklung gGmbH
JCSEE	Joint Committee on Standards for Educational Evaluation
M	Mittelwert
M&E	Monitoring und Evaluierung
N	Nützlichkeit (Evaluierungsstandard)
OECD	Organisation for Economic Cooperation and Development
PEV	Projektevaluierung
PFK	Projektfortschrittskontrolle
SD	Standardabweichung
SMART	Specific, measurable, accepted, realistic, time-bound
STS 08	Stabsstelle 08 (Evaluierung) der GIZ
TOR	Terms of Reference
TZ	Technische Zusammenarbeit
UE	Unabhängige Evaluierung

1 Einleitung

Die Syspons GmbH hat für die Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH eine **Meta-Evaluierung der Projektevaluierungen (PEV)** durchgeführt. Die Meta-Evaluierung wurde von der Stabsstelle Evaluierung der GIZ (STS 08) beauftragt. Das Ziel der Meta-Evaluierung war es, die Qualität der Evaluierungsprozesse und Evaluierungsergebnisse der PEV gemäß **internationaler Evaluierungsstandards** zu untersuchen, um Ansatzpunkte für die Qualitätsverbesserung des PEV-Instruments zu identifizieren. Hierzu wurde die Qualität des PEV-Instruments entlang der Standards Genauigkeit, Durchführbarkeit, Fairness und Nützlichkeit der Deutschen Gesellschaft für Evaluation (DeGEVal) e.V. analysiert und bewertet. Darüber hinaus zielte die Meta-Evaluierung darauf ab, Baselinewerte für die Einführung eines unternehmensweiten Qualitäts-Monitorings des PEV-Instruments zu erheben. Vorbehaltlich einer entsprechenden Entscheidung der Unternehmensführung sollen zukünftig regelmäßige Meta-Evaluierungen der PEV stattfinden, die auf Basis des hier verwendeten Analyserasters die Qualitätsentwicklung der PEV-Prozesse und -Ergebnisse überprüfen.

Meta-Evaluierungen sind laut DeGEVal und anderen Evaluierungsgesellschaften wichtige **Instrumente der Qualitätssicherung**. Seit 2010 sind Meta-Evaluierungen Bestandteil des unabhängigen Qualitätsmanagements von Evaluierungen der GIZ und werden von weiteren Instrumenten ergänzt (u.a. Handreichung und Unterstützungsmaterialien wie annotierte Gliederungen, PEV-Checks, Methoden-Toolbox, etc.). In den vergangenen Jahren (2010 bis 2013) führte die GIZ (inklusive der Vorgängerorganisation InWEnt) sechs Meta-Evaluierungen durch. Diese richteten sich ausschließlich auf die Beurteilung der methodischen Qualität anhand der Evaluierungsberichte und verwendeten jeweils eigene und miteinander nur bedingt vergleichbare Analyseraster. Die Qualität des Evaluierungsprozesses sowie die Nützlichkeit der Ergebnisse wurden hingegen nicht analysiert. Um einem **erweiterten Qualitätsverständnis der GIZ-Evaluierungen** Rechnung zu tragen, bewertet die vorliegende Meta-Evaluierung die in den PEV der GIZ verwirklichte Qualität entlang einschlägiger Standards der internationalen und deutschen Evaluierungspraxis und berücksichtigt dabei sowohl die methodische Qualität als auch die Prozessqualität sowie Nützlichkeit und Nutzung. Die Meta-Evaluierung stützt sich dabei auf zwei zentrale methodische Bausteine. Zum einen ist eine Textanalyse von 70 PEV-Berichten durchgeführt worden, die zwischen Oktober 2014 und September 2015 finalisiert wurden. Zum anderen ist eine standardisierte Online-Befragung von Auftragsverantwortlichen (AV) umgesetzt worden, die im genannten Erhebungszeitraum eine PEV beauftragt haben. Die **Ergebnisse sind repräsentativ** für die im Betrachtungszeitraum durchgeführten PEV.

Dieser Bericht ist wie folgt aufgebaut:

- **Kapitel 2** erläutert die Grundlagen und die Struktur des Analyserasters und gibt einen Überblick über die methodische Vorgehensweise.
- **Kapitel 3** stellt den Evaluierungsgegenstand (PEV) und die Datengrundlage dar.
- **Kapitel 4** analysiert und beurteilt die Ergebnisse der Datenerhebung zur methodischen Qualität.
- **Kapitel 5** analysiert und beurteilt die Ergebnisse der Datenerhebung zur Prozessqualität.
- **Kapitel 6** analysiert und beurteilt die Ergebnisse der Datenerhebung zur Nützlichkeit und Nutzung.
- **Kapitel 7** formuliert die Schlussfolgerungen und die Handlungsempfehlungen des Meta-Evaluierungsteams an die GIZ.
- **Kapitel 8** beinhaltet das Literaturverzeichnis.

Die Anlagen in **Kapitel 9** umfassen das Analyseraster, den Fragebogen der Online-Befragung, die Ergebnisse der Online-Befragung und einen Überblick über die Ergebnisse in den vier DeGEVal-Evaluierungsstandards.

2 Methodik

2.1 Grundlagen und Struktur des Analyserasters

Bei der Entwicklung des Analyserasters wurde darauf geachtet, dass die **folgenden Anforderungen** an Meta-Evaluierungen erfüllt werden, die aus der einschlägigen Literatur abgeleitet wurden:

Abbildung 1: Anforderungen an die Bewertungsinstrumente von Meta-Evaluierungen

Anforderungen nach Widmer (1996, S. 9/10)	Umsetzung der Anforderungen in der Meta-Evaluierung
Methodische Offenheit: Offenheit gegenüber Evaluierungen unterschiedlicher methodologischer Ausrichtungen und verschiedener Evaluierungsansätze	Die Bewertungskriterien und Indikatoren sind bewusst so definiert, dass sie sich nicht auf spezifische Theorien, methodische Vorgehensweisen oder Evaluierungsmethoden stützen. Vielmehr befassen sie sich damit, inwieweit die PEV ihre methodischen Ansätze begründen und sich daran ausrichten. Die tatsächlich gewählten Vorgehensweisen und Evaluierungsmethoden werden dann deskriptiv durch Checklisten abgefragt.
Umfassende Bewertung: Bewertung der gesamten Evaluierung und nicht lediglich Bewertung von Teilaspekten einer Evaluierung	Das erweiterte Qualitätsverständnis der Evaluierung umfasst die Evaluierungsstandards Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit und bewertet somit die Qualität der gesamten Evaluierung. Anders als bisherige Meta-Evaluierungen stützt sich das hier entwickelte Analyseraster sowohl auf die methodische Qualität als auch auf die Prozessqualität und die Nützlichkeit.
Verbreitung, Anerkennung und Akzeptanz: Hohe Akzeptanz im Evaluierungsfeld, Anerkennung durch relevante Fachorganisationen und einschlägige Literatur	Das Analyseraster wurde auf Grundlage nationaler und internationaler Evaluierungsstandards entwickelt. Hierbei wurde abgeglichen, welche Evaluierungsstandards die GIZ in ihren internen Qualitätsinstrumenten (noch nicht) berücksichtigt. International sind verschiedene Kataloge von Evaluierungsstandards entwickelt worden, von denen insbesondere die Standards des Joint Committee on Educational Evaluation (JCSEE) (2000) eine breite Beachtung gefunden haben. Diesen Standards haben sich auch Fachorganisationen wie die DeGEval (2008) angeschlossen, um einen internationalen Erfahrungsaustausch zu vereinfachen (DeGEval 2008, S. 20; 44). Zudem nimmt das Analyseraster Bezug auf die Qualitätsstandards der OECD-DAC (1991; 2010), um ggf. Besonderheiten von Evaluierungen in der Entwicklungszusammenarbeit zu berücksichtigen.
Ausreichender Detaillierungsgrad: Ausreichend differenzierte Formulierung der einzelnen Kriterien	Die verwendeten Bewertungskriterien und ihre Indikatoren werden mit Definitionen hinterlegt, damit die Bewertungsgrundlagen transparent und nachvollziehbar sind.

Quelle: Syspons 2016

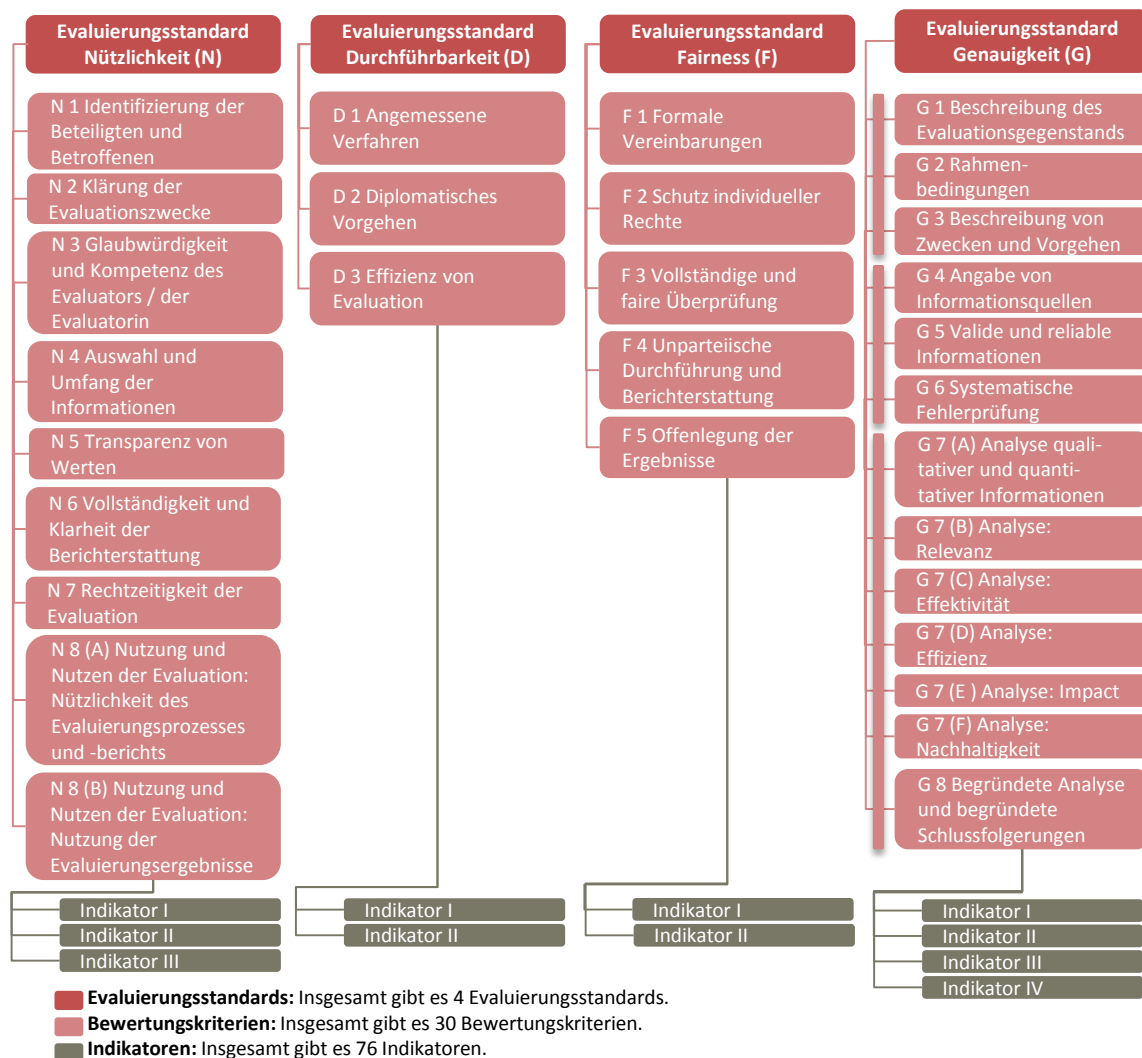
Unter Beachtung dieser Anforderungen ist das Bewertungssystem der Meta-Evaluierung methodisch entlang der folgenden drei Ebenen strukturiert (siehe auch Abbildung 2):

- Auf der ersten Ebene stehen die vier **DeGEval-Evaluierungsstandards** Genauigkeit (methodische Qualität), Durchführbarkeit, Fairness (bilden gemeinsam die Prozessqualität) und Nützlichkeit. Bei der Nützlichkeit wurde auch die Nutzung der Evaluierungsergebnisse erhoben. Jeder Evaluierungsstandard wird durch eine unterschiedliche Anzahl an Bewertungskriterien operationalisiert.

- Die **Bewertungskriterien**, welche ebenfalls weitgehend von der DeGEval übernommen wurden, bilden die zweite Ebene des Analyserasters. Dem Evaluierungsstandard Genauigkeit werden 13 Bewertungskriterien (G 1 bis G 8), der Durchführbarkeit drei Kriterien (D 1 bis D 3), der Fairness fünf Kriterien (F 1 bis F 5) und der Nützlichkeit neun Kriterien (N 1 bis N 8 (B)) zugeordnet.
- Auf der dritten Ebene werden jedem Bewertungskriterium **Indikatoren** zugeordnet, die beschreiben, welche Aspekte gegeben sein müssen, damit ein Kriterium erfüllt wird. Die Anzahl der einem Bewertungskriterium zugeordneten Indikatoren schwankt zwischen einem und sechs Indikatoren.

Die nachfolgende Abbildung gibt einen graphischen Überblick über das Bewertungssystem:

Abbildung 2: Schematische Darstellung des Analyserasters



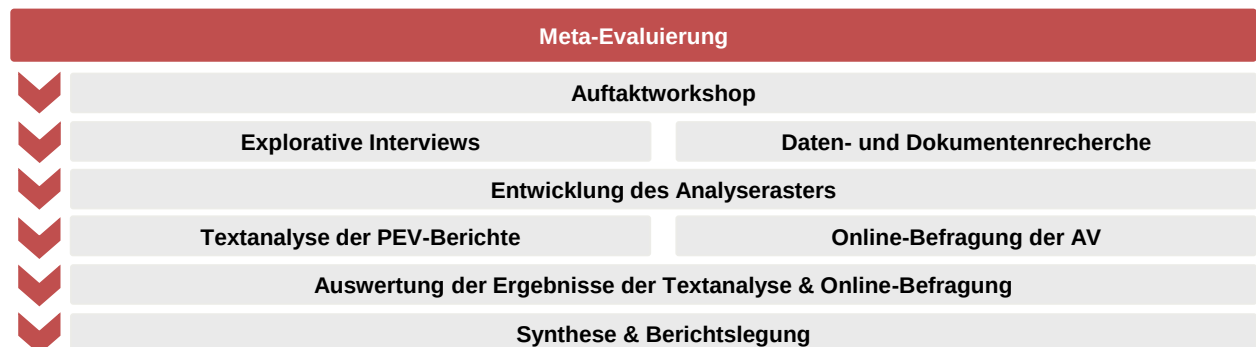
Der vorherige Absatz erklärt den Aufbau dieses Bewertungssystems im Detail.

Quelle: Syspons 2016

2.2 Kurzdarstellung der methodischen Vorgehensweise

Die methodische Vorgehensweise gliederte sich in folgende Schritte:

Abbildung 3: Methodische Vorgehensweise der Meta-Evaluierung



Quelle: Syspons 2016

Aufbauend auf einem Auftragsklärungsgespräch mit der STS 08 in Form eines Workshops wurden acht **explorative Interviews** mit aktuellen und ehemaligen Mitarbeiter/innen der STS 08 geführt. Hierfür wurden Gesprächspartner/innen ausgewählt, (1) die in die Entwicklung der PEV-Unterstützungsmaterialien involviert waren und somit den Einführungsprozess der PEV begleitet haben, (2) die Erfahrungen mit der Durchführung von Qualitäts-Checks der Evaluierungsberichte (PEV-Checks) haben und/ oder (3) die bereits Meta-Evaluierungen in der GIZ gesteuert haben. Alle Gesprächspartner/innen wurden zudem hinsichtlich ihrer Erwartungen befragt. Die Ergebnisse der explorativen Interviews wurden dokumentiert und flossen in die Entwicklung des Analyserasters ein.

Parallel zu den explorativen Interviews wurde ein mehrstufiges **Desk Research** durchgeführt, um die Grundlage für das Analyseraster zu legen. Zunächst wurden anhand der wissenschaftlichen Literatur Anforderungen an Meta-Evaluierungen ausgearbeitet. Darauf aufbauend wurde entlang der Evaluierungsstandards der DeGEval durch eine Textanalyse abgeglichen, welche Bewertungsmaßstäbe sich in anderen internationalen Standards und in den GIZ-internen Qualitätssicherungsinstrumenten wiederfinden. Das Analyseraster im Anhang dieses Berichts dokumentiert die Ergebnisse dieses Vergleichs von Bewertungsmaßstäben.

Das **Analyseraster** für die Meta-Evaluierung wurde in enger Abstimmung mit der STS 08 entwickelt, um eine breite Akzeptanz der Bewertungsgrundlagen in der GIZ und die problemlose Nutzung des Analysesystems in zukünftigen Meta-Evaluierungen sicherzustellen (siehe Analyseraster in Kapitel 9.1).

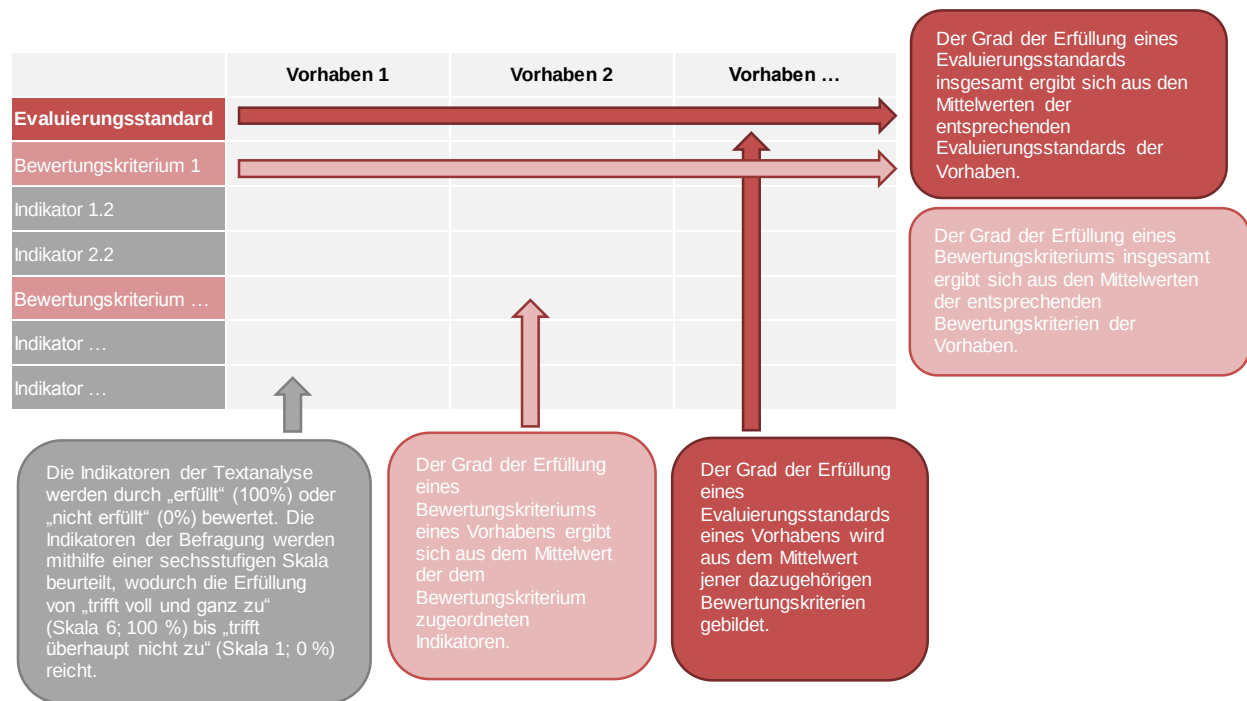
Das Analyseraster liefert die Grundlage für die beiden Methodenbausteine: die Textanalyse der PEV-Berichte und die AV-Onlinebefragung. Für die **Textanalyse der PEV-Berichte** wurden per Zufallsstichprobe 70 von 94 PEV-Berichten ausgewählt, die zwischen Oktober 2014 und September 2015 erstellt wurden. Mit 74 % der Grundgesamtheit liefern die ausgewerteten Evaluierungsberichte damit ein **repräsentatives Bild** der damit erhobenen Bewertungskriterien im Betrachtungszeitraum. Je Bericht wurde ein Auswertungsblatt angelegt, welches die Rahmendaten aufführt und die qualitative und quantitative Bewertung der Indikatoren beinhaltet. Um die Bewertung jedes Indikators transparent und nachvollziehbar zu machen, wurden die Grundlagen der Beurteilung erläutert. Negativ bewertete Indikatoren enthalten eine kurze textliche Begründung. Bei einer positiven Bewertung wurde auf die entsprechende Seitenzahl im Bericht verwiesen.

Neben der Textanalyse der PEV-Berichte bildete eine **standardisierte Online-Befragung** der AV das zweite Erhebungsinstrument der Meta-Evaluierung. Sie ermöglichte die Prüfung von Qualitätskriterien, die sich der Textanalyse entziehen, wie bspw. Fragen zur Nutzung der PEV-Ergebnisse, da diese in den Berichten nicht dokumentiert werden können (siehe Fragebogen in Kapitel 9.2; Ergebnisse der Befragung in Kapitel 9.3). In der Online-Befragung wurden alle 94 AV angeschrieben, die zwischen Oktober 2014 und September 2015 eine PEV beauftragt haben. Der **Rücklauf betrug 79 %** (74 von 94 befragten AV)

und liefert damit **robuste und repräsentative Ergebnisse** (die Fehlerspanne liegt bei 5 % und das Konfidenzniveau bei 90 %) für die entsprechenden Indikatoren. Es muss einschränkend festgehalten werden, dass die AV nur die Perspektiven einer Funktionsgruppe wiedergeben können und nicht die Perspektiven anderer an den PEV beteiligten Funktionsgruppen, wie die Fachplaner/innen.

Die Auswertung der Textanalyse der PEV-Berichte und der Onlinebefragung basierte auf dem folgenden Bewertungssystem:

Abbildung 4: Bewertungssystem



Quelle: Syspons 2016

Dem Bewertungssystem liegt der Anspruch einer vollständigen Zielerreichung zugrunde. Um darauf aufbauend die Bewertungen auf der **Ebene der Evaluierungsstandards und Bewertungskriterien** einheitlich und nachvollziehbar zu gestalten, werden diese in diesem Bericht wie folgt, angelehnt an die Skala der jährlichen Externen Qualitätskontrolle der GIZ¹, angegeben:

- 90–100 % (größtenteils bis vollständig erfüllt)
- 80–89 % (zumeist erfüllt)
- 60–79 % (eher erfüllt)
- < 60 % (bedingt erfüllt).

Neben der Auswertung der PEV-Berichte und der Onlinebefragung anhand des Analyserasters wurden in einem zweiten Schritt mögliche Zusammenhänge zwischen einzelnen Qualitätsaspekten mithilfe von bi- und multivariaten Auswertungsmethoden untersucht. Das Ziel war es hierbei, mögliche Einflussfaktoren auf die methodische Qualität, die Prozessqualität sowie die Nützlichkeit auf Grundlage der vorliegenden Daten zu identifizieren. Dabei wurden auch spezifischere Zusammenhänge auf den dahinter liegenden Ebenen der Bewertungskriterien, Indikatoren und Befragungsisems untersucht. Hierbei wurden deduktiv zunächst **Hypothesen zu allen theoretisch erwarteten Zusammenhängen** formuliert und in Indizes gruppiert, die verschiedene Befragungsisems und Indikatoren zusammenfassen. Die verwendeten Hypothesen sind das Ergebnis der explorativen Interviews und des Austauschs zwischen Meta-

¹ Stern, T., Scheller, O., Freimann, I. (2015). Externe Qualitätskontrolle der GIZ. Ergebnisbericht 2015.

Evaluierungsteam und GIZ STS 08. Diesem theoriegeleiteten Vorgehen unterliegt die Annahme, dass diejenigen Aspekte zusammengefasst werden können, denen ein gemeinsames theoretisches Konstrukt zugrunde liegt. Anschließend wurden die theoretisch erwarteten Zusammenhänge **bi- und multivariat analysiert** und durch Signifikanzprüfungen validiert oder falsifiziert. Hierzu wurde auf Kreuztabellen, Korrelationen und multivariate Regressionen zurückgegriffen. Die signifikanten Einflussfaktoren oder „Stellschrauben“ wurden anschließend näher untersucht, um weitere Wirkungszusammenhänge zu identifizieren. In diesem Bericht werden nur ausgewählte, das heißt: in der Regel signifikante Zusammenhänge erläutert und analysiert, die einen systematischen Einfluss auf die PEV-Qualität haben könnten. Hierbei sind zwei Werte relevant: Die Korrelation der einzelnen Parameter untereinander wurde mit dem Pearson-Korrelationskoeffizienten angegeben. Zudem wurden p-Werte für die statistische Signifikanz erhoben. Liegt ein p-Wert unter $<0,05$, wurde dieser als signifikant angesehen und die entsprechenden Zusammenhänge werden im Bericht analysiert.

Das Meta-Evaluierungsteam ist sich bewusst, dass qualitative Textanalysen immer auch subjektiven Interpretationen einzelner Evaluator/innen unterliegen. Die Ergebnisse der Auswertungen wurden daher, soweit möglich, **trianguliert**². Im Rahmen einer **Forschertriangulation** haben insgesamt vier Personen die Bewertungen der PEV-Berichte vorgenommen, so dass persönliche Bewertungstendenzen verringert werden konnten. Zugleich wurde, um eine vergleichbare Bewertung sicherzustellen, jeder zehnte Bericht durch zwei Evaluator/innen doppelt geprüft. Hierdurch wurde eine Kalibrierung vorgenommen. Die Ergebnisse der Doppelprüfungen (z.B. abweichende Interpretationen von Indikatoren) wurden anschließend im gesamten Team diskutiert und reflektiert. Somit wurde durch den individuellen Austausch über die Bewertungsmaßstäbe die Qualität der Auswertungen erhöht. Zudem wurden der Mittelwert (M) und die Standardabweichung (SD), also die durchschnittliche Abweichung der Eingaben einzelner Evaluator/innen vom Mittelwert eines Bewertungskriteriums, ausgewertet, um die Verlässlichkeit der Ergebnisse zu reflektieren.

Während eine **Datentriangulation** sichergestellt wurde, war eine **Methodentriangulation** nur begrenzt möglich. Zudem konnten nicht alle am PEV-Prozess beteiligten und betroffenen Stakeholder ihre Einschätzung zur Qualität abgeben (u. a. Partnerorganisationen, PEV-Teammitglieder), wodurch die Wahrnehmung der AV in den Mittelpunkt rückt. Eine Ausweitung der Befragung auf weitere Beteiligte und Betroffene war jedoch vor dem Hintergrund von Kosten-Nutzen-Reflexionen nicht möglich. Um diesen Herausforderungen zu begegnen, reflektiert das Meta-Evaluierungsteam die Aussagekraft der Ergebnisse in der Analyse und der Bewertung der Evaluierungsstandards entsprechend.

² Triangulation wird in dieser Meta-Evaluierung definiert als eine Forscherstrategie, in welcher **mindestens drei** Sichtweisen von Forscher/innen, Daten oder Methoden zu einem gleichen Sachverhalt herangezogen werden, um mit den Stärken einer Vorgehensweise die Schwächen einer anderen Vorgehensweise auszugleichen.

3 Evaluierungsgegenstand und Datengrundlage

3.1 Evaluierungsgegenstand

Mit dem Auslaufen des Programms der Unabhängigen Evaluierungen (UE) und dem Anspruch, der Transparenzinitiative IATI beizutreten, wurden im Januar 2014 durch die Einführung der Monitoring und Evaluierung (M&E)-Policy der GIZ die Projektfortschrittskontrollen (PFK) von dem dezentralen Instrument der PEV abgelöst. Dies bringt zwei zentrale Neuerungen mit sich. Zum einen hatten PFK als internes Beratungsinstrument nicht den Anspruch, einen Beitrag zur Rechenschaftslegung und Transparenz zu leisten. Dieser Anspruch verfolgen hingegen die PEV. Zum anderen sollen PEV höheren methodischen und evaluatorischen Ansprüchen gerecht werden als noch PFK. Aufbauend auf diese „Multifunktionalität“ der PEV wurde das Analyseraster der PEV mit seinen vier Evaluierungsstandards konzipiert.

Seitdem sind PEV Bestandteil der M&E-Instrumente in der GIZ. Seit April 2014 werden sie **flächendeckend und verbindlich** in allen Vorhaben im Auftrag des Bundesministeriums für Wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) durchgeführt, die eine Mindestlaufzeit von drei Jahren und ein Mindestvolumen von einer Million Euro haben (GIZ 2013). PEV werden ein Jahr bis sechs Monate vor Ende der Vorhabenslaufzeit durchgeführt. Sind sie als Abschlussevaluierung ohne Folgemaßnahme konzipiert, können sie auch später erfolgen. Falls eine Folgemaßnahme geplant ist, werden PEV in den Prüfprozess von Neuvorhaben integriert (GIZ 2015a).

PEV sollen einem iterativen, prozesshaften Vorgehen entlang von **acht Schritten** folgen. (1) In der Vorbereitungsphase wird der Evaluierungsgegenstand erfasst, wozu der AV vorbereitende Unterlagen bereitstellt. (2) Abgeleitet aus dem Evaluierungsgegenstand und den OECD-DAC-Kriterien werden das Evaluierungsdesign entwickelt und (3) der Ablauf vor Ort geplant. (4) Die Durchführung der PEV umfasst einen Auftaktworkshop, (5) die Datenerhebung und -auswertung (6) sowie ein GIZ-internes Debriefing und (7) anschließend einen Abschlussworkshop mit den Partnern, in welchem die Evaluierungsergebnisse vorgestellt werden. (8) In der Nachbereitungsphase wird der Evaluierungsbericht mit seinen Anhängen - inklusive der Managementanlage und des Kurzberichts - erstellt (GIZ 2015a).

Einhergehend mit der Einführung der PEV wurden **verschiedene Instrumente der Qualitätssicherung** entwickelt. Hierzu gehören:

- eine **Handreichung**, welche als Anlage zum Leitfaden für Programmvorschläge (PV) das Vorgehen für PEV in Kombination mit der Prüfung von Folgemaßnahmen darstellt (GIZ 2015a);
- weitere **Unterstützungsmaterialien**:
 - ein Fahrplan zur PEV, welcher die notwendigen Aufgaben zur Planung, Terminabstimmung und Klärung im Evaluierungsprozess strukturieren soll (GIZ 2015b);
 - ein Muster der Terms of Reference (TOR) für die PEV (GIZ o.D. a);
 - eine Checkliste zur Selbsteinschätzung der Capacity-WORKS-Erfolgsfaktoren. Diese Checkliste liefert Anhaltspunkte dazu, welche Bereiche des Kooperationsmanagements Einfluss auf die Bewertung des Vorhabens nehmen können (GIZ o.D. b);
 - eine Checkliste, in welcher Bewertungsdimensionen mit Analysefragen für die Bewertung der OECD-DAC-Kriterien zusammengefasst werden. Diese Checkliste zeigt zudem auf, wie die Gesamtnote des evaluierten Vorhabens zu berechnen ist (GIZ 2015c);
 - eine (annotierte) Berichtsgliederung des PEV-Berichts und des PEV-Kurzberichtes, welche die wesentlichen Inhalte, Analyseaspekte und Leitfragen zu den einzelnen Kapiteln des Berichts auflistet (GIZ o.D. c; 2015d; o.D. d; 2015e);
 - eine (annotierte) Berichtsgliederung zur internen Managementanlage (GIZ o.D. f; 2015e);

- ein Kommentierungsblatt mit Qualitätsstandards (Mindeststandards und weitere Standards), mit welchem die STS 08 die methodische Qualität der PEV-Berichte prüft und welches auch den PEV-Teams als Hilfsmittel für die Berichtslegung zur Verfügung steht (GIZ 2015g);
- eine Methoden-Toolbox, in welcher Anleitungen und Arbeitshilfen zu analytischen Grundlagen, Evaluierungsdesigns, Evaluierungsmethoden (Interviews, Befragungen, Fokusgruppen, Fallstudien) sowie zu Qualitätskriterien von Indikatoren etc. gesammelt werden (GIZ 2016h);
- Arbeitshilfen zu Indikatoren (GIZ 2014) und zum Wirkungsmodell der GIZ (GIZ 2015f).
- zum Ende des Prozesses: einen **Qualitätscheck** der PEV-Langberichte und Kurzberichte (QS-Check) durch die Stabstelle Evaluierung (STS 08), wobei nur die Kommentare im Kurzbericht verbindlich zu überarbeiten sind.

3.2 Beschreibung der Datengrundlage

Vorhabensdaten

Die **Datengrundlage der Textanalyse** besteht aus 70 Evaluierungsberichten über insgesamt 70 Vorhaben, die zwischen Oktober 2014 und September 2015 finalisiert wurden. Nachfolgend wird ein Überblick über diese Datengrundlage gegeben. Hierbei erfolgt eine regionale, fachliche, finanzielle und zeitliche **Einordnung der evaluierten Vorhaben**.

Die in den PEV evaluierten Vorhaben sind **in allen Regionalbereichen** der GIZ angesiedelt. Es zeigt sich, dass mit einem Drittel der Berichte aus der Stichprobe verhältnismäßig viele der Vorhaben im Regionalbereich B1 (Afrika) liegen, in welchem auch im GIZ-Gesamtportfolio für das BMZ die meisten Vorhaben angesiedelt sind. Im Regionalbereich B2 (Asien/ Pazifik, Lateinamerika/ Karibik), Regionalbereich B3 (Mittelmeer, Europa, Zentralasien) sowie im Regionalbereich B4 (FMB) (Sektorvorhaben), B8, Sonstiges (Globalvorhaben, die nicht aus regionalen Budgets finanziert werden) sind in etwa gleich viele der evaluierten Vorhaben zu verorten³ (siehe Abbildung 5). Zugleich zeigt die Einordnung der Länder mit Risikopotenzial (Mehrfachzuordnung möglich)⁴ für die Lieferfähigkeit der GIZ, dass:

- 41 % (n=29) der Vorhaben sich in Ländern befanden, denen die GIZ krisenbedingt ein hohes Risiko für die Lieferfähigkeit (**Kategorie I**) zuschreibt und von denen wiederum 20 % zu Ländern mit eingeschränkter personeller Lieferfähigkeit zugeordnet wurden;
- 27 % (n=19) der Vorhaben in Ländern mit erhöhter Gefährdung der Lieferfähigkeit (**Kategorie II**) angesiedelt waren;
- 36 % (n=25) der Vorhaben in Ländern mit geringer Gefährdung der Lieferfähigkeit (**Kategorie III**) umgesetzt wurden.

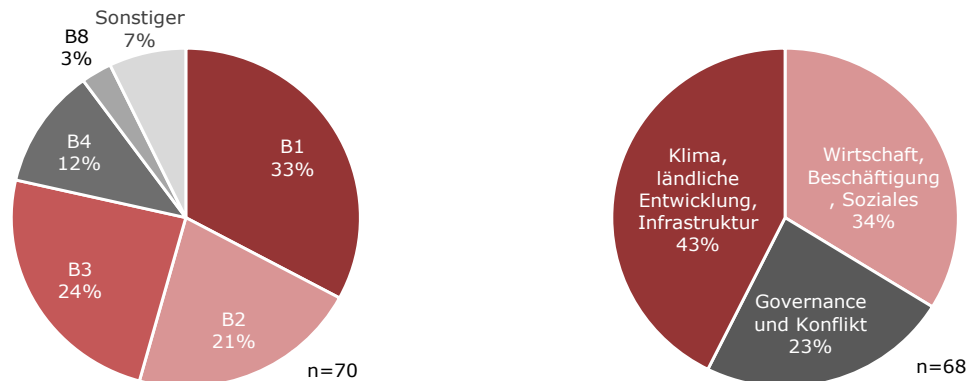
Die evaluierten Vorhaben decken, gemessen an der seit September 2015 bestehenden **fachlichen Gruppierung des Fach- und Methodenbereichs (FMB)**⁵, ein breites Portfolio ab. Sie stammen mehrheitlich (43 %) aus dem Themenbereich **Klima, Ländliche Entwicklung und Infrastruktur** (Klima und Umweltpolitik; Wald, Biodiversität, Landwirtschaft; Ländliche Entwicklung, Ernährungssicherung; Wasser, Abwasser, Abfall; Energie und Verkehr). Weniger Vorhaben (34 %) sind dem Themenbereich **Wirtschaft, Beschäftigung, Soziales** (Bildung; Berufliche Bildung, Arbeitsmarkt; Finanzsystementwicklung, Versicherungen; Gesundheit und Soziale Sicherung; Wirtschaftspolitik und

³ Diese Angaben zu den Regionalbereichen beziehen sich auf die Einordnungen der Vorhaben vor dem Bauhaus 15 Prozess.

⁴ Regionale Vorhaben können in mehreren Länderkategorien eingeordnet werden. Sektor-, Global- und überregionale Vorhaben wurden keiner Länderkategorie zugeordnet.

Privatwirtschaftsförderung; neue gesellschaftspolitische Themen) und (23 %) dem Themenbereich **Governance und Konflikte** (Rechtsstaat und Sicherheit; öffentliche Finanzen und Verwaltung; Demokratie, Politikdialog, Stadt; Frieden und Nothilfe) zuzuordnen (siehe Abbildung 5).

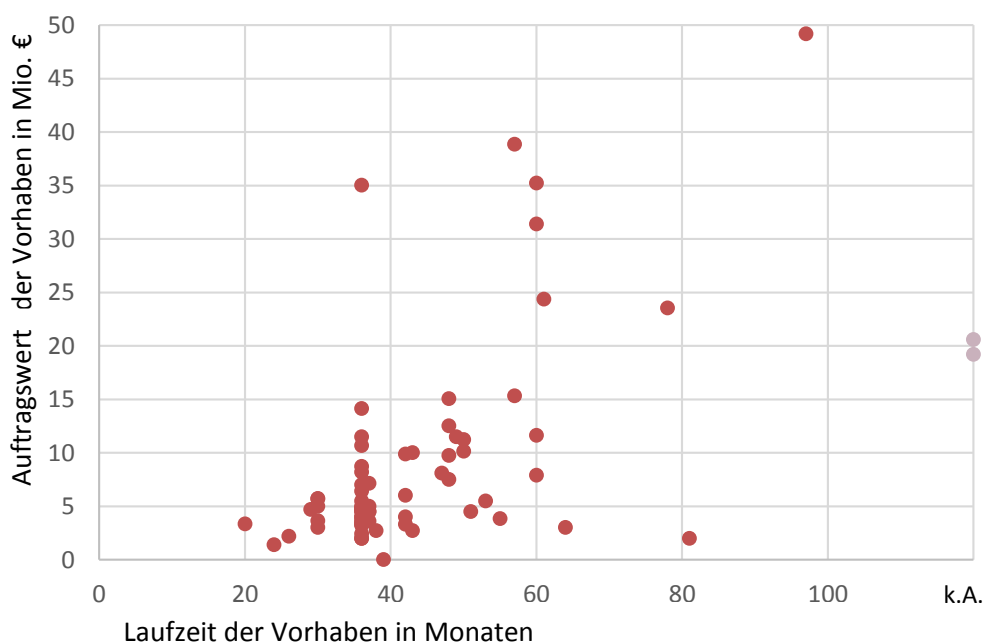
Abbildung 5: Regionalbereiche und fachliche Schwerpunkte der evaluierten Vorhaben



Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

Hinsichtlich des **Auftragswerts** wird deutlich, dass die evaluierten Vorhaben eine breite finanzielle Spannweite abdecken. Während das kleinste Auftragsvolumen bei 1,4 Mio. € liegt, kann das größte Vorhaben auf ein Auftragsvolumen von 49,2 Mio. € zurückgreifen. Das durchschnittliche Auftragsvolumen beträgt circa 9,4 Mio. €. Die gleiche Diversität findet sich bei der **Laufzeit der Vorhaben**, welche von 1,6 Jahren bis 8 Jahren reicht. Die durchschnittliche Laufzeit der evaluierten Vorhaben beträgt 3,6 Jahre (siehe Abbildung 6).

Abbildung 6: Kreuzung des Auftragswerts und der Laufzeit der untersuchten Vorhaben



Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

⁵ Zwei Vorhaben konnten keinem der drei fachlichen Schwerpunkte zugeordnet werden.

In der Gesamtschau sind die Hälfte (n=18) der Vorhaben, zu denen hierzu Angaben verfügbar waren⁶, Neuvorhaben in ihrer ersten **Durchführungsphase**. Die weiteren Vorhaben sind Folgevorhaben in ihrer zweiten (30 %, n=18) oder dritten Phase (20 %, n=12).

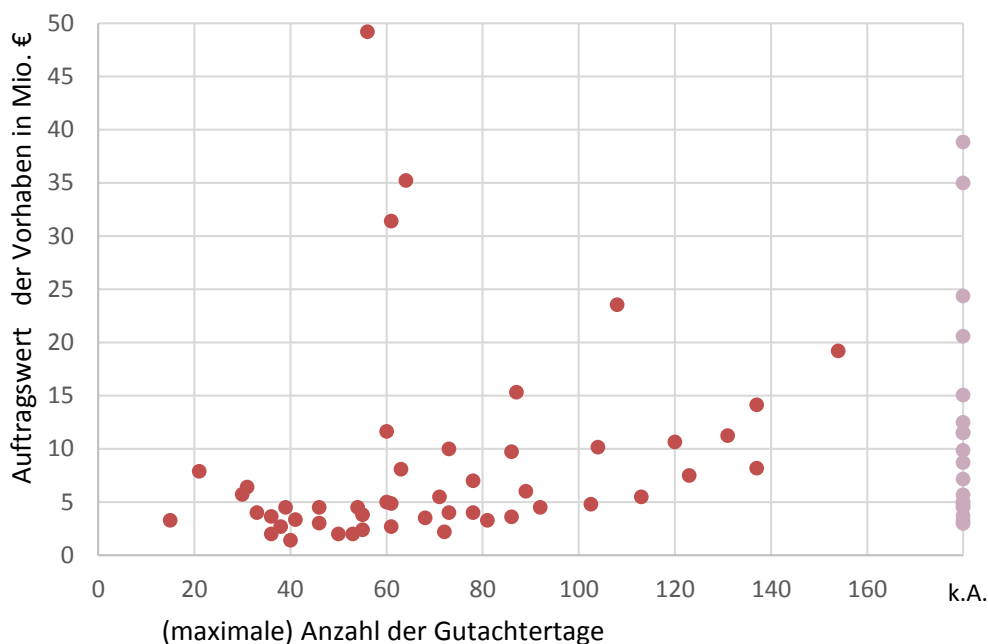
Art der PEV

PEV⁷ sollen einen kritischen Rückblick auf der Grundlage der Bewertung der OECD-DAC-Kriterien ermöglichen. PEV, die zugleich eine **Folgemaßnahme** prüfen, sollen zusätzlich eine Auseinandersetzung mit zukünftigen Planungselementen vornehmen. 87 % (n=61) der vorliegenden Berichte stammen aus PEV, die zugleich eine Folgemaßnahme prüfen. Ist keine Folgemaßnahme geplant, werden PEV als **Abschlussevaluierung** durchgeführt. Dies trifft auf 13 % (n=9) der Berichte der Textanalyse zu.

Rahmendaten der PEV

In den Terms of Reference (TOR) wird die maximale Anzahl der für die PEV zur Verfügung stehenden Gutachtertage aufgelistet. Anhand der TOR kann allerdings nicht abgeleitet werden, wie viele Gutachtertage für die Evaluierung und wie viele Tage für die Planung eines Folgevorhabens angesetzt werden. Durchschnittlich veranschlagen die TOR der evaluierten Vorhaben 10,29 Tage (15 %) für die Vorbereitung, 40,6 Tage (58 %) für die Durchführung und 19,62 Tage (28 %) für die Nachbereitung und damit 70,51 Tage für eine gesamte PEV. Allerdings variieren diese Werte zwischen den evaluierten Vorhaben stark. Die Auswertungen zeigen, dass **ein höherer Auftragswert eines Vorhabens nicht mit einer höheren Anzahl an Gutachtertagen einhergeht**. Dieser Annahme lag die Hypothese zugrunde, dass mit einem höheren Auftragswert auch eine höhere Komplexität der Vorhabensarchitektur einhergeht, die dann in einer Evaluierung entsprechend durch mehr Gutachtertage untersucht werden müsste. Diese Annahme konnte allerdings nicht durch einen signifikanten Zusammenhang bestätigt werden (siehe Abbildung 7). Bislang gibt es hierzu noch keine GIZ-Vorgaben.

Abbildung 7: Kreuzung des Auftragswerts der Vorhaben mit der Anzahl der Gutachtertage



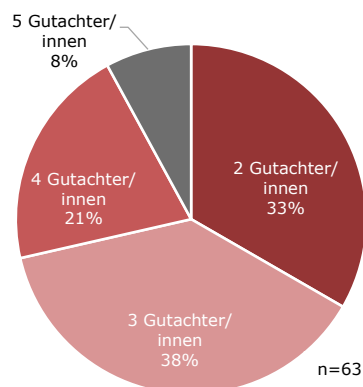
Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

⁶ Nicht aus allen Evaluierungsberichten konnte die Phase des jeweiligen Vorhabens abgeleitet werden. Deswegen verringert sich die entsprechende Grundgesamtheit auf n=60.

⁷ Die Angaben hierzu wurden aus den PEV-Berichten, der Superliste der STS und dem GIZ-Intranet abgeleitet und mit den Angaben der Online-Befragung validiert.

Insgesamt wurden die analysierten PEV von zwei bis fünf Gutachter/innen durchgeführt (siehe Abbildung 8). Größtenteils bestanden die **PEV-Teams sowohl aus internen als auch aus externen Gutachter/innen** (n=54). Nur in Ausnahmen wurden ausschließlich interne Gutachter/innen (n=5) oder ausschließlich externe Gutachter/innen (n=3) eingesetzt.

Abbildung 8: Anzahl der Gutachter/innen



Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

4 Teil I: Ergebnisse zur methodischen Qualität (Genauigkeit)

Die Meta-Evaluierung der PEV im Evaluierungsstandard Genauigkeit ist wie folgt aufgebaut:

Im Kapitel 4.1 werden die Ergebnisse entlang der 13 Bewertungskriterien dargestellt, aus denen sich der Evaluierungsstandard Genauigkeit (G) zusammensetzt. Zur Feststellung der PEV-Qualität in den 13 Bewertungskriterien wurde auf Indikatoren zurückgegriffen. Die Werte der Indikatoren im Evaluierungsstandard Genauigkeit basieren ausschließlich auf der Textanalyse der PEV-Berichte. Die Indikatoren der Textanalyse werden mit „erfüllt“ (100%) oder „nicht erfüllt“ (0%) bewertet. In dem vorliegenden Bericht wird der Erfüllungsgrad der Bewertungskriterien und dazugehörigen Indikatoren im Durchschnitt aller 70 PEV-Berichte dargestellt. Die Ergebnisse sind in der Tabelle im Anhang 9.1 im Überblick dargestellt. **Für weitere Informationen zum Bewertungssystem siehe auch Abbildung 4 in Kapitel 2.2.**

Kapitel 4.2 stellt die Einflussfaktoren auf die methodische Qualität dar, welche auf Grundlage der vorliegenden Daten identifiziert werden konnten.

Kapitel 4.3 enthält schließlich die Bewertung und Schlussfolgerung zur methodischen Qualität durch das Meta-Evaluierungsteam. Dies beinhaltet eine Abwägung der identifizierten Stärken, Schwächen und Einflussfaktoren. Dazu gehört, dass diese Aspekte hinsichtlich vorhandener bzw. fehlender GIZ-Vorgaben reflektiert werden.

4.1 Beschreibung und Analyse des Evaluierungsstandards Genauigkeit

Der Evaluierungsstandard Genauigkeit befasst sich mit der **methodischen Qualität** einer Evaluierung. Die DeGEval definiert den Evaluierungsstandard wie folgt:

„Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt“ (DeGEval 2008).

Dafür formuliert die DeGEval 13 Bewertungskriterien, deren Einhaltung die methodische Qualität einer Evaluierung sicherstellt. Die folgende Abbildung fasst diese Bewertungskriterien übersichtlich zusammen und zeigt deren durchschnittliche Erfüllung, wie sie sich in den jeweils analysierten Indikatoren in den evaluierten PEV darstellt, in Prozent auf:

Abbildung 9: Erfüllungsgrad der Bewertungskriterien im Evaluierungsstandard Genauigkeit

Evaluierungsstandard Genauigkeit (G)	59 %
G 1 Beschreibung des Evaluationsgegenstandes	66 %
G 2 Rahmenbedingungen	74 %
G 3 Beschreibung von Zwecken und Vorgehen	64 %
G 4 Angabe von Informationsquellen	45 %
G 5 Valide und reliable Informationen	56 %
G 6 Systematische Fehlerprüfung	71 %
G 7 (A) Analyse qualitativer und quantitativer Informationen	58 %
G 7 (B) Analyse qualitativer und quantitativer Informationen: Relevanz	65 %
G 7 (C) Analyse qualitativer und quantitativer Informationen: Effektivität	56 %
G 7 (D) Analyse qualitativer und quantitativer Informationen: Effizienz	28 %
G 7 (E) Analyse qualitativer und quantitativer Informationen: Impact	49 %
G 7 (F) Analyse qualitativer und quantitativer Informationen: Nachhaltigkeit	61 %
G 8 Begründete Analyse und begründete Schlussfolgerungen	50 %

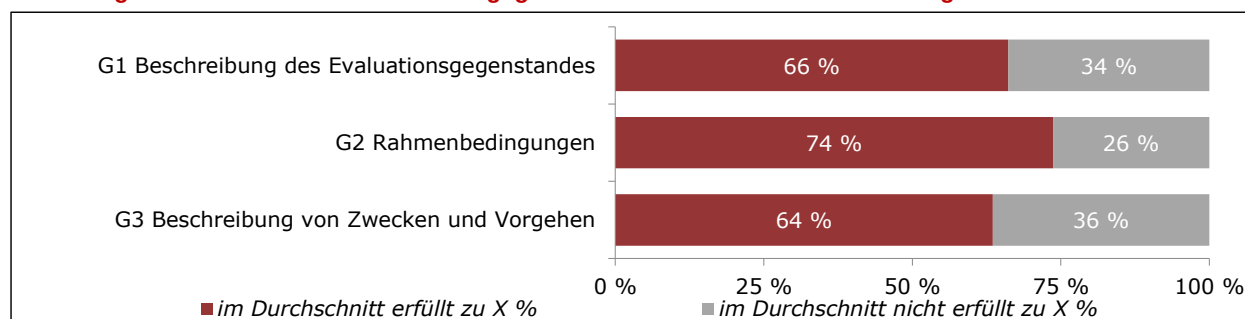
Quelle: Syspons 2016 (Textanalyse der PEV-Berichte und Online-Befragung)

TEIL I: ERGEBNISSE ZUR METHODISCHEN QUALITÄT (GENAUIGKEIT)

Im Durchschnitt aller 70 PEV-Berichte wird der Evaluierungsstandard Genauigkeit laut der Textanalyse **bedingt⁸ erfüllt** (59 %). Dies ist der niedrigste Wert aller vier Evaluierungsstandards.⁹

Die ersten drei Bewertungskriterien zur methodischen Qualität befassen sich mit **übergeordneten methodischen Aspekten**.

Abbildung 10: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien G 1–G 3



Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

G 1 Beschreibung des Evaluationsgegenstandes: Das erste Bewertungskriterium zeigt, ob in den PEV-Berichten die GIZ-Vorhaben klar und genau beschrieben werden. Das Kriterium setzt sich aus drei Indikatoren zusammen. Diese messen, inwiefern eine Definition und Abgrenzung des Evaluierungsgegenstands erfolgt (G 1.1), das Wirkungsmodell dargestellt (G 1.2) und die relevanten Wirkungshypothesen (G 1.3) beschrieben werden.

Wie die Analyse der PEV-Berichte zeigte, erfüllen die PEV im Mittel 66 % der dem Bewertungskriterium G 1 zugeordneten Indikatoren. Das Kriterium ist folglich eher erfüllt (siehe Abbildung 10). So stellen rund zwei Drittel der Berichte (69 %, n=48) den Evaluierungsgegenstand hinreichend deutlich dar (G 1.1). Ein differenzierter Blick zeigt vor allem einen niedrigen Wert bei der Darstellung des finanziellen Budgets, welches lediglich in 24 % der PEV-Berichte (n=17) genannt wird. Obwohl die GIZ vielfach auf ihre langjährige Praxiserfahrung im Capacity-Development hinweist, thematisieren nur 53 % der PEV-Berichte (n=37), wie die Vorhaben auf der individuellen, organisationalen und gesellschaftlichen Ebene zu verorten sind. Höhere Werte finden sich hingegen bei der zeitlichen Abgrenzung (81 %, n=57) und der inhaltlichen Abgrenzung bezüglich der Systemgrenze von Vorhaben (77 %, n=54).

Weiterhin stellen zwar 89 % (n=62) der Berichte das Wirkungsmodell der Vorhaben dar, indem sie zumeist das von den Vorhaben erstellte Wirkungsmodell graphisch integrieren (G 1.2). Allerdings befassen sich deutlich weniger PEV-Berichte auch nachvollziehbar mit den zugrunde gelegten Wirkungsannahmen. 41 % (n=29) der Berichte thematisieren, wo und wie die Wirkungszusammenhänge des Wirkungsmodells zu verorten sind und inwieweit diese plausibel sind (G 1.3).

G 2 Rahmenbedingungen: Im zweiten Bewertungskriterium wird untersucht, ob in den PEV-Berichten der Kontext der GIZ-Vorhaben ausreichend detailliert untersucht worden ist. Eine fundierte Analyse der Rahmenbedingungen soll bei der Interpretation der Analyseergebnisse einen Kontextbezug erlauben. Das Bewertungskriterium wird durch vier Indikatoren belegt. Die ersten drei Indikatoren messen, ob in den PEV-Berichten das evaluierte Vorhaben im Politikkontext (G 2.1), im Entwicklungskontext des Sektors des Partnerlands (G 2.2) und in der Träger- und Partnerstruktur verortet wird (G 2.3). Der vierte Indikator thematisiert, ob in der Bewertung und in den Schlussfolgerungen der PEV-Berichte ein Rückbezug zur Kontextanalyse vorgenommen wird (G 2.4).

Im Mittel der Berichte wird das Bewertungskriterium G 2 mit 74 % eher erfüllt (siehe Abbildung 10). Auf Ebene der dahinterliegenden Indikatoren lassen sich jedoch Schwankungen feststellen. So gelingt den

⁸ Um die Bewertungen auf Ebene der Evaluierungsstandards und Bewertungskriterien einheitlich und nachvollziehbar zu gestalten, werden diese wie folgt angegeben: 90–100 % (größtenteils bis vollständig erfüllt); 80–89 % (zumeist erfüllt); 60–79 % (eher erfüllt); < 60 % (bedingt erfüllt). Siehe hierzu auch Kapitel 2.1.

⁹ Evaluierungsstandards Durchführbarkeit (63 %), Fairness (76 %) und Nützlichkeit (74 %).

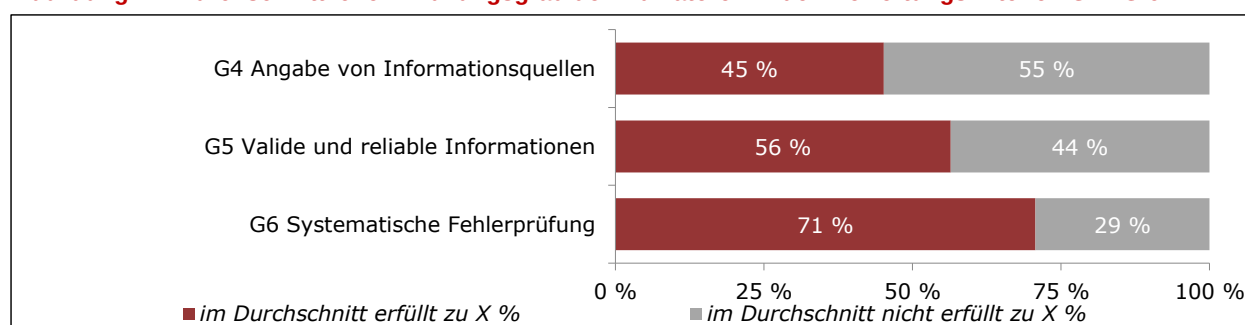
PEV-Berichten weitgehend eine Bezugnahme zum Politikkontext des Partnerlandes (bilaterale Vorhaben), der Partnerregion (regionale Vorhaben) oder zu internationalen Diskursen (Sektorvorhaben) (84 %, n=59) (G 2.1) sowie eine entsprechende sektorale Verortung¹⁰ (84 %, n=47) (G 2.2). Zudem wird in etwas mehr als zwei Dritteln der Berichte (70 %, n=49) deutlich, wie sich die Partnerstruktur zusammensetzt (G 2.3). Allerdings zeigt die Untersuchung ebenfalls, dass bei der Darstellung der PEV-Ergebnisse und Schlussfolgerungen nicht durchgängig auf das Kontextwissen zurückgegriffen wird. Nur zum Teil (60 %, n=42) werden in den Analysen der OECD-DAC-Kriterien die Rahmenbedingungen des Vorhabens aufgegriffen und Bewertungen entsprechend kontextualisiert (G 2.4).

G 3 Beschreibung von Zwecken und Vorgehen: Das dritte Bewertungskriterium erhebt, ob die PEV-Berichte transparent zentrale Informationen zum Zweck, zu den Zielen und zum Vorgehen der Evaluierung darstellen. Diese Transparenz soll es ermöglichen, zu beurteilen, inwieweit die PEV-Berichte den gesetzten Zielen der Evaluierung entsprechen. Das Bewertungskriterium setzt sich aus vier Indikatoren zusammen. Diese messen, inwiefern in den PEV-Berichten Zweck, Ziel und beabsichtigte Verwendung beschrieben (G 3.1, G 3.2) sowie in den Terms of Reference (TOR) Anlass und Evaluierungsfragen der PEV zu finden sind (G 3.3, G 3.4).

Die PEV-Berichte erfüllen im Mittel 64 % dieser vier Indikatoren. Somit ist das Kriterium eher erfüllt (siehe Abbildung 10). Im Detail zeigt sich, dass die TOR den PEV-Berichten fast durchgehend angehängt sind (96 %, n=67) (G 3.3) und den Zweck der Evaluierung transparent machen (87 %, n=61) (G 3.4). Die Berichte selbst hingegen machen ihren Anlass und Zweck (40 %, n=28) (G 3.1) sowie die spezifischen Evaluierungsziele nur bedingt deutlich (31 %, n=22) (G 3.2). Die Ergebnisse und das methodische Vorgehen in den PEV-Berichten werden demnach in geringem Maße vor dem Hintergrund der PEV-Zwecke, -Ziele und beabsichtigten Verwendung reflektiert. Es bleibt daher teils unklar, ob sich die PEV-Teams ausreichend mit den Zielvorstellungen der PEV auseinandergesetzt haben. Dies ist allerdings umso wichtiger, da PEV als Evaluierung mit oder ohne Folgemaßnahme konzipiert sein können und sich demnach die Zwecke maßgeblich unterscheiden können.

Die nächsten drei Bewertungskriterien untersuchen die **Qualität in der methodischen Durchführung**.

Abbildung 11: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien G 4–G 6



Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

G 4 Angabe von Informationsquellen: Das vierte Bewertungskriterium prüft, ob die in den Evaluierungen genutzten Informationsquellen angemessen dokumentiert werden. Hierfür wurde anhand von sechs Indikatoren kontrolliert, ob den PEV-Berichten eine vollständige Liste der genutzten Quellen angehängt ist (G 4.1) sowie ob die Systematik der Dokumentenauswertung (G 4.2) und der Auswahl von Gesprächspartner/innen bzw. die Ziehung von Stichproben deutlich wird (G 4.3). Auch wurden mithilfe der Indikatoren die Verlässlichkeit der Daten des Wirkungsorientierten Monitorings (WOM) (G 4.4), der Baseline Daten (G 4.5) und der Partnerdaten (G 4.6) bewertet und deren Verwendung in der PEV transparent gemacht.

¹⁰ In Bezug auf den Indikator G 2.2 wurden Sektorvorhaben und Globalvorhaben mit „unpassend“ bewertet. Eine entsprechende Kontextualisierung dieser Vorhaben wurde bereits bei Indikator G.2.1 erwartet. Deswegen verringert sich die entsprechende Grundgesamtheit auf n=56.

Im Mittel der Berichte wird das Bewertungskriterium G 4 mit 45 % nur bedingt erfüllt (siehe Abbildung 11). Während eine vollständige Darstellung der Quellen in den PEV-Berichten noch am ehesten gelingt (60 %, n=42) (G 4.1), stellen die PEV-Berichte nicht konsequent dar, wie und warum Dokumente (16 %, n=11) (G 4.2) und Interviewpartner/innen bzw. Befragungsteilnehmer/innen (56 %, n=39) (G 4.3) ausgewählt wurden. Im diesem Kontext zeigte die Analyse der PEV-Berichte auch, dass sich PEV-Teams zwar mehrheitlich mit der Qualität der WOM-Daten eines Vorhabens und den daraus resultierenden Konsequenzen einer möglichen Nutzung in der Evaluierung (71 %, n=50) (G 4.4) befassen, dies jedoch bei Baseline- (36 %, n=25) (G 4.5) und Partnerdaten¹¹ (30 %, n=17) (G 4.6) seltener geschieht. Insgesamt zeigen diese Ergebnisse, dass eine systematische und kritische Auseinandersetzung mit den vorhandenen Informationsgrundlagen durch die PEV nur bedingt erfolgt.

G 5 Valide und reliable Informationen: Als Grundlage einer zuverlässigen Evaluierung sollen die Verfahren zur Datengewinnung so gewählt werden, dass sie die Zuverlässigkeit der erhobenen Daten sicherstellen. Dieses fünfte Bewertungskriterium prüft anhand von zwei Indikatoren¹², inwiefern eine Datentriangulation (G 5.1) und eine Methodentriangulation (G 5.2) in den PEV genutzt wurden, um die Validität der Evaluierungsergebnisse und Schlussfolgerungen zu erhöhen und eine Fehlerüberprüfung zu ermöglichen.

Im Mittel der Berichte wird das Bewertungskriterium G 5 mit 56 % bedingt erfüllt (siehe Abbildung 11). So enthalten 77 % (n=54) der PEV-Berichte Belege für die **Nutzung multipler Daten** zu demselben Sachverhalt, sodass eine Datentriangulation angenommen wurde (G 5.1). Bezüglich der Methodentriangulation trifft dies hingegen nur auf 36 % (n=25) der PEV-Berichte zu. Demnach werden in rund einem Drittel der Berichte mindestens drei Methoden zur Klärung desselben Sachverhaltes kombiniert (G 5.2). Anhaltspunkte für die Nutzung von Triangulationsansätzen fanden sich in den Berichten vor allem in den Kapiteln zur Darstellung der methodischen Vorgehensweise oder konnten aus dem Kapitel zur Bewertung der entwicklungspolitischen Wirksamkeit abgeleitet werden.

G 6 Systematische Fehlerprüfung: Das sechste Bewertungskriterium befasst sich ebenfalls mit Ansätzen der Triangulation, mit denen die gesammelten Informationen systematisch auf Fehler überprüft werden können. Die entsprechenden Indikatoren¹³ operationalisieren demnach, ob die PEV Mechanismen entwickelt haben, um mit widersprüchlichen Ergebnissen umzugehen und fehlerhafte Aussagen zu vermeiden. Der dazugehörige erste Indikator erhebt die Durchführung einer Forschertriangulation (G 6.1), während sich der zweite Indikator mit der Validierung der Ergebnisse mit beteiligten und betroffenen Akteuren befasst (G 6.2).

Diesbezüglich erreicht das Bewertungskriterium G 6 rund 71 % und wird damit eher erfüllt (siehe Abbildung 11). So geben die Berichte teils Anhaltspunkte dafür, dass die Perspektiven im PEV-Team trianguliert (57 %, n=40) (G 6.1) und die PEV-Ergebnisse mit den Partnern geteilt bzw. diskutiert worden sind (84 %, n=59), bspw. durch Abstimmungstermine und Debriefings (G 6.2). Sowohl qualitative als auch quantitative Analysen erfordern diesen iterativen Prozess, in welchem die eigenen Interpretationen und Folgerungen einer Überprüfung unterzogen werden müssen.

Betrachtet man in einer zusätzlichen Auswertung¹⁴ die vier Indikatoren von G 5 und G 6 gemeinsam, so ergibt sich ein durchwachsenes Bild der eingesetzten Triangulationsansätze. Während 16 % der PEV (n=11) auf nur einen Ansatz der Triangulation zurückgreifen, greift rund jede dritte PEV auf zwei oder drei Ansätze und nur jede fünfte PEV (20 %, n=14) auf alle identifizierten Ansätze der Triangulation zurück.

¹¹ Sektorvorhaben und Globalvorhaben wurden hier mit „unpassend“ bewertet, da sie keine Partner in Partnerländern haben. Deswegen verringert sich die entsprechende Grundgesamtheit auf n=56.

¹² Einer Datentriangulation liegt in dieser Meta-Evaluierung der Anspruch zugrunde, dass mindestens drei Daten zu einem Sachverhalt herangezogen werden müssen; einer Methodentriangulation, dass mindestens drei Methoden zu einem Sachverhalt herangezogen werden müssen.

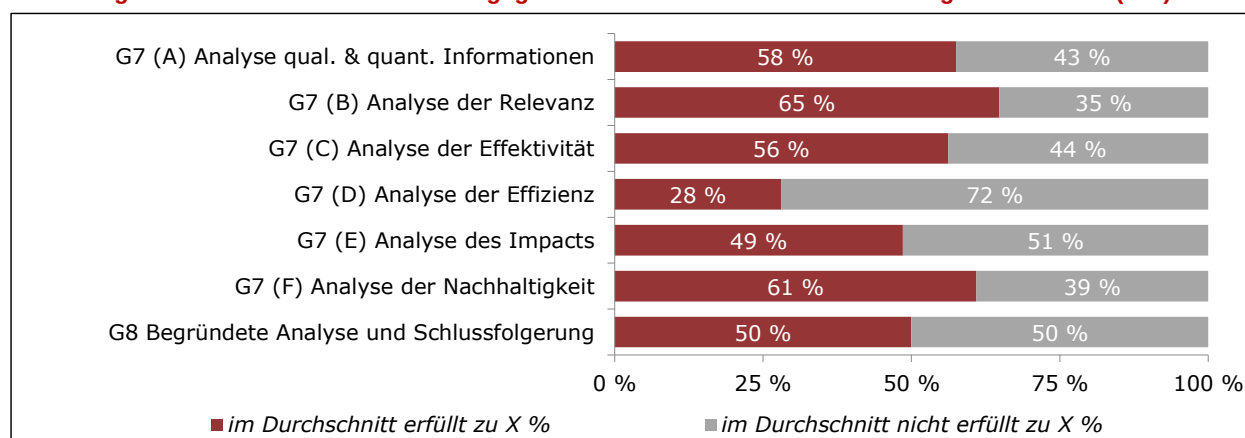
¹³ Einer Forschertriangulation liegt in dieser Meta-Evaluierung der Anspruch zugrunde, dass mindestens die Perspektiven von drei Forschern zu einem Sachverhalt herangezogen werden müssen.

¹⁴ Diese zusätzliche Auswertung fließt nicht in die Bewertung mit ein.

TEIL I: ERGEBNISSE ZUR METHODISCHEN QUALITÄT (GENAUIGKEIT)

Die nachfolgenden sieben Bewertungskriterien im Evaluierungsstandard Genauigkeit befassen sich mit der **Angemessenheit der Analyse und Bewertung** in den PEV-Berichten

Abbildung 12: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien G 7 (A-F)–G 8



Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

G 7 (A): Analyse qualitativer und quantitativer Informationen: Ausgangspunkt einer qualitativ guten Evaluierung ist eine angemessene Auswahl der methodischen Vorgehensweise. Dieser Standard wird im siebten Bewertungskriterium abgebildet und wurde entlang von vier Indikatoren geprüft. Der erste Indikator misst, inwieweit die methodische Vorgehensweise in der PEV die Frage nach der Zuordnungs- und/ oder Beitragsanalyse reflektiert (G 7.1). Weitere Voraussetzungen für eine qualitativ hochwertige Analyse der Informationen sind die Darstellung der Evaluierungsmethoden (G 7.2), die Nutzung einer Methodenvielfalt (G 7.3) und die Diskussion der Vor- und Nachteile der ausgewählten methodischen Vorgehensweise (G 7.4).

Insgesamt wird das Bewertungskriterium G 7 (A) im Durchschnitt der Berichte mit 58 % bedingt erfüllt (siehe Abbildung 12). Zu Beginn steht laut des ersten Indikators (G 7.1) eine Auseinandersetzung mit der Frage, welches „maßgeschneiderte“ methodische Design (theoriebasierter Ansatz, Kontributionsanalyse, experimentelles oder quasi-experimentelles Design) gewählt werden muss, um die im Zentrum stehenden Fragen zu beantworten. Eine solche Auseinandersetzung wird in den betrachteten Berichten nur in Ausnahmen geführt. Lediglich 19 % (n=13) der Berichte begründen hierdurch ihre Vorgehensweise und erfüllen damit den Indikator. Im Detail zeigt sich, dass 13 % der Berichte (n=9) dem Ansatz einer Kontributionsanalyse folgen, indem Plausibilitätsprüfungen genutzt werden, um die Wirkungslogiken zu überprüfen und mögliche Ursachen der beobachteten Veränderungen zu identifizieren. Gleichzeitig begründen 6 % der Berichte (n=4) ihre Vorgehensweise auf der Grundlage eines theoriebasierten Ansatzes. Ein (quasi-)experimentelles Design hingegen konnte in keinem PEV-Bericht identifiziert werden.

Wie die Ergebnisse der Textanalyse weiter zeigen, werden in 81 % (n=57) der Berichte die verwendeten Methoden beschrieben (G 7.2). Deren Stärken und Schwächen legt jede dritte PEV dar (31 %, n=22) (G 7.4). Allerdings kann aus der Textanalyse nicht abgeleitet werden, inwieweit die beschriebenen Methoden (Dokumentenauswertung, Fokusgruppen, qualitative Interviews, quantitative Befragungen, eigene Auswertung der Monitoringdaten) tatsächlich auch in der Durchführung methodischen Qualitätskriterien entsprechen, da die entsprechenden Prüfmissionen nicht vor Ort begleitet wurden.

Die betrachteten PEV greifen fast durchgehend auf mindestens zwei Methoden (99 %, n=69) zurück, wodurch der entsprechende Indikator (G 7.3) fast vollständig erfüllt wird. Allerdings zeigt eine Zusatzauswertung¹⁵, dass die Methodenvielfalt der PEV insgesamt begrenzt ist. Demnach baut rund die Hälfte der PEV (53 %, n=37) ausschließlich auf Interviews und Dokumentenanalysen auf. Gleichzeitig

¹⁵ Diese zusätzliche Auswertung fließt nicht in die Bewertung mit ein.

wenden 37 % (n=19) der PEV drei Methoden und 9 % (n=6) der PEV vier oder mehr Methoden an. In diesem Kontext zeigt eine weitere Zusatzauswertung, dass PEV nur in Ausnahmen (6 %, n=4) qualitative Erhebungen mit Befragungen kombinieren. Konkret setzten zwei PEV von Sektor- bzw. Globalvorhaben und zwei PEV von bilateralen Vorhaben Befragungen ein. Die Gutachtertage reichten in diesen Vorhaben von 61 bis 120. Sie befinden sich damit sowohl über als auch unter dem durchschnittlich eingesetzten Mengengerüst (70,5) einer PEV. Hierbei gehen die Berichte allerdings nur kurz auf die hinter den Befragungen liegende Methodik ein, sodass kaum spezifischere Aussagen zur quantitativen Vorgehensweise in den PEV möglich sind. Allerdings gibt es Anhaltspunkte dafür, dass die Befragungen unterschiedliche Ziele verfolgten. Zum Teil wurden Befragungen eingesetzt, um gezielt Personengruppen in der Breite zu adressieren. So wurden in einem Sektorvorhaben 20 Mitarbeiter/innen der Außenstruktur, in einem anderen Vorhaben Alumni befragt. In den PEV der zwei bilateralen Vorhaben waren die Befragungen den Interviews vorgelagert, um in der Feldphase ausgewählte Aspekte zu vertiefen. Insgesamt wurde die Anzahl der Teilnehmer/innen (mit einer Ausnahme) nicht angegeben, sodass keine Aussagen zur Repräsentativität der erhobenen Daten möglich sind.

G 7 (B): Analyse qualitativer und quantitativer Informationen: Relevanz: Die Analyse und Beurteilung der OECD-DAC-Kriterien ist eine der zentralen Anforderungen an die PEV. Daher wurden diese in dieser Meta-Evaluierung vertiefend untersucht. Da die fünf DAC-Kriterien jeweils verschiedene methodische Anforderungen an PEV-Teams stellen, wird dabei auch nach den fünf Kriterien unterschieden. Das folgende Bewertungskriterium befasst sich zunächst mit der Angemessenheit der methodischen Bearbeitung der Relevanz. Hierzu wurde anhand von drei Indikatoren untersucht, ob die Nachvollziehbarkeit der zugrunde gelegten Rahmenbedingungen gegeben ist (G 7.5) und ob die Mehrdimensionalität der Relevanz (G 7.6) und die strategische Ausrichtung an veränderten Rahmenbedingungen in den PEV-Berichten analysiert wurden (G 7.7).

Insgesamt wird das Bewertungskriterium G 7 (B) im Mittel mit 65 % eher erfüllt. Es ist damit der höchste Wert bei der Bearbeitung der OECD-DAC-Kriterien in den PEV (siehe Abbildung 12). Zwei Drittel der Berichte (69 %, n=48) setzen sich mit der Relevanz der von den Vorhaben identifizierten Kernprobleme auseinander. Diese Auseinandersetzung ist zentral, um die Frage zu beantworten, ob ein einmal konzipiertes Vorhaben sich mehrere Jahre nach seiner Konzeption noch an aktuellen entwicklungspolitischen Fragestellungen orientiert (G 7.5). Zugleich betrachtet die große Mehrzahl der PEV-Berichte (89 %, n=62) die untersuchte Relevanz zumindest zweidimensional. Hierbei wird deutlich, dass fast durchgehend Bezug auf die Grundausrichtung des BMZ genommen wurde, indem Sektorkonzepte oder Länderkonzepte kurz reflektiert wurden. Die Einordnung in Konzepte der Partner und die Kernbedarfe der Zielgruppe erfolgte diesbezüglich ebenfalls größtenteils, wenn auch vergleichsweise seltener (G 7.6). Die Ergebnisse zum letzten Indikator verdeutlichen schließlich, dass nur ein Drittel der PEV-Berichte (37 %, n=26) bei der Bearbeitung der Relevanz die Ausrichtung und Anpassung der Vorhaben an kontextuelle Veränderungen untersuchen (G 7.7).

G 7 (C): Analyse qualitativer und quantitativer Informationen: Effektivität: In diesem Bewertungskriterium wird die fundierte methodische Bearbeitung der Effektivität (Zielerreichung eines Vorhabens) geprüft. Hierzu wurde im ersten Indikator erhoben, ob die Kausalität zwischen Maßnahme und Wirkung kritisch analysiert wurde (G 7.8). Darauf aufbauend wurde in zwei weiteren Indikatoren gemessen, ob die Zielerreichung anhand der Modulzielindikatoren bewertet wurde (G 7.9) und die Qualität der verwendeten Indikatoren ausreichend ist (G 7.10).

Die Berichte erreichen im Bewertungskriterium G 7 (C) im Mittel 56 %, was einer bedingten Erfüllung gleichkommt (siehe Abbildung 12). Auf Ebene der Indikatoren zeigt sich, dass sich die PEV fast durchgehend (91 %, n=64) an den Indikatoren des Modulziels orientieren (G 7.9). Allerdings zeigt die Textanalyse der Berichte auch, dass die Aussagekraft der vorgenommenen Analysen und der Bewertungen der Effektivität kritisch zu betrachten ist. Zum einen hinterfragen weniger als die Hälfte der Berichte (41 %, n=29) explizit oder implizit, inwiefern die beobachteten Wirkungen überhaupt auf das Vorhaben oder aber auf externe Faktoren zurückzuführen sind. Das Kausalitätsproblem wird folglich zumeist nicht adressiert (G 7.8). Zum anderen setzen sich nur etwa ein Drittel (36 %, n=25) der PEV-

Teams in den Berichten mit der Qualität der Indikatoren auseinander und verändern bei identifizierter geringer Qualität auch die entsprechenden Indikatoren. Es wird deutlich, dass PEV-Teams sich zwar oftmals kritisch gegenüber den Vorhabensindikatoren geben und auf Basis einer SMART-Analyse auf Mängel hinweisen. In diesem Kontext wird u. a. kritisiert, dass Indikatoren nicht spezifisch genug für die Messung der Zielerreichung sind (bspw. weil sie mehrstufig formuliert sind). Allerdings führt diese Kritik selten auch zu einer geforderten Überarbeitung bzw. Umformulierung der Indikatoren zum Zwecke einer besseren Wirkungsmessung durch das PEV-Team (G 7.10) – ein Aspekt, der auch in den explorativen Interviews der Meta-Evaluierung kritisiert wurde.

G 7 (D): Analyse qualitativer und quantitativer Informationen: Effizienz: Ausgehend von den Anforderungen an eine methodisch fundierte PEV sollen diese eine Analyse und Bewertung der Effizienz der von ihnen evaluierten Vorhaben vornehmen. Die Einschätzung dieses Bewertungskriteriums basiert auf drei Indikatoren. So wurde untersucht, ob verschiedene Dimensionen der Effizienz analysiert (G 7.11) und die Auswahl von Verfahren der Effizienzmessung begründet (G 7.12) wurden. Es wurde auch analysiert, inwiefern die methodische Bearbeitung der Effizienz die Identifikation von Potenzialen zur Effizienzsteigerung ermöglicht (G 7.13).

Im Mittel der Berichte werden 28 % der Indikatoren erreicht, wodurch das Bewertungskriterium G 7 (D) als bedingt erfüllt angesehen werden kann (siehe Abbildung 12). Das Effizienzkriterium ist im Vergleich der DAC-Kriterien das methodisch am schwächsten bearbeitete Kriterium. Dabei ist zu beachten, dass der erste Indikator (G 7.11) bereits positiv bewertet wurde, wenn die Implementierungs-, Produktions- oder Allokationseffizienz untersucht wurde. Insgesamt nimmt etwas mehr als die Hälfte der Berichte (56 %, n=39) mindestens eine dieser Effizienzanalysen vor. Diesbezüglich setzt sich mehr als jede zweite PEV (56 %, n=39) im Sinne der Implementierungseffizienz zumindest ansatzweise mit der Effizienz der Vorhabensteuerung, der Kooperation mit anderen Gebern oder Unterauftragnehmern auseinander. 10 % der Berichte (n=7) weisen Ansätze einer Analyse der Produktionseffizienz auf, bewerten also die eingesetzten Ressourcen in Relation zu den Outputs. Nur in Ausnahmen (3 %, n=2) setzen PEV im Sinne der Allokationseffizienz Ressourcen und Outcomes zueinander in Bezug. Allgemeine Aussagen, wonach evaluierte Vorhaben mit wenigen Mitteln viele Wirkungen auf Outcome- oder Impactebene erzielen konnten, wurden nicht positiv bewertet, wenn nicht deutlich wurde, woraus und wie diese Aussagen abgeleitet wurden.

In diesem Kontext ist auch die sehr geringe Auseinandersetzung der Berichte mit der Auswahl der Verfahren der Effizienzmessung auffallend. Nur in Ausnahmen legen PEV dar, dass z.B. aufgrund von fehlenden Daten Aspekte der Effizienz nicht erhoben werden konnten. Der entsprechende Indikator (G 7.12) wird damit von allen Indikatoren am geringsten erfüllt (11 %, n=8). Schließlich setzen sich PEV in ihren Berichten auch kaum mit der Frage auseinander, ob Potenziale zur Steigerung der Effizienz systematisch identifiziert werden können (G 7.13: 17 %, n=12).

G 7 (E): Analyse qualitativer und quantitativer Informationen: Impact: Dieses Bewertungskriterium prüft, ob in den PEV-Berichten eine methodisch fundierte Bewertung des Impacts der evaluierten Vorhaben vorgenommen wird. Das Bewertungskriterium basiert auf drei Indikatoren. Im ersten Indikator wurde erhoben, ob bei der Analyse und Bewertung der übergeordneten Wirkungen die Attributionslücke (auch „Zuordnungslücke“ oder „Systemgrenze“) adressiert wurde (G 7.14). Zudem wurde geprüft, inwiefern sich die PEV-Berichte mit der Plausibilität der Impact-Hypothesen auseinandersetzen (G 7.15) und ob die Bewertungsmaßstäbe für die Impact-Bewertung transparent dargestellt werden (G 7.16).

Mit einer Erfüllung von 49 % im Durchschnitt der Berichte wird das Bewertungskriterium G 7 (E) bedingt erfüllt (siehe Abbildung 12). Im Detail zeigt sich auf Ebene der Indikatoren, dass 67 % (n=47) der Berichte die Attributionslücke zwischen den Maßnahmen eines Vorhabens und den übergeordneten Wirkungen berücksichtigen. Diesbezüglich wurden bereits Anhaltspunkte in den Formulierungen positiv bewertet, die auf eine solche Auseinandersetzung hindeuten (z. B. „das Vorhaben leistet einen Beitrag zu“) (G 7.14).

Die Aussagekraft der PEV-Berichte zu den Impacts ist nicht allein aufgrund der Nicht-Erfüllung dieses Indikators eingeschränkt. Sie wird auch durch die Auswertungen des zweiten Indikators beeinträchtigt. So

werden Hypothesen zu den Impacts nicht immer unter dem Gesichtspunkt ihrer Plausibilität reflektiert, sondern teils als gegeben angenommen. Auch wird der Zeitpunkt der Impact-Messung nicht konsequent problematisiert¹⁶ (43 %, n=30) (G 7.15). Zudem wird die Aussagekraft der Impact-Bewertungen in den PEV-Berichten dadurch in Frage gestellt, dass nur jede dritte PEV (36 %, n=25) ihre Bewertungsmaßstäbe transparent macht. Diesbezüglich wird nur sehr selten ein Bezug zu einem ggfs. vorhandenen EZ-Programm genommen (G 7.16).

G 7 (F): Analyse qualitativer und quantitativer Informationen: Nachhaltigkeit: Auch das Bewertungskriterium zur methodischen Bearbeitung der Nachhaltigkeit wird durch drei Indikatoren operationalisiert. Es untersucht, ob die Grenzen der Nachhaltigkeitsmessung beschrieben (G 7.17), die angelegten Ansätze zur Schaffung von Nachhaltigkeit im Vorhaben analysiert (G 7.18) und mindestens zwei Ebenen der Nachhaltigkeit analysiert werden (G 7.19).

Laut der durchschnittlichen Bewertung der PEV-Berichte ist das Bewertungskriterium G 7 (F) mit 61 % eher erfüllt (siehe Abbildung 12). Nur rund jede zweite PEV (47 %, n=33) reflektiert, inwieweit zum Zeitpunkt der Evaluierung eine abschließende Bewertung der Nachhaltigkeit bereits möglich ist (G 7.17). Gleichzeitig setzen sich etwa zwei Drittel der PEV (63 %, n=44) damit auseinander, inwiefern sich die im Vorhaben angelegten Ansätze (z.B. in Form einer Exitstrategie) positiv oder negativ auf die Nachhaltigkeit auswirken können. Hierbei ist auffallend, dass oftmals Aspekte der Effektivität thematisiert werden, weniger aber gezielt Aspekte der Nachhaltigkeit (G 7.18). Im Gegensatz hierzu gelingt es den PEV öfter, die Nachhaltigkeit mehrdimensional zu analysieren (73 %, n=51). Dabei analysieren PEV die institutionelle und die personelle Nachhaltigkeit, vergleichsweise seltener jedoch die finanzielle Nachhaltigkeit der Wirkungen (G 7.19).

G (8): Begründete Analyse und Schlussfolgerungen: Das letzte Bewertungskriterium erhebt, ob Analysen und Schlussfolgerungen der PEV begründet werden. Dabei befassen sich die ersten beiden Indikatoren des Kriteriums damit, ob zwischen Beschreibung, Analyse und Bewertung unterschieden wird (G 8.1) und die Analysen belegt sind (G 8.2). Zudem wurde anhand von Indikatoren erhoben, ob die formulierten Handlungsempfehlungen den methodischen Anforderungen gerecht werden (G 8.3) und nicht-intendierte (positive oder negative) Wirkungen beschrieben werden (G 8.4).

Insgesamt ist das Bewertungskriterium bedingt erfüllt, da im Mittel 50 % der Berichte die entsprechenden Indikatoren erfüllen (siehe Abbildung 12). Hierbei zeigt sich, dass den PEV-Berichten der Dreischritt zwischen Beschreibung, Analyse und Bewertung nur teils (33 %, n=23) gelingt. Diese Trennung gehört zu den grundlegenden Evaluierungsstandards. Sie macht transparent, wie PEV-Teams zu einer Beurteilung gelangen und Schlussfolgerungen herleiten (G 8.1). Laut den Evaluierungsstandards sollen Beschreibungen und Analysen durch Informationsquellen belegt werden, um Ergebnisse nachvollziehen und mögliche Verzerrungen identifizieren zu können. Dies gelingt jeder dritten PEV (G 8.2: 30 %, n=21). Es wird demnach überwiegend nicht deutlich, auf welche Quellen sich die Aussagen in den PEV-Analysen beziehen. Ebenso verdeutlicht die Auswertung des dritten Indikators (G 8.3), dass die Empfehlungen der Evaluierungen in zwei von drei PEVs (64 %, n=45) den erfordernten Qualitätsstandards entsprechen. Kritikpunkte sind hierbei vor allem, dass Empfehlungen nicht immer konsequent aus der Analyse abgeleitet werden, nicht durchgehend spezifisch sind und die Adressaten teils unklar bleiben.

Abschließend wurde geprüft, ob sich die PEV mit möglichen positiven und/ oder negativen nicht-intendierten Wirkungen befassen. Dies trifft auf rund drei Viertel (73 %, n=51) der PEV zu. Dabei beschreiben PEV nicht-intendierte Wirkungen oder schließen diese im Kontext der Analysen der Effektivität oder der Impacts zumindest aus (G 8.4). Inwieweit diesem Ausschluss eine vertiefende Auseinandersetzung mit nicht-intendierten Wirkungen zugrunde liegt, kann auf Grundlage der PEV-Berichte zumeist nicht nachvollzogen werden.

¹⁶ Also die Frage, inwiefern langfristige Wirkungen eines Vorhabens bereits nach zweieinhalb Jahren Implementierung überhaupt gemessen werden können.

4.2 Einflussfaktoren auf die methodische Qualität

Anhand statistischer Analysen wurden Einflussfaktoren auf die methodische Qualität der PEV-Berichte identifiziert und geprüft. Es wurde sowohl der Einfluss der Rahmendaten (z. B. die Vorhabensdaten und die Strukturmerkmale einer PEV) als auch einzelner methodischer Aspekte analysiert, um mögliche Ansatzpunkte (sog. „Stellschrauben“) zur Verbesserung der PEV im Evaluierungsstandard Genauigkeit zu identifizieren (siehe Abbildung 13).

Abbildung 13: Einflussfaktoren Regionalbereich, Umfang der Methoden, Anzahl der Gutachter/innen

Einflussfaktoren		N	M	SD
Regionalbereich	nicht-bilateral	15	51%	0,13
	bilateral	55	62%	0,15
Umfang der Methoden	2 oder weniger	38	54%	0,15
	mehr als 2	32	65%	0,13
Anzahl der Gutachter/innen	2 oder weniger	21	50%	0,14
	mehr als 2	42	65%	0,14

N = Anzahl der Berichte; M = Mittelwert der Berichte der PEV im Genauigkeitsstandard; SD = Standardabweichung
Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

Der erste identifizierte Einflussfaktor auf die methodische Qualität ist die **Art des evaluierten Vorhabens (Regionalbereich)**. Es lassen sich signifikante Unterschiede¹⁷ im Hinblick auf die methodische Qualität zwischen **bilateralen Vorhaben und nicht-bilateralen Vorhaben** feststellen. Die methodische Qualität der bilateralen Vorhaben ist im Durchschnitt besser als bei Sektor- und Globalvorhaben. Demnach erfüllen die bilateralen Vorhaben die Genauigkeitsstandards im Mittel zu 62 %. Nicht-bilaterale Vorhaben, also Sektor- und Globalvorhaben, kommen auf einen geringeren Mittelwert von 51 % (siehe Abbildung 13). Ein detaillierter Blick auf die dahinterliegenden Bewertungskriterien zeigt, dass Sektor- und Globalvorhaben, mit Ausnahme von einem methodischen Kriterium, immer schlechter abschneiden als bilaterale Vorhaben. Als signifikante Faktoren gelten hierbei vor allem die Bewertungskriterien zur Beschreibung von Zwecken und Vorgehen (G 3), zur systematischen Fehlerprüfung (G 7) und zur methodischen Bewertung der Relevanz (G 7 B). Dieses Ergebnis gibt Anhaltspunkte dafür, dass **nicht-bilaterale Vorhaben in den aktuellen Evaluierungen vor größere methodische Herausforderungen** gestellt werden als bilaterale Vorhaben¹⁸. Die Rückmeldungen der AV in der Online-Befragung kommen zu einem ähnlichen Befund. Nach Einschätzung einiger AV aus Sektor- und Globalvorhaben sind PEV stärker auf bilaterale Vorhaben zugeschnitten, welche anderen Erwartungen der Auftraggeber unterliegen und in andere Träger- und Zielgruppenstrukturen eingebettet sind.

Ein signifikanter Unterschied¹⁹ hinsichtlich der methodischen Qualität scheint zudem im **Umfang der Methoden** begründet. Während PEV, die **drei oder mehr Methoden einsetzen**, den Genauigkeitsstandard im Durchschnitt zu 65 % erfüllen, kommen PEV, die mit Desk Research und Interviews lediglich auf zwei Methoden zurückgreifen, im Mittel auf 54 % (siehe Abbildung 13).

Ein weiterer signifikanter Einflussfaktor²⁰ ist die **Anzahl der Gutachter/innen** eines PEV-Teams. So erfüllen PEV-Berichte, die von zwei Gutachter/innen durchgeführt wurden, den Genauigkeitsstandard zu durchschnittlich 50 %. PEV-Berichte hingegen, die von mindestens drei Gutachter/innen durchgeführt wurden, kommen mit 65 % auf einen höheren Mittelwert. Dies legt den Schluss nahe, dass PEV methodisch besser abschneiden, wenn **mindestens drei Gutachter/innen** an ihnen mitwirken (siehe Abbildung 13).

¹⁷ $t_{(68)} = 2,61, p = 0,011$.

¹⁸ Die Kreuztabelle in Abbildung 14 beschäftigt sich vertiefend mit den Einflussfaktoren innerhalb bilateraler und nicht-bilateraler Vorhaben. Sie zeigt hierbei auf, dass die identifizierten Einflussfaktoren auch innerhalb der einzelnen Wirkdimensionen gelten.

¹⁹ $t_{(68)} = 3,33, p = 0,001$. * Dieser Wert leitet sich nicht direkt aus den Indikatoren des Bewertungssystems ab, da Indikator G 7.3 (A) die Methodenvielfalt (im Sinne von: mehr als zwei Methoden) und Indikator G 5.2 die Methodentriangulation (im Sinne von: Dokumentation von mindestens drei Methoden zu einem bestimmten Sachverhalt) messen und somit einen anderen Fokus haben. Aus diesem Grund wurden keine korrelativen Indikatoren aus G herausgerechnet.

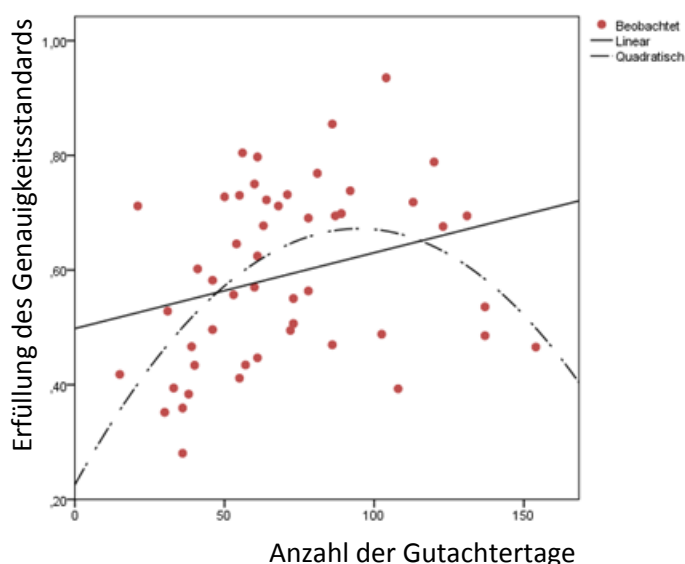
²⁰ $t_{(61)} = 3,88, p < 0,001$.

Gleichzeitig kann die Anzahl der Gutachter/innen als Einflussfaktor nicht unabhängig von dem **Mengengerüst (Gutachtertage)** einer PEV interpretiert werden. Es ist naheliegend, dass die Anzahl der Gutachtertage und der Gutachter/innen sich bedingen – je mehr Gutachter/innen eingesetzt werden, desto mehr Tage stehen einer PEV durchschnittlich auch zur Verfügung. Jedoch verdeutlichte eine Regressionsanalyse, dass die **Anzahl der Gutachter/innen entscheidender** für die Qualität ist als die Anzahl der Gutachtertage. Nichtsdestotrotz gibt die Regressionskurve Anhaltspunkte dafür, dass mit zunehmenden Gutachtertagen **bis zu einem Mengengerüst von circa 86 Tagen** auch die methodische Qualität tendenziell besser wird. Ab diesem Wert ist der Effekt jedoch nicht mehr vorhanden bzw. sogar gegenläufig. Diesbezüglich gilt zu berücksichtigen, dass die Streuung der untersuchten PEV ab diesem Wert verhältnismäßig hoch und die Fallzahlen verhältnismäßig klein sind, sodass hierbei einzelne Vorhaben die Werte stark verändern können (siehe Abbildung 14).

Auch die univariate Varianzanalyse²¹ kommt zu dem Ergebnis, dass es signifikante Unterschiede in der methodischen Qualität zwischen PEV mit unterschiedlichen Mengengerüsten gibt²². Hierzu wurden die PEV in drei Gruppen geteilt: PEV unter 55 Gutachtertagen (n=16); PEV mit 55 bis 80 Gutachtertagen (n=18); PEV mit über 80 Gutachtertagen (n=16). Ein diesbezüglicher statistischer Vergleich weist darauf hin, dass die methodische Qualität in der Gruppe mit den wenigsten Gutachtertagen signifikant schwächer ausfällt als die methodische Qualität der anderen beiden Gruppen.

Insgesamt sollte bei der Interpretation der Ergebnisse zum Einfluss des Mengengerüsts jedoch berücksichtigt werden, dass in den TOR der PEV die Anzahl der Gutachtertage für die Evaluierung und für die Prüfung eines Folgevorhabens **gemeinsam angegeben** werden. Informationen zur Verteilung der Gutachtertage zwischen Evaluierung und Prüfung lagen der Meta-Evaluierung folglich nicht vor.

Abbildung 14: Kreuzung der Genauigkeit und der Anzahl der Gutachtertage



Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

Die Erkenntnisse über den Einfluss des Mengengerüsts auf die methodische Qualität wurden weiter durch die Unterscheidung zwischen Vorbereitung, Durchführung und Nachbereitung vertieft. Dabei zeigte sich, dass vor allem die in die **Nachbereitung einer PEV** investierte Zeit die methodische Qualität einer PEV beeinflusst. Dieser Zusammenhang weist ein signifikantes Niveau auf²³ und ist naheliegend, da die Bewertung der Genauigkeit auf den PEV-Berichten basiert, welche in der Nachbereitungsphase verfasst werden. Gleichzeitig zeigt eine Überprüfung der einzelnen Aspekte der Genauigkeit, dass vor allem die

²¹ Die univariate Varianzanalyse (ANOVA) misst den Einfluss von nur einem Faktor (unabhängige Variable) auf die abhängige Variable.

²² $F_{(2, 47)} = 5,584, p = 0,007$.

²³ $r = 0,37, n = 50, p = 0,01$.

Reflexion der methodischen Qualität (G 4.1–G 6.2) durch das Mengengerüst beeinflusst wird.²⁴ Folglich nimmt ein ausreichend großes Mengengerüst Einfluss auf die Auseinandersetzung mit den Informationsgrundlagen einer PEV und die Verfahren zur Gewinnung valider und reliabler Daten und Informationen.

Die vorangegangenen Analysen deuten darauf hin, dass sich die identifizierten Einflussfaktoren gegenseitig bedingen. Diese Annahme wurde weiter überprüft und in der folgenden Abbildung zusammengefasst (siehe Abbildung 15).

Abbildung 15: Kreuzung zwischen Regionalbereich, Anzahl der Gutachter/innen und Methodenvielfalt

Einflussfaktoren		Anzahl der Gutachter/innen *				Umfang der Methoden			
		2 oder weniger		mehr als 2		2 oder weniger		mehr als 2	
		N	M	N	M	N	M	N	M
Regionalbereich	nicht-bilateral	10	47%	2	63%	10	44%	5	63%
	bilateral	11	53%	40	64%	28	58%	27	66%
Anzahl der Gutachter/innen *	2 oder weniger					15	47%	6	58%
	mehr als 2					20	60%	22	69%

* Bei 4 nicht-bilateralen und 3 bilateralen Vorhaben konnten keine Angaben zu der Anzahl der Gutachter/innen abgeleitet werden. Diese wurden folglich nicht in der Tabelle aufgenommen. N = Anzahl der Berichte; M = Mittelwert im Genauigkeitsstandard

Quelle: Syspons 2016 (Textanalyse der PEV-Berichte)

Zunächst fällt auf, dass die PEV nicht-bilateraler Vorhaben fast ausschließlich von zwei Gutachter/innen durchgeführt wurden, während in den PEV von bilateralen Vorhaben zumeist mehr als zwei Gutachter/innen tätig waren. In den PEV nicht-bilateraler Vorhaben werden zudem eher zwei Methoden eingesetzt, während es in den PEV bilateralen Vorhaben eine breitere Methodenvielfalt gibt. Besteht das PEV-Team aus nur zwei Gutachter/innen, werden etwas öfter zwei Methoden eingesetzt, während die Methodenvielfalt bei PEV mit mehr als zwei Gutachter/innen diverser ist (siehe Abbildung 15).

Gleichzeitig verdeutlicht die Abbildung, dass die **Einflussfaktoren auch ausnahmslos innerhalb der einzelnen Wirkungsdimensionen** erkennbar sind. Die Fallzahlen dabei sind allerdings teils gering. So zeigt sich, dass z.B. der Erfüllungsgrad der Genauigkeit bei bilateralen und nicht nicht-bilateralen Vorhaben niedriger ist, wenn das PEV-Team aus weniger als drei Gutachter/innen besteht (siehe Abbildung 15). Folglich gibt dies Hinweise darauf, dass die identifizierten Einflussfaktoren bestätigt werden können.

Als weitere potenzielle Stellschraube zur Verbesserung der PEV im Evaluierungsstandard Genauigkeit wurde untersucht, ob ein klares **Verständnis des Vorhabens** und die Auseinandersetzung mit der Wirkungslogik eines Vorhabens in der PEV (Index aus: G 1.2, G 1.3, G 8.1) Einfluss auf die methodische Qualität (G insgesamt) haben. Dies wurde bereits in den explorativen Interviews als „Stellschraube“ vermutet und konnte in den Korrelationsanalysen als signifikanter Einflussfaktor bestätigt werden²⁵. Dies bedeutet, dass ein unzureichendes Verständnis des Evaluierungsgegenstands (welches im Rahmen der Vorbereitungsphase einer PEV gewonnen werden sollte), zu einer niedrigeren methodischen Qualität der PEV führt. Als letzter Faktor wurde schließlich analysiert, inwiefern die **Einhaltung des Dreischrittes „Beschreibung, Analyse und Bewertung“** (G 8.2) in den PEV-Berichten Einfluss auf die methodische Qualität (G insgesamt) der PEV hat. Auch dieser Zusammenhang konnte bestätigt werden. Demnach haben PEV-Berichte, die diesem grundlegenden Dreischritt folgen, eine signifikant höhere methodische Qualität.²⁶

²⁴ $r = 0,39$, $n = 50$, $p = 0,005$.

²⁵ $r = 0,439^*$, $n = 70$, $p = 0,00$. *Die korrelativen Indikatoren G 1.2, G 1.3, G.8.1 wurden aus G herausgerechnet.

²⁶ $r = 0,465^*$, $n = 70$, $p = 0,00$. *Der korrelative Indikator G.8.2 wurde aus G herausgerechnet.

Weitere Einflussfaktoren in den **Rahmendaten einer PEV** konnten nicht identifiziert werden. Die durchgeführten Korrelationsanalysen zeigen, dass es **keine systematischen Unterschiede** zwischen Vorhaben aus den unterschiedlichen Gruppierungen des FMB (Themenbereich Klima, Ländliche Entwicklung und Infrastruktur; Themenbereich Wirtschaft, Beschäftigung, Soziales; Themenbereich Governance und Konflikte) gibt. Weiterhin lässt sich hinsichtlich des Abschneidens im Genauigkeitsstandard kein signifikanter Unterschied zwischen den länderbezogenen Regionalbereichen B1, B2, und B3 feststellen. Zugleich konnten in den Korrelationsanalysen keine Anhaltspunkte dafür identifiziert werden, dass Länder mit krisenbedingt hohem Risiko für die Lieferfähigkeit der GIZ signifikant bessere oder schlechtere Werte im Genauigkeitsstandard erhalten als Länder mit geringer oder erhöhter Gefährdung. Schließlich wurden keine methodischen Qualitätsunterschiede zwischen PEV mit oder ohne Anschlussevaluierung festgestellt, wobei bei Letzteren die Fallzahlen zu niedrig sind, um allgemeingültige Aussagen zu treffen.

4.3 Bewertung und Schlussfolgerungen zur methodischen Qualität

Die Ergebnisse der Meta-Evaluierung zum Evaluierungsstandard Genauigkeit verdeutlichen, dass die methodische Qualität, im Vergleich zu den anderen Standards zur Prozessqualität und Nützlichkeit, der zentrale Schwachpunkt der PEV ist.

Hinsichtlich der Stärken fällt auf, dass die PEV bei den formalen Anforderungen an die Berichtslegung und in Bezug auf beschreibende Aspekte, wie die graphische Darstellung des Wirkungsmodells, die Beschreibung des Politik- und Entwicklungskontexts und die Nennung der Methoden, vergleichsweise gut abschneiden. Auch liefern die PEV-Berichte häufig Anhaltspunkte dafür, dass Partner in die Triangulation der Ergebnisse eingebunden und Daten trianguliert wurden. Insgesamt weisen einzelne PEV-Berichte hohe durchschnittliche Werte der methodischen Qualität auf und können als Good Beispiele verwendet werden.

Zugleich verdeutlicht die Meta-Evaluierung nach Ansicht des Meta-Evaluierungsteams tiefgehende methodische **Schwächen** der PEV, die jedoch auch vor dem Hintergrund der relativ kurzen Zeit seit der Einführung des PEV-Instruments diskutiert werden müssen. Die Schwächen sollten auch vor dem Hintergrund den mit der Umstellung von PFK auf PEV einhergehenden veränderten Anforderungen interpretiert werden. So werden an PEV höhere methodische und evaluatorische Anforderungen als noch an PFK gestellt, die sich ggfs. erst noch etablieren müssen.

Grundlegende Evaluierungsstandards, wie die Auseinandersetzung mit Wirkungshypothesen oder die Trennung zwischen Analyse und Bewertung, werden oftmals nicht eingehalten. Dies hat einen signifikanten negativen Einfluss auf die methodische Qualität. Zugleich werden diese Anforderungen jedoch in den Unterstützungsmaterialien zur Berichtslegung der PEV – der annotierten Berichtsgliederung (GIZ 2015d) und der Checklist (GIZ 2015g) – ausführlich formuliert. In diesen Fällen sind die methodischen Herausforderungen der PEV keine Folge unzureichender Vorgaben.

Auffallend ist aus Sicht des Meta-Evaluierungsteams zudem, dass PEV in methodischer Hinsicht schematisch umgesetzt werden. Dies zeigt sich vor allem darin, dass ein ähnlicher **Methodenmix** (Desk Research, Interviews) die PEV dominiert, obwohl in den PEV-Unterstützungsmaterialien der GIZ auch Arbeitshilfen zu weiteren in Evaluierungen verwendeten Methoden gesammelt werden. Gleichzeitig zeigt die Meta-Evaluierung, dass eine breite Methodenvielfalt einen signifikant positiven Einfluss auf die methodische Qualität einer PEV hat. Insgesamt können durch die geringe Vielfalt der Methoden kaum vertiefende Auswertungen erstellt werden. Es lassen sich auf der Grundlage der vier PEV, die Befragungen durchgeführt haben, keine Aussagen dazu formulieren, inwieweit eine qualitativ-quantitative Methodenkombination die Qualität von PEV beeinflusst.

Weitere wichtige Schwächen ergeben sich durch eine wenig **reflektierende Vorgehensweise** in den PEV. So befassen sich PEV oftmals nicht mit den Stärken und Schwächen der angewandten Methoden oder dokumentieren diese Reflexionen zumindest nicht. Es erfolgt häufig keine systematische

Auseinandersetzung damit, wie Gesprächspartner/innen und Dokumente ausgewählt wurden. Auch zeigen sich Optimierungsmöglichkeiten hinsichtlich der Reflexion darin, dass Monitoringdaten, vorhandene Baseline- oder Partnerdaten und die Qualität der Modulzielindikatoren zwar kritisch reflektiert, aber trotz der Einschränkungen dennoch verwendet werden. Bis auf eine Anforderung (Darlegung der Systematik der Dokumentenauswertung) werden alle dieser genannten Aspekte in einzelnen Unterstützungsmaterialien und teils in der Handreichung als Vorgaben aufgegriffen.

Darüber hinaus besteht aus Sicht des Meta-Evaluierungsteams Handlungsbedarf bezüglich der angemessenen methodischen Auseinandersetzung mit den **OECD-DAC-Kriterien**. Die methodische Bearbeitung, wie sie in der jetzigen Form in der Breite sichtbar ist, schränkt die Aussagekraft der Bewertungen der PEV-Teams aus Sicht des Meta-Evaluierungsteams stark ein. Insbesondere drei Aspekte fallen negativ auf:

- 1) Zwar wird die Qualität der Indikatoren der Vorhaben (nach SMART) zumeist von den PEV-Teams kritisiert²⁷, doch wird daraufhin nur sehr selten auch die geforderte Anpassung der Indikatoren vorgenommen. Werden jedoch bspw. Indikatoren zur Messung der Zielerreichung herangezogen, obwohl diese nicht messbar, spezifisch oder relevant sind, ist die Aussagekraft der PEV zur Effektivität zu hinterfragen. Zugleich ist diese Anpassung der Indikatoren eine in der Handreichung und den Unterstützungsmaterialien formulierte Mindestanforderung an PEV (GIZ 2015g). Zugleich ergeben sich Herausforderungen dadurch, dass das Kausalitätsproblem häufig nicht thematisiert wird.
- 2) Die methodische Bearbeitung der Effizienz-Bewertung ist stark ausbaufähig. Dies wurde bereits in anderen Untersuchungen, wie der Meta-Evaluierung Gesundheit²⁸, festgestellt. Allerdings finden sich zwei der in dieser Meta-Evaluierung verwendeten Indikatoren zur Effizienzbewertung nicht in den Vorgaben der GIZ, was die kritischen Erfüllungswerte zum Teil erklären könnte. Gleichzeitig gibt es Anhaltspunkte dafür, dass das grundlegende Verständnis hinter einer Effizienzanalyse im PEV-Team oftmals nicht gegeben ist, z. B. wenn eine scheinbar durchgeführte Produktionseffizienz genannt und dann eine Implementierungseffizienz beschrieben wird. Insgesamt wird Effizienzuntersuchungen in aktuellen (inter-)nationalen Diskussionen eine steigende Bedeutung zugeschrieben. Auch das BMZ ist zuletzt immer stärker an fundierten Effizienz- und Wirtschaftlichkeitsanalysen interessiert.
- 3) Die methodische Auseinandersetzung mit den Möglichkeiten und den Grenzen der Impact-Bewertung ist schwach. Dies schlägt sich vor allem darin nieder, dass PEV-Berichte kaum einen Bezug zu ggfs. vorhandenen EZ-Programmen und den dortigen Messgrößen herstellen. Hingegen ist das BMZ verstärkt daran interessiert, die Beiträge der einzelnen Vorhaben zum EZ-Programm zu messen. Dabei wird der Nutzen der Vorhaben zunehmend danach beurteilt werden, wie diese den EZ-Programmzielen dienen und wie sie mit den weiteren Vorhaben der technischen Zusammenarbeit (TZ) und der finanziellen Zusammenarbeit (FZ) zusammenwirken.

Schließlich zeigt die statistische Überprüfung von Zusammenhängen, dass das Instrument der PEV für die methodischen Ansätze **nicht-bilateraler Vorhaben** (Sektor- und Globalvorhaben) herausfordernd ist. PEV sind in ihrer jetzigen Form stärker auf die Anforderungen von bilateralen Vorhaben zugeschnitten. Aus Sicht des Meta-Evaluierungsteams bestehen die methodischen Herausforderungen bei Sektorvorhaben insbesondere darin, dass das spezielle Wirkungsverständnis und das Kooperationssystem von Sektorvorhaben durch die aus bilateralen Vorhaben abgeleitete Vorgehensweise schwer zu fassen sind. Zugleich steht dies im Kontrast zu der geringeren Anzahl an Gutachtertagen, die Sektorvorhaben dem PEV-Team zur Verfügung stellen. Bislang gibt es keine spezifischen Unterstützungsmaterialien für PEV in Sektor- und Globalvorhaben, um die spezifischen Anforderungen zu fassen.

²⁷ Diese Kritik findet sich auch in anderen Prüfungen wider, unter anderem in der Externen Qualitätskontrolle der GIZ, die jährlich vom BMZ in Auftrag gegeben wird.

²⁸ Schäferhoff, M., Schrade, C., Corhs, T. (SEEK Development), Raetzell, L., Krämer, M. (Rambøll Management Consulting) (2013). GIZ Review Gesundheit: Hauptbericht. Querschnittauswertung mit Metaevaluierung, Effizienzanalyse und Evaluierungssynthese.

TEIL I: ERGEBNISSE ZUR METHODISCHEN QUALITÄT (GENAUIGKEIT)

Die Sicherstellung einer ausreichenden methodischen Qualität durch die **Anzahl der Gutachter/innen und den Umfang des Mengengerüsts** ist ein weiterer Faktor, den es hervorzuheben gilt. Wird an die PEV der Anspruch einer hohen methodischen Qualität gestellt, sind beide oftmals noch zu niedrig.

Im Ergebnis gilt es zu berücksichtigen, inwieweit die Unterstützungsmaterialien der GIZ (siehe auch Kapitel 9.4 im Anhang) die Anforderungen der formulierten Kriterien und Indikatoren bereits aufgreifen. Es zeigt sich, dass die Handreichung und die Unterstützungsmaterialien den Großteil der methodischen Anforderungen, welche dieser Meta-Evaluierung zugrunde gelegt wurden, ebenfalls formulieren. Zugleich gibt es einige Aspekte dieses Bewertungssystems, welche **bislang noch nicht in den**

Unterstützungsmaterialien vorgegeben und dementsprechend in den PEV-Berichten wenig erfüllt werden. Hierzu gehören die Darstellung des Auftragswerts der Vorhaben, die Ableitung der spezifischen Ziele und Zwecke einer Evaluierung, die Darstellung der Systematik der Dokumentenauswertung und die Auseinandersetzung mit ggfs. vorhandenen Partnerdaten. Außerdem wird bislang keine strukturierte Auseinandersetzung mit der methodischen Vorgehensweise hinsichtlich der Zuordnungs- und/ oder Beitragsanalyse gefordert, welche relevant für das Design der Wirkungsmessung ist.

Zugleich hat das Fehlen oder Vorhandensein von methodischen Anforderungen nicht zwingend immer einen negativen oder positiven Einfluss auf die entsprechenden Erfüllungswerte. Einige Aspekte, wie die Überprüfung der Kernprobleme bei der Bearbeitung der Relevanz, werden oftmals erfüllt, ohne dass dies explizit vorgegeben wird. Andere Aspekte, wie die Vorgabe der Auseinandersetzung mit der Qualität der Indikatoren, werden zwar als Mindeststandards in den vorhandenen Unterstützungsmaterialien definiert, dann jedoch oftmals nicht erfüllt. Hier sind also andere Einflussfaktoren anzunehmen.

5 Teil II: Ergebnisse Prozessqualität (Durchführbarkeit und Fairness)

Die Meta-Evaluierung der PEV hinsichtlich der Prozessqualität ist wie folgt aufgebaut:

In den Kapiteln 5.1 und 5.2 werden die Ergebnisse entlang der drei Bewertungskriterien der Durchführbarkeit und der fünf Bewertungskriterien der Fairness analysiert. Zur Feststellung der PEV-Qualität in diesen Bewertungskriterien wurde jeweils auf Indikatoren zurückgegriffen. Die Indikatoren der Durchführbarkeit wurden ausschließlich durch die AV-Befragung erhoben. Die Indikatoren der Fairness basieren ebenfalls größtenteils auf der Befragung und vereinzelt auf der Textanalyse der PEV-Berichte. Die Indikatoren der Textanalyse werden durch „erfüllt“ (100 %) oder „nicht erfüllt“ (0 %) bewertet. Die Indikatoren der Online-Befragung werden mithilfe einer sechsstufigen Skala beurteilt, wodurch die Erfüllung von „trifft voll und ganz zu“ (Skala 6; 100 %) bis „trifft überhaupt nicht zu“ (Skala 1; 0 %) reicht. In dem vorliegenden Bericht wird der Erfüllungsgrad der Bewertungskriterien und dazugehörigen Indikatoren im Durchschnitt der AV-Befragung und der 70 PEV-Berichte dargestellt. Die Ergebnisse sind in der Tabelle im Anhang 9.1 im Überblick dargestellt. **Für weitere Informationen zum Bewertungssystem siehe auch Abbildung 4 in Kapitel 2.2.**

Kapitel 5.3 stellt die Einflussfaktoren auf die Prozessqualität dar, welche auf Grundlage der vorliegenden Daten identifiziert werden konnten.

Kapitel 5.4 enthält schließlich die Bewertung und Schlussfolgerung zur Prozessqualität durch das Meta-Evaluierungsteam. Dies beinhaltet eine Abwägung der identifizierten Stärken, Schwächen und Einflussfaktoren. Dazu gehört, dass diese Aspekte hinsichtlich vorhandener bzw. fehlender GIZ-Vorgaben reflektiert werden.

5.1 Beschreibung und Analyse des Evaluierungsstandards Durchführbarkeit

Der Evaluierungsstandard Durchführbarkeit wird von der DeGEVal wie folgt definiert:

„Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird“ (DeGEVal 2008).

Der Evaluierungsstandard Durchführbarkeit wird anhand von drei Bewertungskriterien gebildet. Die folgende Abbildung fasst diese Bewertungskriterien übersichtlich zusammen und zeigt deren durchschnittliche Erfüllung hinsichtlich der PEV in Prozent auf:

Abbildung 16: Erfüllungsgrad der Bewertungskriterien im Evaluierungsstandard Durchführbarkeit

Evaluierungsstandard Durchführbarkeit (D)	63 %
D 1 Angemessene Verfahren	65 %
D 2 Diplomatisches Vorgehen	61 %
D 3 Effizienz von Evaluation	61 %

Quelle: Syspons 2016 (Online-Befragung der AV)

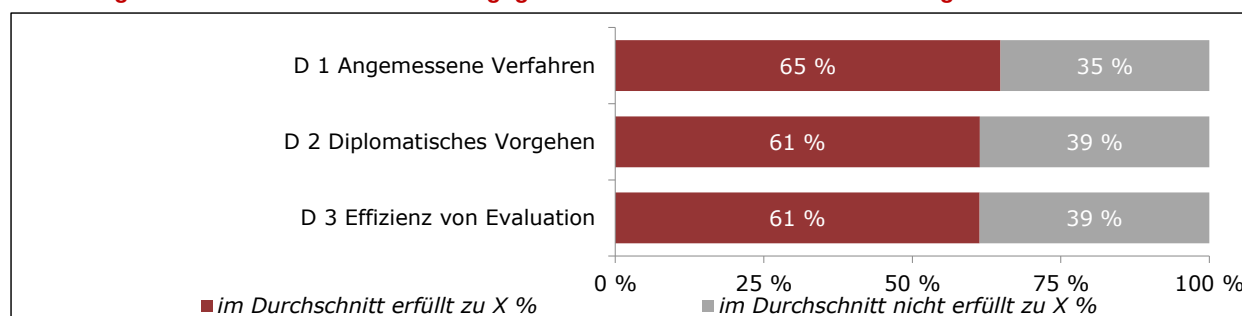
Im Durchschnitt wird der Evaluierungsstandard Durchführbarkeit laut der Online-Befragung eher²⁹ erfüllt (63 %), was nach dem Evaluierungsstandard Genauigkeit (59 %) der geringste Wert ist.³⁰

Die drei Bewertungskriterien des Evaluierungsstandards befassen sich mit der Frage, inwieweit Evaluierungen realistisch, diplomatisch und kostenbewusst geplant und umgesetzt wurden:

²⁹ Um die Bewertungen auf Ebene der Evaluierungsstandards und Bewertungskriterien einheitlich und nachvollziehbar zu gestalten, werden diese wie folgt angegeben: 90–100 % (größtenteils bis vollständig erfüllt); 80–89 % (zumeist erfüllt); 60–79 % (eher erfüllt); < 60 % (bedingt erfüllt). Siehe hierzu auch Kapitel 2.1.

³⁰ Evaluierungsstandards Fairness (76 %) und Nützlichkeit (74 %).

Abbildung 17: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien D 1–D 3



Quelle: Syspons 2016 (Online-Befragung der AV)

D 1 Angemessene Verfahren: Als angemessen gilt ein Evaluierungsverfahren, wenn die Belastungen des Evaluierungsgegenstandes bzw. der Beteiligten und Betroffenen in einem angemessenen Verhältnis zum Nutzen einer Evaluierung stehen. Dies operationalisiert das erste Bewertungskriterium durch drei Indikatoren. Der erste Indikator misst, ob der Zeiteinsatz der Mitarbeiter/innen eines Vorhabens für die Evaluierung im Verhältnis zum Nutzen angemessen war (D 1.1). Der zweite Indikator misst dies für die Partner (D 1.2). Da die Mechanismen der Qualitätskontrolle einer PEV ein wesentlicher Bestandteil von Evaluierungsverfahren sind, befasst sich der dritte Indikator mit der Angemessenheit der Qualitätssicherung (QS) der PEV (D 1.3). Dieser Indikator wurde in der Online-Befragung der AV nur qualitativ über ein Freitextfeld erhoben und wird am Ende dieses Kapitels ausgewertet.

Die beiden ersten Indikatoren erreichen einen mittleren Erfüllungsgrad, wodurch das Bewertungskriterium D 1 mit durchschnittlich 65 % eher erfüllt wird (siehe Abbildung 17). Ein detaillierter Blick verdeutlicht, wie sich dieser Wert zusammensetzt. Demnach bewerten die AV die Angemessenheit des Zeiteinsatzes der Mitarbeiter/innen des Vorhabens im Verhältnis zum Nutzen der PEV mit durchschnittlich 58 % (D 1.1). Die Verteilung der Antworten auf der Befragungsskala variiert hierbei stark und deckt alle Bewertungsmöglichkeiten von „trifft überhaupt nicht zu“ (Skala 1) bis „trifft voll und ganz zu“ (Skala 6) ab. Dies weist auf sehr unterschiedliche Erfahrungen in der Durchführung der PEV hin. Im Vergleich hierzu fällt der Erfüllungsgrad des zweiten Indikators, welcher die Angemessenheit des Zeiteinsatzes der Partner im Vergleich zum Nutzen bewertet, mit durchschnittlich 73 % deutlich positiver aus (D 1.2). Der Vergleich dieser Werte deutet darauf hin, dass die Mitarbeiter/innen eines Vorhabens stärker als die Partner in die Prozessen eingebunden sind oder die Aufwände der Partner weniger im Fokus stehen.

D 2 Diplomatisches Vorgehen: Dieses Bewertungskriterium misst, ob ein diplomatisches Vorgehen in der PEV-Durchführung eingehalten wurde. Diesem Kriterium nach sollen Evaluierungen die Interessen der beteiligten und betroffenen Personen berücksichtigen, um Akzeptanz, Zustimmung und Kooperation in der Durchführung zu erleichtern. Der einzige Indikator dieses Kriteriums erfasst anhand der Befragung der AV, inwieweit der Dialog zwischen den verschiedenen Stakeholdern gefördert wurde (D 2.1 = D 2).

Ausgehend von diesem Indikator ist das Bewertungskriterium D 2 eher erfüllt (61 %) (siehe Abbildung 17). Dieses Ergebnis deutet darauf hin, dass PEV-Teams zwar bereits mehrheitlich, aber noch nicht flächendeckend einen kooperativen Ansatz verfolgen und dabei unterschiedliche Interessensgruppen zusammenbringen. Wenn eine PEV kein diplomatisches Vorgehen wählt, können PEV-Teams Versuche verschiedener Interessensgruppen, die Ergebnisse zu beeinflussen, nicht vermeiden oder antizipieren.

D 3 Effizienz von Evaluation: Das dritte Bewertungskriterium befasst sich mit der Angemessenheit des Aufwands (Budget, Mengengerüst) einer Evaluierung im Verhältnis zum Nutzen der Evaluierung. Demnach wurden zwei Indikatoren gebildet. Diese messen, inwieweit das Budget (D 3.1) und das Mengengerüst (D 3.2) einer Evaluierung als angemessen empfunden werden.³¹

Insgesamt wird das Kriterium D 3 mit durchschnittlich 61 % eher erfüllt (siehe Abbildung 17). Hierbei ist auffällig, dass sich die befragten AV dem Indikator D 3.1 durchschnittlich mit 58 % zustimmten.

TEIL II: ERGEBNISSE ZUR PROZESSQUALITÄT (DURCHFÜHRBARKEIT UND FAIRNESS)

Gemeinsam mit dem Indikator D 1.1 ist dies der niedrigste Wert eines Indikators in dem Evaluierungsstandard Genauigkeit. Zugleich zeigt sich im zweiten Indikator des Kriteriums (D 3.2), dass die Angemessenheit des Mengengerüsts zwar etwas höher und mehrheitlich eher positiv bewertet (durchschnittliche Erfüllung 64 %) wird, jedoch immer noch seitens der AV mit einer großen Varianz eingeschätzt wird.

Im Anschluss an die Fragen zur Durchführbarkeit hatten die AV im Fragebogen die Möglichkeit, **zu den genannten Aspekten** (Miteinsatz, Mengengerüst, finanzielles Budget) **Stellung in einem Freitextfeld zu nehmen**. Bei der Interpretation der Rückmeldungen ist wichtig zu berücksichtigen, dass durch die Integration einer PEV und der Prüfung eines Folgevorhabens zwei Aspekte (Evaluierung und Konzeption) in einem integrierten Modell vereinigt werden. Beide Aspekte sind, darauf deuten auch die Rückmeldungen der AV hin, schwer voneinander zu trennen. Diese Möglichkeit nutzten 31 der teilnehmenden AVs.

Hierbei kamen nur wenige Rückmeldungen, die das **Bewertungskriterium D 1** (Angemessene Verfahren) betreffen. Hierbei verwiesen kritische Stimmen der AV darauf, dass sich die Aufwände für die Vorhaben durch die Einführung des Instruments der PEV erhöht haben. Insbesondere die Vorbereitung einer PEV wird als zeitintensiv, gleichzeitig jedoch als zentral für das Gelingen einer PEV, beschrieben. Ebenfalls nur vereinzelt befassten sich AV in den Rückmeldungen des Freitextfelds mit dem **Bewertungskriterium D 2** (Diplomatisches Vorgehen). Diese Rückmeldungen reflektierten unterschiedliche Erfahrungen. Einerseits wurde von positiven Erfahrungen berichtet, die das frühzeitige Einbinden der Zielgruppen und gesellschaftlich relevanter Organisationen als Gelingensfaktor für die Akzeptanz der Ergebnisse hervorhoben. Andererseits wurde von einem fehlenden Austausch über die PEV-Ergebnisse mit dem BMZ sowie einer fehlenden Berücksichtigung der Partnerperspektiven in den Prüfmmissionen berichtet.

Die meisten der **qualitativen Rückmeldungen der AV** bezogen sich auf das **Bewertungskriterium D 3** (Effizienz von Evaluation), was darauf hindeutet, dass dieses Thema kontrovers diskutiert wird. Es wurde zunächst der Eindruck geteilt, dass die für die Evaluierungen reservierten Budgets den mit einer PEV angestrebten Nutzen nicht zwingend aufwiegen. Sowohl AV kleinvolumiger als auch großvolumiger Vorhaben wiesen auf den aus den knappen Budgets resultierenden Zeitdruck hin, der zulasten der Qualität der PEV geht. Darüber hinaus stand oftmals im Mittelpunkt der Rückmeldungen zum Mengengerüst einer PEV, dass Mengengerüste zu knapp kalkuliert sind, die Vorhaben jedoch kein höheres Budget bereitstellen könnten. Zugleich deuten die Rückmeldungen auf eine Herausforderung von PEV mit Folgemaßnahme hin. So wurde die parallele Durchführung der PEV mit der Konzeption der Folgemaßnahme teils als überfordernd empfunden, da zu wenig Zeit für fachliche Auseinandersetzungen einerseits und die Vorausschau andererseits bleibt.

Abschließend erhob die Befragung qualitative **Einschätzungen der AV zur Angemessenheit der Qualitätssicherung (QS) der PEV**, welche der **Indikator D 3.1** aufgreift. Diese Einschätzungen der AV, werden nachfolgend beschrieben und analysiert.

Die Qualitätssicherung umfasst sowohl die seitens der STS 08 bereitgestellte Handreichungen und die Unterstützungsmaterialien³² (siehe Kapitel 3.1) als auch die QS-Checks der PEV-Berichte durch die STS 08. Vielfach wurden durch die AV die Verständlichkeit und die Qualität der zur Verfügung gestellten Handreichung und Unterstützungsmaterialien thematisiert, wobei sich in ihren Bewertungen ein diverses Bild abzeichnete. Von einigen AV wurden die Hilfsmaterialien als ausreichende Orientierungen angesehen und ihre methodische Breite (z. B. Toolbox) gelobt, die einheitliche Standards gewährleistet. Andere AV empfanden die Hilfsmaterialien als zu komplex und nicht systematisch und ausreichend

³¹ Die Angemessenheit des Zeitaufwands im Verhältnis zum Nutzen wurde bereits an anderer Stelle (Indikator D 3.1) erhoben.

³² In der entsprechenden Frage der Online-Befragung wurden die AV gebeten, mitzuteilen, inwieweit diese mit dem Informationsmaterial, dem Kooperationsprozess und weiteren Aspekten der Qualitätssicherung (u.a. Handreichung, annotierte Berichtsgliederung, PEV-Toolbox, Qualitätscheck der PEV Berichte) zufrieden waren. In der Frage wurde damit nicht differenziert zwischen der Handreichung und einzelnen Unterstützungsmaterialien. Auch die Antworten der AV differenzieren hierzwischen nicht.

genug nutzbar in der Vorbereitung einer PEV. In diesem Kontext kritisierten die AV die mangelnde Konsistenz der Vorgaben bei der Einführung des PEV-Formats, insbesondere die Veränderungen während der Einführung. Zugleich wurde die Flexibilität, die Handreichung und Unterstützungsmaterialien den PEV lassen, um den Kontextbedingungen eines Vorhabens gerecht zu werden, unterschiedlich bewertet. Insgesamt wurde wiederholt der Umfang der Hilfsmaterialien aufgegriffen. AV verwiesen hierbei darauf, dass die Hilfsmaterialien, auch wenn inhaltlich gut, zu umfangreich sind, um als Orientierung für eine PEV dienen zu können. Dies wurde insbesondere als Herausforderung für neue Gutachter/innen angesehen. In diesem Kontext wurde vielfach der Wunsch nach kurzen Übersichten und kompakterem Informationsmaterial geäußert.

Neben den Hilfsmaterialien thematisierten die AV die Qualitätschecks, welche durch die STS 08 durchgeführt werden. Hierbei zeichnet sich ein breites Bild ab, wobei die kritischen Stimmen überwiegen. Einerseits beschreibt eine Gruppe der AV die Qualitätschecks als unterstützend. Andererseits zeigen die Rückmeldungen der AV, dass ein methodisches Feedback zum Bericht durch die STS 08 für die AV nicht immer Priorität hat und die Ergebnisse der Evaluierungen dabei unabhängig von methodischen Schwächen interpretiert werden. Einige der AV thematisieren die zeitliche Relevanz der Qualitätschecks. Hierbei wurden die Rückmeldungen vereinzelt als nicht zeitnah genug beschrieben, und Überarbeitungsschleifen gestalteten sich teils als langwierig. Zudem wurde seitens einiger AVs von nicht-bilateralen Vorhaben die Qualitätssicherung als nicht ausreichend auf die Bedarfe nicht-bilateraler Vorhaben angepasst empfunden.

5.2 Beschreibung und Analyse des Evaluierungsstandards Fairness

Die DeGEval legt dem Evaluierungsstandard Fairness folgende Definition zugrunde:

„Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird“ (DeGEval 2008).

Hierzu formuliert die DeGEval fünf Bewertungskriterien, welche die Fairness einer Evaluierung operationalisieren sollen. Die folgende Abbildung fasst diese Bewertungskriterien übersichtlich zusammen und zeigt deren durchschnittliche Erfüllung hinsichtlich der PEV in Prozent auf:

Abbildung 18: Erfüllungsgrad der Bewertungskriterien im Evaluierungsstandard Fairness

Evaluierungsstandard Fairness (F)	76 %
F 1 Formale Vereinbarungen	73 %
F 2 Schutz individueller Rechte	82 %
F 3 Vollständige und faire Überprüfung	69 %
F 4 Unparteiische Durchführung und Berichterstattung	88 %
F 5 Offenlegung der Ergebnisse	69 %

Quelle: Syspons 2016 (Textanalyse der PEV-Berichte und Online-Befragung der AV)

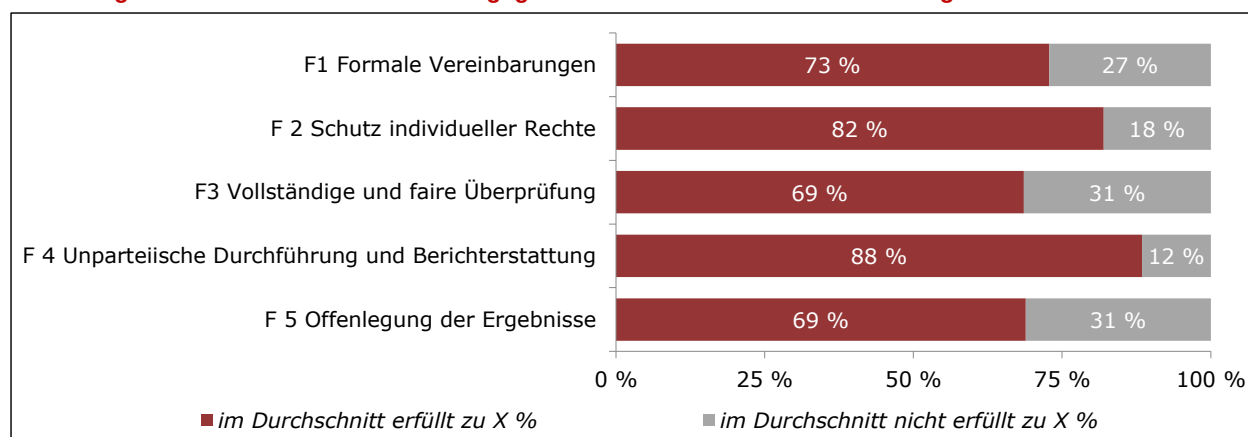
Im Durchschnitt wird der Evaluierungsstandard Fairness laut der Online-Befragung und der Textanalyse eher³³ erfüllt (76 %). Im Vergleich der vier Evaluierungsstandards ist dies der höchste Wert.³⁴

Die fünf Bewertungskriterien des Evaluierungsstandards befassen sich mit der Frage, inwieweit in Evaluierungen respektvoll und fair mit Beteiligten und Betroffenen umgegangen wird:

³³ Um die Bewertungen auf Ebene der Evaluierungsstandards und Bewertungskriterien einheitlich und nachvollziehbar zu gestalten, werden diese wie folgt angegeben: 90–100 % (größtenteils bis vollständig erfüllt); 80–89 % (zumeist erfüllt); 60–79 % (eher erfüllt); < 60 % (bedingt erfüllt). Siehe hierzu auch Kapitel 2.1.

³⁴ Evaluierungsstandards Genauigkeit (59 %), Durchführung (63 %) und Nützlichkeit (74 %).

Abbildung 19: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien F 1–F 5



Quelle: Syspons 2016 (Textanalyse der PEV-Berichte und Online-Befragung der AV)

F 1 Formale Vereinbarungen: Das erste Bewertungskriterium erfasst, ob die Vertragsparteien einer Evaluierung die Vereinbarungen formal festhalten. Dieses Kriterium wird durch einen Indikator gemessen, der erfasst, inwieweit die angehängten Leistungsbeschreibungen der PEV-Berichte die vertraglichen Rahmenbedingungen beschreiben (F 1.1 = F 1).

Im Mittel der PEV-Berichte wird das Bewertungskriterium F 1 eher erfüllt (73 %) (siehe Abbildung 19). Im Detail bedeutet dies, dass die TOR mehrheitlich die Erwartungen, Verantwortlichkeiten, zu erbringenden Leistungen sowie die personellen und zeitlichen Rahmenbedingungen hinreichend detailliert festhalten. Auch wenn die TOR keine formalisierten Ausschreibungen darstellen, da in der Regel bereits feststeht, welche Evaluator/innen beauftragt werden, sollten sie stets die wichtigsten Punkte (was, wie, von wem und wann evaluiert werden soll) beinhalten. Hierbei ist anzumerken, dass die TOR der evaluierten Vorhaben insgesamt wenig individuell auf die zu evaluierenden Vorhaben zugeschnitten sind und sich eng an den Muster-TOR orientieren, die weitgehend übernommen wurden. Damit geht einher, dass anhand der TOR nicht immer deutlich wird, inwieweit es sich bei den Evaluierungen um eine Evaluierung mit oder ohne Folgemaßnahme handelt, da die entsprechenden Textpassagen nicht angepasst werden.

F 2 Schutz individueller Rechte: Zur Fairness gehört es auch, dass Evaluierungen so geplant und durchgeführt werden, dass sie die Sicherheit und Rechte der einbezogenen Akteure und Organisationen schützen. Dieses Bewertungskriterium (F 2) wurde anhand eines Indikators operationalisiert. Der Indikator prüft durch die AV-Befragung, ob der Evaluierungsprozess aus AV-Sicht ethischen Verhaltensregeln folgt, indem PEV-Teams sowohl kultur- als auch gendersensibel vorgehen (F 2.1 = F 2).

Mit durchschnittlich 82 % wird das Bewertungskriterium F 2 zumeist erfüllt (siehe Abbildung 19). Demnach attestieren die AV den PEV-Teams mehrheitlich, ethischen Verhaltensregeln zu folgen. So erhält die Aussage, dass PEV-Teams kultursensibel vorgehen, eine durchschnittliche Zustimmung von 83 %, während das gendersensible Vorgehen der PEV-Teams mit durchschnittlich 82 % bestätigt wird. Nur in vier Ausnahmen haben die AV das Vorgehen der PEV-Teams als wenig sensibel für die Kultur und Religion des Partnerlandes eingeschätzt, wodurch die entsprechenden Evaluierungen möglicherweise einen negativen Einfluss auf den untersuchten Kontext genommen haben.

F 3 Vollständige und faire Überprüfung: Das dritte Bewertungskriterium erhebt, ob PEV die Stärken und die Schwächen des Evaluierungsgegenstandes möglichst vollständig und fair überprüfen und darstellen. Das Kriterium wird mithilfe eines Indikators bewertet, welcher die Ausgewogenheit und Vollständigkeit der Darstellung der Stärken und Schwächen in den PEV-Berichten erfasst (F 3.1 = F 3).

Insgesamt werden in rund 69 % der Berichte (n=48) die Stärken und Schwächen der evaluierten Vorhaben herausgearbeitet. Somit wird das Bewertungskriterium F 3 eher erfüllt (siehe Abbildung 19). Diese Bewertung setzt nicht voraus, dass im Sinne einer ausgewogenen Darstellung der Stärken und

Schwächen diese immer in gleicher Anzahl genannt werden. Vielmehr wird vorausgesetzt, dass sowohl positive als auch negative Aspekte eines Vorhabens herausgearbeitet und bewertet werden.

F 4 Unparteiische Durchführung und Berichterstattung: Das vierte Bewertungskriterium befasst sich damit, ob die Evaluierungsprozesse die unparteiische Position des PEV-Teams erkennen lassen und ob die entsprechenden Bewertungen fair sind. Der dazugehörige Indikator untersucht anhand der Online-Befragung, ob die AV das PEV-Team als unparteiisch einschätzen (F 4.1 = F 4).

Im Mittel der Einschätzungen der AV wird dieses Bewertungskriterium mit 88 % zumeist erfüllt (siehe Abbildung 19). Im Detail spricht sich damit die große Mehrheit der AV dafür aus, dass die PEV-Teams unparteiisch vorgegangen sind und es keine Interessenskonflikte in Bezug auf die evaluierten Vorhaben gab. Dies erhöht die Glaubwürdigkeit der Evaluierungsergebnisse. Insgesamt zeigen die Befragungsergebnisse, dass die AV die Unabhängigkeit auch bei der Teilnahme interner Evaluierungsmitglieder nicht als eingeschränkt empfinden. So gibt es keine systematischen Unterschiede in der Bewertung der Unabhängigkeit, egal ob PEV-Teams aus internen sowie externen oder ausschließlich internen oder ausschließlich externen Evaluators/innen bestanden.

F 5 Offenlegung der Ergebnisse: An PEV wird der Anspruch gestellt, dass die Evaluierungsergebnisse den beteiligten und betroffenen Personen zugänglich gemacht werden. Hiermit befasst sich das fünfte Bewertungskriterium. Für eine Offenlegung müssen die Evaluierungsergebnisse zielgruppengerecht aufbereitet (Indikator F 5.1) und mit beteiligten und betroffenen Akteuren geteilt werden (Indikator F 5.2).

Im Mittel erfüllen 69 % der dem Bewertungskriterium F 5 zugeordneten Indikatoren diese Ansprüche. Das Kriterium ist demnach eher erfüllt (siehe Abbildung 19). Eine Voraussetzung für den Zugang zu Evaluierungsergebnissen ist, dass für die Kommunikation der Ergebnisse entsprechende Formate entwickelt wurden, um die beteiligten und betroffenen Stakeholder einer Evaluierung zu erreichen. Dies thematisiert der erste hinter dem Bewertungskriterium liegende Indikator (F 5.1). Hierbei gilt die Prämisse, dass je nach Nutzergruppe unterschiedliche Verfahren der Informationsverbreitung gewählt werden sollten. Durchschnittlich wird diese zielgruppengerechte Aufbereitung der Ergebnisse durch die AV mit 63 % bewertet. Welche Formate genutzt werden, hat die Befragung nicht erhoben. Darüber hinaus wurden die AV gebeten, zu beurteilen, ob die zentralen Ergebnisse der PEV mit den Beteiligten und Betroffenen geteilt wurden. Dieser Indikator (F 5.2) weist mit einer mittleren Zustimmung von 75 % eine etwas positivere Einschätzung auf. Ein detaillierter Blick auf die zugrunde liegenden Befragungselemente verschafft ein differenzierteres Bild. Demnach werden die Ergebnisse fast durchgehend mit den Partnern geteilt. Sowohl die Einbindung von politischen Partnern als auch von Durchführungspartnern erhält eine hohe Bestätigung (durchschnittliche Erfüllung 89 % und 88 %). Seltener erfolgt ein Austausch mit weiteren Kooperationspartnern (76 %) und Zielgruppen (47 %).

5.3 Einflussfaktoren auf die Prozessqualität

Es wurden keine statistischen Hinweise dafür gefunden, dass andere Evaluierungsstandards oder die Rahmenbedingungen einer PEV einen signifikanten Einfluss auf den Evaluierungsstandard **Durchführbarkeit** und die entsprechenden Bewertungskriterien haben.

Der Evaluierungsstandard **Fairness** und seine Bewertungskriterien wurden nicht als zu erklärende Variablen, sondern nur als Einflussfaktoren auf andere Evaluierungsstandards operationalisiert.

Darüber hinaus konnten punktuell Einflüsse auf die **Fairness** durch die Rahmenbedingungen einer PEV identifiziert werden. Demnach lässt sich ein signifikanter Zusammenhang³⁵ zwischen der Erfüllung des Standards Fairness und der Anzahl der Gutachtertage feststellen. Die Analyse der Mittelwerte weist hier auf Unterschiede zwischen PEV mit weniger als 55 Gutachtertage (n=12) einerseits und PEV mit 55 bis 80 Gutachtertage (n=14) und mehr als 80 Gutachtertage (n=12) andererseits hin. Demnach schneidet

³⁵ $r = 0,34$, $n = 38$, $p = 0,035$.

die erste Gruppe im Evaluierungsstandard Fairness durchschnittlich um 16 % schlechter ab als die anderen beiden Gruppen.

Folglich haben PEV mit ausreichend Gutachtertage einen signifikant positiven Einfluss darauf, dass die Interessen und Bedürfnisse der betroffenen Personen und Gruppen berücksichtigt und die Evaluierungen entsprechend ausgerichtet werden. Ein weiterer signifikanter Faktor auf den Erfüllungsgrad der Fairness sind ausreichende Gutachtertage in der Vorbereitung und Nachbereitung³⁶ einer Evaluierung. Als Erklärung hierfür kann herangezogen werden, dass in der Vorbereitung einer PEV die spätere Einbindung relevanter Stakeholder angelegt und in der Nachbereitung abgeschlossen wird. Insgesamt kann diese Meta-Evaluierung nur Einflussfaktoren untersuchen, die mit der vorhandenen Datenbasis untersucht werden können, wobei auch andere Faktoren außerhalb des Untersuchungsbereichs einen Einfluss nehmen können. So kann z. B. die Einbeziehung von Stakeholdern nicht nur durch das PEV-Team, sondern auch durch die zeitliche Unterstützung der Mitarbeiter/innen eines Vorhabens gefördert werden.

5.4 Bewertung und Schlussfolgerungen zur Prozessqualität

Nach Ansicht des Meta-Evaluierungsteams fällt die Prozessqualität der PEV divers aus. Während Aspekte der Verfahren der Evaluierung (Evaluierungsstandard Durchführbarkeit), ausgehend von dem gesamten Erfüllungsgrad des Bewertungskriteriums, Schwächen aufweisen, ist der faire Umgang der PEV-Teams mit beteiligten und betroffenen Stakeholder (Evaluierungsstandard Fairness) positiver zu bewerten. Beides soll nachfolgend getrennt voneinander beurteilt werden.

Die Meta-Evaluierung zeigt, dass die Evaluierungsverfahren und die Effizienz der Evaluierungen (**Durchführbarkeit**) seitens der AV kontrovers wahrgenommen werden. Der Zeiteinsatz, den politische Träger und Durchführungsorganisationen aufbringen, erweist sich im Verhältnis zum Nutzen der PEV noch am ehesten als angemessen und als **Stärke** des Standards Durchführbarkeit. Hinsichtlich der Unterstützungsmaterialien, welche PEV-Teams zur Verfügung gestellt werden, weisen die Ergebnisse der Meta-Evaluierung darauf hin, dass trotz bestehender Schwächen hinsichtlich des Umfangs, der Flexibilität und Komplexität, die Inhalte der Hilfsmaterialien zumeist als hilfreich hervorzuheben sind.

Insgesamt kommt das Meta-Evaluierungsteam zu dem Schluss, dass die Kriterien im Evaluierungsstandard zwar mehrheitlich seitens der AV eher positiv eingeschätzt, jedoch aber wegen der insgesamt kontroversen Einschätzungen und des Erfüllungsgrads insgesamt auch kritisch reflektiert werden sollten. Die Angemessenheit der Verfahren für die Mitarbeiter/innen scheint hierbei eine zentrale **Schwäche**. In der Wahrnehmung rund jedes dritten AV wiegt bei der Durchführung einer PEV der Nutzen nicht die damit verbundenen Aufwände und Kosten auf. Aus Sicht des Meta-Evaluierungsteams schwächen diese Rückmeldungen die Akzeptanz des Instruments der PEV bei den AV. Insbesondere die zu erbringenden Leistungen einer PEV mit Folgevorhaben lassen sich aus AV-Sicht nicht zwingend mit dem intendierten Nutzen rechtfertigen. Im Ergebnis ist aus Sicht des Meta-Evaluierungsteams auffallend, dass es bislang keine institutionalisierten Richtwerte in der GIZ gibt, wie viele Gutachtertage oder welches Budget AV durchschnittlich für Evaluierungen veranschlagen sollten.

Aus den Analysen zur Durchführbarkeit lässt sich schlussfolgern, dass die Unterstützungsmaterialien und -leistungen der PEV noch nicht genug wirken. Als Schwäche konnte identifiziert werden, dass diese oftmals als zu umfassend (insbesondere für neue Gutachter/innen) oder in der Qualitätssicherung als zu aufwendig wahrgenommen werden. Zugleich scheinen die Unterstützungsmaterialien und -leistungen weniger passgenau auf nicht-bilaterale Vorhaben zugeschnitten zu sein, wobei diese Aussage nur auf den qualitativen Rückmeldungen der AV beruht.

³⁶ Anzahl der Gutachtertage für die Vorbereitung: $r = 0,36$, $n = 38$, $p = 0,029$. Anzahl der Gutachtertage für die Nachbereitung: $r = 0,35$, $n = 38$, $p = 0,029$.

TEIL II: ERGEBNISSE ZUR PROZESSQUALITÄT (DURCHFÜHRBARKEIT UND FAIRNESS)

Möglichst valide und reliable Ergebnisse, wie sie durch eine methodisch fundierte Evaluierung angestrebt werden, ziehen entsprechende Kosten nach sich. Dieser Annahme folgend gibt die Meta-Evaluierung Anhaltspunkte dafür, dass die Standards Genauigkeit (methodische Qualität) und Durchführbarkeit (Prozessqualität), deren Erfüllung in der Breite gering ist, in einem Spannungsverhältnis zueinander stehen könnten. Folglich steht die Diskussion im Fokus, wie zukünftig die methodische Qualität erhöht werden kann, ohne die negativen Wahrnehmungen zu Aufwänden und Kosten zu verschärfen.

Darüber hinaus verdeutlicht die Meta-Evaluierung, dass der Umgang der PEV-Teams mit beteiligten und betroffenen Stakeholdern (**Fairness**) zumeist fair erfolgt. So erreicht der Standard Fairness im Vergleich zu den weiteren Standards eine hohe Erfüllungsrate und weist damit auf eine zufriedenstellende Qualität hin. Die **Stärken** der PEV zeigen sich hierbei in der ethisch korrekten Durchführung. Diese umfasst eine kultur- und gendersensible Vorgehensweise sowie eine unparteiische Umsetzung. Damit weisen die Ergebnisse darauf hin, dass die PEV zumeist den Grundsätzen von Integrität, Menschenrechten und Unparteilichkeit entspricht. Darüber hinaus bestehen Anhaltspunkte dafür, dass eine ausreichende Anzahl an Gutachtertagen einen signifikant positiven Einfluss auf eine faire Evaluierung hat.

Gleichzeitig verdeutlicht die Meta-Evaluierung auch **Herausforderungen** hinsichtlich der Fairness. Dies kann entlang des Evaluierungsprozesses verdeutlicht werden. So werden die Vereinbarungen der Evaluierung selten ausreichend spezifisch in den TOR festgehalten, sodass den Evaluator/innen die geforderten Leistungen und Zeitpläne nicht immer verständlich gemacht werden. Optimierungspotenzial besteht hinsichtlich der zielgruppenspezifischen Aufbereitung der Evaluierungsergebnisse. Das Meta-Evaluierungsteam weist auf die hohe Bedeutung hin, die einer je nach Adressatengruppe angepassten Kommunikation der Ergebnisse zukommt – dies wird bislang noch nicht ausreichend in Wert gesetzt. Abschließend ist insgesamt festzuhalten, dass sich die insbesondere die Unterstützungsmaterialien aus Sicht des Meta-Evaluierungsteams ausreichend mit Vorgaben oder Orientierungspunkten einer fairen Evaluierung befassen. Sowohl die M&E-Policy der GIZ (GIZ 2013) als auch die Methodentoolbox (GIZ 2015h) und punktuell die Handreichung zum methodischen Vorgehen (GIZ 2015a) greifen die diskutierten Aspekte auf.

6 Teil III: Ergebnisse zur Nützlichkeit

Die Meta-Evaluierung der PEV hinsichtlich des Evaluierungsstandards Nützlichkeit ist wie folgt aufgebaut:

Im Kapitel 6.1 werden die Ergebnisse entlang der acht Bewertungskriterien dargestellt, aus denen sich der Evaluierungsstandard zusammensetzt. Zur Feststellung der PEV-Qualität in den acht Bewertungskriterien wurde auf Indikatoren zurückgegriffen. Die Indikatoren der Nützlichkeit wurden größtenteils durch die AV-Befragung erhoben. Nur vereinzelt gründet die Indikatorenbewertung auf die Textanalyse der PEV-Berichte. Die Indikatoren der Textanalyse werden durch „erfüllt“ (100 %) oder „nicht erfüllt“ (0 %) bewertet. Die Indikatoren der Online-Befragung werden mithilfe einer sechsstufigen Skala beurteilt, wodurch die Erfüllung von „trifft voll und ganz zu“ (Skala 6; 100 %) bis „trifft überhaupt nicht zu“ (Skala 1; 0 %) reicht. In dem vorliegenden Bericht wird der Erfüllungsgrad der Bewertungskriterien und dazugehörigen Indikatoren im Durchschnitt der AV-Befragung und der 70 PEV-Berichte dargestellt. Die Ergebnisse sind in der Tabelle im Anhang 9.1 im Überblick dargestellt. **Für weitere Informationen zum Bewertungssystem siehe auch Abbildung 4 in Kapitel 2.2.**

Kapitel 6.2 stellt die Einflussfaktoren auf die Qualität der Nützlichkeit dar, welche auf Grundlage der vorliegenden Daten identifiziert werden konnten.

Kapitel 6.3 enthält schließlich die Bewertung und Schlussfolgerung zur Qualität der Nützlichkeit durch das Meta-Evaluierungsteam. Dies beinhaltet eine Abwägung der identifizierten Stärken, Schwächen und Einflussfaktoren. Dazu gehört, dass diese Aspekte hinsichtlich vorhandener bzw. fehlender GIZ-Vorgaben reflektiert werden.

6.1 Beschreibung und Analyse des Evaluierungsstandards Nützlichkeit

Hinsichtlich des Evaluierungsstandards Nützlichkeit definiert die DeGEval:

„Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an den geklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet“ (DeGEval 2008).

Dafür formuliert die DeGEval acht Bewertungskriterien, deren Einhaltung die Qualität der Nützlichkeit und des Nutzens sicherstellen. Die folgende Abbildung fasst diese Bewertungskriterien übersichtlich zusammen und zeigt deren durchschnittliche Erfüllung hinsichtlich der PEV in Prozent auf:

Abbildung 20: Erfüllungsgrad der Bewertungskriterien im Evaluierungsstandard Nützlichkeit

Evaluierungsstandard Nützlichkeit (N)	74%
N 1 Identifizierung der Beteiligten und Betroffenen	83 %
N 2 Klärung der Evaluationszwecke	83 %
N 3 Glaubwürdigkeit und Kompetenz des Evaluators / der Evaluatorin	57 %
N 4 Auswahl und Umfang der Informationen	84 %
N 5 Transparenz von Werten	71 %
N 6 Vollständigkeit und Klarheit der Berichterstattung	86 %
N 7 Rechtzeitigkeit der Evaluation	81 %
N 8 (A) Nutzung & Nutzen der Evaluation: Nützlichkeit des Evaluierungsprozesses und -berichts	73 %
N 8 (B) Nutzung & Nutzen der Evaluation: Nutzung der Evaluierungsergebnisse	65 %

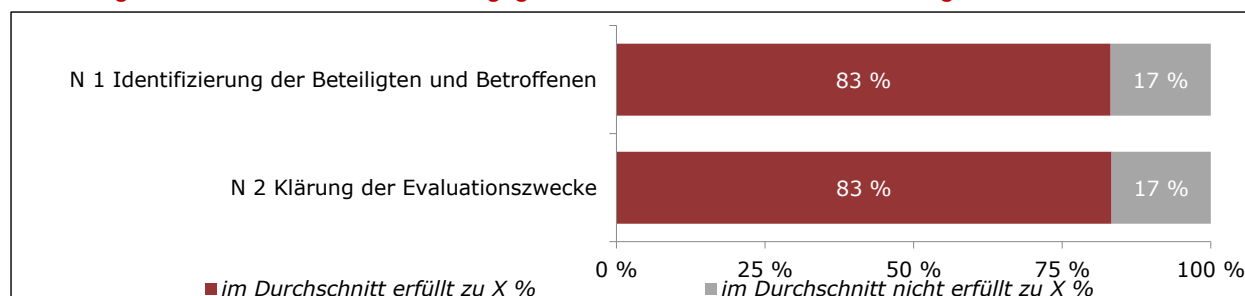
Quelle: Syspons 2016 (Textanalyse der PEV-Berichte und Online-Befragung)

TEIL III: ERGEBNISSE ZUR NÜTZLICHKEIT

Im Durchschnitt wird der Evaluierungsstandard eher³⁷ erfüllt (74 %). Damit erzielt die Nützlichkeit nach dem Evaluierungsstandard Fairness (Erfüllungsgrad 76 %) die höchste Qualität.³⁸

Die ersten zwei Bewertungskriterien befassen sich mit der Vorbereitung der PEV:

Abbildung 21: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien N 1–N 2



Quelle: Syspons 2016 (Online-Befragung der AV)

N 1 Identifizierung der Beteiligten und Betroffenen: Das erste Bewertungskriterium der Nützlichkeit betrachtet, inwieweit der Kreis der beteiligten und betroffenen Stakeholder einer PEV identifiziert wurde. Dies soll gewährleisten, dass bereits in der Vorbereitung einer PEV die Nützlichkeit gefördert werden kann. Gemessen wurde das Bewertungskriterium auf der Grundlage von zwei Indikatoren. Der erste Indikator erfasst, inwieweit die relevanten Stakeholder systematisch identifiziert wurden (N 1.1). Der zweite Indikator erhebt, inwieweit diese Stakeholder in die Evaluierung einbezogen wurden (N 1.2).

Beide Indikatoren erreichen einen hohen Erfüllungsgrad, wodurch das Bewertungskriterium N 1 mit einem Durchschnitt von 83 % zumeist erfüllt wird (siehe Abbildung 21). Laut der Rückmeldungen der AV werden die Interessen bestehender und potentieller Stakeholder im Mittel zu 81 % frühzeitig und systematisch identifiziert (N 1.1) und die relevanten Stakeholder, soweit möglich, in die Evaluierung einbezogen (N 1.2: 85 %). Ein vertiefender Blick auf die Antworten der AV zeigt, dass hierbei mehrheitlich sehr hohe Zustimmungswerte vergeben wurden und AV nur in Ausnahmen (vier AV) zu einer kritischen Einschätzung tendierten. Die starke Einbindung der relevanten Stakeholder in der Vorbereitungsphase der PEV erhöht die Wahrscheinlichkeit, dass die PEV-Ergebnisse von diesen Stakeholdern akzeptiert werden und damit nützlich sind. In einer Zusatzfrage zu den beiden Indikatoren wurde in der Online-Befragung ermittelt, auf welche Stakeholder sich die AV in ihren Antworten beziehen.³⁹ Zumeist betrifft dies Durchführungsorganisationen (96 %, n=67), politische Träger (94 %, n=62) und weitere Kooperationspartner (94 %, n=68). Etwas seltener sind hiermit auch Auftraggeber (85 %, n=57) und Zielgruppen (75 %, n=50) gemeint.

N 2 Klärung der Evaluationszwecke: Aufbauend auf der Einbindung der Stakeholder sollte der Zweck der Evaluierung mit ihnen abgesprochen werden, damit die Fragestellungen und Interessen einer breiten Akteursgruppe berücksichtigt werden können. Diesen Faktor untersucht das zweite Bewertungskriterium, welches mit einem Indikator operationalisiert wurde (N 2.1 = N 2).

Im Mittel der Antworten der AV wird das Bewertungskriterium N 2 mit ebenfalls 83 % zumeist erfüllt (siehe Abbildung 21). Durch den verbreiteten Ansatz, die Informationsbedarfe der bestehenden und potenziellen Stakeholder frühzeitig zu ermitteln und daraufhin die Zwecke der Evaluierung mit diesen abzustimmen, wird in den PEV die Grundlage dafür gelegt, dass die Evaluierungsergebnisse beachtet und konstruktiv aufgenommen werden.

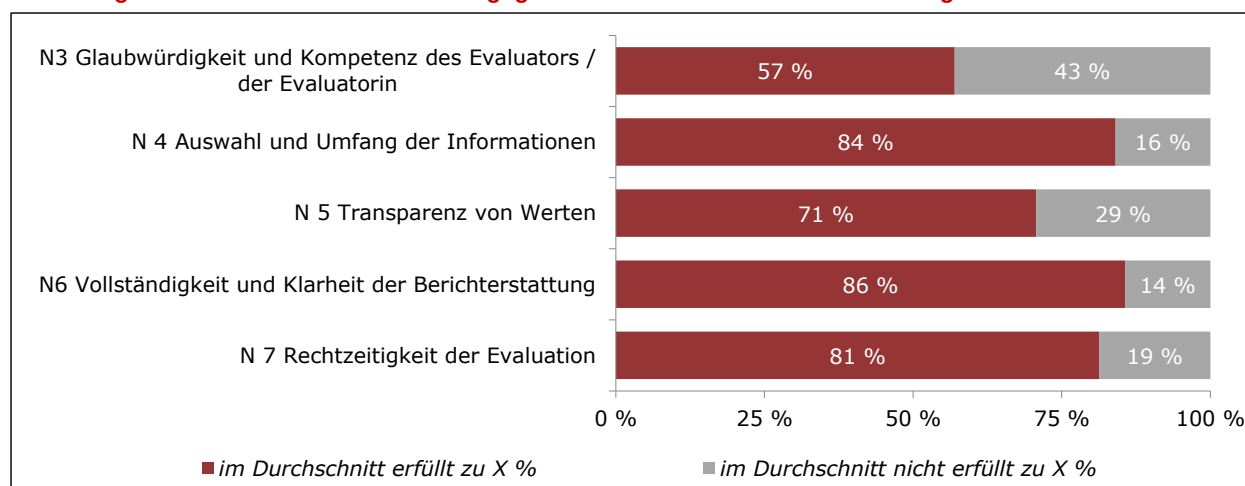
Die folgenden fünf Bewertungskriterien thematisieren die Durchführung und Nachbereitung der PEV:

³⁷ Um die Bewertungen auf Ebene der Evaluierungsstandards und Bewertungskriterien einheitlich und nachvollziehbar zu gestalten, werden diese wie folgt angegeben: 90–100 % (größtenteils bis vollständig erfüllt); 80–89 % (zumeist erfüllt); 60–79 % (eher erfüllt); < 60 % (bedingt erfüllt). Siehe hierzu auch Kapitel 2.1.

³⁸ Evaluierungsstandard Genauigkeit (59 %) und Durchführung (63 %).

³⁹ Angaben nachfolgend jeweils für die AV, die hierzu nicht „keine Angaben“ oder „nicht zutreffend“ angekreuzt haben. Dadurch ändert sich jeweils die Grundgesamtheit (n).

Abbildung 22: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien N 3–N 7



Quelle: Syspons 2016 (Textanalyse der PEV-Berichte und Online-Befragung der AV)

N 3 Glaubwürdigkeit und Kompetenz des Evaluators/ der Evaluatorin: Damit die Ergebnisse einer Evaluierung auf eine hohe Akzeptanz stoßen, sollte das PEV-Team glaubwürdig und kompetent sein. Die beiden diesem Bewertungskriterium zugrunde liegenden Indikatoren beruhen auf der Einschätzung der Kompetenzen des PEV-Teams durch die AV (N 3.1) und der Auseinandersetzung der TOR einer PEV mit den geforderten Kompetenzen (N 3.2). Sie beziehen sich damit auf unterschiedliche Aspekte.

Im Mittel ergibt sich eine bedingte Erfüllung des Bewertungskriteriums N 3 von 57 % (siehe Abbildung 22), wobei die beiden dahinterliegenden Indikatoren **sehr unterschiedlich** ausfallen. Insgesamt schätzen die AV die Kompetenz der PEV-Teams als hoch ein, was eine durchschnittliche Erfüllung des Indikators N 3.1 von 86 % verdeutlicht. Es wurde nicht nach der Einschätzung der Kenntnisse einzelner Gutachter/innen gefragt, sondern danach, inwieweit durch eine geeignete personelle Zusammensetzung des PEV-Teams die geforderten Kompetenzen abgedeckt wurden. Eine vertiefende Betrachtung der dahinterliegenden Befragungsisems verdeutlicht leichte Unterschiede der Einschätzungen der Kompetenzbereiche. So schätzen die AV die fachlich-sektorale Kompetenz des PEV-Teams (durchschnittlich 90 %) am höchsten ein. Ebenfalls bestätigen die AV die Prozesskompetenz hinsichtlich des GIZ-Auftragsmanagements und der Steuerung einer PEV (86 %) sowie die methodische Kompetenz (85 %) und die Beratungskompetenz (82 %) der PEV-Teams.

Im Gegensatz hierzu thematisiert der zweite Indikator des Bewertungskriteriums (N 3.2) die Auseinandersetzung mit den Kompetenzen der PEV-Teams in den TOR, welche den Evaluierungsberichten angehängt sind. In 29 % (n=20) der Berichte wird in den TOR beschrieben, welche Kompetenzen das PEV-Team aufweisen sollte.

N 4 Auswahl und Umfang der Informationen: Die Nützlichkeit einer Evaluierung wird gestärkt durch eine hinreichende Informativität der PEV. Dieses vierte Bewertungskriterium wurde mit einem Indikator belegt (N 4.1 = N 4). Dieser erfasst, inwieweit die PEV-Berichte die wesentlichen Fragestellungen einer Evaluierung umfassen und dadurch nutzbare Informationen für die beabsichtigte Verwendung liefern.

Hierbei zeigt sich, dass das Bewertungskriterium N 4 laut der Online-Befragung mit durchschnittlich 84 % zumeist erfüllt wird (siehe Abbildung 22). Die hohen Zustimmungswerte der AV weisen darauf hin, dass sich die PEV in der Breite auf jene Aspekte konzentrieren, die zur Beantwortung der Evaluierungsfragen relevant sind.

N 5 Transparenz von Werten: Das fünfte Bewertungskriteriums misst, inwiefern die Werte von Beteiligten und Betroffenen bzw. des PEV-Teams in den PEV-Berichten transparent gemacht werden. Diesem Kriterium liegt die Annahme zugrunde, dass die Werte dieser Personen und Gruppen Einfluss auf die Bewertungen in den PEV-Berichten haben und sich dadurch auch auf die Nützlichkeit auswirken. Das Kriterium baut auf zwei Indikatoren auf. Diese erfassen, ob in den PEV-Berichten Hinweise darauf

bestehen, dass der Auseinandersetzung mit den Werten des PEV-Teams (N 5.1) sowie der Beteiligten und Betroffenen (N 5.2) Raum bei der Diskussion um Evaluierungsergebnisse gegeben wurde. Diese beiden Indikatoren wurden durch die Textanalyse erhoben.

Im Durchschnitt der Berichte wird das Bewertungskriterium N 5 mit 71 % eher erfüllt (siehe Abbildung 22). Diesbezüglich wird in rund 57 % (n=40) der PEV-Berichte deutlich, dass ein Austausch im PEV-Team zu Evaluierungsergebnissen erfolgt (N 5.1), wie zum Beispiel im Rahmen von Synthesetreffen. Zugleich wird anhand von 84 % (n=59) der Berichte deutlich, dass für die Interpretation der Ergebnisse Partner und/ oder Zielgruppen einbezogen wurden (N 5.2). Auch diese Ergebnisse deuten an, dass durch den Austausch mit Partnern und/ oder Zielgruppen in den PEV ein Mechanismus besteht, durch den die Werte bei der Interpretation von Evaluierungsergebnissen transparent gemacht werden kann. Beide hier analysierten Indikatoren deuten den Grad der Werte-Transparenz jedoch nur an, da allein anhand der Berichte nicht nachvollzogen werden kann, welchen tatsächlichen Einfluss die Wertevorstellungen im Interpretationsvorgang einer Evaluierung eingenommen haben.

N 6 Vollständigkeit und Klarheit der Berichterstattung: Das sechste Bewertungskriterium der Nützlichkeit befasst sich mit der Frage, inwieweit die Evaluierungsberichte alle wesentlichen Informationen beinhalten (N 6). Der einzige Indikator dieses Kriteriums untersucht, inwieweit sich die Berichte an den inhaltlichen und formalen Vorgaben der annotierten Berichtsgliederung der GIZ orientieren (N 6.1 = N 6).

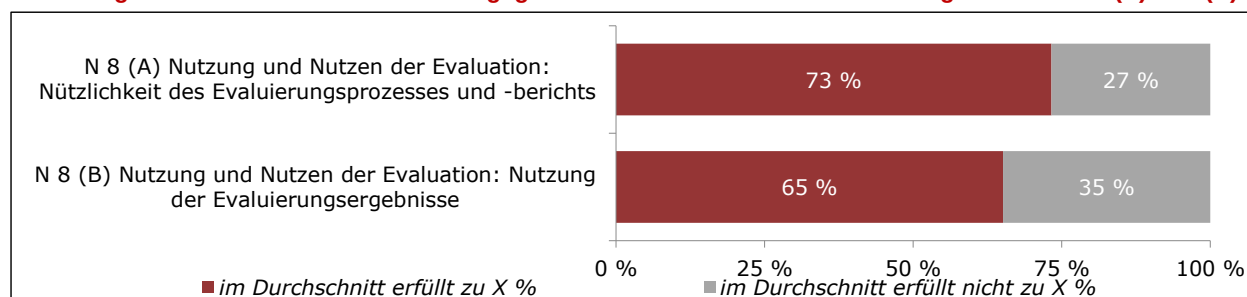
Die Berichte erreichen hierbei im Mittel 86 % (n=60), womit das Kriterium zumeist erfüllt wird (siehe Abbildung 22). Hierbei gilt zu beachten, dass der Indikator bereits dann negativ bewertet wurde, wenn bspw. obligatorische Kapitel oder Anhänge fehlten oder nur sehr verkürzt dargestellt wurden.

N 7 Rechtzeitigkeit der Evaluation: Evaluierungen sollen laut dem siebten Bewertungskriterium so rechtzeitig begonnen und abgeschlossen werden, dass ihre Ergebnisse in Entscheidungsprozessen berücksichtigt werden können. Die beiden diesbezüglichen Indikatoren messen, inwieweit der Evaluierungsprozess rechtzeitig durchgeführt wurde (N 7.1) und die Ergebnisse einer PEV rechtzeitig vorlagen (N 7.2).

Das Bewertungskriterium N 7 wurde im Mittel der AV-Befragung zumeist (81 %) erfüllt (siehe Abbildung 22). Die AV-Befragung zeigt, dass der Evaluierungsprozess überwiegend zu einem für die Nutzung der Erkenntnisse günstigen Zeitpunkt umgesetzt wurde (N 7.1: 85 %). Die Bereitstellung des PEV-Berichts zu einem für die Nutzung passenden Zeitpunkt erfolgte hingegen seltener (N 7.2: 78 %). Der leichte Unterschied zwischen den Werten ist ein Anhaltspunkt dafür, dass die Berichte trotz der Rechtzeitigkeit des Prozesses nicht immer zeitgerecht vorliegen und somit für die Beantwortung steuerungsrelevanter Fragen der evaluierten Vorhaben einer PEV noch besser in Wert gesetzt werden können.

Die letzten beiden Bewertungskriterien thematisieren die Nutzung und den Nutzen der PEV:

Abbildung 23: Durchschnittlicher Erfüllungsgrad der Indikatoren in den Bewertungskriterien N 8 (A)–N 8 (B)



Quelle: Syspons 2016 (Online-Befragung der AV)

N 8 (A) Nützlichkeit des Evaluierungsprozesses und -berichts: In der Planung, Durchführung und Berichterstattung einer PEV soll sichergestellt werden, dass diese zur Kenntnis genommen und genutzt

werden. Das Kriterium basiert auf zwei Indikatoren. Der erste Indikator erhebt die Nützlichkeit des Evaluierungsprozesses (N 8.1), der zweite Indikator die Nützlichkeit des Berichts (N 8.2).

Das entsprechende Bewertungskriterium N 8 (A) wird nach der Ansicht der AV im Durchschnitt eher erfüllt (73 %) (siehe Abbildung 23). Im Detail zeigt sich, dass der Evaluierungsprozess von Seiten der AV durchschnittlich zu 72 % als nützlich erachtet wird (N 8.1). Die Auswertung der einzelnen Items der Befragung zeigt, wie sich dieser Wert zusammensetzt. Demnach wird die Vorbereitung durchschnittlich zu 66 %, die Durchführung zu 78 % und die Nachbereitung zu 69 % als nützlich empfunden. Die Durchführung ist damit aus Sicht der AV der nützlichste Schritt des Evaluierungsprozesses, während der Nutzen aus Vor- und Nachbereitung nachgelagert ist. Zugleich wird die Nützlichkeit des Berichts, so der zweite Indikator (N 8.2), durchschnittlich mit 74 % und damit ähnlich nützlich wie die Durchführung bewertet.

N 8 (B): Nutzung der Evaluierungsergebnisse: Neben der Nützlichkeit einer Evaluierung kommt der Inwertsetzung der Ergebnisse (Nutzung) eine zentrale Bedeutung zu. Da die Nutzung von PEV-Ergebnissen auf verschiedenen Ebenen erfolgen kann, wurde dieses Bewertungskriterium mit sechs Indikatoren hinterlegt und auf Grundlage der AV-Befragung untersucht. Diese Indikatoren erfassen, inwieweit die Ergebnisse auf der GIZ-internen Projekt-/ Programmebene (N 8.3), der GIZ-internen Sektor-/ Regionalebene (N 8.4), der Ebene des Kooperationssystems von Vorhaben (N 8.5), des erweiterten Kooperationssystems von Vorhaben (N 8.6), für den sonstigen Austausch (N 8.7) und für die Verbesserung einer Folgemaßnahme (N 8.8⁴⁰) genutzt werden.

Mit durchschnittlich 65 % gilt das Bewertungskriterium N 8 (B) als eher erfüllt (siehe Abbildung 23). Die sechs Indikatoren erreichen jedoch einen unterschiedlich hohen Erfüllungsgrad.⁴¹

Den höchsten Nutzungswert (durchschnittliche Erfüllung 83 %) erreicht die PEV auf der GIZ-internen Projekt- und Programmebene (N 8.3). Dieses zentrale Nutzungserwarten an eine PEV wird folglich erfüllt. Ein differenzierter Blick auf die Befragungssitems der Online-Befragung zeigt, wie sich die Ergebnisse der Befragung zusammensetzen. Demnach werden die Ergebnisse der PEV vor allem während der Planung des Angebots einer Folgemaßnahme (95 %) und des Austausches im Projektteam (93 %) genutzt. Ebenfalls häufig werden die Ergebnisse unter Steuerungsgesichtspunkten genutzt, so für die Steuerung des Folgevorhabens (85 %) und des laufenden Vorhabens (80 %). Hingegen verdeutlicht die Befragung, dass AV die Evaluierungsergebnisse vergleichsweise seltener zur Förderung des Austausches über die Ergebnisse im Landesbüro und innerhalb der Landesportfolios (61 %) nutzen. Aus diesem Grund erfolgt die Nutzung der Ergebnisse vorrangig innerhalb der Vorhaben.

Die Evaluierungsergebnisse sollen auch in sektorale und regionale Austausch- und Lernprozesse integriert werden, um einen Nutzen für die GIZ-internen Sektor-/ Regionalebene zu generieren (N 8.4). Im Vergleich mit allen anderen Nutzungsebenen erreicht dieser Indikator den niedrigsten Erfüllungswert (durchschnittliche Erfüllung 41 %). Ein differenzierter Blick auf die Befragung gibt dabei Aufschluss über die Art der Nutzung. So werden die PEV-Ergebnisse am ehesten in die Fach- und Regionalbereiche eingespielt, um eine Reflexion zu Verbesserungen der fachlichen und regionalen Ansätzen vorzunehmen (50 %), und im unternehmensweiten Wissensmanagements in Wert gesetzt (45 %). In der Fachverbandsarbeit im Sektor und in der Produkterstellung guter Beispiele (21 %) werden die Ergebnisse hingegen nur punktuell genutzt. Auf dieser Nutzungsebene ist auffällig, dass insbesondere AV von Sektorvorhaben, die Beratung für Sektorthemen leisten, der Nutzung der Ergebnisse in der Fachverbandsarbeit geringe Bedeutung zuschreiben.

Im dritten Indikator wurde erhoben, ob die Evaluierungsergebnisse und Empfehlungen der PEV auf der Ebene des Kooperationssystems (Partner, Zielgruppen) eingebracht werden (N 8.5). Im Mittel kommt dieser Indikator auf eine Erfüllung von 66 %. Dieser Wert setzt sich aus verschiedenen in der Befragung

⁴⁰ In der Online-Befragung war dies eine Filterfrage. Nur AV, die eine Projektevaluierung mit Folgemaßnahme beauftragt haben, konnten hierzu antworten.

⁴¹ Für detaillierte Ergebnisse der Online-Befragung, siehe auch Ergebnisse der Online-Befragung in Kapitel 9.3.

erhobenen Nutzungsmöglichkeiten zusammen. Demnach werden die Ergebnisse einer PEV am ehesten für die direkte Beratung der Counterparts in den Partnerorganisationen (70 %) und die Gestaltung politischer Reformprozesse (67 %) genutzt. Laut der AV eignen sich die PEV-Ergebnisse weniger häufig, um die Erreichbarkeit der Zielgruppe zu erhöhen oder die Organisationsentwicklung von Partnerorganisationen zu adressieren. Diese Arten der Nutzung werden laut AV-Rückmeldung zu 63 % bzw. zu 59 % erfüllt.

Im Rahmen der systematischen Nutzung von Ergebnissen wurde anhand des vierten Indikators untersucht, wie diese im Austausch mit Auftraggebern und anderen Gebern in Wert gesetzt werden (N 8.6). Im Durchschnitt wird dieser Indikator zu 59 % erfüllt. Im Einzelnen zeichnen die hinter diesem Wert liegenden Befragungssitems ein diverses Bild. So werden die PEV-Ergebnisse vergleichsweise häufig für Fach- und Strategiediskussionen mit dem BMZ genutzt (72 %), wobei nicht erhoben wurde, ob sich dies auf die Länder- oder Sektorreferate oder die WZ-Referent/innen bezieht. Dagegen spielt die Nutzung der Ergebnisse im Austausch mit anderen Gebern (49 %) und anderen Auftraggebern (38 %) eine vergleichsweise untergeordnete Rolle. Dies kann auch durch die Verankerung der PEV im Auftragsverfahren des BMZ begründet werden. Schließlich wurde erhoben, inwieweit die Ergebnisse dem Austausch mit weiteren Akteuren dienen (N 8.7). Dies bewerteten die AV in der Befragung mit einer durchschnittlichen Erfüllung von 56 %.

Zuletzt wurde in der Befragung ein vertiefender Blick darauf geworfen, inwieweit die Ergebnisse der Evaluierung in einer Folgemaßnahme⁴² genutzt wurden (N 8.8). Insgesamt wird der Indikator mit durchschnittlich 77 % erfüllt. Im Einzelnen geht die durchschnittliche Einschätzung der AV zu der Nutzung der in den PEV evaluierten OECD-DAC-Kriterien auseinander. Die vergleichsweise hohe Nutzung der Ergebnisse der Analyse der Relevanz einer PEV (85 %) ist hierbei auffällig. Die Analysen der Effektivität und der Nachhaltigkeit werden durchschnittlich zu 78 %, die Analysen zum Impact zu 74 % und zur Effizienz zu 71 % als nutzbar empfunden. Folglich wird den Berichtsteilen zu den OECD-DAC-Kriterien ein unterschiedlicher Nutzen zugesprochen.

6.2 Einflussfaktoren auf die Nützlichkeit

Anhand von statistischen Analysen wurden potenzielle Einflussfaktoren auf die Nützlichkeit und den Nutzen der PEV getestet. In einem ersten Schritt wurde geprüft, wodurch die **Nützlichkeit des Prozesses** (N 8.1) beeinflusst wird. Die Ergebnisse hierzu weisen darauf hin, dass das diplomatische Vorgehen der PEV-Teams (D 2.1) Einfluss auf die Nützlichkeit des Prozesses nimmt⁴³. Damit ist gemeint, dass PEV, denen es gelingt, unterschiedliche Interessensgruppen zusammenzubringen, eine höhere Nützlichkeit erreichen. Die Gründe für diesen signifikanten Zusammenhang sind vermutlich darin zu suchen, dass eine ausgewogene Berücksichtigung der Interessensgruppen die Akzeptanz und Zustimmung in der Durchführung der Evaluierung sichert. Dies deckt sich mit den qualitativen Rückmeldungen der AV, die teils auf den Austausch mit der Zielgruppe, den Partnern und dem BMZ als Gelingensbedingung für eine hohe Nützlichkeit hinweisen.

Weiterhin wurden Zusammenhänge identifiziert, die die **Nützlichkeit des Berichts** (N 8.2) beeinflussen. Die Meta-Evaluierung ergab, dass die Nützlichkeit des PEV-Berichts signifikant von der Rechtzeitigkeit des Evaluierungsprozesses (N 7.1) und der Verfügbarkeit von PEV-Ergebnissen (N 7.2) abhängt.⁴⁴ Diese Einschätzung spiegelt sich auch in den qualitativen Rückmeldungen der AV wider. Dabei wurde in einigen Fällen darauf hingewiesen, dass die Nützlichkeit der PEV-Berichte eingeschränkt war, weil die Berichte erst bei bereits angelaufener Folgemaßnahme vorlagen. Auch wurde vereinzelt darauf verwiesen, dass der Zeitpunkt der PEV in einem Spannungsverhältnis zu Erkenntnis- und

⁴² In der Online-Befragung war dies eine Filterfrage. Nur AV, die eine Projektevaluierung mit Folgemaßnahme beauftragt haben, konnten hierzu antworten.

⁴³ $r = 0,343$, $n = 69$, $p = 0,004$.

⁴⁴ $r = 0,337$, $n = 68$, $p = 0,005$.

Steuerungsinteressen der AV stehen kann, wenn die PEV bis zu ein Jahr vor Vorhabensende die Wirkungen eines dreijährigen Vorhabens in seiner ersten Phase evaluieren sollen.

Neben der Nützlichkeit wurde auch untersucht, wodurch die **Nutzung** (N 8.3–N 8.7) einer PEV beeinflusst wird. Hierbei wurde zunächst ein naheliegender Zusammenhang bestätigt. So nimmt die Nützlichkeit des Evaluierungsprozesses und -berichts (N8 (A)) einen Einfluss auf die Nutzung der PEV-Ergebnisse⁴⁵. Zugleich wird eine PEV, welche als informativ wahrgenommen wird und die die wesentlichen Fragestellungen der Evaluierung abdeckt (N 4.1), als nützlich bewertet.⁴⁶

Des Weiteren weisen die Analysen darauf hin, dass zwischen der systematischen Beteiligung von relevanten Stakeholdern und den Nutzungsmöglichkeiten der PEV ein positiver Zusammenhang besteht. Werden also relevante Vorhabens-Stakeholder bei der Vorbereitung der Evaluierung frühzeitig eingebunden (N 1.1, N 1.2)⁴⁷ und wird der Evaluierungszweck mit diesen ausgehandelt (N 2.1)⁴⁸, erhöht sich tendenziell auch die Nutzung einer PEV. In diesem Sinne zeigte sich auch, dass diplomatisches Vorgehen (D 2.1) in PEV aufgrund des in den Evaluierungen etablierten Dialogs mit den Interessensgruppen die Nutzung der PEV signifikant erhöhen kann.⁴⁹ In diesem Kontext konnte auch die Annahme bestätigt werden, dass die zielgruppengerechte Aufbereitung und das Teilen der Evaluierungsergebnisse mit den relevanten Stakeholdern (F 5.1, F 5.2) eine Voraussetzung dafür sind, dass Ergebnisse einer PEV in Wert gesetzt werden können.⁵⁰

Bei der statistischen Untersuchung wurde weiterhin bestätigt, dass zwischen den Einschätzungen der AV zur **Nutzung** der Ergebnisse (N 8.3–N 8.7) und der Einschätzungen der AV zur Angemessenheit des Aufwandes im Verhältnis zum Nutzen (D 3.1, D 3.2) ein signifikanter Zusammenhang besteht.⁵¹ Dieser Zusammenhang verdeutlicht, dass bei einem hohen wahrgenommenen Nutzen der PEV durch die AV auch das finanzielle Budget und das Mengengerüst einer PEV als angemessen eingeschätzt werden.

Schließlich falsifizierte die Untersuchung einige der untersuchten vermuteten Zusammenhänge. Demnach beeinflusst die Genauigkeit – also die methodische Qualität – die Einschätzung der AV hinsichtlich der Nützlichkeit einer Evaluierung nicht signifikant. Hierbei konnten auch keine systematisch unterschiedlichen Bewertungen in Bezug auf die methodische Qualität der einzelnen OECD-DAC-Kriterien und die entsprechende Einschätzung der AV zur Verwendung der Ergebnisse der OECD-DAC-Bewertungen festgestellt werden. Auch schätzen die AV bilateraler und die AV nicht-bilateraler Vorhaben den Nutzen von PEV nicht signifikant unterschiedlich ein.

6.3 Bewertung und Schlussfolgerungen zur Nützlichkeit

Die Analyse verdeutlicht die Stärken und Schwächen der PEV hinsichtlich ihrer Nützlichkeit und Nutzungsintensität.

Aus Sicht des Meta-Evaluierungsteams liegt eine der **Stärken** der PEV in der Einbeziehung relevanter Stakeholder. Es gelingt den PEV-Teams größtenteils, die beteiligten und betroffenen Stakeholder in die Evaluierung einzubinden und die PEV entsprechend auszurichten. Dabei sollte positiv hervorgehoben werden, dass die Evaluierungszwecke oftmals gemeinsam diskutiert werden. Dies schafft die Grundlage dafür, dass die PEV auch für die beteiligten und betroffenen Stakeholder nützlich sind. Die Analysen hierzu zeigen, dass der Austausch mit den Stakeholdern im Evaluierungsprozess einen systematischen Einfluss auf die Nützlichkeit der PEV nimmt und somit eine wichtige „Stellschraube“ für die Erbringung bzw. weitere Erhöhung der Nützlichkeit darstellt.

⁴⁵ $r = 0,366$, $n = 71$, $p = 0,002$.

⁴⁶ $r = 0,303$, $n = 74$, $p = 0,009$.

⁴⁷ $r = 0,299$, $n = 73$, $p = 0,010$.

⁴⁸ $r = 0,274$, $n = 73$, $p = 0,019$.

⁴⁹ $r = 0,399$, $n = 73$, $p < 0,001$.

⁵⁰ $r = 0,277$, $n = 74$, $p = 0,017$.

⁵¹ $r = 0,51$, $n = 74$, $p = 0,003$.

Eine weitere Stärke liegt aus Sicht des Meta-Evaluierungsteams in der hohen Nützlichkeit der PEV für die GIZ-interne Projekt- und Programmebene, wobei sich dies vor allem auf das evaluierte Vorhaben beschränkt, in dem die PEV für Lern- und Reflexionsprozesse verwendet werden. Damit erfüllt ein Großteil der PEV eine der zentralen Nutzenerwartungen, die an das Instrument gestellt werden.

Schließlich bewertet das Meta-Evaluierungsteam auch die Erfüllung formaler Kriterien als eine Stärke der PEV. Wie sich hierzu aus Sicht der AV zeigte, gelingt es den PEV-Berichten zumeist, die Standards hinsichtlich der Klarheit sowie der Informationstiefe und -breite der Berichterstattung zu erfüllen. Zugleich sind die hohen Zustimmungswerte dahingehend, dass Evaluierungsprozess und -bericht zu einem für die Nutzung günstigen Zeitpunkt umgesetzt wurden, positiv hervorzuheben. Diese hohe Rechtzeitigkeitsbewertung ist zugleich eine der „Stellschrauben“, die die spätere Nutzung beeinflusst und bedingt auch, dass alle in dem Prozess einer Evaluierung angelegten Schritte zeitgemäß erfüllt werden müssen.

Schwächen zeigten sich nach Ansicht des Meta-Evaluierungsteams bezüglich der Kompetenzen der PEV-Teams. Obwohl die AV die methodische Kompetenz (Evaluierungskompetenz) im PEV-Team positiv beurteilen, steht diese Einschätzung im Gegensatz zu den Ergebnissen der Meta-Evaluierung im Bereich methodischer Qualität. So zeigt sich, dass der Evaluierungsstandard der Genauigkeit, welcher sich mit der methodischen Qualität einer Evaluierung befasst, mit durchschnittlich 59 % nur bedingt erfüllt wird. Hier liegt eine Diskrepanz vor, die gegebenenfalls damit zusammenhängt, dass AV die methodische Kompetenz von PEV-Teams nicht adäquat einschätzen oder der methodischen Qualität nicht ausreichend Aufmerksamkeit schenken. Insgesamt sollte zudem aus Sicht des Meta-Evaluierungsteams kritisch reflektiert werden, dass die Ergebnisse zur methodischen Qualität einer Evaluierung keinen Einfluss auf die Einschätzungen zur Nützlichkeit haben. Demnach werden Evaluierungen weitgehend unabhängig davon, ob sie methodisch valide, robust und fundiert sind, als nützlich bewertet. Der Stellenwert der methodischen Qualität ist hierbei folglich gering. In der hohen Nutzung bei gleichzeitig geringer Validität und Robustheit der Daten zeigt sich aus Sicht des Meta-Evaluierungsteams ein **Risiko**.

Zugleich zeigt ein weiterer Einflussfaktor die Nützlichkeit und den Nutzen einer PEV-Herausforderung auf. Demnach beeinflusst das diplomatische Vorgehen der PEV-Teams sowohl die Nützlichkeit des Prozesses als auch die Nutzung der Evaluierungsergebnisse positiv. Allerdings schneiden die PEV hinsichtlich ihres diplomatischen Vorgehens eher durchwachsen ab. Die damit verbunden erhöhte Wahrscheinlichkeit, Einflussversuche von Interessensgruppen nicht immer zu antizipieren, kann sich negativ auf die Nutzung der Evaluierung auswirken.

Schließlich belegen die Ergebnisse der Analysen aus Sicht des Meta-Evaluierungsteams ein **mögliches Spannungsfeld** zwischen den Evaluierungsstandards Durchführbarkeit und Nützlichkeit. Entscheidend ist in diesem Zusammenhang, dass, wenn der Nutzen einer PEV erhöht wird, auch die mit ihr verbundenen Aufwände weniger stark ins Gewicht fallen. Dies ist eine wichtige Erkenntnis, wenn man Nutzung und Durchführbarkeit gemeinsam diskutiert.

Neben den diskutierten Stärken und Schwächen befasste sich die Meta-Evaluierung auch mit der Frage, inwieweit die Unterstützungsmaterialien und Handreichungen der GIZ Aspekte der Nützlichkeit einer Evaluierung thematisieren (siehe auch Kapitel 9.4 im Anhang). Demnach hat sich gezeigt, dass die Nützlichkeit einer PEV in Vorbereitung, Durchführung und Nachbereitung sowie die Nutzung der Evaluierungsergebnisse insbesondere in der GIZ-Policy für Monitoring und Evaluierung (2013) und in der Methoden-Toolbox (2015h) in der Breite thematisiert werden. Die dortigen Diskussionen zur Prozessqualität und zur Berichts- und Produktqualität sowie zum Lernen aus Evaluierungen greifen nach Ansicht des Meta-Evaluierungsteams die wesentlichen hier diskutierten Bewertungskriterien bereits auf.

7 Schlussfolgerungen und Handlungsempfehlungen

Die Schlussfolgerungen und Handlungsempfehlungen leiten sich aus den Analyseergebnissen ab. In der Darstellung der Handlungsfelder wird unterschieden zwischen methodischer Qualität, Prozessqualität und Nützlichkeit. Alle Handlungsempfehlungen richten sich an die GIZ.

7.1 Handlungsfeld 1: Methodische Qualität

Den Schlussfolgerungen und Empfehlungen zur Steigerung der methodischen Qualität liegt die Erkenntnis der Meta-Evaluierung zugrunde, dass die methodische Qualität der PEV in der Breite schwach ist und vor dem Hintergrund des GIZ-Qualitätsanspruchs, alle DEGEVAL Standards vollständig zu erfüllen, ein wesentliches Handlungsfeld für die Weiterentwicklung des PEV-Instruments darstellt.

Schlussfolgerung 1.1: Die methodische Qualität der Analysen der PEV ist durch einen geringen Reflexionsgrad und eine geringe Methodenvielfalt gekennzeichnet. PEV stellen nur in Ausnahmen ihre methodische Vorgehensweise mit Hinweis auf die Beantwortung von Evaluierungsfragen oder die Besonderheiten des Vorhabens begründet dar. Zugleich werden die Vor- und Nachteile der gewählten Methoden nur teilweise reflektiert. Dies führt dazu, dass eine eingeschränkte Auswahl an Evaluierungsmethoden eingesetzt wird, die sich zumeist auf qualitative Interviews und Daten- und Dokumentenanalysen konzentriert. Quantitative Befragungen werden hingegen fast nie eingesetzt. Dabei zeigte die Meta-Evaluierung, dass sich die Anwendung von mindestens drei Methoden positiv auf die methodische Gesamtqualität auswirkt. Der geringe Reflexionsgrad schlägt sich auch in den Analysen der OECD-DAC-Kriterien nieder. So schneiden die PEV bezüglich der Einschätzung der Kausalität zwischen Maßnahmen und Wirkungen (Effektivität), der Bewertung der Hypothesen zu den langfristigen Wirkungen (Impact) und der Beschreibung der Grenzen der Messung der Nachhaltigkeit (Nachhaltigkeit) schlecht ab. Hierbei zeigen sich aus Sicht des Meta-Evaluierungsteams auch die Auswirkungen einer unter Umständen unzureichenden Auftragsklärung zwischen AV und PEV-Teams, was sich auch in der fehlenden formal verankerten Inception-Phase niederschlägt, in der das Evaluierungsdesign abgewogen wird und die in den früheren Unabhängigen Evaluierungen (UE) standardmäßig durchgeführt wurde. Hierauf haben bereits andere Untersuchungen, so die Meta-Evaluierung Gesundheit (2013) hingewiesen.

Empfehlung: Die Handreichung und die Unterstützungsmaterialien der PEV greifen bereits eine Vielzahl der Vorgaben der Evaluierungsstandards auf. Allerdings sollten die Unterstützungsmaterialien der PEV auch die Auseinandersetzung und Reflexion mit methodischen Entscheidungen der PEV als Analysefragen und zu bewertende Aspekte benennen. Andernfalls bleibt die Nachvollziehbarkeit der Analysen der PEV eingeschränkt.

Empfehlung: Die Unterstützungsmaterialien der PEV sollten stärker auf die Bedeutung der methodischen Vielfalt der PEV für die methodische Qualität einer PEV hinweisen. Gelingt es, die Anzahl der angewendeten Methoden in der PEV zu fördern, könnte dies einen signifikant positiven Einfluss auf die methodische Qualität einer Evaluierung haben.

Empfehlung: Die in der Meta-Evaluierung identifizierten Einflussfaktoren für die methodische Qualität sollten in der Auftragsklärung einer PEV thematisiert werden. Es sollte reflektiert werden, inwieweit diese in der Konzeption und Durchführung der PEV umgesetzt oder begründet ausgeschlossen werden.

Schlussfolgerung 1.2: Die insgesamt geringe methodische Qualität der PEV widerspricht der positiven Einschätzung zur Methodenkompetenz der PEV-Teams durch die AV. Zugleich werden Evaluierungsergebnisse unabhängig von ihrer methodischen Qualität genutzt. Insgesamt sollte kritisch reflektiert werden, dass die Ergebnisse zur methodischen Qualität einer PEV, die durchschnittlich auf eine bedingte Erfüllung hinweisen, keinen Einfluss auf die Einschätzungen zur Nützlichkeit durch die AV haben. Zudem liegt eine Diskrepanz zwischen der bedingten Erfüllung des Genauigkeitsstandards und der Einschätzung der Methodenkompetenz der PEV-Teams durch die AV vor. Dies könnte

gegebenfalls damit zusammenhängen, dass die AV die methodische Kompetenz von PEV-Teams nicht adäquat einschätzen oder der methodischen Qualität nicht ausreichend Aufmerksamkeit schenken. Gleichzeitig werden in den PEV grundlegende Standards, wie die Auseinandersetzung mit Wirkungshypothesen, die Belegung von Analysen mit Quellenangaben und die Trennung zwischen Analyse und Bewertung, nicht konsequent eingehalten. In der hohen Nutzung bei gleichzeitig geringer Validität und Robustheit der Daten liegt nach Ansicht des Meta-Evaluierungsteams ein Risiko.

Empfehlung: Die methodische Qualität könnte durch den Aufbau eines Expert/innen-Pools unterstützt werden, auch um die AV bei der Wahl geeigneter externer Evaluator/innen mit Evaluierungskompetenz zu unterstützen. Bei dieser Entscheidung sollten die bisherigen Erfahrungen mit dem Kosten-Nutzen-Verhältnis abgewogen werden, die mit dem Aufbau des Monitoring-Expert/innen-Pools verbunden sind. In der Auswahl der Expert/innen könnte bewusst auf Profile gesetzt werden, die ein breites Spektrum sowohl qualitativer als auch quantitativer Methoden abdecken, um die Methodenvielfalt zu fördern. Die AV könnten gleichzeitig durch Profilbeschreibungen unterstützt werden, die aufzeigen, welche Kenntnisse Evaluations-Expert/innen mitbringen sollten.

Empfehlung: Zugleich sollten die TOR deutlicher machen, welche Kompetenzprofile gesucht werden, und darauf aufbauend auch die Verantwortung der Evaluierungs-Expert/innen im PEV-Prozess stärken.

Empfehlung: Da Mitarbeiter/innen des FMB in der Regel Teil der PEV-Teams sind, könnten Investitionen in deren Training und Weiterbildung eine Möglichkeit sein, positive Impulse zur methodischen Qualität der PEV zu liefern. Die Trainingsschwerpunkte sollten auf der Auseinandersetzung mit grundlegenden Evaluierungsstandards liegen, wie zum Beispiel der Trennung zwischen Beschreibung, Analyse und Bewertung oder der Reflexion von Wirkungshypothesen. Damit könnte eine Sensibilisierung für grundlegende methodische Aspekte der PEV erreicht werden, ohne zugleich die Verantwortung der Evaluierungs-Expert/innen im PEV-Prozess zu unterminieren.

Schlussfolgerung 1.3: Es gibt keine Orientierungspunkte für die Anzahl der Gutachter/innen-Tage, die im Mengengerüst einer PEV geplant werden sollen, und für die Anzahl der Mitglieder der PEV-Teams. Die Anzahl der Gutachter/innen-Tage einer PEV beeinflusst die methodische Qualität der PEV. PEV unter 55 Gutachtertage fallen methodisch schwächer aus als PEV mit mehr Gutachtertage. Dies gilt unabhängig von dem Auftragsvolumen eines Vorhabens. Zugleich ist methodische Qualität von PEV mit zunehmenden Gutachtertage bis zu einem Mengengerüst von circa 86 Tagen tendenziell besser. Ab diesem Wert ist der Effekt jedoch nicht mehr vorhanden bzw. sogar gegenläufig. Diesbezüglich gilt zu berücksichtigen, dass die Streuung der untersuchten PEV ab diesem Wert verhältnismäßig hoch und die Fallzahlen verhältnismäßig klein sind. Ein ähnlicher Trend lässt sich für die Größe des PEV-Teams erkennen. Die methodische Qualität einer PEV ist höher, wenn mehr als zwei Gutachter/innen an der Evaluierung teilnehmen.

Empfehlung: Es wird empfohlen, Orientierungspunkte für den Ressourceneinsatz einer PEV zu formulieren. Dies betrifft sowohl die Anzahl der Gutachter/innen (mehr als zwei Gutachter/innen) als auch die Anzahl der Gutachter/innen-Tage (mehr als rund 55 Tage). Hierbei sollte berücksichtigt werden, dass diese Orientierungspunkte größtenteils von PEV mit Folgevorhaben abgeleitet wurden. Mit der Erhöhung des Mengengerüsts steigen jedoch auch die zeitlichen und finanziellen Aufwände der Vorhaben. Vor dem Hintergrund der identifizierten teils kritischen Wahrnehmung der AV zu den Aufwänden der PEV, baut sich hier ein Spannungsverhältnis auf. Die Diskussion darüber, wie die methodische Qualität erhöht werden kann ohne negative Wahrnehmungen zu Aufwänden und Kosten zu verschärfen, sollte GIZ-intern geführt werden (siehe hierzu auch Empfehlungen zur Schlussfolgerung 2.4).

Schlussfolgerung 1.4: Die Analyse und Bewertung einzelner OECD-DAC-Kriterien, insbesondere Effizienz, Impact und Effektivität, weisen weitgehende Schwächen auf. Die damit einhergehende geringe Validität der OECD-DAC-Analysen bei gleichzeitiger hoher Nutzung ist ein Risiko. Diese Kriterien sind zugleich von unternehmensstrategischer Bedeutung. Es ist auffallend, dass PEV in ihren Impact-Untersuchungen kaum Bezug auf EZ-Programme nehmen und diesen somit eine geringe Bedeutung zuschreiben. Bewertungen des Impacts sind nicht nachvollziehbar, wenn die Attributionslücke

nicht behandelt oder Vorhaben mit Indikatoren auf Output-Ebene weitreichende Impacts zugeschrieben werden. Zugleich schneidet die Untersuchung der Effizienz als schwächstes OECD-DAC-Kriterium ab. Es besteht bei den PEV-Teams scheinbar teils kein klares Verständnis davon, was „Effizienz“ bedeutet und wie sie gemessen werden kann. Schließlich ist bei den Effektivitätsanalysen auffällig, dass zwar die Qualität der Indikatoren der Vorhaben (nach SMART) zumeist von den PEV-Teams kritisiert wird, doch wird daraufhin nur sehr selten auch die geforderte Anpassung der Indikatoren vorgenommen. Dies schränkt die Aussagekraft der vorgenommenen Analysen ein.

Empfehlung: Es wird empfohlen, in den Unterstützungsmaterialien und in den Trainings für PEV-Teams die Bedeutung der Einbettung der Vorhaben in EZ-Programme hervorzuheben. Dies könnte einen positiven Einfluss auf die Auseinandersetzung mit den Bewertungsmaßstäben zur Beurteilung des Beitrags der Vorhaben zu Impacts haben. Dies ist vor dem Hintergrund relevant, dass das BMZ daran interessiert ist, in Zukunft verstärkt die Beiträge der Vorhaben zu EZ-Programmzielen zu messen.

Empfehlung: Es sollte verstärkt in die Qualität der Effizienzuntersuchungen investiert werden. Demnach sollten in der Handreichung zu den OECD-DAC-Kriterien die Erwartungen an eine Effizienzanalyse geschärft werden, u. a. um darzulegen, welche Effizienzuntersuchung (Implementierungs-, Produktions- oder Allokationseffizienz) mit welcher Datengrundlage möglich ist. Dadurch kann das Bewusstsein der PEV-Teams für die begründete Auswahl von Methoden und Verfahren der Effizienzmessung und deren Grenzen geschärft werden.

Empfehlung: Es sollte im Rahmen einer PEV anhand ausgewählter Vorhaben eine Analyse der Allokationseffizienz durchgeführt werden, um zu verdeutlichen, wie in der Effizienzanalyse eines Vorhabens die GIZ-Ressourcen und Wirkungen zueinander in Bezug gesetzt werden können. Hierzu eignen sich ggfs. kofinanzierte GIZ-Vorhaben, welche die benötigten Datengrundlagen aufgrund der Vorgaben zu Kofinanzierungen bereits aufbereitet haben. Hierdurch könnten aktuelle (inter-)nationale Diskussionen zum Nachweis von „Value for Money“ in der EZ und im BMZ aktiv aufgegriffen werden.

Empfehlung: Es sollte eine Diskussion mit Gutachter/innen des FMB, die zuletzt PEV durchgeführt haben, darüber erfolgen, warum als nicht-SMART identifizierte Indikatoren oftmals nicht umformuliert werden. Zugleich könnten Unterstützungsmaterialien der PEV Beispiele aufzeigen, wie Indikatoren umformuliert werden, ohne deren Inhalte und damit die Ansprüche der Vorhaben zu verändern.

Schlussfolgerung 1.5: Das Instrument der PEV ist eher auf bilaterale als auf nicht-bilaterale Vorhaben zugeschnitten, wodurch Letztere methodisch signifikant schlechter abschneiden.

Derzeit ist die PEV auf das spezifische Wirkungsverständnis und das Verständnis des Kooperationssystems von bilateralen Vorhaben ausgerichtet. Es wird jedoch zugleich auf Sektorvorhaben angewandt, die andere methodische Anforderungen aufweisen. Die PEV berücksichtigt aus Sicht des Meta-Evaluierungsteams nicht ausreichend, dass Sektorvorhaben nicht auf der Outcome-Ebene gesteuert werden können, was nicht zuletzt daran liegt, dass der Partner zugleich der Auftraggeber ist. Zudem schneiden PEV von Sektorvorhaben schlechter ab, weil sie eine geringere Anzahl an Gutachtertagen zur Verfügung stellen.

Empfehlung: Die zentralen Handreichungen zur methodischen Vorgehensweise und zur annotierten Berichtsgliederung sollten durch spezifische Hinweise auf die Wirkungsanforderungen von Sektorvorhaben ergänzt werden. Dies sollte mit einer Diskussion darüber verbunden werden, welche Ebenen der Evaluierung (Output, Outcome) bei Sektorvorhaben im Fokus stehen können und wie diese methodisch evaluiert werden müssen, um valide und robuste Ergebnisse zu erzielen.

7.2 Handlungsfeld 2: Prozessqualität

Ausgehend von der kontroversen Einschätzung der AV zu Verfahren und Effizienz von PEV (Durchführbarkeit) und den positiven Ergebnissen eines fairen Umgangs in einer PEV (Fairness), wurden die folgenden Schlussfolgerungen und Handlungsempfehlungen entwickelt.

Schlussfolgerung 2.1: Die Handreichung und die Unterstützungsleistungen der STS 08 konnten bislang als Teil des PEV-Prozesses in der Breite nicht den gewünschten Effekt auf die methodische Qualität der PEV-Berichte entfalten. Hinsichtlich der Hilfsmaterialien, die den PEV-Teams zur Verfügung gestellt werden, weisen die Ergebnisse der Meta-Evaluierung auf einen Überarbeitungsbedarf hin. Zwar wurde festgestellt, dass die Inhalte der Handreichungen die meisten methodischen Anforderungen enthalten und auch teils als hilfreich angesehen werden. Allerdings werden die Handreichung und Unterstützungsmaterialien, insbesondere aus der Perspektive neuer PEV-Mitglieder und AV, gleichzeitig oftmals als zu „überladen“ und aufgrund der Vielzahl der Materialien als unübersichtlich wahrgenommen.

Empfehlung: Hierzu könnte die zentrale Handreichung weiterentwickelt werden (Starter-Kit PEV, FAQ PEV o. Ä.), dabei entlang des PEV-Umsetzungsprozesses strukturiert werden und die Unterstützungsmaterialien an geeigneten Stellen verlinkt. Gleichzeitig könnte in dieser zentralen Handreichung stärker verdeutlicht werden, welche Handreichungen verbindlich oder begleitend (optional) sind und welche Handreichungen insbesondere für neue PEV-Mitglieder und AV relevant sind.

Schlussfolgerung 2.2: Die Qualitäts-Checks konnten in der Breite nicht den gewünschten Effekt auf den PEV-Prozess entfalten. Da die PEV-Checks nur vorübergehend als Pilotierung eingeführt wurden, steht eine Reflexion ihrer Wirksamkeit an. Bislang sind die Qualitäts-Checks als eine „end-of-pipe“-Qualitätssicherung angelegt. Damit hängt die Wirksamkeit der Qualitäts-Checks stark von effektiven und zeitnahen Überarbeitungsprozessen ab. Zugleich können Überarbeitungsschleifen den Prozess in die Länge ziehen und eine zeitnahe Rückmeldung durch die STS 08 erschweren. Zugleich können sie nur rückwirkend Rückmeldungen zur methodischen Qualität der PEV-Berichte geben, haben aber keinen Einfluss auf die Auswahl und Umsetzung der methodischen Vorgehensweise und der Methoden während der Durchführung einer PEV.

Empfehlung: Es wird eine Diskussion zur Weiterführung der PEV-Checks empfohlen. Sollte sich dafür entschieden werden, müsste eine zeitnahe Prozessgestaltung der Qualitäts-Checks garantiert werden, um den Evaluierungsprozess nicht in die Länge zu ziehen. Sollte sich dagegen entschieden werden, müssten andere Mechanismen etabliert werden. Dies könnte zum Beispiel ein Help Desk der STS 08 oder anderer Organisationseinheiten sein, ggfs. mit Sprechstundenzeiten, an das sich PEV-Teams bei methodischen Rückfragen wenden können.

Schlussfolgerung 2.3: Ausgehend von dem dieser Meta-Evaluierung zugrunde gelegten Analyseraster greifen die Grundlagendokumente der PEV nicht alle Qualitätskriterien auf. Als Teil des Prozesses der Qualitätskontrolle einer Evaluierung wird an die Grundlagendokumente der Anspruch gerichtet, dass sie die international akzeptierten Qualitätskriterien operationalisieren.

In Bezug auf die methodische Qualität werden bislang die Darstellung des Auftragswerts der Vorhaben, die Ableitung der Ziele und Zwecke einer Evaluierung, die Darstellung der Systematik der Dokumentenauswertung und die Auseinandersetzung mit Partnerdaten nicht gefordert. Zugleich ist bislang keine Reflexion der methodischen Vorgehensweise hinsichtlich der Zuordnungs- und/ oder Beitragsanalyse sowie einiger Indikatoren bezüglich der OECD-DAC-Kriterien gefordert. Zudem gibt es bisher noch keine Orientierungspunkte (Aufwände, etc.) für die Durchführbarkeit einer PEV.

Empfehlung: Die genannten Lücken sollten in den Grundlagendokumente geschlossen werden, indem die fehlenden Standards aufgenommen oder begründet ausgeschlossen werden. Dadurch würde es gelingen, die zentralen Handreichungen kohärent zum internationalen Qualitätsverständnis einer Evaluierung weiterzuentwickeln.

Schlussfolgerung 2.4: Die Aufmerksamkeit wird in der Vorhabensrealität stark auf die verursachten Aufwände einer PEV gelegt und weniger auf ihre damit verbundene methodische Qualität. Die Angemessenheit der Verfahren wird insbesondere in Bezug auf die zeitlichen Aufwände für die Mitarbeiter/innen der evaluierten Vorhaben kontrovers eingeschätzt, weniger in Bezug auf die Aufwände der Partner. Zugleich wird insbesondere der finanzielle Aufwand einer PEV kontrovers seitens

der AV eingeschätzt. Insgesamt fokussieren sich die Einschätzungen hierzu primär auf eine Kostendiskussion, weniger auf die Kosten-Nutzen-Relation. Dies spiegelt sich darin wieder, dass Einschätzungen zur Durchführbarkeit nicht den Nutzen einer PEV beeinflussen.

Empfehlung: In Abstimmung mit dem BMZ sollte eine Diskussion über die M&E-Aufwände im Rahmen einer eigenen Budgetlinie erfolgen. Dies könnte dem Eindruck mancher AV entgegenwirken, dass die Aufwände für M&E aus dem Budget der Vorhaben entnommen werden. Durch eine eigene Budgetlinie könnte sichergestellt werden, dass genügend Ressourcen für die PEV zur Verfügung stehen (relevant für die Qualität), und zugleich der zum Teil negativen Wahrnehmung der Kosten und Aufwände entgegengewirkt werden.

Empfehlung: Die unternehmensstrategische Bedeutung von M&E könnte durch verbindliche und international akzeptierte Orientierungswerte gesteigert werden, die von anderen Gebern bereits genannt werden (i.d.R. 3 bis 5 % des Projektbudgets). Hierzu können bereits bestehende Erfahrungen der GIZ mit den Vorgaben anderer Geber herangezogen werden.

7.3 Handlungsfeld 3: Nützlichkeit

Das Kriterium der Nützlichkeit schneidet vergleichsweise gut ab. Die folgenden Schlussfolgerungen und Handlungsempfehlungen sollen die Nützlichkeit und Nutzung der PEV-Ergebnisse weiter verbessern.

Schlussfolgerung 3.1: Die Einbindung von beteiligten und betroffenen Stakeholdern wirkt sich signifikant positiv auf die Nutzung der Evaluierungsergebnisse aus. Die Ergebnisse der Meta-Evaluierung dokumentieren, dass relevante Stakeholder größtenteils eingebunden und PEV entsprechend ausgerichtet werden. Auch die Evaluierungsergebnisse werden oftmals mit beteiligten und betroffenen Stakeholdern geteilt. Allenfalls kann die Einbindung von Zielgruppen noch verbessert werden. Die Bedeutsamkeit zeigt sich darin, dass der Austausch mit relevanten Stakeholdern einen systematischen Einfluss auf die Nützlichkeit einer PEV nimmt und somit eine „Stellschraube“ hierfür ist.

Empfehlung: Es wird empfohlen, auf diesen Stärken aufzubauen und spezifische Kommunikationsformate für einzelne Stakeholdergruppen zu entwickeln, die bislang noch nicht ausreichend adressiert wurden (z. B. Zielgruppen), um die zielgruppengerechte Aufbereitung der Evaluierungsergebnisse weiter zu verbessern. Diese Kommunikationsformate sollten sich an dem jeweiligen Interessensbedarf und den jeweiligen Anforderungen der unterschiedlichen Akteure orientieren.

Empfehlung: Um die Einbindung der Stakeholder weiter zu stärken, sollte die Sensibilisierung für partizipative Ansätze stärker als bisher in die Handreichungen und die Unterstützungsmaterialien der PEV aufgenommen werden. Bislang finden sich Ansatzpunkte hierfür eher in der Methoden-Toolbox, sollten aber prominenter in anderen Handreichungen platziert werden.

Schlussfolgerung 3.2: Die Nutzung der PEV-Ergebnisse gelingt auf der GIZ-internen Sektor-/Regionalebene vor allem vorhabenübergreifend bislang noch nicht ausreichend. Bisher werden die Ergebnisse kaum für die Fachverbundsarbeit im Sektor und nur teilweise für die Reflexion zur Verbesserung in den fachlichen und regionalen Ansätzen sowie das unternehmensweite Wissensmanagement in Wert gesetzt. Die Nutzung auf der sektoralen und regionalen Ebene schneidet von allen Nutzungsdimensionen am schlechtesten ab. Es bleibt folglich die Chance ungenutzt, die Ergebnisse des Instruments der PEV in einem breiten Kontext zu nutzen und sich über ein Vorhaben hinaus auszutauschen.

Empfehlung: Es sollte mehr Aufmerksamkeit auf die Nutzung der Evaluierungsergebnisse auf der sektoralen und regionalen Ebene gelegt werden. Hierbei sollte geprüft werden, inwieweit sich bereits vorhandene Formate (z. B. Fachverbunds- und Lernveranstaltungen) eignen oder inwieweit neue Formate des regionalen und sektoralen Austausches zu Evaluierungsergebnissen sinnvoll sind.

Empfehlung: Es wird empfohlen, dass die annotierte Berichtsgliederung der PEV-Berichte bei den zentralen Empfehlungen unter dem Aspekt „Lessons Learned für die GIZ als Gesamtorganisation“ zentrale regionale und sektorale Erkenntnisse festhält. Diese könnten anschließend zentral ausgewertet und in den Austauschformaten diskutiert werden.

Schlussfolgerung 3.3: Die Interessen der **Partnerorganisationen** sollten konsequent in Vorbereitung, Durchführung und Nachbereitung verankert werden. Die Meta-Evaluierung zeigt, dass Interessen von Partnern in der Konzeption berücksichtigt und die relevanten Stakeholder weitgehend in die Vorbereitung einbezogen werden. Zugleich setzen sich die PEV-Berichte bei den Rahmenbedingungen mit den Kontextfaktoren im Partnerland auseinander. Diese Aspekte werden zugleich noch nicht ausreichend bei den Analysen der OECD-DAC Kriterien in Wert gesetzt. Trotz dessen werden die Ergebnisse der PEV wiederum mit den Partnern geteilt.

Empfehlung: Es wird empfohlen, die vorhandenen Stärken (Partnerdialog) weiter in Wert zu setzen und, um die Relevanz der Ergebnisse für Partnerländer und Partnerorganisationen zu erhöhen, die Kontextfaktoren in den Analysen stärker zu berücksichtigen. Es sollte keinen Bruch in der Kette der Partnerorientierung zwischen Vorbereitung, Durchführung und Nachbereitung einer PEV geben.

8 Literaturverzeichnis

DeGEval (2008). Standards für Evaluation. 4. unveränderte Auflage. Mainz.

GIZ (2015 a). Handreichung zum methodischen Vorgehen bei der Projektevaluierung und Prüfung von TZ-Folgemaßnahmen im Auftrag des BMZ. Anlage 2 zum Leitfaden zur Vorbereitung von TZ-Maßnahmen im Auftrag des BMZ. Bonn und Eschborn.

GIZ (2015 b). Fahrplan Projektevaluierung.

GIZ (2015 c). Anleitung für die Nutzung der Checkliste für die Bewertung der OECD/DAC-Kriterien.

GIZ (2015 d). Annotierte Berichtsgliederung / Projektevaluierungsbericht.

GIZ (2015 e). Annotierte Berichtsgliederung / Projektevaluierung Kurzbericht.

GIZ (2015 f). Annotierte Berichtsgliederung / Managementanlage zum Projektevaluierungsbericht (intern).

GIZ (2015 g). Kommentierungsblatt für den Qualitätscheck von Projektevaluierungen.

GIZ (2015 h). Methoden-Toolbox. Bonn und Eschborn.

GIZ (2015 i). Das Wirkungsmodell der GIZ. Eine Arbeitshilfe. Bonn und Eschborn.

GIZ (2014). Indikatoren. Eine Arbeitshilfe. Bonn und Eschborn.

GIZ (2013). Policy für Monitoring und Evaluierung der GIZ. Bonn und Eschborn.

GIZ (o.D. a). Terms of Reference für Projektevaluierung und Prüfung einer Folgemaßnahme.

GIZ (o.D. b). Selbsteinschätzung nach Capacity WORKS.

GIZ (o.D. c). Berichtsgliederung / Projektevaluierungsbericht.

GIZ (o.D. d). Berichtsgliederung / Projektevaluierung Kurzbericht.

GIZ (o.D. e). Berichtsgliederung / Managementanlage zum Projektevaluierungsbericht (intern).

JCSEE (2000). Handbuch der Evaluationsstandards. Die Standards des „Joint Committee on Standards for Educational Evaluation. 2. Auflage. Opladen: Leske + Budrich.

OECD-DAC (2010). DAC-Reihe Leitlinien und Grundsatztexte. Qualitätsstandards für die Entwicklungs evaluierung.

OECD-DAC (1991). Principles for the Evaluation of Development Assistance.

Schäferhoff, M., Schrade, C., Corhs, T. (SEEK Development), Raetzell, L., Krämer, M. (Rambøll Management Consulting) (2013). GIZ Review Gesundheit: Hauptbericht. Querschnittauswertung mit Metaevaluierung, Effizienzanalyse und Evaluierungssynthese.

Stern, T., Scheller, O., Freimann, I. (2015). Externe Qualitätskontrolle der GIZ. Ergebnisbericht 2015.

Widmer, T. (1996). Meta-Evaluation. Kriterien zur Bewertung von Evaluationen. Bern, Stuttgart, Wien: Verlag Paul Haupt.

9 Anlagen

- Überblick über die Ergebnisse in den DeGEval-Evaluierungsstandards
- Ergebnisse der Online-Befragung
- Fragebogen der Online-Befragung
- Analyseraster

9.1 Überblick über die Ergebnisse in den DeGEval-Evaluierungsstandards

Ergebnisse im Evaluierungsstandard Genauigkeit

Für weitere Informationen zur Struktur des Analyserasters und zum Bewertungssystem siehe auch Kapitel 2.1 und Abbildung 3 in Kapitel 2.2.

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		Nein		Ja		Gesamtsumme		M
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	
G 1 Beschreibung des Evaluationsgegenstandes	Der Evaluationsgegenstand soll klar und genau beschrieben und dokumentiert werden, so dass er eindeutig identifiziert werden kann.							66%
G 1.1 Der Evaluierungsgegenstand wird genau beschrieben.	Der Indikator ist dann erfüllt, wenn die Evaluierung auf einer genauen Definition und Abgrenzung des Evaluierungsgegenstands basiert. Dies ist dann gegeben, wenn mindestens vier der folgenden sechs Aspekte beschrieben sind: zeitliche Abgrenzung (Phase), aufgewendete Mittel, regionale Abgrenzung, inhaltliche Abgrenzung (Mehrebenenansatz, Systemgrenze, CD Ebenen). (Textanalyse der PEV-Berichte)	22	31%	48	69%	70	0	69%
Checklist:	zeitliche Abgrenzung (Phase)	13	19%	57	81%	70	0	
	aufgewendete Mittel	53	76%	17	24%	70	0	
	regionale Abgrenzung	17	24%	53	76%	70	0	
	inhaltliche Abgrenzung (Mehrebenenansatz)	19	27%	51	73%	70	0	
	inhaltliche Abgrenzung (Systemgrenze)	16	23%	54	77%	70	0	
	inhaltliche Abgrenzung (CD Ebenen)	33	47%	37	53%	70	0	
G 1.2 Das Wirkungsmodell wird dargestellt.	Der Indikator ist dann erfüllt, wenn die Wirkungslogik der Maßnahme des aktuellen Moduls graphisch dargestellt oder textlich eindeutig beschrieben wird, so dass die Systemgrenzen des Vorhabens deutlich werden. Falls für die evaluierte Maßnahme kein Wirkungsmodell vorliegt, muss dies nachträglich (graphisch oder textlich) dargestellt werden. (Textanalyse der PEV-Berichte)	8	11%	62	89%	70	0	89%
G 1.3 Die relevanten Wirkungshypothesen werden dargestellt.	Der Indikator ist dann erfüllt, wenn die Wirkungslogik der Maßnahme anhand eines Narrativs dargestellt wird. Die Wirkungshypothesen müssen begründete Vermutungen über Zusammenhänge innerhalb des Wirkungsmodells darstellen. Es muss deutlich werden, was unter welchen Umständen und warum erwartet werden kann. Jede Hypothese muss zudem empirisch überprüfbar und falsifizierbar sein. Als Mindeststandard gilt, dass zumindest die Einordnung des Modulziels auf der Outcome Ebene und des Programmziels (falls vorhanden) auf der Impact Ebene und die entsprechenden Wirkungsannahmen (kausale Annahmen bei Outcome und zumindest plausible Annahmen bei Impact) auf diese Zielsetzungen hin deutlich werden sollen. (Textanalyse der PEV-Berichte)	41	59%	29	41%	70	0	41%

Bewertungskriterien / Indikatoren / Deskriptoren	Definition	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		Nein		Ja		Gesamtsumme		M
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	
G 2 Rahmenbedingungen	Der Kontext des Evaluationsgegenstandes soll ausreichend detailliert untersucht und analysiert werden.							74%
G 2.1 Der Evaluierungsgegenstand wird im Politikkontext des Partnerlandes verortet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsbericht den Politikkontext beschreibt: Beschreibung von Strategien, Leitlinien oder Zielen des Partnerlandes. Kenntnisse des Politikkontexts sind erforderlich, um insbesondere die Gelingensbedingungen und die Wechselwirkungen der Maßnahme realistisch einzuschätzen. (Textanalyse der PEV-Berichte)	11	16%	59	84%	70	0	84%
G 2.2 Der Evaluierungsgegenstand wird im Entwicklungskontext des Sektors im Partnerland verortet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsbericht den Entwicklungskontext im Sektor beschreibt: sozioökonomische, politische oder kulturelle Rahmenbedingungen im Sektor. Kenntnisse des Entwicklungskontexts sind erforderlich, um insbesondere die Gelingensbedingungen und die Wechselwirkungen der Maßnahme realistisch einzuschätzen. (Textanalyse der PEV-Berichte)	9	16%	47	84%	56	14	84%
G 2.3 Der Evaluierungsgegenstand wird in der Träger- und Partnerstruktur verortet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsbericht die Träger- und der Partnerstruktur beschreibt: Kenntnisse der Träger- und Partnerstruktur gehören zum institutionellen Kontexts einer Maßnahme und sind erforderlich, um die Gelingensbedingungen und die Wechselwirkungen der Maßnahme realistisch einzuschätzen. (Textanalyse der PEV-Berichte)	21	30%	49	70%	70	0	70%
G 2.4 In der Bewertung und in den Schlussfolgerungen wird ein Rückbezug auf die Kontextanalyse vorgenommen.	Der Indikator ist dann erfüllt, wenn deutlich wird, wie der untersuchte Kontext die Bewertung der Evaluierung beeinflusst. Der Einfluss der Kontextfaktoren auf die Ergebnisse der Evaluierung wird transparent gemacht. Dies hilft auch, eine mögliche Übertragbarkeit der Evaluierungsergebnisse auf andere Maßnahmen einschätzen zu können. (Textanalyse der PEV-Berichte)	28	40%	42	60%	70	0	60%
G 3 Beschreibung von Zwecken und Vorgehen	Gegenstand, Zwecke, Fragestellungen und Vorgehen der Evaluation, einschließlich der angewandten Methoden, sollen genau dokumentiert und beschrieben werden, so dass sie identifiziert und eingeschätzt werden können.							64%
G 3.1 Anlass, Zweck und beabsichtigte Verwendung der Evaluierung werden transparent beschrieben.	Der Indikator ist dann erfüllt, wenn deutlich wird, was der Anlass und Zweck (Zwischen- oder Schlussequalierung) der Evaluierung sind und wer bzw. wie die Ergebnisse verwendet soll/en (Nutzer/ Nutzung). (Textanalyse der PEV-Berichte)	42	60%	28	40%	70	0	40%
G 3.2 Die spezifischen Ziele der Evaluierung werden deutlich.	Der Indikator ist dann erfüllt, wenn transparent dargestellt ist, was mit der Evaluierung erreicht werden soll (bspw. Überprüfung von Relevanz, Zielerreichung, Nachhaltigkeit, Handlungsempfehlungen für Folgemodul, Übertragbarkeit auf andere Kontexte, etc.). Die spezifischen Evaluierungsziele müssen aus den TOR abgeleitet sein. (Textanalyse der PEV-Berichte)	48	69%	22	31%	70	0	31%
G 3.3 Die TOR einschließlich der Evaluierungsfragen der Prüfmission befinden sich im Anhang des Prüfberichts.	Der Indikator ist dann erfüllt, wenn die TOR mit den Evaluierungsfragen ein Anhang des Evaluierungsberichts sind. So kann beurteilt werden, inwieweit sich die Prüfmission ausreichend mit den Evaluierungsfragen auseinander gesetzt hat. (Textanalyse der PEV-Berichte)	3	4%	67	96%	70	0	96%
G 3.4 Anhand der TOR lassen sich Aussagen zum Zweck der Evaluierung ableiten.	Der Indikator ist dann erfüllt, wenn die TOR den Anlass und Zweck der Evaluierung klar definieren. (Textanalyse der PEV-Berichte)	9	13%	61	87%	70	0	87%

Bewertungskriterien / Indikatoren / Deskriptoren	Definition	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		Nein		Ja		Gesamtsumme		M
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	
G 4 Angabe von Informationsquellen	Die im Rahmen einer Evaluation genutzten Informationsquellen sollen hinreichend genau dokumentiert werden, damit die Verlässlichkeit und Angemessenheit der Informationen eingeschätzt werden kann.							45%
G 4.1 Der Bericht umfasst eine vollständige Liste der Gesprächspartner/innen und sonstigen Informationsquellen.	Der Indikator ist dann erfüllt, wenn die verwendeten Informationsquellen beschrieben werden, so dass die Aussagen im Bericht evidenzbasiert unterlegt sind. Falls Anonymität und das Recht auf Vertraulichkeit der einzelnen Gesprächspartner/innen die Darstellung der Informationsquellen einschränken, wird entsprechend darauf hingewiesen. (Textanalyse der PEV-Berichte)	28	40%	42	60%	70	0	60%
G 4.2 Der Bericht stellt transparent dar, welcher Systematik der Dokumentenauswertung zugrunde lag.	Der Indikator ist dann erfüllt, wenn die Auswertung der verwendeten Dokumente beschrieben werden. Hierdurch wird deutlich, welche Dokumente warum ausgewertet wurden und welches System der Dokumentenauswertung zugrunde lag. (Textanalyse der PEV-Berichte)	59	84%	11	16%	70	0	16%
G 4.3 Die Auswahl von Gesprächspartner bzw. die Ziehung von Stichproben geschieht systematisch.	Der Indikator ist dann erfüllt, wenn die Auswahl von Gesprächspartner/-innen (bzw. Stichproben bei quantitativen Analysen) nach klar definierten und transparenten Kriterien geschieht (z.B. Reflektion der Erwartungen an die Aussagefähigkeit der TN, Zusammensetzung der TN, Rolle der Prüfmission bei der Auswahl der TN), und die Auswirkungen der Auswahl auf die Repräsentativität der Ergebnisse erläutert wird. Da im Rahmen einer Prüfmission nicht jeder befragt werden kann, der zur Grundgesamtheit gehört, muss eine begründete Auswahl aus der Grundgesamtheit vorgenommen werden. (Textanalyse der PEV-Berichte)	31	44%	39	56%	70	0	56%
G 4.4 Aufbauend auf die Bewertung der Verlässlichkeit der Daten des WOM, werden diese entweder verwendet oder ausgeschlossen.	Der Indikator ist dann erfüllt, wenn zunächst eine kritische Bewertung der Verlässlichkeit der Daten des WOM erfolgt und die Daten des WOM dann, ausgehend von der Bewertung, entweder verwendet oder, falls nicht oder nur eingeschränkt auf die Daten zurück gegriffen wird, begründet ausgeschlossen werden. (Textanalyse der PEV-Berichte)	20	29%	50	71%	70	0	71%
G 4.5 Aufbauend auf die Bewertung der Verlässlichkeit der Baseline Daten, werden diese entweder verwendet oder ausgeschlossen.	Der Indikator ist dann erfüllt, wenn zunächst eine kritische Bewertung der Verlässlichkeit der Baseline Daten erfolgt und die Baseline Daten dann, ausgehend von der Bewertung, entweder verwendet oder, falls nicht oder nur eingeschränkt auf die Daten zurück gegriffen wird, begründet ausgeschlossen werden. (Textanalyse der PEV-Berichte)	45	64%	25	36%	70	0	36%
G 4.6 Aufbauend auf die Bewertung der Verlässlichkeit der Partnerdaten, werden diese entweder verwendet oder ausgeschlossen.	Der Indikator ist dann erfüllt, wenn zunächst eine kritische Bewertung der Verlässlichkeit der Partnerdaten erfolgt und die Partnerdaten dann, ausgehend von der Bewertung, entweder verwendet oder, falls nicht oder nur eingeschränkt auf die Daten zurück gegriffen wird, begründet ausgeschlossen werden. (Textanalyse der PEV-Berichte)	39	70%	17	30%	56	14	30%

Bewertungskriterien / Indikatoren / Deskriptoren	Definition	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		Nein		Ja		Gesamtsumme		M
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	
G 5 Valide und reliable Informationen	Die Verfahren zur Gewinnung von Daten sollen so gewählt oder entwickelt und dann eingesetzt werden, dass die Zuverlässigkeit der gewonnenen Daten und ihre Gültigkeit bezogen auf die Beantwortung der Evaluationsfragestellungen nach fachlichen Maßstäben sichergestellt sind. Die fachlichen Maßstäbe sollen sich an den Gütekriterien quantitativer und qualitativer Sozialforschung orientieren.							56%
G 5.1 Es wird eine Datentriangulation durchgeführt.	Der Indikator ist dann erfüllt, wenn Daten / Informationen zu einem gleichen Sachverhalt durch die Einbeziehung verschiedener Akteure erhoben werden. (Textanalyse der PEV-Berichte)	16	23%	54	77%	70	0	77%
G 5.2 Es wird eine Methodentriangulation durchgeführt.	Der Indikator ist dann erfüllt, wenn Informationen zu einem gleichen Sachverhalt durch verschiedene Methoden der Datensammlung erhoben werden. (Textanalyse der PEV-Berichte)	45	64%	25	36%	70	0	36%
G 6 Systematische Fehlerprüfung	Die in einer Evaluation gesammelten, aufbereiteten, analysierten und präsentierten Informationen sollen systematisch auf Fehler geprüft werden.							71%
G 6.1 Es wird eine Forschertriangulation durchgeführt.	Der Indikator ist dann erfüllt, wenn das PEV-Team seine Erfahrungen und Interpretationen / Perspektiven trianguliert. Hinweise auf die systematische Überprüfung der gesammelten Informationen sind z.B. Synthesetreffen, die dokumentiert wurden, oder die Darstellung auf Unstimmigkeiten der Mitglieder des PEV-Teams, die entsprechend diskutiert wurden. (Textanalyse der PEV-Berichte)	30	43%	40	57%	70	0	57%
G 6.2 Es wird eine Triangulation der Ergebnisse mit den Beteiligten und Betroffenen durchgeführt.	Der Indikator ist dann erfüllt, wenn eine Triangulation der Ergebnisse mit den Beteiligten und Betroffenen erfolgt. So soll eine Fehlerüberprüfung der Analyseergebnisse erfolgen. Hinweise auf die systematische Überprüfung der gesammelten Informationen sind z.B. gemeinsame Syntheseworkshops oder die Kommentierung des Berichts durch Beteiligte und Betroffene. (Textanalyse der PEV-Berichte)	11	16%	59	84%	70	0	84%

Bewertungskriterien / Indikatoren / Deskriptoren	Definition	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		Nein		Ja		Gesamtsumme		
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	M
G 7 (A) Analyse qualitativer und quantitativer Informationen	Qualitative und quantitative Informationen einer Evaluation sollen nach fachlichen Maßstäben angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation effektiv beantwortet werden können.							58%
G 7.1 (A) Die methodische Vorgehensweise beantwortet die Frage nach der Zuordnungs- und/ oder Beitragsanalyse.	Der Indikator ist dann erfüllt, wenn die methodische Vorgehensweise mit Hinweis auf die Beantwortung von Evaluierungsfragen, Bedingungen der Untersuchung und/ oder Besonderheiten des Evaluierungsgegenstandes begründet dargestellt wird. So kann bspw. eine Attributionsanalyse mit dem Hinweis auf den Evaluierungsgegenstand begründet ausgeschlossen werden. (Textanalyse der PEV-Berichte)	57	81%	13	19%	70	0	19%
Checklist:	keine begründete Vorgehensweise	13	19%	57	81%	70	0	
	experimentelles Design	70	100%	0	0%	70	0	
	quasi-experimentelles Design	70	100%	0	0%	70	0	
	Ansatz einer Kontributionsanalyse	61	87%	9	13%	70	0	
	theoriebasierter Ansatz	66	94%	4	6%	70	0	
G 7.2 (A) Die Evaluierungsmethoden werden dargestellt.	Der Indikator ist dann erfüllt, wenn die verwendeten Methoden beschrieben werden, so dass die Aussagen evidenzbasiert unterlegt sind. (Textanalyse der PEV-Berichte)	13	19%	57	81%	70	0	81%
G 7.3 (A) Es gibt eine Methodenvielfalt.	Der Indikator ist dann erfüllt, wenn mehr als eine Methode verwendet wird. (Textanalyse der PEV-Berichte)	1	1%	69	99%	70	0	99%
Checklist:	Dokumentaueswertung	1	1%	69	99%	70	0	
	Fokusgruppen	51	73%	19	27%	70	0	
	Interviews	0	0%	70	100%	70	0	
	Befragung	66	94%	4	6%	70	0	
	eigene Auswertung der Monitoringdaten durch die Prüfmission	53	76%	17	24%	70	0	
G 7.4 (A) Die Vor- und Nachteile der gewählten Methoden werden dargestellt.	Der Indikator ist dann erfüllt, wenn die Methodenauswahl nachvollziehbar ist. Diese Nachvollziehbarkeit spiegelt sich in einer Klärung darüber, welche Vorteile die ausgesuchte Methode hat und welche möglichen Nachteile vorhanden sind und wie damit umgegangen werden soll, wider. (Textanalyse der PEV-Berichte)	48	69%	22	31%	70	0	31%

Bewertungskriterien / Indikatoren / Deskriptoren	Definition	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		Nein		Ja		Gesamtsumme		
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	M
G 7 (B) Analyse qualitativer und quantitativer Informationen: Relevanz	Die methodische Bearbeitung der Relevanzbewertung ist angemessen.							65%
G 7.5 (B) Die Nachvollziehbarkeit der zugrunde gelegten Rahmenbedingungen und Kernprobleme der Maßnahme ist gegeben.	Der Indikator ist dann erfüllt, wenn bei der Bearbeitung der Relevanz die Rahmenbedingungen und Kernbedarfe, auf die die Maßnahme beruht, überprüft wird. (Textanalyse der PEV-Berichte)	22	31%	48	69%	70	0	69%
G 7.6 (B) Die Mehrdimensionalität der Relevanz wird analysiert.	Der Indikator ist dann erfüllt, wenn bei der Bearbeitung der Relevanz entlang der Kernbedarfe von Zielgruppe und Partner sowie der Ziele des Auftraggebers analysiert wird. (Textanalyse der PEV-Berichte)	8	11%	62	89%	70	0	89%
G 7.7 (B) Die strategische Ausrichtung der Maßnahme an veränderten Rahmenbedingungen werden analysiert.	Der Indikator ist dann erfüllt, wenn die strategische Ausrichtung einer Maßnahme an Veränderungen in den Rahmenbedingungen analysiert werden. (Textanalyse der PEV-Berichte)	44	63%	26	37%	70	0	37%
G 7 (C) Analyse qualitativer und quantitativer Informationen: Effektivität	Die methodische Bearbeitung der Effektivitätsbewertung ist angemessen.							56%
G 7.8 (C) Die Kausalität zwischen Maßnahmen und Wirkungen wird differenziert analysiert und eingeschätzt.	Der Indikator ist dann erfüllt, wenn beobachtete Zusammenhänge zwischen Maßnahmen und Wirkungen nicht ohne Weiteres als ursächliche Zusammenhänge interpretiert werden. Der Einfluss von Drittvariablen muss ausgeschlossen werden. Korrelationen dürfen nicht mit Kausalitäten gleichgesetzt werden. (Textanalyse der PEV-Berichte)	41	59%	29	41%	70	0	41%
G 7.9 (C) Die Zielerreichung wird anhand von Modulindikatoren bewertet.	Der Indikator ist dann erfüllt, wenn deutlich wird, welche Indikatoren zur Bewertung der Zielerreichung als Grundlage genommen wurden. Wurden während der Projektplanung für einzelne Aspekte von Effektivität keine Indikatoren definiert, müssen diese Indikatoren nachträglich gebildet werden. (Textanalyse der PEV-Berichte)	6	9%	64	91%	70	0	91%
G 7.10 (C) Die verwendeten Indikatoren zur Messung und Beurteilung der Zielerreichung haben eine hohe Qualität.	Der Indikator ist dann erfüllt, wenn ersichtlich ist, dass sich die Prüfer mit der Qualität der Indikatoren auseinandergesetzt haben (zum Beispiel durch die Nutzung von SMART oder CREAM-Kriterien). Wird eine geringe Qualität festgestellt, müssen neue bzw. veränderte Indikatoren für die Evaluation genutzt werden. (Textanalyse der PEV-Berichte)	45	64%	25	36%	70	0	36%

Bewertungskriterien / Indikatoren / Deskriptoren	Definition	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		Nein		Ja		Gesamtsumme		
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	M
G 7 (D) Analyse qualitativer und quantitativer Informationen: Effizienz	Die methodische Bearbeitung der Effizienzbewertung ist angemessen.							28%
G 7.11 (D) Die verschiedenen Ebenen der Effizienz einer Maßnahme werden analysiert.	Der Indikator ist dann erfüllt, wenn im Evaluierungsbericht dargestellt ist, welche Ebenen der Effizienzmessung analysiert werden und die Auswahl begründet wird. Mögliche Ebenen sind: die Implementierungseffizienz (Analyse der Strukturen und Prozesse der Vorhabensumsetzung), der Produktionseffizienz (Verhältnis Input zu Output) und der Allokationseffizienz (Verhältnis von Input zu Outcome). (Textanalyse der PEV-Berichte)	31	44%	39	56%	70	0	56%
Checklist:	Implementierungseffizienz	31	44%	39	56%	70	0	
	Produktionseffizienz,	63	90%	7	10%	70	0	
	Allokationseffizienz	68	97%	2	3%	70	0	
G 7.12 (D) Die Auswahl von Methoden und Verfahren der Effizienzmessung wird begründet.	Der Indikator ist dann erfüllt, wenn im Evaluierungsbericht die Auswahl der Methoden und Verfahren zur Messung der Effizient begründet wird. (Textanalyse der PEV-Berichte)	62	89%	8	11%	70	0	11%
Checklist:	Level 2 methods	62	89%	8	11%	70	0	
	Level 1 methods	67	96%	3	4%	70	0	
	Descriptive methods	70	100%	0	0%	70	0	
G 7.13 (D) Die Bearbeitung der Effizienz ermöglicht die Identifikation von Potenzialen zur Effizienzsteigerung.	Der Indikator ist dann erfüllt, wenn die verwendeten Methoden und Verfahren implizit oder explizit Komponenten/Teile der Maßnahme mit Alternativen vergleichen, um Verbesserungspotenziale aufzuzeigen. (Textanalyse der PEV-Berichte)	58	83%	12	17%	70	0	17%

Bewertungskriterien / Indikatoren / Deskriptoren	Definition	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		Nein		Ja		Gesamtsumme		
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	M
G 7 (E) Analyse qualitativer und quantitativer Informationen: Impact	Die methodisch Bearbeitung der Impactbewertung ist angemessen.							49%
G 7.14 (E) In der Analyse und Beurteilung der übergeordneten Wirkungen wird die Attributionslücke analysiert.	Der Indikator ist dann erfüllt, wenn im Evaluierungsbericht bei der Analyse und Beurteilung der übergeordneten Wirkungen die Zuordnungs- bzw. Attributionslücke reflektiert wird. Es wird kein direkter und monokausaler Zusammenhang zwischen Maßnahme und übergeordneten Zielen hergestellt. (Textanalyse der PEV-Berichte)	23	33%	47	67%	70	0	67%
G 7.15 (E) Die Plausibilität der Hypothesen zu den intendierten langfristigen Wirkungen wird bewertet.	Der Indikator ist dann erfüllt, wenn sich die Evaluation mit der Plausibilität der Hypothesen zu übergeordneten langfristigen Wirkungen auseinandersetzt. (Textanalyse der PEV-Berichte)	40	57%	30	43%	70	0	43%
G 7.16 (E) Bewertungsmaßstäbe zur Analyse und Beurteilung des Beitrages der Maßnahme zu übergeordneten Wirkungen sind transparent dargestellt.	Der Indikator ist dann erfüllt, wenn beschreiben ist, anhand welcher Bewertungsmaßstäbe der Beitrag des Vorhabens zu übergeordneten entwicklungspolitischen Wirkungen bewertet wird. Diese Bewertungsmaßstäbe können entweder selbst vom Prüfteam entwickelt (zum Beispiel (Proxi-)Indikatoren) oder vorhandene Maßstäbe (zum Beispiel aus Teil A des PV) herangezogen werden. (Textanalyse der PEV-Berichte)	45	64%	25	36%	70	0	36%
G 7 (F) Analyse qualitativer und quantitativer Informationen: Nachhaltigkeit	Die methodisch Bearbeitung der Nachhaltigkeitsbewertung ist angemessen.							61%
G 7.17 (F) Die Grenzen der Messung der Nachhaltigkeit werden beschrieben.	Der Indikator ist dann erfüllt, wenn bei der Bearbeitung der Nachhaltigkeit beschrieben wird, inwiefern zum Zeitpunkt der Evaluierung eine ergebnisorientierte Analyse und Beurteilung der Nachhaltigkeit von Effekten überhaupt möglich ist. (Textanalyse der PEV-Berichte)	37	53%	33	47%	70	0	47%
G 7.18 (F) Die angelegten Ansätze zur Schaffung von Nachhaltigkeit werden analysiert.	Der Indikator ist dann erfüllt, wenn in der Analyse und Beurteilung der Nachhaltigkeit die angelegten Ansätze zur Schaffung von Nachhaltigkeit berücksichtigt werden. Zum Beispiel wird evaluiert, inwiefern die Maßnahme eine Exit- oder Nachhaltigkeitsstrategie besitzt, Ansätze für die Anschlussfähigkeit der Maßnahme betreibt oder Aktivitäten zur Verantwortungsübergabe (Ownership) an den Partner betreibt. Die Grenzen der Messung werden dann bei der Analyse und Beurteilung eingehalten. (Textanalyse der PEV-Berichte)	26	37%	44	63%	70	0	63%
G 7.19 (F) Es werden mindestens zwei Ebenen der Nachhaltigkeit im Rahmen einer Prognose analysiert.	Der Indikator ist dann erfüllt, wenn der Mehrdimensionalität von Nachhaltigkeit im Rahmen einer Prognose Rechnung getragen wird und mindestens Nachhaltigkeit auf zwei Ebenen betrachtet wird (finanzielle, institutionelle, personelle, technologische, soziale, ökologische Nachhaltigkeit) überprüft. (Textanalyse der PEV-Berichte)	19	27%	51	73%	70	0	73%

Bewertungskriterien / Indikatoren / Deskriptoren	Definition	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		Nein		Ja		Gesamtsumme		M
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	
G 8 Begründete Analyse und begründete Schlussfolgerung	Die in einer Evaluation gezogenen Folgerungen sollen ausdrücklich begründet werden, damit die Adressaten und Adressatinnen diese einschätzen können.							50%
G 8.1 Es wird zwischen Beschreibung, Analyse und Bewertung unterschieden.	Der Indikator ist dann erfüllt, wenn der Dreischritt Beschreibung – Analyse – Bewertung aus evaluatorischer Einschätzung zumeist eingehalten wird. Das heißt, es wird nicht direkt von der Beschreibung zur Bewertung gesprungen bzw. die Analyse und Bewertung werden nicht vermischt. Der Dreischritt soll mindestens bei zwei der folgenden drei Aspekte eingehalten werden: Wirkungsmodell, Bewertung der OECD/DAC-Kriterien, Schlussfolgerungen. (Textanalyse der PEV-Berichte)	47	67%	23	33%	70	0	33%
Checklist:	Wirkungsmodell	47	67%	23	33%	70	0	
	Bewertung der OECD/ DAC-Kriterien	44	63%	26	37%	70	0	
	Schlussfolgerungen	45	64%	25	36%	70	0	
G 8.2 Beschreibungen und Analysen werden belegt.	Der Indikator ist dann erfüllt, wenn Beschreibung und Analyse im Bericht mit relevanten Zahlen, Daten und Fakten untermauert werden. Für alle Zahlen, Daten und Fakten, werden Quellen angegeben. Demnach sind bei der Beschreibung und Analyse von Ergebnissen der Datenerhebung die jeweiligen Informationsquellen ersichtlich. Hierbei werden die Daten, die von der Evaluierungsmission erhoben wurden, als solche gekennzeichnet. (Textanalyse der PEV-Berichte)	49	70%	21	30%	70	0	30%
G 8.3 Empfehlungen werden aus der Analyse abgeleitet und sind spezifisch, realistisch und termingebunden. Sie richten sich an die wesentlichen Nutzer und ihre Umsetzung ist messbar.	Der Indikator ist dann erfüllt, wenn für alle Handlungsempfehlungen ein Ziel bzw. ein Zweck definiert worden ist. Die Empfehlungen müssen realistisch sein und sich aus den Schlussfolgerungen ableiten. Die Realisierbarkeit aller Handlungsempfehlungen muss sichergestellt sein. Z.B indem Handlungsempfehlungen in Zusammenarbeit mit denjenigen entwickelt werden, die sie umsetzen müssen. Alle Handlungsempfehlungen müssen einen Zeitraum bzw. einen Zeitpunkt für die Umsetzung benennen und die Umsetzung muss empirisch erhoben werden können. (Textanalyse der PEV-Berichte)	25	36%	45	64%	70	0	64%
G 8.4 Die positiven und/ oder negativen nicht-intendierten Wirkungen der Maßnahme werden beschrieben.	Der Indikator ist dann erfüllt, wenn sich die Evaluierung mit möglichen positiven und/ oder negativen nicht-intendierten Wirkungen befasst. Hiermit soll gewährleistet werden, dass sich die Evaluierung möglicher Hinweise auf nicht-intendierte Wirkung bewusst ist. (Textanalyse der PEV-Berichte)	19	27%	51	73%	70	0	73%
Genauigkeit insgesamt								59%

Ergebnisse im Evaluierungsstandard Durchführbarkeit (Prozessqualität I)

Für weitere Informationen zur Struktur des Analyserasters und zum Bewertungssystem siehe auch Kapitel 2.1 und Abbildung 3 in Kapitel 2.2.

Zur Übersicht wurde in der nachfolgenden Tabelle zusammengefasst, welche der zu den Indikatoren gehörenden Befragungsisems der Online-Befragung durch die AV „**eher mit nein**“ (Bewertung von 1 bis 3 Punkten auf einer Skala von 1 (trifft überhaupt nicht zu) bis 6 (trifft voll und ganz zu)) und „**eher mit ja**“ (Bewertung von 4 bis 6 Punkten) bewertet wurden. Diese Werte bilden allerdings nicht die in der Tabelle angegebenen Mittelwerte zur „**Erfüllung im Durchschnitt**“ ab. Diese Werte wurde anhand des Antwortverhaltens der AV differenzierter abgeleitet: Bewertung 6: 100 %; Bewertung 5: 80 %; Bewertung 4: 60 %; Bewertung 3: 40 %; Bewertung 2: 20 %; Bewertung 1: 0 %.

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		(eher) Nein		(eher) Ja		Gesamtsumme		
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	M
D 1 Angemessene Verfahren	Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass Belastungen des Evaluationsgegenstandes bzw. der Beteiligten und Betroffenen in einem angemessenen Verhältnis zum erwarteten Nutzen der Evaluation stehen.							65%
D 1.1 Das Verfahren der Evaluierung wird als angemessen für die Mitarbeiter/-innen des Vorhabens bezeichnet.	Der Indikator ist erfüllt, wenn laut Auftraggeber der Zeiteinsatz der Mitarbeiter/-innen des Vorhabens im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen war. (AV-Befragung)	27	36%	47	64%	74	0	58%
D 1.2 Das Verfahren der Evaluierung wird als angemessen für die Partner des Vorhabens bezeichnet.	Der Indikator ist erfüllt, wenn laut Auftraggeber der Zeiteinsatz der Politischen Träger und Durchführungsorganisationen im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen war. (AV-Befragung)	10	14%	60	86%	70	4	73%
D 1.3 Die Qualitätskontrolle der Evaluierung wird als angemessen bezeichnet.	Der Indikator ist dann erfüllt, wenn der Auftraggeber mit den in Mechanismen der Qualitätskontrolle (Backstopping und Unterstützungsangebote der STS 08, QS Check der Evaluierungsberichte) zufrieden war. (AV-Befragung)	Dieser Indikator wurde nur qualitativ erhoben.						
D 2 Diplomatisches Vorgehen	Evaluationen sollen so geplant und durchgeführt werden, dass eine möglichst hohe Akzeptanz der verschiedenen Beteiligten und Betroffenen in Bezug auf Vorgehen und Ergebnisse der Evaluation erreicht werden kann.							61%
D 2.1 Die Evaluierung verfolgt einen kooperativen Ansatz und bringt unterschiedliche Interessensgruppen zusammen.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die Evaluierung verschiedene Interessen ausgewogen berücksichtigt hat. Hierzu hat sie den Dialog zwischen den verschiedenen Stakeholdern gesucht. (AV-Befragung)	23	32%	50	68%	73	1	61%
D 3 Effizienz von Evaluation	Der Aufwand für Evaluation soll in einem angemessenen Verhältnis zum Nutzen der Evaluation stehen.							61%
D 3.1 Das Budget der Evaluierung wird als angemessen bezeichnet.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass das eingesetzte finanzielle Budget im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen war. (AV-Befragung)	18	24%	56	76%	74	0	58%
D 3.2 Das Mengengerüst der Evaluierung wird als angemessen bezeichnet.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass das Mengengerüst (Anzahl der Tage des PEV-Teams aus den TOR) im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen war. (AV-Befragung)	28	38%	46	62%	74	0	64%
Durchführbarkeit insgesamt								63%

Ergebnisse im Evaluierungsstandard Fairness (Prozessqualität II)

Für weitere Informationen zur Struktur des Analyserasters und zum Bewertungssystem siehe auch Kapitel 2.1 und Abbildung 3 in Kapitel 2.2.

Zur Übersicht wurde in der nachfolgenden Tabelle zusammengefasst, welche der zu den Indikatoren gehörenden Befragungssitems der Online-Befragung durch die AV „**eher mit nein**“ (Bewertung von 1 bis 3 Punkten auf einer Skala von 1 (trifft überhaupt nicht zu) bis 6 (trifft voll und ganz zu)) und „**eher mit ja**“ (Bewertung von 4 bis 6 Punkten) bewertet wurden. Diese Werte bilden allerdings nicht die in der Tabelle angegebenen Mittelwerte zur „**Erfüllung im Durchschnitt**“ ab. Diese Werte wurde anhand des Antwortverhaltens der AV differenzierter abgeleitet: Bewertung 6: 100 %; Bewertung 5: 80 %; Bewertung 4: 60 %; Bewertung 3: 40 %; Bewertung 2: 20 %; Bewertung 1: 0 %.

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		(eher) Nein		(eher) Ja		Gesamtsumme		M
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	
F 1 Formale Vereinbarungen	Die Pflichten des PEV-Teams (was, wie, von wem, wann getan werden soll) sollen schriftlich festgehalten werden, damit die Parteien verpflichtet sind, alle Bedingungen dieser Vereinbarung zu erfüllen oder aber diese neu auszuhandeln.							73%
F 1.1 Die angehängte Leistungsbeschreibung des Evaluierungsberichtes beschreibt die vertraglichen Rahmenbedingungen.	Der Indikator ist dann erfüllt, wenn die angehängte Leistungsbeschreibung beschreibt was, wie, von wem und wann evaluiert werden soll (siehe auch Indikator 3.3, der das warum - also die Ziele in den ToRs prüft). (Textanalyse der PEV-Berichte)	19	27%	51	73%	70	0	73%
F 2 Schutz individueller Rechte	Evaluationen sollen so geplant und durchgeführt werden, dass Sicherheit, Würde und Rechte der in eine Evaluation einbezogenen Personen geschützt werden.							82%
F 2.1 Der Evaluierungsprozess folgt ethischen Verhaltensregeln.	Der Indikator ist dann erfüllt, wenn durch den Auftraggeber bestätigt wird, dass das PEV-Team kultursensibel (Kultur, Religion) und gendersensibel vorgegangen ist. (AV-Befragung)							82%
Befragungseitems der Online-Befragung:	Kultursensibles Vorgehen	4	6%	59	94%	63	11	
	Gendersensibles Vorgehen	5	7%	65	93%	70	4	
F 3 Vollständige und faire Überprüfung	Evaluationen sollen die Stärken und die Schwächen des Evaluationsgegenstandes möglichst vollständig und fair überprüfen und darstellen, so dass die Stärken weiter ausgebaut und die Schwachpunkte behandelt werden können.							69%
F 3.1 Die Bewertung der Maßnahme spiegelt seine Stärken und Schwächen wider.	Der Indikator ist dann erfüllt, wenn die Stärken und Schwächen der Maßnahme so vollständig wie möglich erfasst und ausgewogen dargestellt sind. (Textanalyse der PEV-Berichte)	22	31%	48	69%	70	0	69%
F 4 Unparteiische Durchführung und Berichterstattung	Die Evaluation soll unterschiedliche Sichtweisen von Beteiligten und Betroffenen auf Gegenstand und Ergebnisse der Evaluation in Rechnung stellen. Berichte sollen ebenso wie der gesamte Evaluationsprozess die unparteiische Position des Evaluationsteams erkennen lassen. Bewertungen sollen fair und möglichst frei von persönlichen Gefühlen getroffen werden.							88%
F 4.1 Die Auswahl des PEV-Teams ermöglichte eine unparteiische Durchführung und Berichterstattung der Evaluierung.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass das PEV-Team unparteiisch war und damit keine Interessenskonflikte in Kontext des evaluierten Vorhabens hat. (AV-Befragung)	5	7%	66	93%	71	3	88%

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		(eher) Nein		(eher) Ja		Gesamtsumme		
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	M
F 5 Offenlegung der Ergebnisse	Die Evaluationsergebnisse sollen allen Beteiligten und Betroffenen soweit wie möglich zugänglich gemacht werden.							69%
F 5.1 Die zentralen Evaluierungsergebnisse werden zielgruppengerecht aufbereitet.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass für die Kommunikation der Ergebnisse an die beteiligten und betroffenen Stakeholder spezifische Formate entwickelt entwickelt wurden. (AV-Befragung)	23	32%	49	68%	72	2	63%
F 5.2 Die zentralen Evaluierungsergebnisse werden mit den Beteiligten und Betroffenen geteilt.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die zentralen Ergebnisse mit den beteiligten und betroffenen Stakeoldern geteilt wurden. (AV-Befragung)							75%
Befragungsite der Online-Befragung:	Politischer Träger	5	7%	64	93%	69	5	
	Durchführungsorganisationen	6	9%	63	91%	69	5	
	Weitere Kooperationspartner	14	20%	55	80%	69	5	
	Zielgruppen	32	51%	31	49%	63	11	
Fairness insgesamt								76%

Ergebnisse im Evaluierungsstandard Nützlichkeit

Für weitere Informationen zur Struktur des Analyserasters und zum Bewertungssystem siehe auch Kapitel 2.1 und Abbildung 3 in Kapitel 2.2.

Zur Übersicht wurde in der nachfolgenden Tabelle zusammengefasst, welche der zu den Indikatoren gehörenden Befragungsisems der Online-Befragung durch die AV „**eher mit nein**“ (Bewertung von 1 bis 3 Punkten auf einer Skala von 1 (trifft überhaupt nicht zu) bis 6 (trifft voll und ganz zu)) und „**eher mit ja**“ (Bewertung von 4 bis 6 Punkten) bewertet wurden. Diese Werte bilden allerdings nicht die in der Tabelle angegebenen Mittelwerte zur „**Erfüllung im Durchschnitt**“ ab. Diese Werte wurde anhand des Antwortverhaltens der AV differenzierter abgeleitet: Bewertung 6: 100 %; Bewertung 5: 80 %; Bewertung 4: 60 %; Bewertung 3: 40 %; Bewertung 2: 20 %; Bewertung 1: 0 %.

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		(eher) Nein		(eher) Ja		Gesamtsumme		
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	M
N 1 Identifizierung der Beteiligten und Betroffenen	Die am Evaluationsgegenstand beteiligten oder von ihm betroffenen Personen bzw. Personengruppen sollen identifiziert werden, damit deren Interessen geklärt und so weit wie möglich bei der Anlage der Evaluation berücksichtigt werden können.							83%
N 1.1 Die potentiell beteiligten und betroffenen Personen wurden systematisch und frühzeitig identifiziert.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die Interessen bestehender und potentieller Stakeholder (z.B. politische Träger, DO, weitere Kooperationspartner, Zielgruppe, Auftraggener) in Bezug auf die Evaluierung systematisch und frühzeitig identifiziert wurden. (AV-Befragung)	5	7%	67	93%	72	2	81%
N 1.2 Die potentiell beteiligten und betroffenen Personen wurden, wann immer möglich, einbezogen.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die bestehenden und potentiellen Stakeholder in Bezug auf die Evaluierung, wann immer möglich, einbezogen wurden. (AV-Befragung)	4	5%	69	95%	73	1	85%
N 2 Klärung der Evaluationszwecke	Es soll deutlich bestimmt sein, welche Zwecke mit der Evaluation verfolgt werden, so dass die Beteiligten und Betroffenen Position dazu beziehen können und das Evaluationsteam einen klaren Arbeitsauftrag verfolgen kann.							83%
N 2.1 Der Zweck der Evaluierung wurde frühzeitig ausgehandelt.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die Zwecke der Evaluierung frühzeitig mit bestehenden und potentiellen Stakeholdern abgesprochen wurden. (AV-Befragung)	8	11%	65	89%	73	1	83%
N 3 Glaubwürdigkeit und Kompetenz des Evaluators / der Evaluatorin	Wer Evaluationen durchführt, soll persönlich glaubwürdig sowie methodisch und fachlich kompetent sein, damit bei den Evaluationsergebnissen ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird.							57%
N 3.1 Das PEV-Team wird als kompetent und dadurch glaubwürdig eingeschätzt.	Der Indikator ist dann erfüllt, wenn das PEV-Team laut Auftraggeber die folgenden Kompetenzen abgedeckt hat: fachliche Kompetenz (im Sektor), methodische Kompetenz (Evaluierungsmethoden), Beratungskompetenz (z.B. Capacity WORKS, Capacity Development), Prozesskompetenz (GIZ-Auftragsmanagement, Steuerung einer PEV). (AV-Befragung)							86%
Befragungssitems der Online-Befragung:	Fachliche Kompetenz (im Sektor)	2	3%	71	97%	73	1	
	Methodische Kompetenz (Evaluierungsmethoden)	3	4%	71	96%	74	0	
	Beratungskompetenz (z.B. Capacity WORKS, Capacity Development)	6	8%	67	92%	73	1	
N 3.2 Die ToR beschreiben das Profil, welches das Evaluationsteam abdecken sollte.	Prozesskompetenz (GIZ-Auftrags-management, Steuerung einer PEV)	4	5%	70	95%	74	0	29%
	Der Indikator ist dann erfüllt, wenn anhand der Beschreibung des gesuchten Profils des PEV-Teams in den ToR hervorgeht, dass sich der Auftraggeber mit der Glaubwürdigkeit des Evaluationsteams befasst hat. Hierzu sollten die folgenden Kompetenzen in dem gesuchten Profil des Evaluationsteams in den ToR abgedeckt werden: fachliche Kompetenz (im Sektor), methodische Kompetenz (Evaluierungsmethoden), Beratungskompetenz (z.B. Capacity WORKS, Capacity Development), Prozesskompetenz (GIZ-Auftragsmanagement, Steuerung einer PEV). (Textanalyse der PEV-Berichte)	50	71%	20	29%	70	0	

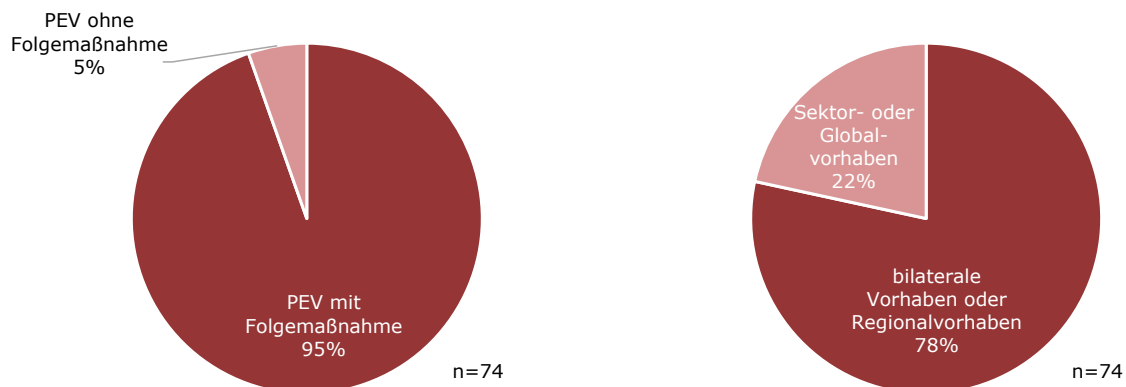
Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		(eher) Nein		(eher) Ja		Gesamtsumme		M
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	
N 4 Auswahl und Umfang der Informationen	Auswahl und Umfang der erfassten Informationen sollen die Behandlung der zu untersuchenden Fragestellungen zum Evaluationsgegenstand ermöglichen und gleichzeitig den Informationsbedarf des Auftraggebers und anderer Adressaten und Adressatinnen berücksichtigen.							84%
N 4.1 Die Evaluierung ist hinreichend informativ.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die in der Evaluierung gewonnenen Informationen die wesentlichen vereinbarten Fragestellungen der Evaluierung abdecken. (AV-Befragung)	4	5%	70	95%	74	0	84%
N 5 Transparenz von Werten	Die Perspektiven und Annahmen der Beteiligten und Betroffenen, auf denen die Evaluation und die Interpretation der Ergebnisse beruhen, sollen so beschrieben werden, dass die Grundlagen der Bewertungen klar ersichtlich sind.							71%
N 5.1 Es wird eine Forschertriangulation durchgeführt.	Der Indikator ist dann erfüllt, wenn das PEV-Team seine Erfahrungen und Interpretationen / Perspektiven trianguliert. Hinweise auf die systematische Überprüfung der gesammelten Informationen sind z.B. Synthesetreffen, die dokumentiert wurden, oder die Darstellung auf Unstimmigkeiten der Mitglieder des PEV-Teams, die entsprechend diskutiert wurden. (Textanalyse der PEV-Berichte)	30	43%	40	57%	70	0	57%
N 5.2 Es wird eine Triangulation der Ergebnisse mit den Beteiligten und Betroffenen durchgeführt.	Der Indikator ist dann erfüllt, wenn eine Triangulation der Ergebnisse mit den Beteiligten und Betroffenen erfolgt. So soll eine Fehlerüberprüfung der Analyseergebnisse erfolgen. Hinweise auf die systematische Überprüfung der gesammelten Informationen sind z.B. gemeinsame Syntheseworkshops oder die Kommentierung des Berichts durch Beteiligte und Betroffene. (Textanalyse der PEV-Berichte)	11	16%	59	84%	70	0	84%
N 6 Vollständigkeit und Klarheit der Berichterstattung	Evaluationsberichte sollen alle wesentlichen Informationen zur Verfügung stellen, leicht zu verstehen und nachvollziehbar sein.							86%
N 6.1 Die annotierte Berichtsgliederung wird angewandt.	Der Indikator ist dann erfüllt, wenn die inhaltlichen und formalen Vorgaben der annotierten Berichtsgliederung eingehalten werden. (Textanalyse der PEV-Berichte)	10	14%	60	86%	70	0	86%
N 7 Rechtzeitigkeit der Evaluation	Evaluationsvorhaben sollen so rechtzeitig begonnen und abgeschlossen werden, dass ihre Ergebnisse in anstehende Entscheidungsprozesse bzw. Verbesserungsprozesse einfließen können.							81%
N 7.1 Der Evaluierungsprozess wird rechtzeitig durchgeführt.	Der Indikator ist dann erfüllt, wenn der Evaluierungsprozess zu einem für die Nutzung der Erkenntnisse günstigen Zeitpunkt durchgeführt wurde. (AV-Befragung)	6	8%	67	92%	73	1	85%
N 7.2 Evaluationsergebnisse stehen rechtzeitig zur Verfügung.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass der Bericht zu einem für die Nutzung passenden Zeitpunkt vorlag. (AV-Befragung)	9	13%	63	88%	72	2	78%

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		(eher) Nein		(eher) Ja		Gesamtsumme		M
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	
N 8 (A) Nutzung und Nutzen der Evaluation: Nützlichkeit des Evaluierungsprozess und -berichts	Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation aufmerksam zur Kenntnis zu nehmen und ihre Ergebnisse zu nutzen.							73%
N 8.1 Der Evaluierungsprozess (Planung, Durchführung, Nachbereitung) wird als nützlich bewertet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsprozess dem Nutzenerwarten des Auftraggebers entspricht. Dies ist dann der Fall, wenn der Evaluierungsprozess einen Beitrag zum Lernen geleistet hat. (AV-Befragung)							72%
Befragungsisems der Online-Befragung:	Vorbereitung der PEV	22	32%	47	68%	69	5	
	Durchführung der PEV	8	11%	62	89%	70	4	
	Follow-Up im Nachgang der PEV	13	20%	51	80%	64	10	
N 8.2 Der Evaluierungsbericht wird als nützlich bewertet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsbericht dem Nutzenerwarten des Auftraggebers entspricht. Dies ist dann der Fall, wenn der Bericht einen Beitrag zum Lernen geleistet hat. (AV-Befragung)	13	19%	55	81%	68	6	74%
N 8 (B) Nutzung und Nutzen der Evaluation: Nutzung der Evaluierungsergebnisse	Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation aufmerksam zur Kenntnis zu nehmen und ihre Ergebnisse zu nutzen.							65%
N 8.3 Die Ergebnisse der Evaluierung werden genutzt für die GIZ-interne Projekt-/ Programmebene.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für den Austausch zwischen den Mitarbeiter/-innen, den Austausch im Landesbüro, die Steuerung des laufenden Vorhabens, die Planung des Angebots und die Steuerung der Folgephase. (AV-Befragung)							83%
Befragungsisems der Online-Befragung:	Austausch zwischen den Mitarbeiter/-innen	2	3%	69	97%	71	3	
	Austausch im Landesbüro	21	32%	45	68%	66	8	
	Steuerung des laufenden Vorhabens	8	11%	64	89%	72	2	
	Planung des Angebots	2	3%	68	97%	70	4	
	Steuerung der Folgephase	5	8%	58	92%	63	11	
N 8.4 Die Ergebnisse der Evaluierung werden genutzt für die GIZ-interne Sektor-/ Regionalebene.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für die Fachverbandsarbeit im Sektor, die Reflexion zu Verbesserungen in den fachlichen und regionalen Ansätzen und das unternehmensweite Wissensmanagement. (AV-Befragung)							41%
Befragungsisems der Online-Befragung:	Fachverbandsarbeit im Sektor	43	78%	12	22%	55	19	
	Reflexion zu Verbesserungen in den fachlichen und regionalen Ansätzen	29	48%	32	52%	61	13	
	Unternehmensweites Wissensmanagement	28	50%	28	50%	56	18	

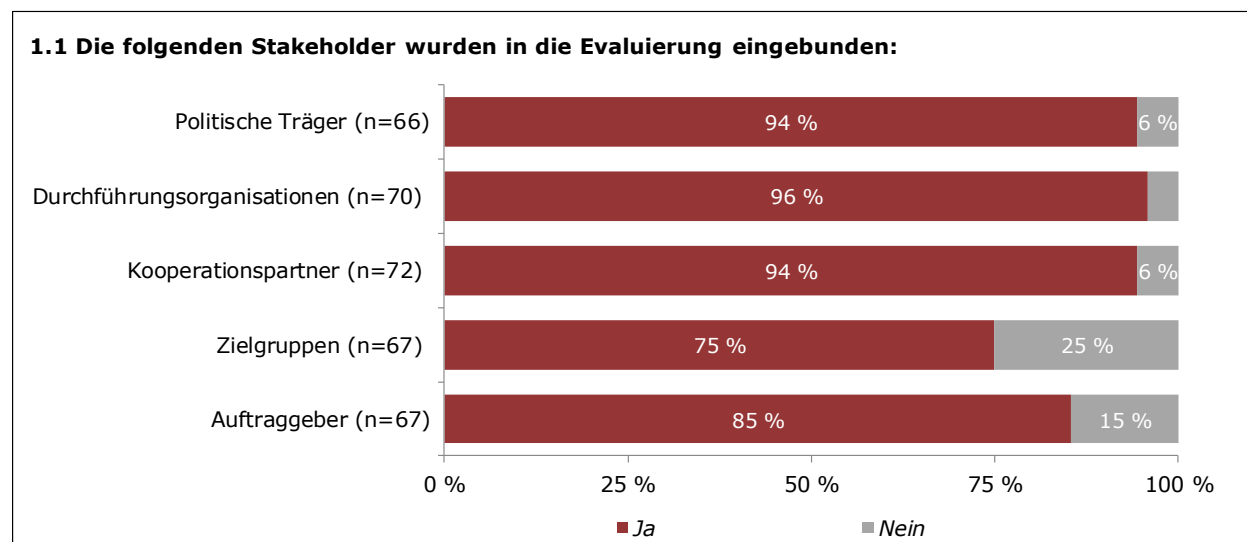
Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Indikatoren / Deskriptoren						Erfüllung im Durchschnitt
		(eher) Nein		(eher) Ja		Gesamtsumme		
		Anzahl	%	Anzahl	%	Anzahl	Fehlend	M
N 8.5 Die Ergebnisse der Evaluierung werden genutzt für die Ebene des Kooperationssystems.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für die direkte Beratung der Counterparts in der Partnerorganisation (personelle Ebene), die Organisationsentwicklung von Partnerorganisationen (organisationale Ebene), die Gestaltung politischer Reformprozesse (Mikro-, Meso-, Makroebene) und die Verbesserung der Erreichung der Zielgruppen. (AV-Befragung)							66%
Befragungsisems der Online-Befragung:	Direkte Beratung der Counterparts in der Partnerorganisation	15	24%	48	76%	63	11	
	Organisationsentwicklung von Partnerorganisationen	21	34%	40	66%	61	13	
	Gestaltung politischer Reformprozesse (Mikro-, Meso-, Makroebene)	15	24%	48	76%	63	11	
	Verbesserung der Erreichung der Zielgruppen	19	31%	43	69%	62	12	
N 8.6 Die Ergebnisse der Evaluierung werden genutzt für die Ebene des erweiterten Kooperationssystems.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für Fach- und Strategiediskussionen mit dem BMZ, Fach- und Strategiediskussionen mit anderen Auftraggebern (außer BMZ) und die Geberabstimmung im Sektor. (AV-Befragung)							59%
	Fach- und Strategiediskussionen mit dem BMZ	15	23%	51	77%	66	8	
	Fach- und Strategiediskussionen mit anderen Auftraggebern (außer BMZ)	27	60%	18	40%	45	29	
	Geberabstimmung im Sektor	27	48%	29	52%	56	18	
N 8.7 Die Ergebnisse der Evaluierung werden genutzt für Sonstiges.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für den Austausch mit weiteren Akteuren. (AV-Befragung)	27	43%	36	57%	63	11	56%
N 8.8 Die Ergebnisse der Evaluierung werden genutzt für die Verbesserung der Folgemaßnahme	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die Ergebnisse der Evaluierung einen positiven Einfluss haben auf die Relevanz, Effektivität, Effizienz, Nachhaltigkeit und den Impact der Folgemaßnahme. (AV-Befragung)							77%
Befragungsisems der Online-Befragung:	Relevanz der Folgemaßnahme	5	8%	61	92%	66	8	
	Effektivität der Folgemaßnahme	8	12%	58	88%	66	8	
	Effizienz der Folgemaßnahme	14	21%	53	79%	67	7	
	Beiträge zu den langfristigen Wirkungen	15	22%	52	78%	67	7	
	Nachhaltigkeit der Folgemaßnahme	12	18%	53	82%	65	9	
Nützlichkeit insgesamt								74%

9.2 Ergebnisse der Online-Befragung

Einstieg



1. Beteiligung von Stakeholdern



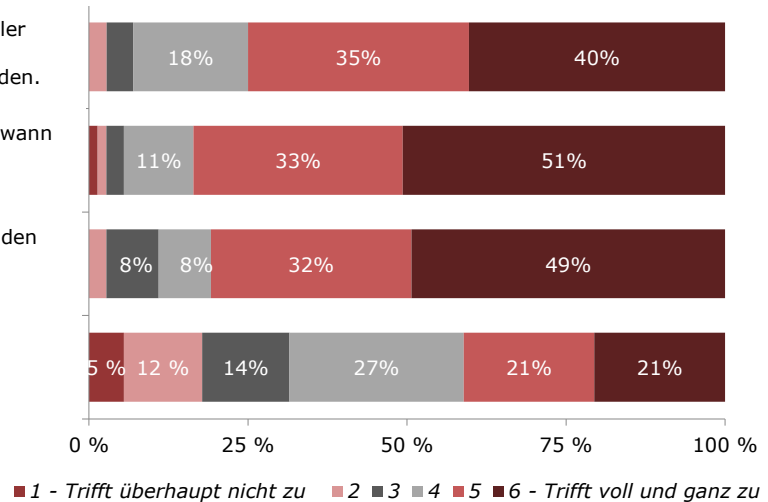
1.2 Inwiefern wurde darauf geachtet, dass die Interessen dieser Stakeholder in der Evaluierung berücksichtigt wurden? Im PEV-Prozess wurde darauf geachtet, dass...

... die Interessen bestehender und potentieller Stakeholder in Bezug auf die Evaluierung systematisch und frühzeitig identifiziert wurden.
(n=72)

... bestehende und potentielle Stakeholder, wann immer möglich, einbezogen wurden.
(n=73)

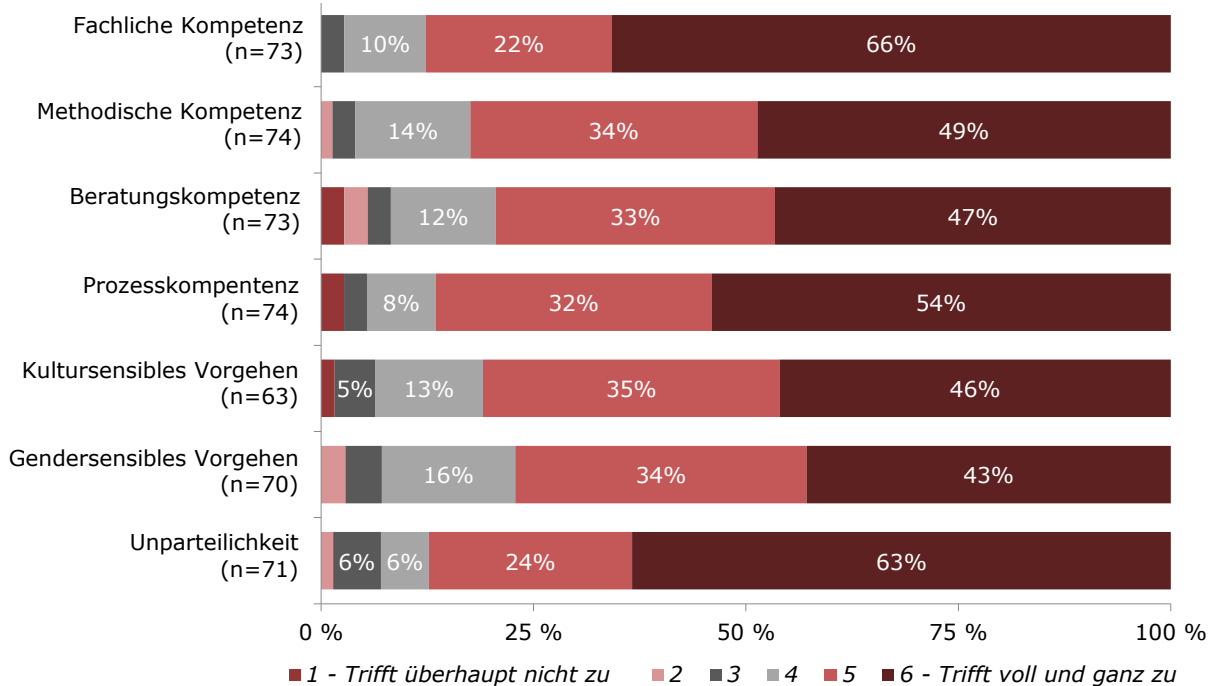
... die Zwecke der Evaluierung mit bestehenden und potentiellen Stakeholdern frühzeitig im Evaluierungsprozess abgesprochen wurden.
(n=73)

... der Dialog zwischen den verschiedenen Stakeholder gefördert wurde.
(n=73)



2. PEV-Team

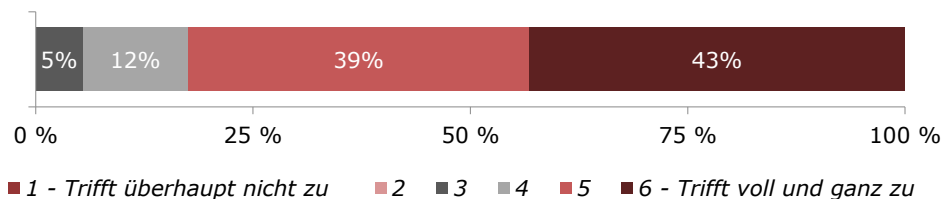
2.1 Das PEV-Team hat die folgenden Kompetenzen abgedeckt:



3. Informationsgehalt und Nützlichkeit

3.1 Die in der Evaluation gewonnenen Informationen...

... decken die wesentlichen vereinbarten Fragestellungen der Evaluierung ab. (n=74)

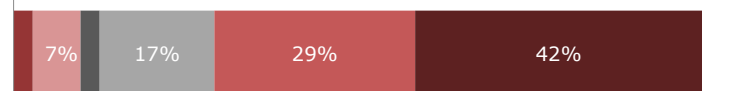


3.2 Grundlagen für die Nutzung der Evaluierungsergebnisse:

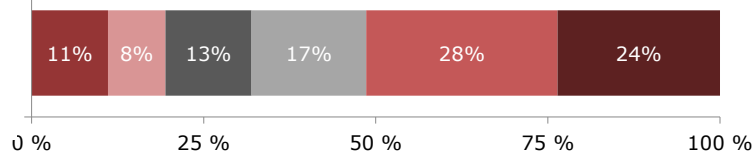
Der Evaluierungsprozess wurde zu einem für die Nutzung der Erkenntnisse günstigen Zeitpunkt durchgeführt. (n=73)



Der Evaluierungsbericht lag zu einem für die Nutzung passenden Zeitpunkt vor. (n=72)



Für die Kommunikation der Ergebnisse an die beteiligten und betroffenen Stakeholder wurden spezifische Formate entwickelt. (n=72)



3.3 Die Evaluierungsergebnisse wurden mit den beteiligten und betroffenen Stakeholdern geteilt:

Politischer Träger (n=69)



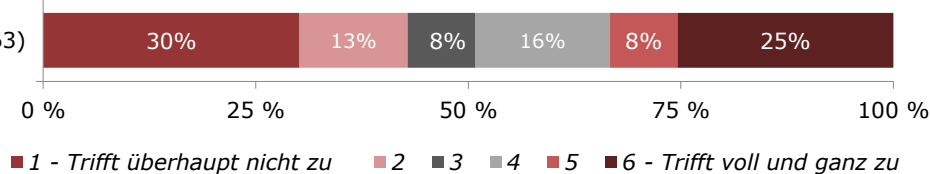
Durchführungsorganisation (n=69)



Weitere Kooperationspartner (n=69)

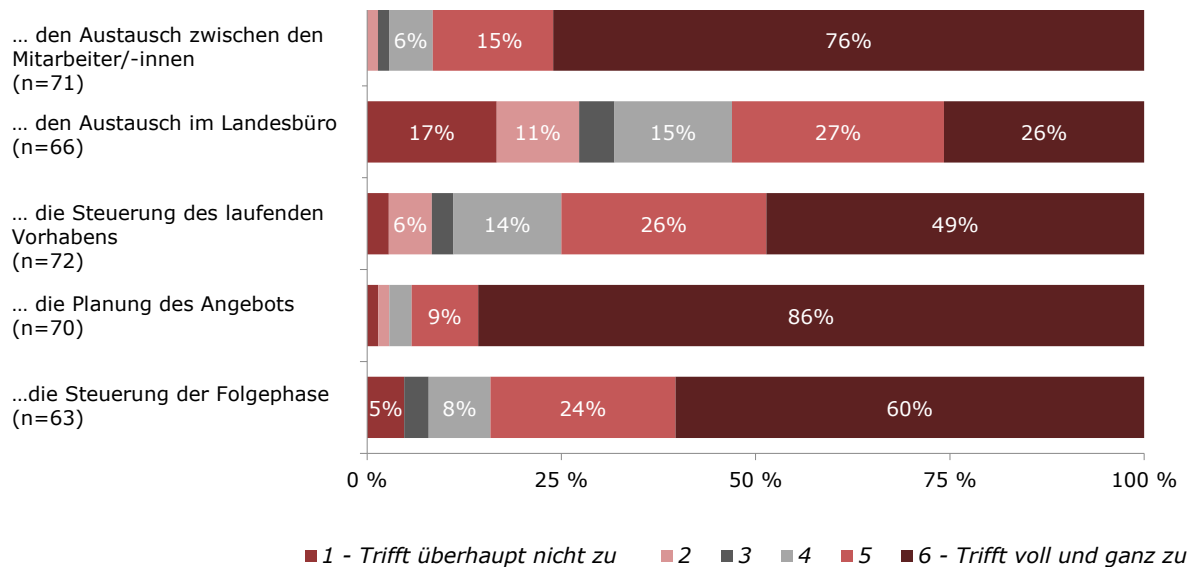


Zielgruppen (n=63)

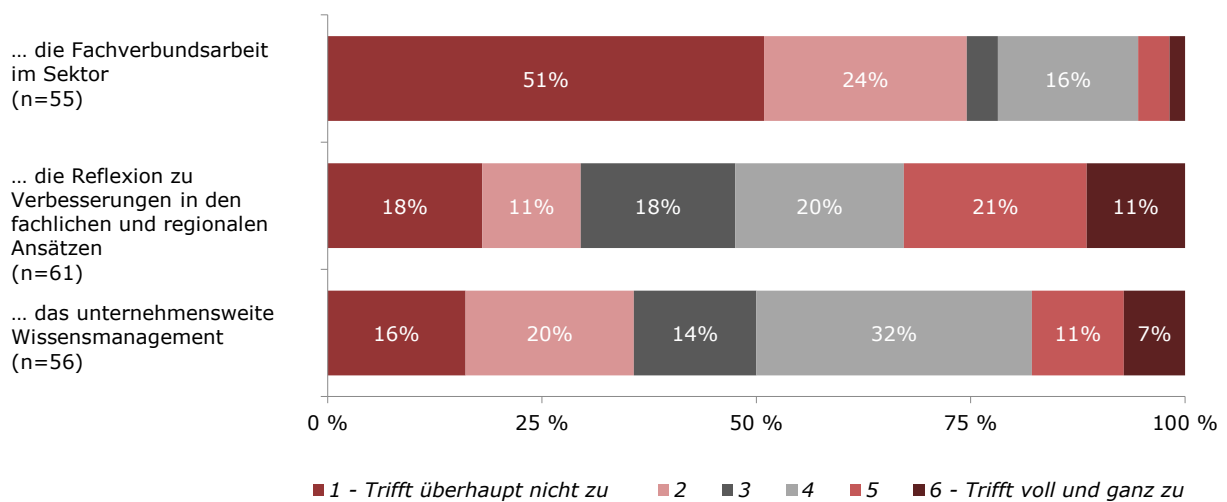


4. Nutzung

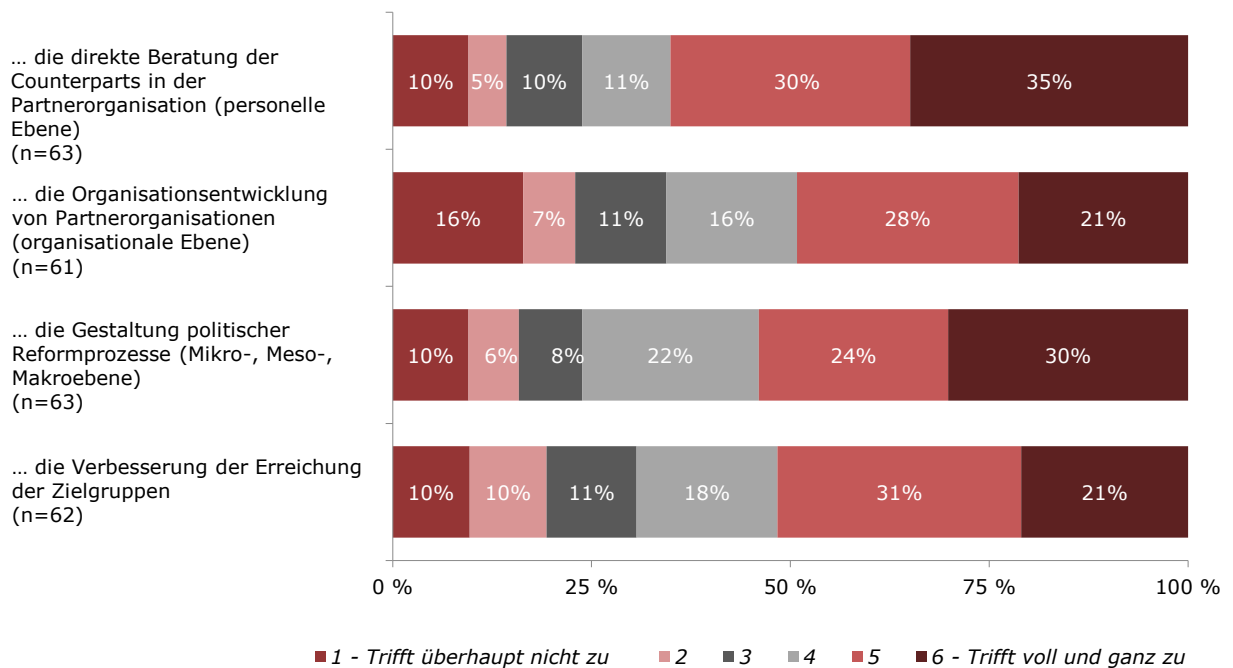
4.1 a) Die Ergebnisse der Evaluierung wurden genutzt für die GIZ-interne Projekt-/ Programmebene:



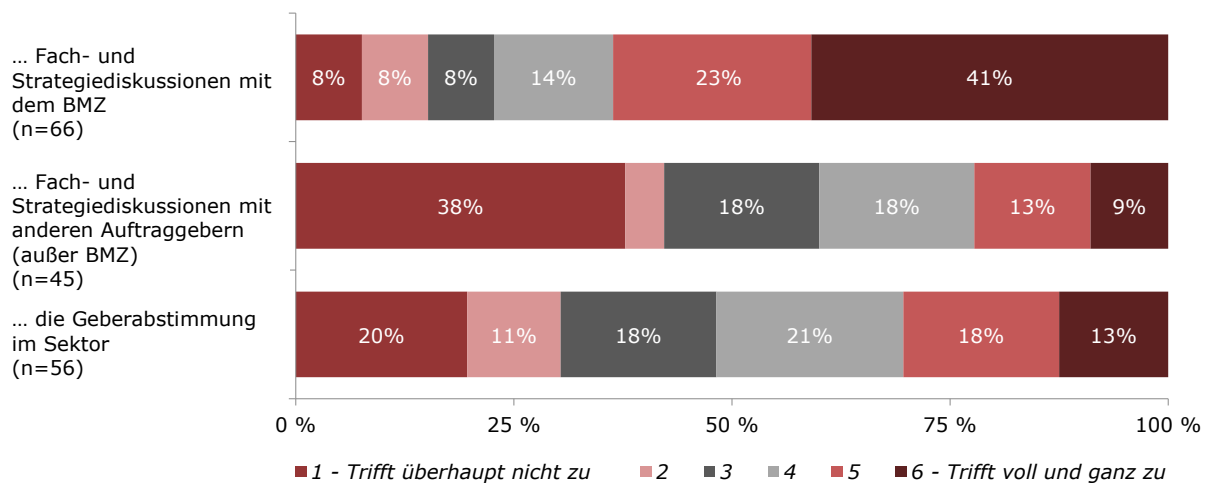
4.1 b) Die Ergebnisse der Evaluierung wurden genutzt für die GIZ-interne Sektor-/ Regionalebene:



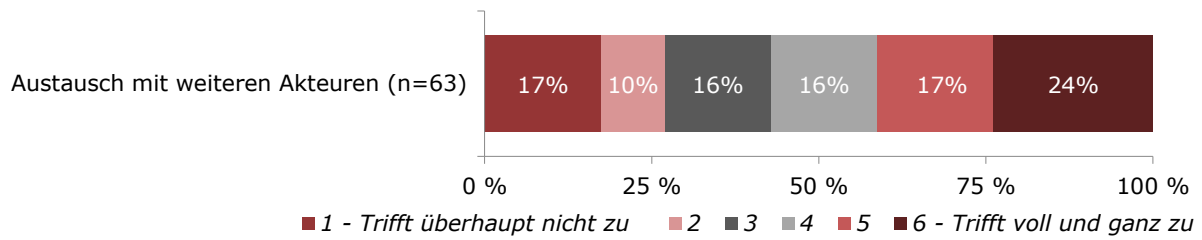
4.1 c) Die Ergebnisse der Evaluierung wurden genutzt für die Ebene des Kooperationsystems:



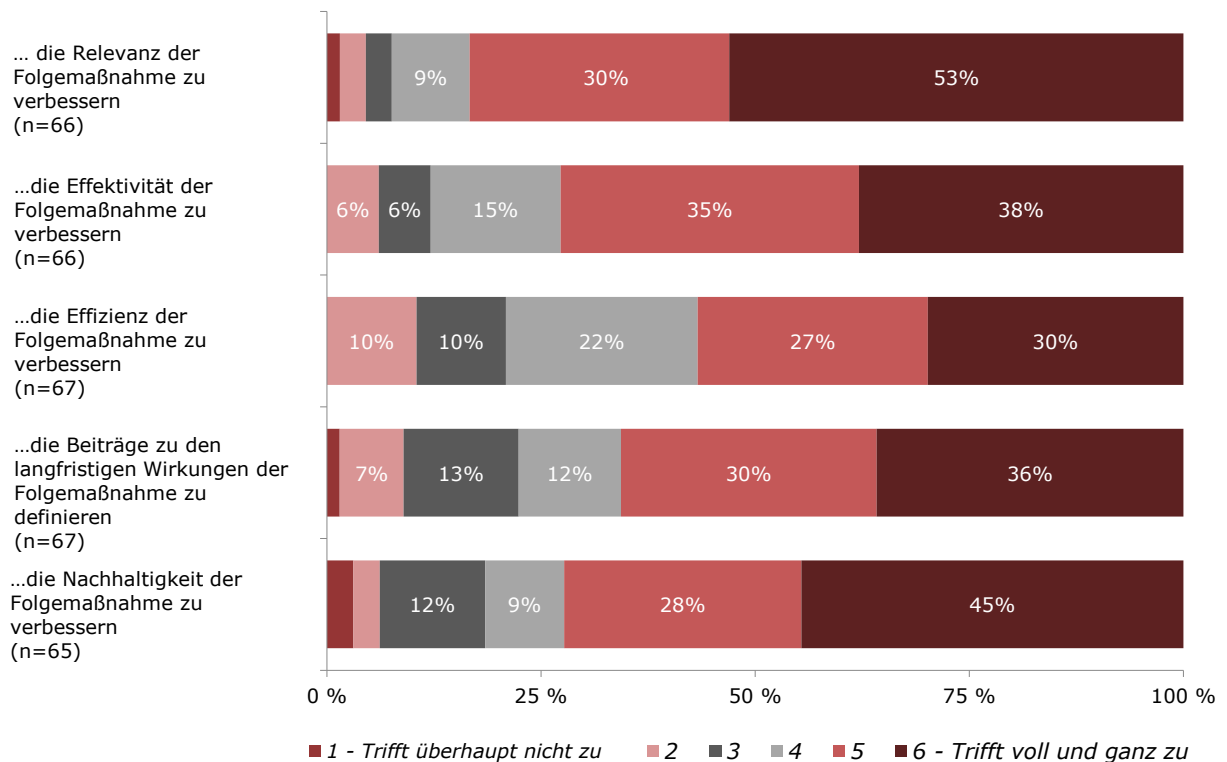
4.1 d) Die Ergebnisse der Evaluierung wurden genutzt für die Ebene des erweiterten Kooperationsystems:



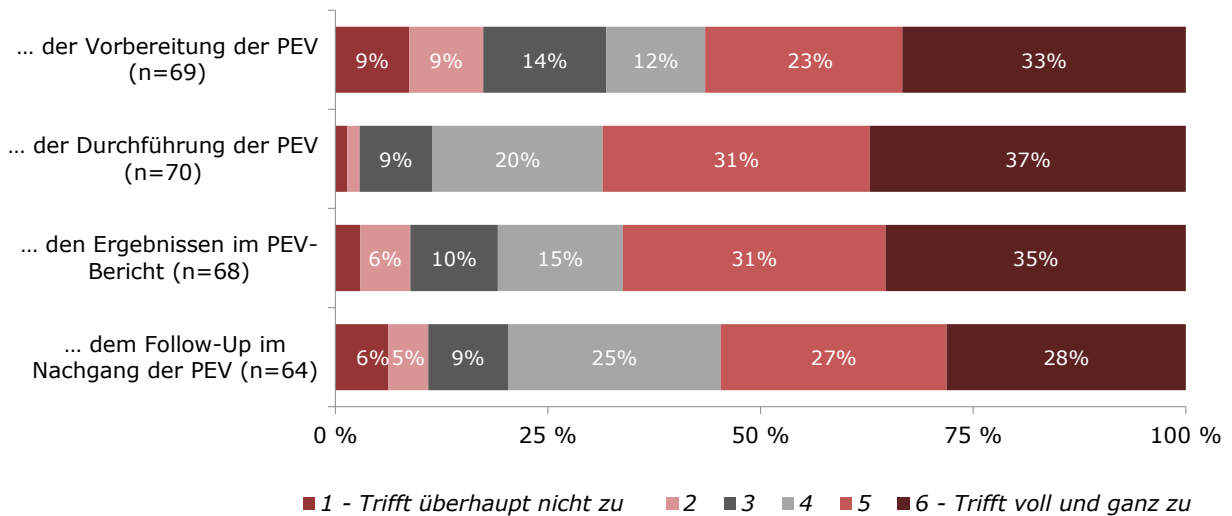
4.1 e) Die Ergebnisse der Evaluierung wurden genutzt für Sonstiges:



4.2 Die Ergebnisse der Evaluierung konnten genutzt werden, um...

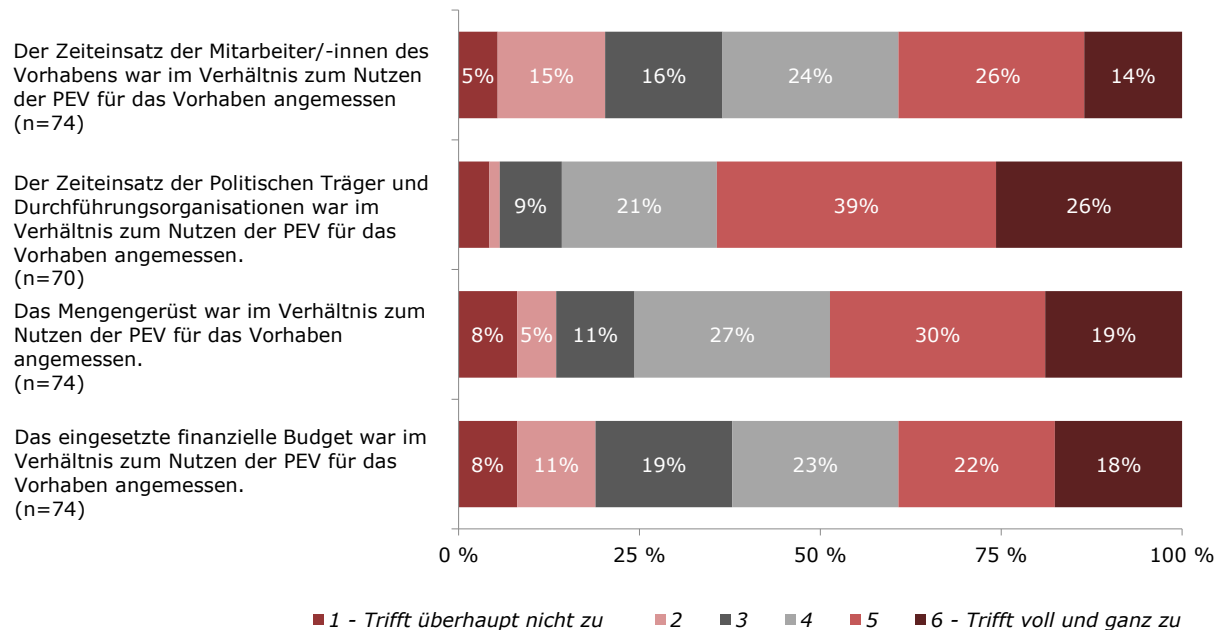


4.3 Für oben genannte Nutzung wurden vor allem Erkenntnisse genutzt aus...



5. Effizienz der PEV

5.1 Bitte bewerten Sie die Angemessenheit des Aufwandes der PEV im Verhältnis zum Nutzen für das Vorhaben:



9.3 Fragebogen der Online-Befragung

Einstieg

Zu Beginn der Befragung möchten wir Sie bitten, uns die Art der PEV zu nennen, die in ihrem Vorhaben durchgeführt wurde:

- ☐ Projektevaluierung ohne Folgemaßnahme
- ☐ Projektevaluierung mit Folgemaßnahme

Zudem möchten wir Sie bitten, die Art des evaluierten Vorhabens anzugeben:

- ☐ bilaterales Vorhaben oder Regionalvorhaben
- ☐ Sektorvorhaben oder Globalvorhaben

1. Beteiligung von Stakeholdern

Bitte wählen Sie aus, inwieweit die folgenden Angaben zutreffen (Mehrfachauswahl möglich).		ja	nein	Nicht zutreffend	Keine Angabe
N 1.1 N 1.2 N 2.1 D 2.1	1.1 Die folgenden Vorhabens-Stakeholder wurden in die Evaluierung eingebunden: • Politischer Träger • Durchführungsorganisationen • Weitere Kooperationspartner • Zielgruppen • Auftraggeber	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐

Bitte wählen Sie auf einer Skala von 1 bis 6 aus, inwieweit die folgenden Aussagen hinsichtlich der Projektevaluierung (PEV), die Sie beauftragt haben, zutreffen.		Trifft überhaupt nicht zu			Trifft voll und ganz zu			Keine Angabe
		1	2	3	4	5	6	
N 1.1 N 1.2 N 2.1 D 2.2	1.2 Inwiefern wurde darauf geachtet, dass die Interessen dieser Stakeholder in der Evaluierung berücksichtigt wurden? Im PEV-Prozess wurde darauf geachtet, dass... • ... die Interessen bestehender und potentieller Stakeholder in Bezug auf die Evaluierung systematisch und frühzeitig identifiziert wurden. • ... bestehende und potentielle Stakeholder, wann immer möglich, einbezogen wurden. • ... die Zwecke der Evaluierung mit bestehenden und potentiellen Stakeholdern frühzeitig im Evaluierungsprozess abgesprochen wurden. • ... der Dialog zwischen den verschiedenen Stakeholder gefördert wurde.	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐

2. PEV-Team

Bitte wählen Sie auf einer Skala von 1 bis 6 aus, inwieweit die folgenden Aussagen hinsichtlich der Projektevaluierung (PEV), die Sie beauftragt haben, zutreffen.		Trifft überhaupt nicht zu			Trifft voll und ganz zu			Keine Angabe
		1	2	3	4	5	6	
N 3.1	2.1 Das PEV-Team hat die folgenden Kompetenzen abgedeckt:							
	• Fachliche Kompetenz (im Sektor)	☐	☐	☐	☐	☐	☐	☐
	• Methodische Kompetenz (Evaluierungsmethoden)	☐	☐	☐	☐	☐	☐	☐
	• Beratungskompetenz (z.B. Capacity WORKS, Capacity Development)	☐	☐	☐	☐	☐	☐	☐
	• Prozesskompetenz (GIZ-Auftragsmanagement, Steuerung einer PEV)	☐	☐	☐	☐	☐	☐	☐
F 2.1	• Kultursensibles Vorgehen (Kultur, Religion)	☐	☐	☐	☐	☐	☐	☐
	• Gendersensibles Vorgehen	☐	☐	☐	☐	☐	☐	☐
F 4.1	• Unparteilichkeit (keine Interessenskonflikte)	☐	☐	☐	☐	☐	☐	☐

3. Informationsgehalt und Nützlichkeit

Bitte wählen Sie auf einer Skala von 1 bis 6 aus, inwieweit die folgenden Aussagen hinsichtlich der Projektevaluierung (PEV), die Sie beauftragt haben, zutreffen.		Trifft überhaupt nicht zu			Trifft voll und ganz zu			Keine Angabe
		1	2	3	4	5	6	
N 4.1	3.1 Die in der Evaluierung gewonnenen Informationen... ... decken die wesentlichen vereinbarten Fragestellungen der Evaluierung ab.	☐	☐	☐	☐	☐	☐	☐
N 7.1	3.2 Grundlagen für die Nutzung der Evaluierungsergebnisse: Der Evaluierungs<u>prozess</u> wurde zu einem für die Nutzung der Erkenntnisse günstigen Zeitpunkt durchgeführt.	☐	☐	☐	☐	☐	☐	☐
N 7.2	Der Evaluierungs<u>bericht</u> lag zu einem für die Nutzung passenden Zeitpunkt vor.	☐	☐	☐	☐	☐	☐	☐
F 5.1	Für die Kommunikation der Ergebnisse an die beteiligten und betroffenen Stakeholder wurden spezifische Formate entwickelt.	☐	☐	☐	☐	☐	☐	☐
F 5.2	3.3 Die Evaluierungsergebnisse wurden mit den beteiligten und betroffenen Stakeholdern geteilt: - Politischer Träger - Durchführungsorganisationen - Weitere Kooperationspartner - Zielgruppen	☐	☐	☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐	☐	☐

4. Nutzung

Bitte wählen Sie auf einer Skala von 1 bis 6 aus, inwieweit die folgenden Aussagen hinsichtlich der Projektevaluierung (PEV), die Sie beauftragt haben, zutreffen.		Trifft überhaupt nicht zu			Trifft voll und ganz zu			Keine Angabe
		1	2	3	4	5	6	
N 8.3	4.1 a) Die Ergebnisse der Evaluierung wurden genutzt für die GIZ-interne Projekt-/ Programmebene: ... den Austausch zwischen den Mitarbeiter/-innen ... den Austausch im Landesbüro ... die Steuerung des laufenden Vorhabens ... die Planung des Angebots <i>[Filter: Projektevaluierung mit Folgemaßnahme]</i> ...die Steuerung der Folgephase <i>[Filter: Projektevaluierung mit Folgemaßnahme]</i>	☐	☐	☐	☐	☐	☐	☐
N 8.4	4.1 b) Die Ergebnisse der Evaluierung wurden genutzt für die GIZ-interne <i>Sektor-/ Regionalebene</i> : ... die Fachverbandsarbeit im Sektor ... die Reflexion zu Verbesserungen in den fachlichen und regionalen Ansätzen ... das unternehmensweite Wissensmanagement	☐	☐	☐	☐	☐	☐	☐
N 8.5	4.1 c) Die Ergebnisse der Evaluierung wurden genutzt für die <i>Ebene des Kooperationssystems</i> : ... die direkte Beratung der Counterparts in der Partnerorganisation (personelle Ebene) ... die Organisationsentwicklung von Partnerorganisationen (organisationale Ebene) ... die Gestaltung politischer Reformprozesse (Mikro-, Meso-, Makroebene) ... die Verbesserung der Erreichung der Zielgruppen	☐	☐	☐	☐	☐	☐	☐
N 8.6	4.1 d) Die Ergebnisse der Evaluierung wurden genutzt für die <i>Ebene des erweiterten Kooperationssystems</i> : ... Fach- und Strategiediskussionen mit dem BMZ ... Fach- und Strategiediskussionen mit anderen Auftraggebern (außer BMZ) ... die Geberabstimmung im Sektor	☐	☐	☐	☐	☐	☐	☐
N 8.7	4.1 e) Die Ergebnisse der Evaluierung wurden genutzt für <i>Sonstiges</i> : Austausch mit weiteren Akteuren	☐	☐	☐	☐	☐	☐	☐
N 8.8 <i>[Filter: Projek tevalui]</i>	4.2 Die Ergebnisse der Evaluierung konnten genutzt werden, um... ...die <i>Relevanz</i> der Folgemaßnahme zu verbessern (Tun wir das Richtige?) ...die <i>Effektivität</i> der Folgemaßnahme zu verbessern (Wie erreichen wir unsere Ziele?)	☐	☐	☐	☐	☐	☐	☐

erung mit Folge maßna hme]	• ...die <i>Effizienz</i> der Folgemaßnahme zu verbessern (Wie gelingt eine wirtschaftliche Zielerreichung?) • ...die Beiträge zu den <i>langfristigen Wirkungen</i> der Folgemaßnahme zu definieren (Wie tragen wir zur Erreichung übergeordneter entwicklungspolitischer Ziele bei?) • ...die <i>Nachhaltigkeit</i> der Folgemaßnahme zu verbessern (Wie können die Wirkungen von Dauer sein?)	☐ ☐ ☐	☐ ☐ ☐	☐
N 8.1 N 8.1 N 8.2 N 8.1	4.3 Für oben genannte Nutzung wurden vor allem Erkenntnisse genutzt aus... • ... der Vorbereitung der PEV • ... der Durchführung der PEV • ... den Ergebnissen im PEV-Bericht • ... dem Follow-Up im Nachgang der PEV	☐ ☐ ☐	☐ ☐ ☐	☐

5. Effizienz der PEV

Bitte wählen Sie auf einer Skala von 1 bis 6 aus, inwieweit die folgenden Aussagen hinsichtlich der Projektevaluierung (PEV), die Sie beauftragt haben, zutreffen.		Trifft überhaupt nicht zu 1	2	3	4	5	Trifft voll und ganz zu 6	Keine Angabe
D 1.1 D 1.2 D 3.2 D 3.1	5.1 Bitte bewerten Sie die Angemessenheit des Aufwandes der PEV im Verhältnis zum Nutzen für das Vorhaben:							
	• Der Zeiteinsatz der Mitarbeiter/-innen des Vorhabens war im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen.	☐	☐	☐	☐	☐	☐	☐
	• Der Zeiteinsatz der Politischen Träger und Durchführungsorganisationen war im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen.	☐	☐	☐	☐	☐	☐	☐
	• Das Mengengerüst (Anzahl der Tage des PEV-Teams aus den TOR) war im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen.	☐	☐	☐	☐	☐	☐	☐
	• Das eingesetzte finanzielle Budget war im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen.	☐	☐	☐	☐	☐	☐	☐

Nachfolgend haben Sie die Möglichkeit zu den genannten Aspekten (Miteinsatz, Mengengerüst, finanzielles Budget) sowie weiteren Aspekten, die für Sie hinsichtlich der Effizienz der PEV relevant sind, Stellung zu nehmen. **Bitte machen Sie keine personenbezogenen Angaben.**

[Freitextfeld]

Abschluss

Eine an Ihren Bedarfen orientierte Unterstützung zur Qualitätssicherung des Evaluationsprozesses ist für das Unternehmen wichtig. Hierzu können Sie uns gerne mitteilen, inwieweit Sie mit dem Informationsmaterial, dem Kooperationsprozess und weiteren Aspekten der Qualitätssicherung (u.a. Handreichung, annotierte Berichtsgliederung, PEV-Toolbox, Qualitätscheck der PEV Berichte) zufrieden sind. **Bitte machen Sie keine personenbezogenen Angaben.**

[Freitextfeld]

9.4 Analyseraster

Ergebnisse im Evaluierungsstandard Genauigkeit (methodische Qualität)

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
G 1 Beschreibung des Evaluationsgegenstandes	Der Evaluationsgegenstand soll klar und genau beschrieben und dokumentiert werden, so dass er eindeutig identifiziert werden kann.							
G 1.1 Der Evaluierungsgegenstand wird genau beschrieben.	Der Indikator ist dann erfüllt, wenn die Evaluierung auf einer genauen Definition und Abgrenzung des Evaluierungsgegenstands basiert. Dies ist dann gegeben, wenn mindestens vier der folgenden sechs Aspekte beschrieben sind: zeitliche Abgrenzung (Phase), aufgewendete Mittel, regionale Abgrenzung, inhaltliche Abgrenzung (Mehrebenenansatz, Systemgrenze, CD Ebenen). (Textanalyse der PEV-Berichte)	X		X		X	X	
Checklist:	zeitliche Abgrenzung (Phase)	X		X		X	X	
	aufgewendete Mittel							
	regionale Abgrenzung	X		X		X	X	
	inhaltliche Abgrenzung (Mehrebenenansatz)	X				X		
	inhaltliche Abgrenzung (Systemgrenze)	X		X		X		
	inhaltliche Abgrenzung (CD Ebenen)	X		X		X		
G 1.2 Das Wirkungsmodell wird dargestellt.	Der Indikator ist dann erfüllt, wenn die Wirkungslogik der Maßnahme des aktuellen Moduls graphisch dargestellt oder textlich eindeutig beschrieben wird, so dass die Systemgrenzen des Vorhabens deutlich werden. Falls für die evaluierte Maßnahme kein Wirkungsmodell vorliegt, muss dies nachträglich (graphisch oder textlich) dargestellt werden. (Textanalyse der PEV-Berichte)	X		X		X	X	
G 1.3 Die relevanten Wirkungshypothesen werden dargestellt.	Der Indikator ist dann erfüllt, wenn die Wirkungslogik der Maßnahme anhand eines Narrativs dargestellt wird. Die Wirkungshypothesen müssen begründete Vermutungen über Zusammenhänge innerhalb des Wirkungsmodells darstellen. Es muss deutlich werden, was unter welchen Umständen und warum erwartet werden kann. Jede Hypothese muss zudem empirisch überprüfbar und falsifizierbar sein. Als Mindeststandard gilt, dass zumindest die Einordnung des Modulziels auf der Outcome Ebene und des Programmziels (falls vorhanden) auf der Impact Ebene und die entsprechenden Wirkungsannahmen (kausale Annahmen bei Outcome und zumindest plausible Annahmen bei Impact) auf diese Zielsetzungen hin deutlich werden sollen. (Textanalyse der PEV-Berichte)	X		X		X	X	

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
G 2 Rahmenbedingungen	Der Kontext des Evaluationsgegenstandes soll ausreichend detailliert untersucht und analysiert werden.							
G 2.1 Der Evaluierungsgegenstand wird im Politikkontext des Partnerlandes verortet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsbericht den Politikkontext beschreibt: Beschreibung von Strategien, Leitlinien oder Zielen des Partnerlandes. Kenntnisse des Politikkontexts sind erforderlich, um insbesondere die Gelingensbedingungen und die Wechselwirkungen der Maßnahme realistisch einzuschätzen. (Textanalyse der PEV-Berichte)	X						
G 2.2 Der Evaluierungsgegenstand wird im Entwicklungskontext des Sektors im Partnerland verortet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsbericht den Entwicklungskontext im Sektor beschreibt: sozioökonomische, politische oder kulturelle Rahmenbedingungen im Sektor. Kenntnisse des Entwicklungskontexts sind erforderlich, um insbesondere die Gelingensbedingungen und die Wechselwirkungen der Maßnahme realistisch einzuschätzen. (Textanalyse der PEV-Berichte)	X						
G 2.3 Der Evaluierungsgegenstand wird in der Träger- und Partnerstruktur verortet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsbericht die Träger- und der Partnerstruktur beschreibt: Kenntnisse der Träger- und Partnerstruktur gehören zum institutionellen Kontexts einer Maßnahme und sind erforderlich, um die Gelingensbedingungen und die Wechselwirkungen der Maßnahme realistisch einzuschätzen. (Textanalyse der PEV-Berichte)							
G 2.4 In der Bewertung und in den Schlussfolgerungen wird ein Rückbezug auf die Kontextanalyse vorgenommen.	Der Indikator ist dann erfüllt, wenn deutlich wird, wie der untersuchte Kontext die Bewertung der Evaluierung beeinflusst. Der Einfluss der Kontextfaktoren auf die Ergebnisse der Evaluierung wird transparent gemacht. Dies hilft auch, eine mögliche Übertragbarkeit der Evaluierungsergebnisse auf andere Maßnahmen einschätzen zu können. (Textanalyse der PEV-Berichte)							
G 3 Beschreibung von Zwecken und Vorgehen	Gegenstand, Zwecke, Fragestellungen und Vorgehen der Evaluation, einschließlich der angewandten Methoden, sollen genau dokumentiert und beschrieben werden, so dass sie identifiziert und eingeschätzt werden können.							
G 3.1 Anlass, Zweck und beabsichtigte Verwendung der Evaluierung werden transparent beschrieben.	Der Indikator ist dann erfüllt, wenn deutlich wird, was der Anlass und Zweck (Zwischen- oder Schlussevaluierung) der Evaluierung sind und wer bzw. wie die Ergebnisse verwendet soll/en (Nutzer/ Nutzung). (Textanalyse der PEV-Berichte)							X
G 3.2 Die spezifischen Ziele der Evaluierung werden deutlich.	Der Indikator ist dann erfüllt, wenn transparent dargestellt ist, was mit der Evaluierung erreicht werden soll (bspw. Überprüfung von Relevanz, Zielerreichung, Nachhaltigkeit, Handlungsempfehlungen für Folgemodul, Übertragbarkeit auf andere Kontexte, etc.). Die spezifischen Evaluierungsziele müssen aus den TOR abgeleitet sein. (Textanalyse der PEV-Berichte)							
G 3.3 Die TOR einschließlich der Evaluierungsfragen der Prüfmission befinden sich im Anhang des Prüfberichts.	Der Indikator ist dann erfüllt, wenn die TOR mit den Evaluierungsfragen ein Anhang des Evaluierungsberichts sind. So kann beurteilt werden, inwieweit sich die Prüfmission ausreichend mit den Evaluierungsfragen auseinander gesetzt hat. (Textanalyse der PEV-Berichte)	X						
G 3.4 Anhand der TOR lassen sich Aussagen zum Zweck der Evaluierung ableiten.	Der Indikator ist dann erfüllt, wenn die TOR den Anlass und Zweck der Evaluierung klar definieren. (Textanalyse der PEV-Berichte)				X		X	X

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
G 4 Angabe von Informationsquellen	Die im Rahmen einer Evaluation genutzten Informationsquellen sollen hinreichend genau dokumentiert werden, damit die Verlässlichkeit und Angemessenheit der Informationen eingeschätzt werden kann.							
G 4.1 Der Bericht umfasst eine vollständige Liste der Gesprächspartner/innen und sonstigen Informationsquellen.	Der Indikator ist dann erfüllt, wenn die verwendeten Informationsquellen beschrieben werden, so dass die Aussagen im Bericht evidenzbasiert unterlegt sind. Falls Anonymität und das Recht auf Vertraulichkeit der einzelnen Gesprächspartner/innen die Darstellung der Informationsquellen einschränken, wird entsprechend darauf hingewiesen. (Textanalyse der PEV-Berichte)	X						
G 4.2 Der Bericht stellt transparent dar, welcher Systematik der Dokumentenauswertung zugrunde lag.	Der Indikator ist dann erfüllt, wenn die Auswertung der verwendeten Dokumente beschrieben werden. Hierdurch wird deutlich, welche Dokumente warum ausgewertet wurden und welches System der Dokumentenauswertung zugrunde lag. (Textanalyse der PEV-Berichte)							
G 4.3 Die Auswahl von Gesprächspartner bzw. die Ziehung von Stichproben geschieht systematisch.	Der Indikator ist dann erfüllt, wenn die Auswahl von Gesprächspartner/-innen (bzw. Stichproben bei quantitativen Analysen) nach klar definierten und transparenten Kriterien geschieht (z.B. Reflektion der Erwartungen an die Aussagefähigkeit der TN, Zusammensetzung der TN, Rolle der Prüfmision bei der Auswahl der TN), und die Auswirkungen der Auswahl auf die Repräsentativität der Ergebnisse erläutert wird. Da im Rahmen einer Prüfmision nicht jeder befragt werden kann, der zur Grundgesamtheit gehört, muss eine begründete Auswahl aus der Grundgesamtheit vorgenommen werden. (Textanalyse der PEV-Berichte)	X		X		X	X	
G 4.4 Aufbauend auf die Bewertung der Verlässlichkeit der Daten des WOM, werden diese entweder verwendet oder ausgeschlossen.	Der Indikator ist dann erfüllt, wenn zunächst eine kritische Bewertung der Verlässlichkeit der Daten des WOM erfolgt und die Daten des WOM dann, ausgehend von der Bewertung, entweder verwendet oder, falls nicht oder nur eingeschränkt auf die Daten zurück gegriffen wird, begründet ausgeschlossen werden. (Textanalyse der PEV-Berichte)	X						
G 4.5 Aufbauend auf die Bewertung der Verlässlichkeit der Baseline Daten, werden diese entweder verwendet oder ausgeschlossen.	Der Indikator ist dann erfüllt, wenn zunächst eine kritische Bewertung der Verlässlichkeit der Baseline Daten erfolgt und die Baseline Daten dann, ausgehend von der Bewertung, entweder verwendet oder, falls nicht oder nur eingeschränkt auf die Daten zurück gegriffen wird, begründet ausgeschlossen werden. (Textanalyse der PEV-Berichte)	X				X		
G 4.6 Aufbauend auf die Bewertung der Verlässlichkeit der Partnerdaten, werden diese entweder verwendet oder ausgeschlossen.	Der Indikator ist dann erfüllt, wenn zunächst eine kritische Bewertung der Verlässlichkeit der Partnerdaten erfolgt und die Partnerdaten dann, ausgehend von der Bewertung, entweder verwendet oder, falls nicht oder nur eingeschränkt auf die Daten zurück gegriffen wird, begründet ausgeschlossen werden. (Textanalyse der PEV-Berichte)							X

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
G 5 Valide und reliable Informationen	Die Verfahren zur Gewinnung von Daten sollen so gewählt oder entwickelt und dann eingesetzt werden, dass die Zuverlässigkeit der gewonnenen Daten und ihre Gültigkeit bezogen auf die Beantwortung der Evaluationsfragestellungen nach fachlichen Maßstäben sichergestellt sind. Die fachlichen Maßstäbe sollen sich an den Gütekriterien quantitativer und qualitativer Sozialforschung orientieren.							
G 5.1 Es wird eine Datentriangulation durchgeführt.	Der Indikator ist dann erfüllt, wenn Daten / Informationen zu einem gleichen Sachverhalt durch die Einbeziehung verschiedener Akteure erhoben werden. (Textanalyse der PEV-Berichte)			X		X	X	X
G 5.2 Es wird eine Methodentriangulation durchgeführt.	Der Indikator ist dann erfüllt, wenn Informationen zu einem gleichen Sachverhalt durch verschiedene Methoden der Datensammlung erhoben werden. (Textanalyse der PEV-Berichte)			X		X	X	X
G 6 Systematische Fehlerprüfung	Die in einer Evaluation gesammelten, aufbereiteten, analysierten und präsentierten Informationen sollen systematisch auf Fehler geprüft werden.							
G 6.1 Es wird eine Forschertriangulation durchgeführt.	Der Indikator ist dann erfüllt, wenn das PEV-Team seine Erfahrungen und Interpretationen / Perspektiven trianguliert. Hinweise auf die systematische Überprüfung der gesammelten Informationen sind z.B. Synthesetreffen, die dokumentiert wurden, oder die Darstellung auf Unstimmigkeiten der Mitglieder des PEV-Teams, die entsprechend diskutiert wurden. (Textanalyse der PEV-Berichte)			X				
G 6.2 Es wird eine Triangulation der Ergebnisse mit den Beteiligten und Betroffenen durchgeführt.	Der Indikator ist dann erfüllt, wenn eine Triangulation der Ergebnisse mit den Beteiligten und Betroffenen erfolgt. So soll eine Fehlerüberprüfung der Analyseergebnisse erfolgen. Hinweise auf die systematische Überprüfung der gesammelten Informationen sind z.B. gemeinsame Syntheseworkshops oder die Kommentierung des Berichts durch Beteiligte und Betroffene. (Textanalyse der PEV-Berichte)			X				

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
G 7 (A) Analyse qualitativer und quantitativer Informationen	Qualitative und quantitative Informationen einer Evaluation sollen nach fachlichen Maßstäben angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation effektiv beantwortet werden können.							
G 7.1 (A) Die methodische Vorgehensweise beantwortet die Frage nach der Zuordnungs- und/ oder Beitragsanalyse.	Der Indikator ist dann erfüllt, wenn die methodische Vorgehensweise mit Hinweis auf die Beantwortung von Evaluierungsfragen, Bedingungen der Untersuchung und/ oder Besonderheiten des Evaluierungsgegenstandes begründet dargestellt wird. So kann bspw. eine Attributionsanalyse mit dem Hinweis auf den Evaluierungsgegenstand begründet ausgeschlossen werden. (Textanalyse der PEV-Berichte)						X	
Checklist:	keine begründete Vorgehensweise						X	
	experimentelles Design						X	
	quasi-experimentelles Design						X	
	Ansatz einer Kontributionsanalyse						X	
	theoriebasierter Ansatz						X	
G 7.2 (A) Die Evaluierungsmethoden werden dargestellt.	Der Indikator ist dann erfüllt, wenn die verwendeten Methoden beschrieben werden, so dass die Aussagen evidenzbasiert unterlegt sind. (Textanalyse der PEV-Berichte)	X		X		X	X	
G 7.3 (A) Es gibt eine Methodenvielfalt.	Der Indikator ist dann erfüllt, wenn mehr als eine Methode verwendet wird. (Textanalyse der PEV-Berichte)			X		X	X	X
Checklist:	Dokumentaenauswertung						X	
	Fokusgruppen						X	
	Interviews						X	
	Befragung						X	
	eigene Auswertung der Monitoringdaten durch die Prüfmission							
G 7.4 (A) Die Vor- und Nachteile der gewählten Methoden werden dargestellt.	Der Indikator ist dann erfüllt, wenn die Methodenauswahl nachvollziehbar ist. Diese Nachvollziehbarkeit spiegelt sich in einer Klärung darüber, welche Vorteile die ausgesuchte Methode hat und welche möglichen Nachteile vorhanden sind und wie damit umgegangen werden soll, wider. (Textanalyse der PEV-Berichte)	X				X	X	

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
G 7 (B) Analyse qualitativer und quantitativer Informationen: Relevanz	Die methodische Bearbeitung der Relevanzbewertung ist angemessen.							
G 7.5 (B) Die Nachvollziehbarkeit der zugrunde gelegten Rahmenbedingungen und Kernprobleme der Maßnahme ist gegeben.	Der Indikator ist dann erfüllt, wenn bei der Bearbeitung der Relevanz die Rahmenbedingungen und Kernbedarfe, auf die die Maßnahme beruht, überprüft wird. (Textanalyse der PEV-Berichte)							
G 7.6 (B) Die Mehrdimensionalität der Relevanz wird analysiert.	Der Indikator ist dann erfüllt, wenn bei der Bearbeitung der Relevanz entlang der Kernbedarfe von Zielgruppe und Partner sowie der Ziele des Auftraggebers analysiert wird. (Textanalyse der PEV-Berichte)	X	X					
G 7.7 (B) Die strategische Ausrichtung der Maßnahme an veränderten Rahmenbedingungen werden analysiert.	Der Indikator ist dann erfüllt, wenn die strategische Ausrichtung einer Maßnahme an Veränderungen in den Rahmenbedingungen analysiert werden. (Textanalyse der PEV-Berichte)		X					
G 7 (C) Analyse qualitativer und quantitativer Informationen: Effektivität	Die methodische Bearbeitung der Effektivitätsbewertung ist angemessen.							
G 7.8 (C) Die Kausalität zwischen Maßnahmen und Wirkungen wird differenziert analysiert und eingeschätzt.	Der Indikator ist dann erfüllt, wenn beobachtete Zusammenhänge zwischen Maßnahmen und Wirkungen nicht ohne Weiteres als ursächliche Zusammenhänge interpretiert werden. Der Einfluss von Drittvariablen muss ausgeschlossen werden. Korrelationen dürfen nicht mit Kausalitäten gleichgesetzt werden. (Textanalyse der PEV-Berichte)			X		X	X	
G 7.9 (C) Die Zielerreichung wird anhand von Modulindikatoren bewertet.	Der Indikator ist dann erfüllt, wenn deutlich wird, welche Indikatoren zur Bewertung der Zielerreichung als Grundlage genommen wurden. Wurden während der Projektplanung für einzelne Aspekte von Effektivität keine Indikatoren definiert, müssen diese Indikatoren nachträglich gebildet werden. (Textanalyse der PEV-Berichte)	X	X			X	X	
G 7.10 (C) Die verwendeten Indikatoren zur Messung und Beurteilung der Zielerreichung haben eine hohe Qualität.	Der Indikator ist dann erfüllt, wenn ersichtlich ist, dass sich die Prüfer mit der Qualität der Indikatoren auseinandergesetzt haben (zum Beispiel durch die Nutzung von SMART oder CREAM-Kriterien). Wird eine geringe Qualität festgestellt, müssen neue bzw. veränderte Indikatoren für die Evaluation genutzt werden. (Textanalyse der PEV-Berichte)	X		X		X	X	

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
G 7 (D) Analyse qualitativer und quantitativer Informationen: Effizienz	Die methodische Bearbeitung der Effizienzbewertung ist angemessen.							
G 7.11 (D) Die verschiedenen Ebenen der Effizienz einer Maßnahme werden analysiert.	Der Indikator ist dann erfüllt, wenn im Evaluierungsbericht dargestellt ist, welche Ebenen der Effizienzmessung analysiert werden und die Auswahl begründet wird. Mögliche Ebenen sind: die Implementierungseffizienz (Analyse der Strukturen und Prozesse der Vorhabensumsetzung), der Produktionseffizienz (Verhältnis Input zu Output) und der Allokationseffizienz (Verhältnis von Input zu Outcome). (Textanalyse der PEV-Berichte)	X	X					
Checklist:	Implementierungseffizienz	X	X					
	Produktionseffizienz,	X	X					
	Allokationseffizienz	X	X					
G 7.12 (D) Die Auswahl von Methoden und Verfahren der Effizienzmessung wird begründet.	Der Indikator ist dann erfüllt, wenn im Evaluierungsbericht die Auswahl der Methoden und Verfahren zur Messung der Effizienz begründet wird. (Textanalyse der PEV-Berichte)							
Checklist:	Level 2 methods							
	Level 1 methods							
	Descriptive methods							
G 7.13 (D) Die Bearbeitung der Effizienz ermöglicht die Identifikation von Potenzialen zur Effizienzsteigerung.	Der Indikator ist dann erfüllt, wenn die verwendeten Methoden und Verfahren implizit oder explizit Komponenten/Teile der Maßnahme mit Alternativen vergleichen, um Verbesserungspotenziale aufzuzeigen. (Textanalyse der PEV-Berichte)							

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
G 7 (E) Analyse qualitativer und quantitativer Informationen: Impact	Die methodisch Bearbeitung der Impactbewertung ist angemessen.							
G 7.14 (E) In der Analyse und Beurteilung der übergeordneten Wirkungen wird die Attributionslücke analysiert.	Der Indikator ist dann erfüllt, wenn im Evaluierungsbericht bei der Analyse und Beurteilung der übergeordneten Wirkungen die Zuordnungs- bzw. Attributionslücke reflektiert wird. Es wird kein direkter und monokausaler Zusammenhang zwischen Maßnahme und übergeordneten Zielen hergestellt. (Textanalyse der PEV-Berichte)	X						
G 7.15 (E) Die Plausibilität der Hypothesen zu den intendierten langfristigen Wirkungen wird bewertet.	Der Indikator ist dann erfüllt, wenn sich die Evaluation mit der Plausibilität der Hypothesen zu übergeordneten langfristigen Wirkungen auseinandersetzt. (Textanalyse der PEV-Berichte)		X					
G 7.16 (E) Bewertungsmaßstäbe zur Analyse und Beurteilung des Beitrages der Maßnahme zu übergeordneten Wirkungen sind transparent dargestellt.	Der Indikator ist dann erfüllt, wenn beschreiben ist, anhand welcher Bewertungsmaßstäbe der Beitrag des Vorhabens zu übergeordneten entwicklungspolitischen Wirkungen bewertet wird. Diese Bewertungsmaßstäbe können entweder selbst vom Prüftteam entwickelt (zum Beispiel (Proxi-)Indikatoren) oder vorhandene Maßstäbe (zum Beispiel aus Teil A des PV) herangezogen werden. (Textanalyse der PEV-Berichte)	X	X	X		X	X	
G 7 (F) Analyse qualitativer und quantitativer Informationen: Nachhaltigkeit	Die methodisch Bearbeitung der Nachhaltigkeitsbewertung ist angemessen.							
G 7.17 (F) Die Grenzen der Messung der Nachhaltigkeit werden beschrieben.	Der Indikator ist dann erfüllt, wenn bei der Bearbeitung der Nachhaltigkeit beschrieben wird, inwiefern zum Zeitpunkt der Evaluierung eine ergebnisorientierte Analyse und Beurteilung der Nachhaltigkeit von Effekten überhaupt möglich ist. (Textanalyse der PEV-Berichte)							
G 7.18 (F) Die angelegten Ansätze zur Schaffung von Nachhaltigkeit werden analysiert.	Der Indikator ist dann erfüllt, wenn in der Analyse und Beurteilung der Nachhaltigkeit die angelegten Ansätze zur Schaffung von Nachhaltigkeit berücksichtigt werden. Zum Beispiel wird evaluiert, inwiefern die Maßnahme eine Exit- oder Nachhaltigkeitsstrategie besitzt, Ansätze für die Anschlussfähigkeit der Maßnahme betreibt oder Aktivitäten zur Verantwortungsübergabe (Ownership) an den Partner betreibt. Die Grenzen der Messung werden dann bei der Analyse und Beurteilung eingehalten. (Textanalyse der PEV-Berichte)		X					
G 7.19 (F) Es werden mindestens zwei Ebenen der Nachhaltigkeit im Rahmen einer Prognose analysiert.	Der Indikator ist dann erfüllt, wenn der Mehrdimensionalität von Nachhaltigkeit im Rahmen einer Prognose Rechnung getragen wird und mindestens Nachhaltigkeit auf zwei Ebenen betrachtet wird (finanzielle, institutionelle, personelle, technologische, soziale, ökologische Nachhaltigkeit) überprüft. (Textanalyse der PEV-Berichte)							

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
G 8 Begründete Analyse und begründete Schlussfolgerung	Die in einer Evaluation gezogenen Folgerungen sollen ausdrücklich begründet werden, damit die Adressaten und Adressatinnen diese einschätzen können.							
G 8.1 Es wird zwischen Beschreibung, Analyse und Bewertung unterschieden.	Der Indikator ist dann erfüllt, wenn der Dreischritt Beschreibung – Analyse – Bewertung aus evaluatorischer Einschätzung zumeist eingehalten wird. Das heißt, es wird nicht direkt von der Beschreibung zur Bewertung gesprungen bzw. die Analyse und Bewertung werden nicht vermischt. Der Dreischritt soll mindestens bei zwei der folgenden drei Aspekte eingehalten werden: Wirkungsmodell, Bewertung der OECD/DAC-Kriterien, Schlussfolgerungen. (Textanalyse der PEV-Berichte)					X	X	
Checklist:	Wirkungsmodell	X				X		
	Bewertung der OECD/ DAC-Kriterien					X		
	Schlussfolgerungen					X		
G 8.2 Beschreibungen und Analysen werden belegt.	Der Indikator ist dann erfüllt, wenn Beschreibung und Analyse im Bericht mit relevanten Zahlen, Daten und Fakten untermauert werden. Für alle Zahlen, Daten und Fakten, werden Quellen angegeben. Demnach sind bei der Beschreibung und Analyse von Ergebnissen der Datenerhebung die jeweiligen Informationsquellen ersichtlich. Hierbei werden die Daten, die von der Evaluierungsmission erhoben wurden, als solche gekennzeichnet. (Textanalyse der PEV-Berichte)	X				X	X	
G 8.3 Empfehlungen werden aus der Analyse abgeleitet und sind spezifisch, realistisch und termingebunden. Sie richten sich an die wesentlichen Nutzer und ihre Umsetzung ist messbar.	Der Indikator ist dann erfüllt, wenn für alle Handlungsempfehlungen ein Ziel bzw. ein Zweck definiert worden ist. Die Empfehlungen müssen realistisch sein und sich aus den Schlussfolgerungen ableiten. Die Realisierbarkeit aller Handlungsempfehlungen muss sichergestellt sein. Z.B indem Handlungsempfehlungen in Zusammenarbeit mit denjenigen entwickelt werden, die sie umsetzen müssen. Alle Handlungsempfehlungen müssen einen Zeitraum bzw. einen Zeitpunkt für die Umsetzung benennen und die Umsetzung muss empirisch erhoben werden können. (Textanalyse der PEV-Berichte)					X	X	
G 8.4 Die positiven und/ oder negativen nicht-intendierten Wirkungen der Maßnahme werden beschrieben.	Der Indikator ist dann erfüllt, wenn sich die Evaluierung mit möglichen positiven und/ oder negativen nicht-intendierten Wirkungen befasst. Hiermit soll gewährleistet werden, dass sich die Evaluierung möglicher Hinweise auf nicht-intendierte Wirkung bewusst ist. (Textanalyse der PEV-Berichte)	X	X					

Ergebnisse im Evaluierungsstandard Durchführbarkeit (Prozessqualität I)

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
D 1 Angemessene Verfahren	Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass Belastungen des Evaluationsgegenstandes bzw. der Beteiligten und Betroffenen in einem angemessenen Verhältnis zum erwarteten Nutzen der Evaluation stehen.							
D 1.1 Das Verfahren der Evaluierung wird als angemessen für die Mitarbeiter/-innen des Vorhabens bezeichnet.	Der Indikator ist erfüllt, wenn laut Auftraggeber der Zeiteinsatz der Mitarbeiter/-innen des Vorhabens im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen war. (AV-Befragung)							
D 1.2 Das Verfahren der Evaluierung wird als angemessen für die Partner des Vorhabens bezeichnet.	Der Indikator ist erfüllt, wenn laut Auftraggeber der Zeiteinsatz der Politischen Träger und Durchführungsorganisationen im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen war. (AV-Befragung)							
D 1.3 Die Qualitätskontrolle der Evaluierung wird als angemessen bezeichnet.	Der Indikator ist dann erfüllt, wenn der Auftraggeber mit den in Mechanismen der Qualitätskontrolle (Backstopping und Unterstützungsangebote der STS 08, QS Check der Evaluierungsberichte) zufrieden war. (AV-Befragung)					X	X	X
D 2 Diplomatisches Vorgehen	Evaluationen sollen so geplant und durchgeführt werden, dass eine möglichst hohe Akzeptanz der verschiedenen Beteiligten und Betroffenen in Bezug auf Vorgehen und Ergebnisse der Evaluation erreicht werden kann.							
D 2.1 Die Evaluierung verfolgt einen kooperativen Ansatz und bringt unterschiedliche Interessensgruppen zusammen.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die Evaluierung verschiedene Interessen ausgewogen berücksichtigt hat. Hierzu hat sie den Dialog zwischen den verschiedenen Stakeholdern gesucht. (AV-Befragung)							X
D 3 Effizienz von Evaluation	Der Aufwand für Evaluation soll in einem angemessenen Verhältnis zum Nutzen der Evaluation stehen.							
D 3.1 Das Budget der Evaluierung wird als angemessen bezeichnet.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass das eingesetzte finanzielle Budget im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen war. (AV-Befragung)							
D 3.2 Das Mengengerüst der Evaluierung wird als angemessen bezeichnet.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass das Mengengerüst (Anzahl der Tage des PEV-Teams aus den TOR) im Verhältnis zum Nutzen der PEV für das Vorhaben angemessen war. (AV-Befragung)							

Ergebnisse im Evaluierungsstandard Fairness (Prozessqualität II)

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
F 1 Formale Vereinbarungen	Die Pflichten des PEV-Teams (was, wie, von wem, wann getan werden soll) sollen schriftlich festgehalten werden, damit die Parteien verpflichtet sind, alle Bedingungen dieser Vereinbarung zu erfüllen oder aber diese neu auszuhandeln.							
F 1.1 Die angehängte Leistungsbeschreibung des Evaluierungsberichtes beschreibt die vertraglichen Rahmenbedingungen.	Der Indikator ist dann erfüllt, wenn die angehängte Leistungsbeschreibung beschreibt was, wie, von wem und wann evaluiert werden soll (siehe auch Indikator 3.3, der das warum - also die Ziele in den ToRs prüft). (Textanalyse der PEV-Berichte)				X			
F 2 Schutz individueller Rechte	Evaluationen sollen so geplant und durchgeführt werden, dass Sicherheit, Würde und Rechte der in eine Evaluation einbezogenen Personen geschützt werden.							
F 2.1 Der Evaluierungsprozess folgt ethischen Verhaltensregeln.	Der Indikator ist dann erfüllt, wenn durch den Auftraggeber bestätigt wird, dass das PEV-Team kultursensibel (Kultur, Religion) und gendersensibel vorgegangen ist. (AV-Befragung)						X	X
F 3 Vollständige und faire Überprüfung	Evaluationen sollen die Stärken und die Schwächen des Evaluationsgegenstandes möglichst vollständig und fair überprüfen und darstellen, so dass die Stärken weiter ausgebaut und die Schwachpunkte behandelt werden können.							
F 3.1 Die Bewertung der Maßnahme spiegelt seine Stärken und Schwächen wider.	Der Indikator ist dann erfüllt, wenn die Stärken und Schwächen der Maßnahme so vollständig wie möglich erfasst und ausgewogen dargestellt sind. (Textanalyse der PEV-Berichte)					X	X	
F 4 Unparteiische Durchführung und Berichterstattung	Die Evaluation soll unterschiedliche Sichtweisen von Beteiligten und Betroffenen auf Gegenstand und Ergebnisse der Evaluation in Rechnung stellen. Berichte sollen ebenso wie der gesamte Evaluationsprozess die unparteiische Position des Evaluationsteams erkennen lassen. Bewertungen sollen fair und möglichst frei von persönlichen Gefühlen getroffen werden.							
F 4.1 Die Auswahl des PEV-Teams ermöglichte eine unparteiische Durchführung und Berichterstattung der Evaluierung.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass das PEV-Team unparteiisch war und damit keine Interessenskonflikte in Kontext des evaluierten Vorhabens hat. (AV-Befragung)			X			X	X
F 5 Offenlegung der Ergebnisse	Die Evaluationsergebnisse sollen allen Beteiligten und Betroffenen soweit wie möglich zugänglich gemacht werden.							
F 5.1 Die zentralen Evaluierungsergebnisse werden zielgruppengerecht aufbereitet.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass für die Kommunikation der Ergebnisse an die beteiligten und betroffenen Stakeholder spezifische Formate entwickelt wurden. (AV-Befragung)							X
F 5.2 Die zentralen Evaluierungsergebnisse werden mit den Beteiligten und Betroffenen geteilt.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die zentralen Ergebnisse mit den beteiligten und betroffenen Stakeholdern geteilt wurden. (AV-Befragung)			X				X

Ergebnisse im Evaluierungsstandard Nützlichkeit

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
N 1 Identifizierung der Beteiligten und Betroffenen	Die am Evaluationsgegenstand beteiligten oder von ihm betroffenen Personen bzw. Personengruppen sollen identifiziert werden, damit deren Interessen geklärt und so weit wie möglich bei der Anlage der Evaluation berücksichtigt werden können.							
N 1.1 Die potentiell beteiligten und betroffenen Personen wurden systematisch und frühzeitig identifiziert.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die Interessen bestehender und potentieller Stakeholder (z.B. politische Träger, DO, weitere Kooperationspartner, Zielgruppe, Auftraggeber) in Bezug auf die Evaluierung systematisch und frühzeitig identifiziert wurden. (AV-Befragung)							X
N 1.2 Die potentiell beteiligten und betroffenen Personen wurden, wann immer möglich, einbezogen.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die bestehenden und potentiellen Stakeholder in Bezug auf die Evaluierung, wann immer möglich, einbezogen wurden. (AV-Befragung)						X	X
N 2 Klärung der Evaluationszwecke	Es soll deutlich bestimmt sein, welche Zwecke mit der Evaluation verfolgt werden, so dass die Beteiligten und Betroffenen Position dazu beziehen können und das Evaluationsteam einen klaren Arbeitsauftrag verfolgen kann.							
N 2.1 Der Zweck der Evaluierung wurde frühzeitig ausgehandelt.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die Zwecke der Evaluierung frühzeitig mit bestehenden und potentiellen Stakeholdern abgesprochen wurden. (AV-Befragung)						X	X
N 3 Glaubwürdigkeit und Kompetenz des Evaluators / der Evaluatorin	Wer Evaluationen durchführt, soll persönlich glaubwürdig sowie methodisch und fachlich kompetent sein, damit bei den Evaluationsergebnissen ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird.							
N 3.1 Das PEV-Team wird als kompetent und dadurch glaubwürdig eingeschätzt.	Der Indikator ist dann erfüllt, wenn das PEV-Team laut Auftraggeber die folgenden Kompetenzen abgedeckt hat: fachliche Kompetenz (im Sektor), methodische Kompetenz (Evaluierungsmethoden), Beratungskompetenz (z.B. Capacity WORKS, Capacity Development), Prozesskompetenz (GIZ-Auftragsmanagement, Steuerung einer PEV). (AV-Befragung)				X		X	X
N 3.2 Die ToR beschreiben das Profil, welches das Evaluationsteam abdecken sollte.	Der Indikator ist dann erfüllt, wenn anhand der Beschreibung des gesuchten Profils des PEV-Teams in den ToR hervorgeht, dass sich der Auftraggeber mit der Glaubwürdigkeit des Evaluationsteams befasst hat. Hierzu sollten die folgenden Kompetenzen in dem gesuchten Profil des Evaluationsteams in den ToR abgedeckt werden: fachliche Kompetenz (im Sektor), methodische Kompetenz (Evaluierungsmethoden), Beratungskompetenz (z.B. Capacity WORKS, Capacity Development), Prozesskompetenz (GIZ-Auftragsmanagement, Steuerung einer PEV). (Textanalyse der PEV-Berichte)				X		X	

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
N 4 Auswahl und Umfang der Informationen	Auswahl und Umfang der erfassten Informationen sollen die Behandlung der zu untersuchenden Fragestellungen zum Evaluationsgegenstand ermöglichen und gleichzeitig den Informationsbedarf des Auftraggebers und anderer Adressaten und Adressatinnen berücksichtigen.							
N 4.1 Die Evaluierung ist hinreichend informativ.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die in der Evaluierung gewonnenen Informationen die wesentlichen vereinbarten Fragestellungen der Evaluierung abdecken. (AV-Befragung)						X	
N 5 Transparenz von Werten	Die Perspektiven und Annahmen der Beteiligten und Betroffenen, auf denen die Evaluation und die Interpretation der Ergebnisse beruhen, sollen so beschrieben werden, dass die Grundlagen der Bewertungen klar ersichtlich sind.							
N 5.1 Es wird eine Forschertriangulation durchgeführt.	Der Indikator ist dann erfüllt, wenn das PEV-Team seine Erfahrungen und Interpretationen / Perspektiven trianguliert. Hinweise auf die systematische Überprüfung der gesammelten Informationen sind z.B. Synthesetreffen, die dokumentiert wurden, oder die Darstellung auf Unstimmigkeiten der Mitglieder des PEV-Teams, die entsprechend diskutiert wurden. (Textanalyse der PEV-Berichte)			X				
N 5.2 Es wird eine Triangulation der Ergebnisse mit den Beteiligten und Betroffenen durchgeführt.	Der Indikator ist dann erfüllt, wenn eine Triangulation der Ergebnisse mit den Beteiligten und Betroffenen erfolgt. So soll eine Fehlerüberprüfung der Analyseergebnisse erfolgen. Hinweise auf die systematische Überprüfung der gesammelten Informationen sind z.B. gemeinsame Syntheseworkshops oder die Kommentierung des Berichts durch Beteiligte und Betroffene. (Textanalyse der PEV-Berichte)			X				
N 6 Vollständigkeit und Klarheit der Berichterstattung	Evaluationsberichte sollen alle wesentlichen Informationen zur Verfügung stellen, leicht zu verstehen und nachvollziehbar sein.							
N 6.1 Die annotierte Berichtsgliederung wird angewandt.	Der Indikator ist dann erfüllt, wenn die inhaltlichen und formalen Vorgaben der annotierten Berichtsgliederung eingehalten werden. (Textanalyse der PEV-Berichte)	X				X		
N 7 Rechtzeitigkeit der Evaluation	Evaluationsvorhaben sollen so rechtzeitig begonnen und abgeschlossen werden, dass ihre Ergebnisse in anstehende Entscheidungsprozesse bzw. Verbesserungsprozesse einfließen können.							
N 7.1 Der Evaluierungsprozess wird rechtzeitig durchgeführt.	Der Indikator ist dann erfüllt, wenn der Evaluierungsprozess zu einem für die Nutzung der Erkenntnisse günstigen Zeitpunkt durchgeführt wurde. (AV-Befragung)			X				X
N 7.2 Evaluationsergebnisse stehen rechtzeitig zur Verfügung.	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass der Bericht zu einem für die Nutzung passenden Zeitpunkt vorlag. (AV-Befragung)			X				X

Bewertungskriterien / Indikatoren / Deskriptoren	Definitionen	Quellen zur Ableitung des Indikators						
		Annotierte Gliederung	Checklist OECD	Method. Vorgehen PEV	Muster TOR	Checklist QS	Methoden Toolbox	M&E Policy
N 8 (A) Nutzung und Nutzen der Evaluation: Nützlichkeit des Evaluierungsprozess und -berichts	Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation aufmerksam zur Kenntnis zu nehmen und ihre Ergebnisse zu nutzen.							
N 8.1 Der Evaluierungsprozess (Planung, Durchführung, Nachbereitung) wird als nützlich bewertet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsprozess dem Nutzenerwartungen des Auftraggebers entspricht. Dies ist dann der Fall, wenn der Evaluierungsprozess einen Beitrag zum Lernen geleistet hat. (AV-Befragung)						X	X
N 8.2 Der Evaluierungsbericht wird als nützlich bewertet.	Der Indikator ist dann erfüllt, wenn der Evaluierungsbericht dem Nutzenerwartungen des Auftraggebers entspricht. Dies ist dann der Fall, wenn der Bericht einen Beitrag zum Lernen geleistet hat. (AV-Befragung)							
N 8 (B) Nutzung und Nutzen der Evaluation: Nutzung der Evaluierungsergebnisse	Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation aufmerksam zur Kenntnis zu nehmen und ihre Ergebnisse zu nutzen.							
N 8.3 Die Ergebnisse der Evaluierung werden genutzt für die GIZ-interne Projekt-/ Programmebene.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für den Austausch zwischen den Mitarbeiter/-innen, den Austausch im Landesbüro, die Steuerung des laufenden Vorhabens, die Planung des Angebots und die Steuerung der Folgephase. (AV-Befragung)						X	X
N 8.4 Die Ergebnisse der Evaluierung werden genutzt für die GIZ-interne Sektor-/ Regionalebene.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für die Fachverbandsarbeit im Sektor, die Reflexion zu Verbesserungen in den fachlichen und regionalen Ansätzen und das unternehmensweite Wissensmanagement. (AV-Befragung)						X	X
N 8.5 Die Ergebnisse der Evaluierung werden genutzt für die Ebene des Kooperationssystems.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für die direkte Beratung der Counterparts in der Partnerorganisation (personelle Ebene), die Organisationsentwicklung von Partnerorganisationen (organisationale Ebene), die Gestaltung politischer Reformprozesse (Mikro-, Meso-, Makroebene) und die Verbesserung der Erreichung der Zielgruppen. (AV-Befragung)						X	X
N 8.6 Die Ergebnisse der Evaluierung werden genutzt für die Ebene des erweiterten Kooperationssystems.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für Fach- und Strategiediskussionen mit dem BMZ, Fach- und Strategiediskussionen mit anderen Auftraggebern (außer BMZ) und die Geberabstimmung im Sektor. (AV-Befragung)						X	X
N 8.7 Die Ergebnisse der Evaluierung werden genutzt für Sonstiges.	Der Indikator ist dann erfüllt, wenn laut Auftraggeber die Ergebnisse der Evaluierung genutzt werden für den Austausch mit weiteren Akteuren. (AV-Befragung)						X	X
N 8.8 Die Ergebnisse der Evaluierung werden genutzt für die Verbesserung der Folgemaßnahme	Der Indikator ist dann erfüllt, wenn der Auftraggeber bestätigt, dass die Ergebnisse der Evaluierung einen positiven Einfluss haben auf die Relevanz, Effektivität, Effizienz, Nachhaltigkeit und den Impact der Folgemaßnahme. (AV-Befragung)							X

WISSEN WAS WIRKT
WISSEN WAS WIRKT
WISSEN WAS WIRKT
WISSEN WAS WIRKT
WISSEN WAS WIRKT
WISSEN WAS WIRKT
WISSEN WAS WIRKT

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Sitz der Gesellschaft
Bonn und Eschborn

Friedrich-Ebert-Allee 40
53113 Bonn
T +49 228 44 60-1877
F +49 228 44 60-2877

Dag-Hammarskjöld-Weg 1-5
65760 Eschborn
T +49 61 96 79-1408
F +49 61 96 79-80 1408

E evaluierung@giz.de
I www.giz.de/monitoring