



18. Evaluierungsbericht 2023–2024

Begleiten. Bewerten. Lernen.



Vorwort

Liebe Leserinnen und Leser,

in einer Zeit, die von Volatilität und geopolitischen Unsicherheiten geprägt ist, stehen wir sowohl in Deutschland als auch in unseren Partnerländern vor großen Herausforderungen. Die Bundestagswahl und die kürzlich abgeschlossenen Koalitionsverhandlungen haben gezeigt, dass die Entwicklungszusammenarbeit stärker denn je ins öffentliche Bewusstsein gerückt ist und ihr Nutzen für Deutschland hinterfragt wird. Ein Trend, den wir in unseren europäischen Partnerländern und nicht zuletzt in den USA beobachten können.

Der zunehmend kritischen öffentlichen Debatte über die Wirksamkeit der Entwicklungszusammenarbeit wollen wir mit Fakten und guten Argumenten begegnen. Der zweijährliche Evaluierungsbericht erscheint seit mehr als drei Jahrzehnten. Seitdem hat die Messung unserer Wirkungen und deren transparente Darstellung in der öffentlichen Debatte noch einmal deutlich an Bedeutung gewonnen. In der aktuellen Weltlage sind sie ein wichtiger Schlüssel, um das Vertrauen in unsere Maßnahmen zu stärken und die Relevanz der Entwicklungszusammenarbeit zu unterstreichen.

„Begleiten. Bewerten. Lernen.“ ist das Motto des Evaluierungsberichts. Das Kapitel „Begleiten“ präsentiert die Arbeit des im Jahr 2023 gegründeten Development Impact Labs – ein Instrument, bei dem wir den Fokus noch stärker auf die Wirkungsketten unserer Finanzierungen legen. In einer Übersicht über die weltweit laufenden Impact Evaluierungen kommen viele Kolleginnen und Kollegen der KfW Entwicklungsbank zu Wort und teilen ihre Erfahrungen. Außerdem gibt es spannende Einblicke in Impact Evaluierungen von Vorhaben aus Vietnam, Tansania und Madagaskar. Ich als Vorständin freue mich zu sehen, wie viel hier in so kurzer Zeit gelungen ist.



Christiane Laibach

Mitglied des Vorstandes der KfW Bankengruppe, zuständig für Entwicklungszusammenarbeit sowie Export- und Projektfinanzierung.

1991



Wir schauen zurück auf eine lange Historie: Seit 1991 veröffentlicht die KfW Entwicklungsbank ihre Evaluierungsergebnisse.

Im Kapitel „Bewerten“ finden Sie eine Weltkarte mit Stimmen operativer Kolleginnen und Kollegen, die für eine Ex-post-Evaluierung im Einsatz waren, sowie einen Streifzug durch die wichtigsten Evaluierungsergebnisse der letzten zwei Jahre. Ex-post-Evaluierungen sind unser langjähriges Instrument, um öffentlich Rechenschaft abzulegen und für Transparenz zu sorgen. Besonders hervorzuheben sind die aufschlussreichen Lessons Learned aus Evaluierungen von Vorhaben im Finanz-, Energie- und Bildungssektor.

Wir wollen uns kontinuierlich verbessern. Daher zeigt uns das letzte Kapitel – „Lernen“ –, was wir aus diesen Ergebnissen lernen können. Diesmal liegt der Fokus auf einer Evaluierung der Kooperation zwischen der technischen und finanziellen Entwicklungszusammenarbeit. Die Ergebnisse verdeutlichen an mehreren Stellen, wie gut wir gemeinsam die Vorhaben in unseren Partnerländern voranbringen, aber auch, wo es noch Entwicklungspotenziale gibt. Zuletzt gibt es dann einen besonders aktuellen Beitrag: Jochen Kluge, Leiter der Evaluierungsabteilung, und Estelle Raimondo, Head of Methods in der Evaluierungsabteilung der Weltbank diskutieren, wie Künstliche Intelligenz in der Evaluierung von Entwicklungszusammenarbeit genutzt werden kann, und welche Risiken und Chancen sie uns bietet.

Ich lade Sie ein, diesen Bericht zu erkunden und die vielfältigen Erkenntnisse zu entdecken, die wir aus unseren Evaluierungen gewonnen haben. Nutzen Sie die Ergebnisse über unsere Wirkungen für Gespräche im beruflichen, aber auch im privaten Umfeld. Ich bin überzeugt, dass jede und jeder Einzelne von uns einen Beitrag zur aktuellen Debatte leisten kann.

Herzlichst,

Christiane Lebel

2025



Inhalt



Begleiten

6-17

Impact Evaluierungs-Champions weltweit	8-9
KfW Development Impact Lab: Wirkungen messen und verstehen	10-11
Impact Evaluierungen in der Praxis	12-17
Erste Ergebnisse aus dem Simiyu Climate Resilience Projekt in Tansania	12-13
Zwischen Biodiversitätsschutz und wirtschaftlicher Entwicklung in Madagaskar	14-15
Eine Wetterversicherung für die Landwirtschaft in Vietnam	16-17



Lernen

32-43

Künstliche Intelligenz: Chancen und Risiken für die Bereitstellung von Evidenz	34-41
Kooperation in der deutschen Entwicklungszusammenarbeit	42-43



Bewerten

18-31

Ex-post-Evaluierungs-Champions weltweit

20-21

Ex-post-Evaluierungen 2023-2024

22-31

Streifzug durch Sektoren und Regionen

22-25

Die Erfolgsquote der Vorhaben

26-27

Drei Sektoren im Fokus: Lessons Learned

28-31



Methodik

44-48

Ex-post-Evaluierung: Von der Grundgesamtheit bis zur Erfolgsquote

46-47

FZ Evaluierung stellt sich vor

48

Evaluierungsbericht 2023-2024: Begleiten



Von Anfang an dabei

Mit Beginn des Vorhabens beginnt häufig auch die Impact Evaluierung. Gemeinsam mit den Partnern werden die Wirkungen der Vorhaben untersucht.



Impact Evaluierungs-Champions weltweit

Die Anzahl der Impact Evaluierungen von Projekten der Finanziellen Zusammenarbeit (FZ) wächst kontinuierlich. Diese Evaluierungen werden gemeinsam mit zumeist wissenschaftlichen externen Partnern durchgeführt und von der Evaluierungsabteilung der KfW fachlich eng begleitet – in einigen Fällen auch eigenständig realisiert. So wird sichergestellt, dass die Ergebnisse für laufende und zukünftige FZ-Vorhaben, aber auch darüber hinaus, relevant sind. Die Mitwirkung unserer Projektpartner sowie der lokalen Fachkräfte der KfW vor Ort, zusammen mit den KfW-Projektmanagerinnen und -managern, ist entscheidend für den Erfolg der Evaluierungen und bedeutet das ein oder andere Mal auch einen Sprung in ungewohnte Gewässer. Hier teilen einige der KfW Impact-Champions ihre Erfahrungen und Beweggründe.



„Die Impact Evaluierung von FZ Evaluierung und Wissenschaftlerinnen und Wissenschaftlern der Universität Groningen unterstützte das Projektteam schon früh im Projektzyklus bei der Diskussion von Wirkungshypothesen und Interventionsstrategien. Zudem half sie dabei, eine systematische und transparente Strategie zur Auswahl der Projektteilnehmenden zu entwickeln: In Honduras werden diese nun zufällig aus den offiziellen Listen der zurückgekehrten Migrantinnen und Migranten ausgewählt und dann durch das Sozialministerium telefonisch zur Projektteilnahme eingeladen. Wir stehen noch am Anfang der Evaluierung, aber die datenorientierte Begleitung hat schon jetzt spannende Impulse geliefert.“

Elisabeth Hoffmann und Liliana Monteiro,
Berufliche und soziale Integration
in Honduras, Guatemala und El Salvador,
mit der Universität Groningen



„Es ist spannend, durch die Begleitevaluierung mehr über die tatsächlichen Wirkungen des Vorhabens zu erfahren. Insbesondere den Einfluss der Maßnahmen auf die Gesundheit und Bildung/Beschäftigung von Frauen und Kindern erheben zu können, ist viel wert und motiviert uns als Projektteam zusätzlich. Auch von Partnerseite wird die Evaluierung sehr positiv wahrgenommen. Das Vorhaben profitiert durch die Option, durch die Wirkungsevaluierung schon während der Umsetzung des Vorhabens Ergebnisse der Haushaltsbefragungen zu erhalten und diese ggf. in der Projekt(um)steuerung nutzen zu können.“

Freya von Negenborn,
Trinkwasserversorgung in Tansania,
mit der KfW-Evaluierungsabteilung

 Multisektorale UN-Vorhaben / Fragilität

 Bildung

 Finanzwesen

 Energie

 Landwirtschaft und Naturressourcenschutz

 Gesundheit

 Wasser



„Die Tracer- und Impact-Studie ermittelte die kurz- bis mittelfristigen Wirkungen unseres Skills-Vorhabens auf die Teilnehmenden, insbesondere deren Beschäftigung, Vermittelbarkeit, Lebensstil und psychosoziales Wohlbefinden. Die Studie bestand aus drei Datenerhebungen zwischen 2022 und 2023 unter Teilnehmenden und Nichtteilnehmenden sowie einer Mischung aus quantitativen und qualitativen Erhebungsmethoden. Ergänzt wurde dies durch Einzelinterviews, Gruppendiskussionen und Umfragen unter Arbeitgebern. American Institutes for Research (AIR), UNICEF, FZ Evaluierung und das KfW-Projektteam arbeiteten in der Umsetzung der Studie gut und eng zusammen. Die Ergebnisse wurden mit dem Bundesministerium für wirtschaftliche Zusammenarbeit (BMZ) und auf Plattformen geteilt und stellen einen wichtigen Meilenstein der Impact-Messung im non-formalen Berufsbildungskontext dar.“

Anne Wessendorf und Leonard Dlubatz,
Berufsbildung im Libanon,
mit AIR



„Ich bin begeistert, dass sich Igloo, als schnell wachsendes InsurTech, und BlueOrchard, als Impact Fund Manager, mit großem Engagement auf eine mehrjährige Impact Evaluierung eingelassen haben. Im Bereich der Klimarisikobewertungen gibt es bisher nur sehr wenige wissenschaftlich fundierte Erkenntnisse über entwicklungspolitische Wirkungen. Die Studie soll helfen, weitere Ansätze der Klimarisikofinanzierung, gerade auch unter dem Global Shield against Climate Risks, besser zu gestalten.“

Stefan Hirche,
Klimarisikoversicherung in Vietnam,
mit dem RWI Essen



„Kann eine verbesserte Trinkwasserversorgung den sozialen Zusammenhalt zwischen Geflüchteten und lokaler Bevölkerung steigern? Wir schreiben dies häufig ohne klaren Nachweis in unseren Wirkungsmatrizen. Meine Kollegin von FZ Evaluierung und ich überprüfen dies datenbasiert in einer Impact Evaluierung gemeinsam mit Professor Anselm Hager der Humboldt-Universität zu Berlin. Angesichts der Herausforderungen in einem fragilen Kontext ist das nicht einfach, vor allem wenn das akademische Anspruchsniveau nicht kompromittiert werden soll: Kämpfe, Überflutungen, Geheimdienste, die sich Gedanken über die Datenerhebung machen. Um dennoch ein robustes Ergebnis zu erzielen, wurden wir kreativ und haben die ersten beiden Baseline-Datenerhebungen mittlerweile erfolgreich abgeschlossen. Die Antwort auf unsere Evaluierungsfrage wird es allerdings erst nach der letzten Datenerhebung in etwa 4 Jahren geben.“

Daniela Henrike Klau-Panhans,
Trinkwasserversorgung in Somalia, Sudan, Äthiopien,
mit der HU Berlin



„Die Impact Evaluierung ermöglicht es uns, die Wechselwirkungen zwischen Naturschutz und den sozioökonomischen Bedingungen der angrenzenden Gemeinden umfassender zu verstehen und sogar zu quantifizieren. Diese Erkenntnisse sind für unsere operative Arbeit von entscheidender Bedeutung, da der Erfolg unserer Maßnahmen maßgeblich davon abhängt, ob die lokale Bevölkerung den Naturschutz unterstützt. Die enge Zusammenarbeit mit unseren lokalen Partnern und die Einbeziehung aller Perspektiven in das Evaluierungsdesign haben die Relevanz und zukünftige Nützlichkeit der Ergebnisse sicher erheblich gesteigert.“

Sebastian Manthey,
Biodiversitätsschutz in Madagaskar,
mit dem IRD

KfW Development Impact Lab

Wirkungen messen und verstehen

Das KfW Development Impact Lab steuert seit 2023 als Teil von FZ Evaluierung die Impact Evaluierungen (IE) von Vorhaben der Finanziellen Zusammenarbeit (FZ). Hauptaufgabe des Labs ist es, wichtige operative Fragestellungen zu Projektwirkungen mit modernen quantitativen Methoden und Daten zu beantworten. Häufig wird dies in Zusammenarbeit mit der Wissenschaft umgesetzt; das Impact Lab fungiert hier als Brücke zwischen den wissenschaftlichen und den praxisnahen Zielen einer IE.

Impact Evaluierungen ergänzen Ex-post-Evaluierungen (EPEs) von KfW-Vorhaben. Während EPEs einige Jahre nach Projektende entlang der sechs OECD-DAC Kriterien Relevanz, Kohärenz, Effizienz, Effektivität, Impact, Nachhaltigkeit den Projekterfolg beurteilen (siehe Kapitel Bewerten ab Seite 18) fokussieren Impact Evaluierungen auf die Messung von kausalen Projektwirkungen (Attribution) und begleiten ein FZ-Vorhaben idealerweise von Anfang an und über die gesamte Laufzeit. Aufgrund des größeren Aufwands werden IE gezielt dort eingesetzt, wo das Erkenntnisinteresse besonders groß ist (siehe Weltkarte der Impact Evaluierungs-Champions, S. 8).

Das Ziel jeder Impact Evaluierung ist es, die Wirkungen eines Vorhabens kausal zu messen. Das heißt: Wirkungen eindeutig einem FZ-Vorhaben zuzuordnen bzw. auszuschließen, dass Wirkungen auch ohne das FZ-Vorhaben zustande gekommen wären – zum Beispiel aufgrund von Wirtschaftswachstum oder veränderter

klimatischer Bedingungen. Methodisch vergleichen IE dazu die tatsächlich erreichten Ergebnisindikatoren in der Teilnehmegruppe – d.h. jene Personen, Unternehmen, oder Regionen auf die das FZ-Vorhaben abzielt – mit denen einer Vergleichsgruppe, d.h. ähnliche Personen, Unternehmen, Regionen, die nicht Teil des Vorhabens sind („no action“ Szenario).

Die Methodik ist vielfältig: Es werden sowohl Haushaltsbefragungen (siehe Beispiel Vietnam auf S. 16) durchgeführt als auch Satelliten- oder andere Sekundärdaten (siehe Beispiel Madagaskar auf S. 14) analysiert. Die Analysen beantworten nicht nur quantitativ wie groß die kausalen Wirkungen sind, sondern auch warum sie (ggf. nicht) eingetroffen sind und welche Bevölkerungsschicht am stärksten davon profitierte. Das KfW Development Impact Lab entwickelt hierzu für jede IE eines FZ-Vorhabens das bestmögliche methodische Design, welches an die Gegebenheiten des Vorhabens angepasst wird („form follows function“) und den Projektablauf unterstützend begleiten kann. So wird sichergestellt, dass Erkenntnisse aus der IE sowohl für das laufende als auch für künftige Vorhaben von Nutzen sein können.



Die Impact Evaluierungen nutzen digitale Lösungen für die Datenerhebung - im Feld oder per Satellit.

Das KfW Development Impact Lab von 2023-2024 in Zahlen



Zusammenarbeit mit sieben
Universitäten sowie fünf
Forschungsinstituten



19 Impact Evaluierungen
in allen Sektoren der
finanziellen Zusammenarbeit



Messung von **Wirkung**
pro eingesetztem Euro



Kooperation mit
Weltbank, UNICEF, WFP,
FAO, AFD und FCDO bei der
Implementierung
der Impact Evaluierungen



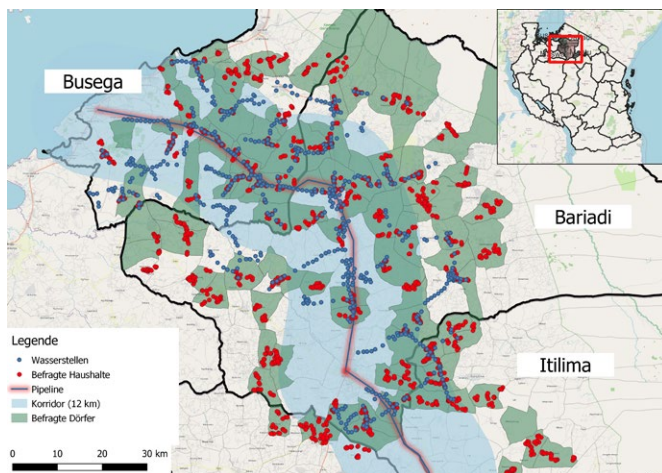
**Nutzung moderner
Methoden**
Ökonometrische Modelle,
Machine Learning und
Qualitative Methoden

Impact Evaluierung

Erste Ergebnisse aus dem Simiyu Climate Resilience Projekt in Tansania

Das Simiyu Climate Resilience Projekt (SCRP) wird gemeinsam finanziert durch die Bundesregierung über die KfW, den Green Climate Fund sowie die tansanische Regierung. Mit dem Projekt soll die Resilienz von 495.000 Menschen in der ländlichen, trockenen Simiyu-Region im Norden Tansanias gegenüber den Auswirkungen des Klimawandels gestärkt werden: Durch die Verbesserung der Wasserversorgung, der Sanitärversorgung und durch die Förderung klimaangepasster Landwirtschaft.

Kern des Projekts ist der Bau einer 100 km langen Wasserpipeline, die Wasser aus dem Lake Victoria aufbereitet, in das Inland leitet und der Bevölkerung zur Verfügung stellt: Sauber, sicher und ganzjährig. Die Erwartungen an das SCRП sind hoch. Das Wasser soll zuverlässig bei den Menschen ankommen sowie von diesen genutzt und bezahlt werden können. Die Wasserpipeline soll so Krankheiten und die Belastung von Frauen und Kindern verringern, die bisher weite Wege zurücklegen müssen, um Wasser – von meist schlechter Qualität – zu holen.

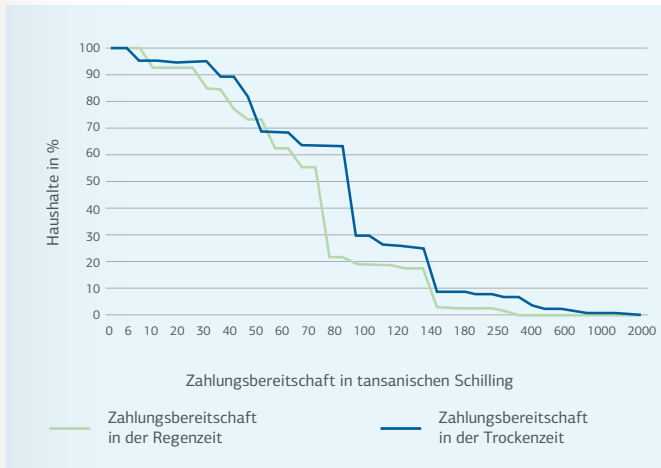


Die Karte zeigt die Projektregion Simiyu, den geplanten Bau der Pipeline und der öffentlichen Wasserpumpen (blaue Punkte), die im Rahmen der Untersuchung zufällig ausgewählten Dörfer (grüne Flächen) und die Standorte der interviewten Haushalte (rote Punkte). Rund zwei Drittel der Dörfer befinden sich innerhalb des Versorgungskorridors des Projektes (hellblau-schraffierte Flächen). Das andere Drittel liegt außerhalb des Korridors.

Die Impact Evaluierung, die vom KfW Development Impact Lab der KfW Entwicklungsbank durchgeführt wird, begleitet das Vorhaben seit seinen Anfängen. Im Jahr 2024, vor Beginn der Bauarbeiten, fuhr ein Interviewer-Team des tansanischen Forschungspartners „ideas in Action“ (iAct) in 90 zufällig ausgewählte Dörfer der Simiyu-Region und befragte 90 Dorfvorsteher und rund 1.800 Haushalte. Im Mittelpunkt der Befragung stand der individuelle Wasserkonsum und die Zahlungsbereitschaft für sauberes Wasser. Außerdem führten Wasseranalytikerinnen und Wasseranalytiker mit tragbaren Laboren an 270 Wasserquellen Proben durch, um deren chemische und physische Parameter zu bestimmen.

Die Ergebnisse unterstützen die Umsetzung des Vorhabens mit Informationen über Absatzmengen und Zahlungsbereitschaft der zukünftigen Konsumentinnen und Konsumenten, die insbesondere für die Tarifgestaltung und damit den nachhaltigen Betrieb der geschaffenen Anlagen von hoher Relevanz sind. So wird überprüft, ob die Prognosen des Vorhabens mit den Realitäten der Menschen übereinstimmen. Die bisherigen Ergebnisse zeigen beispielsweise, dass Haushalte im Durchschnitt weniger als 20 Liter pro Tag und Kopf für Trinken, Kochen und Waschen konsumieren. Häufig gehen Projektplanungen von deutlich höheren Werten aus. Die detaillierte Datenerhebung ermittelt und visualisiert entscheidungsrelevante Parameter wie Zahlungsbereitschaft, Krankheitsinzidenz oder Wasserqualität auf Dorfebene: Damit lassen sich z.B. im Projektgebiet Orte identifizieren, die beim Anschluss an das Wassersystem priorisiert werden können, um die Wirksamkeit des Vorhabens weiter zu erhöhen.

Die Methodik ist vielfältig: Die Ergebnisse belegen auch die Relevanz und potenzielle Wirkung des Vorhabens. Die Mehrheit der befragten Dorfvorsteher nennt eine unzureichende Wasserversorgung als wesentliches Entwicklungshemmnis. Wasser ist äußerst knapp in der Region, vor allem während der insgesamt rund acht Monate andauernden Trockenperiode. Die Menschen vor Ort verbringen enorm viel Zeit mit dem Wasserholen: In der Trockenzeit durchschnittlich drei Stunden pro Tag. Dies geht zu Lasten der Schulbildung, denn in mehr als einem Fünftel der befragten Haushalte sind Kinder und Jugendliche – vor allem Mädchen – für das Wasserholen zuständig. Jede vierte Familie berichtet dabei, dass ein oder mehrere ihrer Kinder regelmäßig Unterrichtszeit verpassen, weil sie Wasser holen müssen.

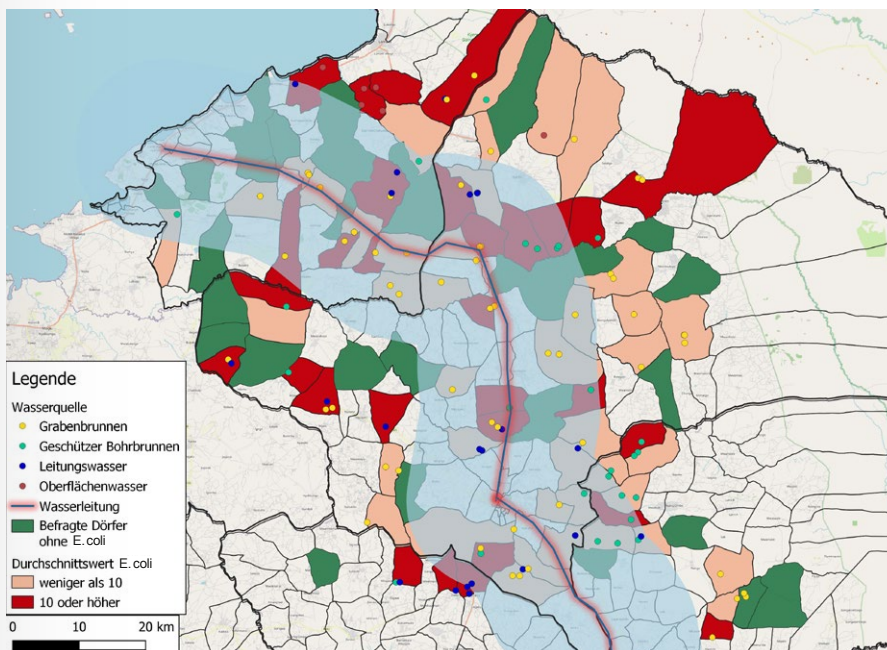


Darstellung der Zahlungsbereitschaft für sauberes Wasser in der Simiyu-Region.

Zahlungsbereitschaft für sauberes Wasser: Ein Schwerpunkt der Untersuchung war die Ermittlung der Zahlungsbereitschaft der Menschen in der Simiyu-Region für sauberes Wasser. Die Ergebnisse in der Abbildung zeigen, dass mehr als 90 % der Befragten bereit wären, für sauberes Wasser zu zahlen. In der Grafik sind die Nachfragekurven der Haushalte für einen 20-Liter-Eimer sicheres Wasser in der Trocken- und der Regenzeit eingezeichnet. Der fallende Verlauf der Kurven zeigt, dass der Anteil der Haushalte, welcher einen bestimmten Preis bezahlen würde, sinkt, wenn der Wasserpreis steigt. Auch sind die Menschen bereit, in der Trockenzeit mehr zu zahlen als in der Regenzeit. In der Trockenperiode haben die Menschen eine sehr hohe Zahlungsbereitschaft, was die Bedeutung von Wasser unterstreicht: Sie sind bereit, für ihren Bedarf an sauberem Wasser durchschnittlich 12 % der gesamten Haushaltsausgaben zu investieren.



Das 17-jährige Mädchen in Kigoma, Tansania, macht sich auf den Weg zum Wassersammeln: Der Weg zum See dauert 20 Minuten hin und 20 Minuten zurück und das Wasser schmeckt salzig.



Übersicht der kontaminierten Wasserquellen.

Überschreitung von erlaubten E. coli-Werten: In den Dörfern wurden Wasserproben an insgesamt 272 von den Menschen genutzten Wasserquellen entnommen. Dabei kamen mobile Wasserlabore zum Einsatz. Die Abbildung zeigt jene Wasserquellen, in denen die E. coli-Konzentration in der Probe deutlich über dem nationalen Schwellenwert liegt: 54 % der Wasserquellen zeigen zu schlechte Werte.

Escherichia coli ist eine Bakterienart, die im Darm von Menschen und Tieren vorkommt. Einige Stämme von E. coli sind harmlos, während andere pathogene Stämme Krankheiten verursachen können. E. coli wird häufig als Indikator für die Wasserqualität verwendet, da das Vorhandensein dieser Bakterien auf eine mögliche Kontamination mit Fäkalien hinweisen kann. Die Ergebnisse legen nahe, dass eine Wasserversorgung per Pipeline mit sicherem, sauberem Wasser Krankheitsfälle in der Projektregion reduzieren wird.

Impact Evaluierung

Zwischen Biodiversitätsschutz und wirtschaftlicher Entwicklung in Madagaskar

Madagaskar ist ein globaler Biodiversitäts-Hotspot.

Allerdings gehört das Land zu den ärmsten Ländern der Welt und belegte im Jahr 2024 Platz 177 von 193 laut UNDP's Index der menschlichen Entwicklung; die WHO prognostiziert zudem ein Bevölkerungswachstum von 70 % bis 2050. Bei hoher Armut und steigender Bevölkerung ist es nicht verwunderlich, dass die Nachfrage nach landwirtschaftlichen Flächen steigt und einen enormen Nutzungsdruck auf die Wälder ausübt. Der ökologische Schutz und die wirtschaftliche Entwicklung scheinen hier miteinander zu konkurrieren.

Die Impact Evaluierung untersucht Spannungen und Synergien zwischen Umweltschutz und den wirtschaftlichen Bedürfnissen der lokalen Gemeinden. Das projektnahe Forschungsprojekt BETSAKA (eng.: Biodiversity-Economic Tradeoff and Synergy Assessments for Conservation Areas) betrachtet die Entwicklungen von Naturschutzwäldern in Madagaskar von 2000 bis 2024, welche durch die Finanzielle Zusammenarbeit (FZ) und die französische Entwicklungsagentur Agence Française de Développement (AFD) unterstützt wurden. BETSAKA beantwortet datenbasiert vier bisher wenig untersuchte Fragen, die für die KfW als einer der weltweit größten Naturschutz-Financiers zentral sind:

- Reduziert die KfW-Schutzgebietsfinanzierung Abholzung in Madagaskar?
- Begrenzt die Finanzierung Brände in und um diese Gebiete?
- Verbessert sie die Lebensbedingungen der Anwohnenden?
- Inwiefern werden diese Wirkungen durch Finanzierung, Governance-Systeme und Managementpraktiken der Schutzgebiete beeinflusst?

Diese Fragen werden in einer gemeinsamen Impact Evaluierung mit mehreren internationalen Partnern untersucht: Die Evaluierungsabteilung der KfW Entwicklungsbank, die Evaluierungsabteilung der AFD sowie das French National Research Institute for Sustainable Development (IRD). Hinzu kommt die Expertise der Universität Antananarivo, der Madagaskar National Parks (MNP) und der Stiftung für Schutzgebiete und Biodiversität (FAPBM).

Die Impact Evaluierung nutzt mehrere methodische Instrumente, um die Auswirkungen von Schutzgebieten auf Waldverlust, Waldbrände und sozioökonomische Bedingungen von Anrainergemeinden ganzheitlich zu analysieren. Dies umfasst zum einen die Nutzung von georeferenzierten Satellitendaten von Schutzgebieten, um deren tatsächliche Waldflächenentwicklung mit den Entwicklungen in identischen, nicht-geschützten Gebieten

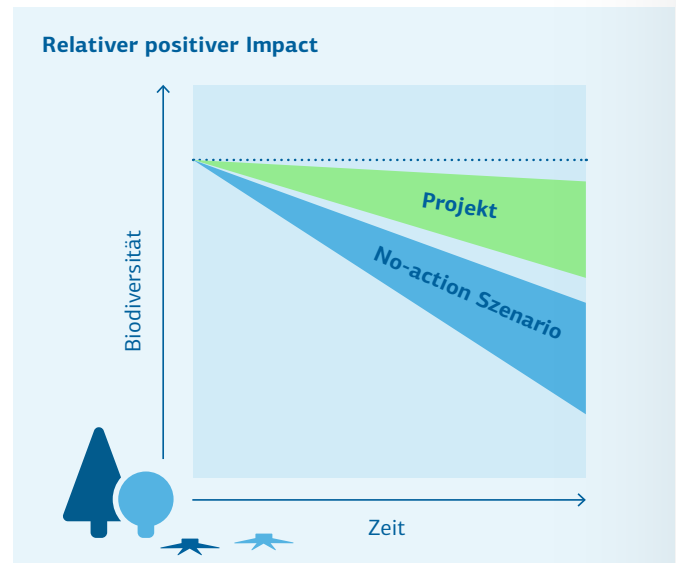


Illustration der typischen Wirkung von Schutzgebietsförderung auf Waldflächen: Signifikant mehr Walderhalt im Vergleich zum „no-action“ Szenario.

zu vergleichen („no-action“ Szenario). Außerdem werden sozio-ökonomische Haushaltsbefragungen durchgeführt, um die Wirkung der Schutzgebietsvorhaben auf die ansässigen Menschen zu bestimmen. Flankiert werden die beiden quantitativen Ansätze von Experteninterviews, um die Governance-Systeme der Schutzgebiete und deren Auswirkungen qualitativ besser einschätzen zu können.

Die Ergebnisse der Impact Evaluierung stärken die Projektumsetzung neben der Wirkungsmessung selbst auch durch die Bereitstellung von Daten und methodischen Tools zur Überwachung und Bewertung des Engagements der KfW in Madagaskar. Zum Beispiel hat das BETSAKA-Team eine offizielle Aktualisierung der Schutzgebietsgrenzen bei der World Database on Protected Areas (WDPA) erwirkt und somit Wissenschaft und Praxis Zugriff auf aktualisierte Daten gewähren können. Dies hilft der madagassischen Regierung und deutschen Entwicklungszusammenarbeit ganz konkret bei der zielgerichteten Planung weiterer Vorhaben. Die finalen Ergebnisse zur Wirksamkeit von Naturschutzpolitiken sind darüber hinaus nicht nur für Madagaskar und die KfW, sondern auch für andere Länder und Geber von Bedeutung.



Teilnehmerinnen und Teilnehmer eines Workshops zum BETSAKA Projekt an der Universität Antananarivo.

Ony Rakotoarisoa, Generaldirektorin der Madagaskar National Parks (MNP):



„Das BETSAKA-Projekt bietet einen großen Mehrwert hinsichtlich der Managementeffektivität und -effizienz der von den MNP verwalteten Schutzgebiete. Es ermöglicht uns, die entscheidenden Parameter zu identifizieren, auf die wir in unserem Management achten müssen, damit die Ökosystemdienstleistungen unserer Parks und Reservate sich positiv auf die sozioökonomischen Bedingungen der Anwohner auswirken und effektiv zur Erhaltung der Biodiversität in Madagaskar beitragen.“



Antje Göllner-Scholz, Referentin für Wirtschaftliche Zusammenarbeit (BMZ) in Madagaskar:

„Die außergewöhnliche Flora und Fauna Madagaskars steht durch den Verlust von Waldflächen und Biodiversität vor gravierenden Herausforderungen. Diese Verluste haben nicht nur lokale, sondern auch globale Auswirkungen auf das Klima und gefährden die Widerstandsfähigkeit der Wälder. Die deutsche Entwicklungszusammenarbeit unterstützt die madagassischen Nationalparks und deren Anrainer dabei, den Balanceakt zwischen dem Schutz der Biodiversität und der ökonomischen Entwicklung zu bewältigen und Synergien zu nutzen. Die BETSAKA-Evaluierung spielt eine zentrale Rolle, um unsere Erfolge zu verifizieren, Wissenslücken zu schließen und wertvolle Erkenntnisse für die Zukunft zu gewinnen.“



Ingrid Dallmann, Managerin für Impact Evaluierung, Agence Française de Développement (AFD):

„Die Zusammenarbeit im BETSAKA-Projekt beruht auf einer Idee des gegenseitigen Lernens: Sie vereint die Kompetenzen von Biodiversitätsexperten (MNP, FAPBM) mit denen der akademischen Welt (IRD, Universität Antananarivo) und den Entwicklungsbanken (AFD, KfW). Diese Teamarbeit ermöglicht es uns, unsere Bewertung umfassend durchzuführen, die Kapazitäten unserer Partner zu stärken – z.B. durch Austausch zu geo-räumlicher Folgenabschätzung oder die Unterstützung von Dissertationen. Außerdem lässt sie uns mit einer gemeinsamen und dadurch einflussreicheren Stimme sprechen. Die Zusammenarbeit in einem solchen Evaluierungsprojekt ist eine einzigartige Erfahrung, die ich meinen Kolleginnen und Kollegen an anderen Institutionen wärmstens empfehlen kann.“

Impact Evaluierung

Eine Wetterversicherung für die Landwirtschaft in Vietnam

Extreme Wetterereignisse werden als Folge des Klimawandels häufiger und mit größerer Intensität auftreten. Auch in Vietnam sind kleine landwirtschaftliche Betriebe zunehmend Starkregen, Tropenstürmen und Dürren ausgesetzt. Ohne wirksame Anpassung kann dies ihre Lebensgrundlage zerstören und ihr Armutsrisiko erhöhen. Wetterversicherungen sind für sie ein vielversprechendes Instrument zur Absicherung gegen Wetterrisiken.

Ein derartiges Versicherungsprodukt wurde vom singapurischen FinTech Start-up Igloo entwickelt und durch das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) über die KfW Entwicklungsbank und BlueOrchard Finance, einem Impact Investment Manager, gefördert: Kleine landwirtschaftliche Betriebe können sich durch eine Versicherungspolice gegen konkrete Wetterrisiken absichern. Pro Hektar Land kostet eine Police für 30 Tage 50.000 vietnamesische Dong (rund 2 Euro). Die versicherten Reisbäuerinnen und Reisbauern erhalten innerhalb von fünf Tagen eine Auszahlung falls starke Regenfälle auftreten. Da eine aufwändige individuelle Überprüfung der Schäden wegfällt, kann eine solche indexbasierte Wetterversicherung deutlich kostengünstiger als herkömmliche Agrarversicherungen angeboten werden.

Trotz des überzeugend konzipierten Anpassungsinstruments haben in Vietnam zunächst weniger Kleinbäuerinnen und Kleinbauern als erhofft eine Wetterversicherung erworben. Dies stellt die politischen Verantwortlichen vor Herausforderungen.

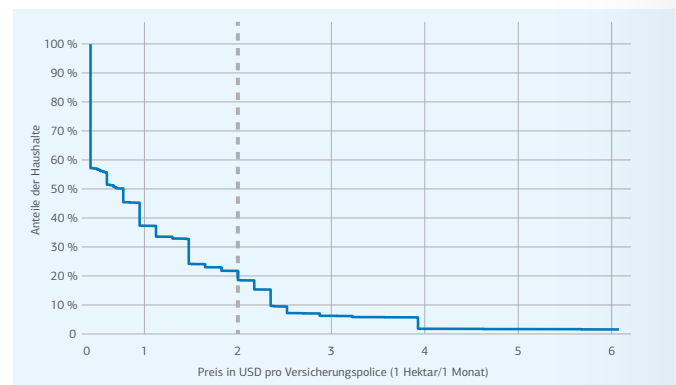
Die Impact Evaluierung begleitet das Projekt und beantwortet folgende Fragen:

- Wie hoch ist die Zahlungsbereitschaft der kleinen landwirtschaftlichen Betriebe für eine indexbasierte Wetterversicherung?
- Welche Faktoren erklären die bisher niedrige Nachfrage?
- Sind Prämienrabatte und Trainings wirksame Kanäle, um die Nachfrage nach einer Versicherung zu erhöhen?

Die Rigorose Wirkungsevaluierung einer Trigger-basierten landwirtschaftlichen Versicherung in Vietnam, wird durch Wissenschaftlerinnen und Wissenschaftler des RWI – Leibniz Institut für Wirtschaftsforschung e.V. umgesetzt. Gefördert wird das Evaluierungsprojekt von der KfW Entwicklungsbank und vom Deutschen Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval) aus Mitteln des Bundesministeriums für wirtschaftliche Zusammenarbeit und Entwicklung. Eine erste repräsentative Haushaltsbefragung unter rund 3.000 Kleinbetrieben in Vietnams Mekong Delta ermittelte u.a. die Zahlungsbereitschaft (Willingness-to-Pay) der Bäuerinnen und Bauern für eine indexbasierte Versicherung gegen Starkregen.

Erste Ergebnisse der Studie zeigen, dass die Zahlungsbereitschaft höher ist als erwartet. Ein bedeutender Anteil der befragten Kleinbetriebe hat Interesse an der Versicherung von Igloo: 56,8 % gaben eine positive Zahlungsbereitschaft an und 17,8 % nannten sogar eine Zahlungsbereitschaft über dem Marktpreis der Versicherung - ein überraschendes Ergebnis.

Auf Basis dieser Befragungsergebnisse ist es möglich, zwei Maßnahmen zur Steigerung der Nachfrage in zufällig ausgewählten Haushalten zu testen. Die erste Gruppe erhält einen Prämienrabatt: Einen Gutschein, mit dem kleine landwirtschaftliche Betriebe Igloos Versicherung gegen Starkregen zu vergünstigten Preisen erwerben können. Auch der Grad der Vergünstigung wird per Zufallsverfahren bestimmt und beträgt zwischen 10 % und 90 % Rabatt auf den regulären Marktpreis. Der zweiten Gruppe werden Trainings zum Thema Klimaresilienz und finanzielle Bildung angeboten.



Zahlungsbereitschaft für Igloos Versicherung gegen Starkregen.

Wie stärken die Ergebnisse der Impact Evaluierung die Vorhabenumsetzung? Erkenntnisse über die Zahlungsbereitschaft für eine Wetterversicherung helfen den Projektpartnern dabei, Hindernisse in der Nachfrage zu identifizieren und Strategien zu entwickeln, um diese Hindernisse zu adressieren. Die konkreten Kosten und die Wirksamkeit zweier potenzieller Strategien – Prämienrabatte und Trainings – testet die Evaluierung direkt bei den potenziellen Versicherungsnehmern, um die operative Umsetzung des FZ-Vorhabens durch Igloo zu unterstützen. Gleichzeitig sind die im Projekt adressierten Fragen relevant für viele Wetterversicherungen weltweit, auch über den vietnamesischen Kontext hinaus.



Digitale Datensammlung bei einem Reisbauern in Vietnam.



Die Bäuerinnen und Bauern stimmen der Teilnahme an der Evaluierung nach einer kurzen Vorstellung der Studie und des Datenschutzes zu.



„Die Wirkungsevaluierung des Projekts könnte für Igloo potenziell von großem Wert sein. Sie ermöglicht es Igloo, den Nutzen von Wetterindexversicherungen für landwirtschaftliche Betriebe zu demonstrieren, auch um Unterstützung von der Regierung für dieses Programm zu gewinnen. Wir hoffen es auszuweiten, um Millionen von Landwirten zu erreichen.“

Mahesh Joshi, BlueOrchard Finance Ltd, Head of Asia Private Equity Investment

Das Evaluierungsteam



Prof. Dr. Kati Krähnert, RWI und Ruhr-Universität Bochum

„Unser Evaluierungsprojekt ist ein gutes Beispiel dafür, wie rigorose wissenschaftliche Methoden Politikakteurinnen und -akteuren eine fundiertere Entscheidungsgrundlage bieten können.“



Dr. Lukas Mogge, RWI

„Unsere Forschung zeigt auf, welche Faktoren kleine landwirtschaftliche Betriebe dazu bewegen, Klimaanpassungsmaßnahmen umzusetzen. Damit können wir dazu beitragen, Hindernisse in der Klimaanpassung zu überwinden.“



Dr. Emmanuel Nshakira Rukundo, RWI

„Spannend ist, dass unsere Analysen Neues zum Stand der wissenschaftlichen Forschung beitragen. Bislang ist die Zahlungsbereitschaft für eine Wetterversicherung nicht gut verstanden.“

Evaluierungsbericht 2023-2024: Bewerten



Vorhaben unabhängig überprüfen

Etwa die Hälfte der Vorhaben werden nach Abschluss durch FZ-Evaluierung unter die Lupe genommen.



Ex-post-Evaluierungs-Champions weltweit

Die evaluierten Vorhaben der Finanziellen Zusammenarbeit (FZ) verteilen sich über nahezu die ganze Welt. Im Auftrag der unabhängigen Evaluierungsabteilung führen Mitarbeiterinnen und Mitarbeiter der KfW Entwicklungsbank Evaluierungen durch – selbstverständlich nur, wenn sie zuvor weder direkt noch indirekt in das jeweilige Vorhaben involviert waren. Ziel ist es, über den Tellerrand zu schauen und aus Evaluierungen für laufende und zukünftige Vorhaben zu lernen. Hier teilen einige von ihnen ihre Erfahrungen.



„Bei der Evaluierung habe ich sehr interessante Einblicke in die komplexen politischen und gesellschaftlichen Herausforderungen in Bosnien-Herzegowina bekommen. Beeindruckt hat mich die große Professionalität der lokalen Kooperationspartner bei der Bewältigung der Folgen der Jahrhundertflut 2014. Das Fluthilfeprojekt wurde ohne politische Einflussnahme schnell und zielgerichtet in der Föderation Bosnien und Herzegowina als auch in der Republika Srpska umgesetzt. Ich habe schon einige Evaluierungen durchgeführt und dabei immer sehr viel gelernt. Ich kann daher allen Kolleginnen und Kollegen in der FZ empfehlen, eine Evaluierung durchzuführen.“

Dr. Matthias Schlund,
Beseitigung von Flutschäden in Bosnien-Herzegowina



„Bei der Evaluierung dieses Biodiversitätsvorhabens kam ich in ein Land, dessen allgemeine Rahmenbedingungen sich seit der Projektumsetzung leider verschlechtert hatten (Wirtschaft, Coronafolgen und Drogenkriminalität). Vor diesem Hintergrund war ich überrascht, wie gut und mit welchen geringen Mitteln die damals umgesetzten Strukturen weiterhin genutzt werden und noch immer zu den angestrebten Schutzziele beitragen.“

Ingo Baum,
Stärkung von Schutzgebieten in Ecuador



„Besonders beeindruckt hat mich die gender-transformative Konzeption des Vorhabens, welches in seinen Kommunikationskampagnen die sexuelle Selbstbestimmung von Frauen, aber auch die Verantwortung der männlichen Partner in den Fokus rückt. Der Projektträger stellte zudem die kontinuierliche Wirkungsmessung in den Vordergrund: Dank einer Vielzahl begleitender Studien, die im Rahmen des Vorhabens durchgeführt wurden, konnte die Ex-post-Evaluierung auf einer starken Datengrundlage aufbauen.“

Hannah Linnemann,
Familienplanung und HIV-Prävention in Côte d'Ivoire

● Multisektorale UN-Vorhaben / Fragilität

● Bildung

● Finanzwesen

● Energie

● Landwirtschaft und Ressourcenschutz

● Gesundheit

● Wasser

● Dezentralisierung und Governance

● Sonstiges

„Es war erfreulich zu beobachten, dass der Projektträger auch nach der Implementierung des Projekts von dessen Grundkonzept überzeugt ist und es weiter ausbauen möchte. Allerdings kam es während der Projektimplementierung zu erheblichen Verzögerungen, die auf verschiedene Faktoren wie Überflutungen, Hackerangriffe und Zollprobleme zurückzuführen sind. Aus diesen Erfahrungen konnten wir viel für zukünftige Projekte lernen. Besonders interessant war für mich der Konflikt, der für die serbische Energieversorgung besteht: Trotz des Ausbaus erneuerbarer Energien wird Kohle voraussichtlich noch lange als Energieträger genutzt werden. Daher ist es so entscheidend, die Kohlenutzung so effizient wie möglich zu gestalten.“

Aaron Müller,
Regulierung von Kohleenergie in Serbien



„Besonders beeindruckt haben mich die tiefen Emotionen und die Dankbarkeit, die bei meinen Gesprächen mit den Begünstigten zum Ausdruck kamen. Durch Investitionen in agroforstliche Maßnahmen wurde es indigenen indischen Familien in ländlichen Gebieten ermöglicht, ihr Leben neu zu gestalten und auch künftigen Generationen bessere Perspektiven zu bieten. Die Evaluierungserfahrung hat mich in meiner Überzeugung bestärkt, dass unsere Vorhaben Großartiges bewirken können.“

Theresa Schneider,
Förderung der Landwirtschaft in Indien



„Wir haben den besonderen Charakter des Vorhabens auch aus einer kritischen Perspektive bestätigt. Die Berufsbildungsinstitute sind hervorragend ausgestattet, und das engagierte Lehr- und Managementpersonal hat erfolgreich praktische Inhalte in den Lehrplan integriert. Außerdem wurde eine enge Zusammenarbeit mit der Privatwirtschaft aufgebaut, die innerbetriebliche Weiterbildung für das Lehrpersonal fördert und den Auszubildenden durch Praktika den Zugang zu lokalen Unternehmen erleichtert. Die hohe Beschäftigungsquote von über 80 % der Absolventinnen und Absolventen spricht für den Erfolg des Vorhabens. Die gemeinsame Evaluierung hat uns viel Freude bereitet und zu einem interessanten fachlichen Austausch geführt. Unsere unterschiedlichen Erfahrungen und Einschätzungen haben die Qualität des Evaluierungsberichts aus unserer Sicht deutlich verbessert.“

Annelise Umutyana und Nicole Turad,
Technische und berufliche Bildung in Vietnam



„Die mehrfach bekundete Wertschätzung von verschiedenen Begünstigten machte die hohe Relevanz dieses Ernährungssicherungsvorhabens deutlich. Dabei kam besonders zum Ausdruck, wie zentral der Zugang zu grundlegender Versorgung in Krisenzeiten ist. Durch die Ausgabe von Lebensmittelgutscheinen konnte ein schneller, unkomplizierter und wirkungsvoller Beitrag zur Ernährungssicherung geleistet werden – ein entscheidender Erfolgsfaktor in der pandemiebedingten Ausnahmesituation. Dank der Durchführung der Evaluierung im Tandem konnten wir unsere Eindrücke und Sichtweisen alleine immer wieder diskutieren und somit auch Aspekte herausarbeiten, die uns alleine möglicherweise nicht aufgefallen wären.“

Maren Linda Verspohl und Marc-André Hensel,
Corona-Soforthilfe zur Ernährungssicherung in Südafrika

Ex-post-Evaluierungen

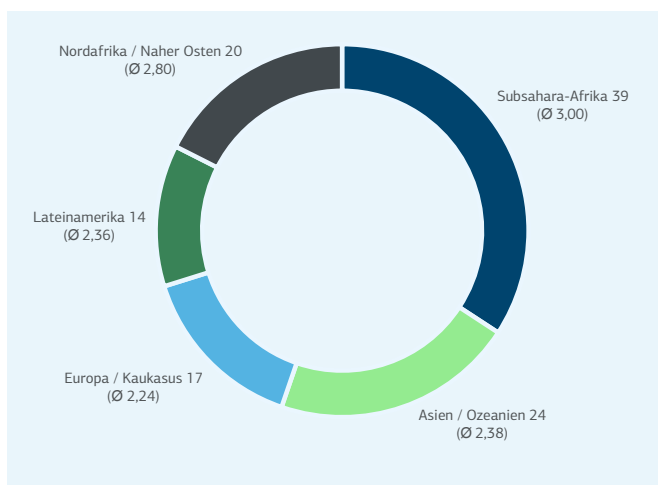
Streifzug durch Sektoren und Regionen

Wo lagen die Schwerpunkte der Evaluierung in den Jahren 2023 und 2024? Was sind die wichtigsten Ergebnisse? Im Folgenden ein Überblick über die evaluierten Vorhaben nach Sektoren und Regionen, ihre Bewertungen und die Lessons Learned, die aus den Evaluierungen abgeleitet werden konnten. Die einzelnen Evaluierungsberichte sind in der [IDEaL-App](#) abrufbar.

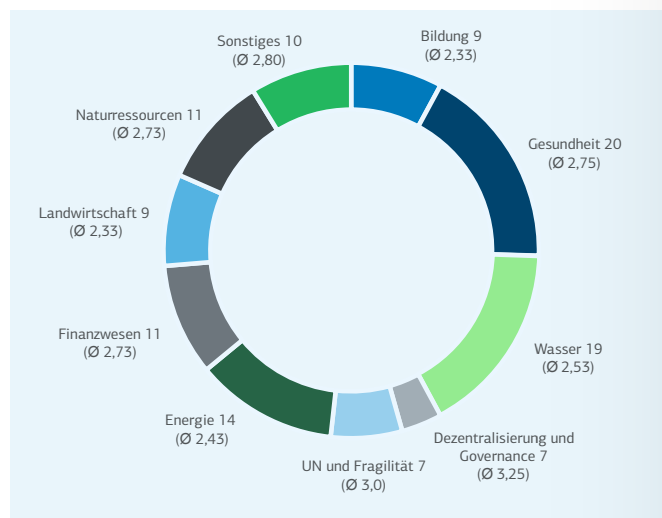
Im Zeitraum 2023-2024 wurden 114 Vorhaben der Finanziellen Zusammenarbeit (FZ) evaluiert, ermittelt als Zufallsstichprobe aus allen abgeschlossenen FZ-Vorhaben und repräsentativ für das FZ-Portfolio. Weitere 26 FZ-Vorhaben wurden im Rahmen dieser 114 Ex-post-Evaluierungen mitbetrachtet, da sie eng mit den Vorhaben der Stichprobe verknüpft waren (siehe Kapitel Methodik, S. 44).

Der größte Anteil der 114 evaluierten Vorhaben befand sich in Subsahara-Afrika (39), gefolgt von Asien/Ozeanien (24) und Nordafrika/Naher Osten (20). Auf die Regionen Europa/Kaukasus (17) und Lateinamerika (14) entfielen leicht weniger Vorhaben. Insgesamt zeigt die regionale Verteilung der Evaluierungen Kontinuität über die Zeit: Schon in den Vorjahren entfiel etwas über die Hälfte auf den afrikanischen Kontinent und den Nahen Osten. Bei der sektoralen Betrachtung sind die Sektoren Gesundheit, Wasser und Energie jene Sektoren mit der größten Anzahl an evaluierten Vorhaben.

Auf der in den Ex-post-Evaluierungen verwendeten standardisierten OECD-DAC-Bewertungsskala von 1 (sehr erfolgreich) bis 6 (gänzlich erfolglos) haben Vorhaben aus Europa/Kaukasus die besten durchschnittlichen Bewertungen ($\bar{x}$ 2,24), dicht gefolgt von Lateinamerika ($\bar{x}$ 2,36) und Asien/Ozeanien ($\bar{x}$ 2,38). Die Region Subsahara-Afrika weist im regionalen Vergleich die niedrigste Durchschnittsbewertung auf, ist mit $\bar{x}$ 3,0 aber insgesamt gesehen immer noch erfolgreich.



Anzahl evaluerter Vorhaben und Durchschnittsbewertungen nach Regionen.



Anzahl evaluerter Vorhaben und Durchschnittsbewertungen nach Sektoren

Im sektoralen Blick schneiden die Sektoren Bildung, Landwirtschaft (je $\bar{x}$ 2,33), und Energie ($\bar{x}$ 2,43) besonders gut ab, während die durchschnittlichen niedrigste Bewertung in den Sektoren Dezentralisierung und Governance ($\bar{x}$ 3,25) sowie UN und Fragilität ($\bar{x}$ 3,0) erzielt wurden.



Vorhaben aus dem Sektor Bildung schließen überdurchschnittlich ab. Hier: Schülerinnen und Schüler der Technischen Berufsschule in Porto Novo, Benin, werden in Landvermessung ausgebildet und nutzen die von der FZ geförderte Ausstattung von Laboren und Werkstätten.

Der Naturressourcenschutz glänzt mit kreativen Lösungen, steht aber auch vor Herausforderungen

Ein Schwerpunkt der im Schnitt mit 2,73 bewerteten Vorhaben im Naturressourcenschutz lag in der Förderung von Schutzgebieten: Neun der elf Vorhaben sollten ausgewählte Schutzgebiete in Ländern Afrikas, Lateinamerikas und Asiens stärken. Zu den gängigen Maßnahmen gehörten die Erstellung von Schutzgebietsmanagementplänen, der Bau von Kontrollposten, Verwaltungsgebäuden oder touristischer Infrastruktur sowie die Finanzierung von Trainings und Fortbildungen und die Förderung der in oder an den Grenzen von Schutzgebieten lebenden Bevölkerung. In einigen Vorhaben wurden flankierend nationale Naturschutzfonds mit Kapital ausgestattet.

Die evaluierten Vorhaben wiesen eine hohe Relevanz auf, da die ausgewählten Zielregionen vielfach zu den Biodiversitäts-Hotspots gehörten und die angedachten Maßnahmen gut geeignet waren, die vorherrschenden Kernprobleme zu adressieren.

Viele der geförderten Schutzgebiete befinden sich in Ländern mit niedrigem Einkommen und sind chronisch unterfinanziert. In diesen Kontexten ist das Engagement der FZ und anderer Geber essenziell, um einen effektiven Betrieb und somit die gewünschten Schutzwirkungen zu erzielen. Als im Zuge der COVID-Pandemie die Zuschüsse aus öffentlichen Haushaltsmitteln gekürzt wurden und andere Einnahmequellen, beispielsweise aus dem Tourismus, vollständig wegbrachen, konnte die FZ durch ihre Unterstützung den fortlaufenden Betrieb sicherstellen und damit einen bedeutsamen Beitrag zur Stabilisierung leisten. Die Evaluierungsergebnisse zeigen zudem, dass das Problem einer unzureichenden Grundfinanzierung durch die Kapitalerträge aus Naturschutzfonds erheblich und nachhaltig gemindert werden kann. Dabei wird das Fondvermögen am Kapitalmarkt angelegt und die erwirtschafteten Erträge können für Betriebskosten der Schutzgebiete genutzt werden. In drei Evaluierungen (*Madagaskar*, *Ecuador* und *Côte d'Ivoire*) stellte sich dieser Ansatz als ein erfolgskritischer Faktor heraus.

Naturschutzkonzepte können nur dann nachhaltig verankert werden, wenn diese gemeinsam mit der lokalen Bevölkerung umgesetzt werden und deren Interessen aufnehmen. Viele der evaluierten Vorhaben beförderten eine stärkere Integration der Bevölkerung in das Parkmanagement (*Brasilien*, *Ecuador* und *Namibia*) und finanzieren Maßnahmen zur Förderung der ansässigen Bevölkerung. Die Evaluierungsergebnisse zeigen allerdings, dass solche Maßnahmen häufig nur einen geringen Teil des



Waldschutzgebiet nahe Khovaling, Tadschikistan: Dieser Bauer und seine Frau haben mit Unterstützung eines KfW-Förderprogramms über 1.000 Bäume und ca. 5.000 Setzlinge angebaut. Aus den Erträgen haben sie Weizen gepflanzt und eine Baumschule gegründet.

Gesamtbudgets ausmachen (*Côte d'Ivoire* und *CCAD*) und erforderliche Budgetreduzierungen während der Umsetzungsphase zu Lasten der Anrainerförderung gehen können (*Madagaskar* und *Ecuador*).

Ein Vorhaben im *Südkaucasus* fällt besonders positiv auf. In diesem Vorhaben wurden mit ausgewählten Gemeinden, die in besonders wichtigen Ökokorridoren schützenswerter Naturräume liegen, Naturschutzverträge geschlossen. Die Verträge legen Land- und Ressourcenschutzmaßnahmen fest, für deren Umsetzung die betroffenen Gemeinden jährliche Ausgleichszahlungen aus einem Fonds erhalten. Die Zahlungen haben zu einer spürbaren Verbesserung der Einkommens- und Lebenssituation bei gleichzeitiger Zunahme der Wildtierpopulationen geführt. Das Vorhaben ist ein Musterbeispiel für einen holistischen und höchst partizipativen Ansatz.

Eine Herausforderung für die Durchführung und die Evaluierung der Vorhaben stellten zum Teil unzureichende Daten zum Zustand der Schutzgebiete, ihrer Biodiversität und Artenvielfalt, dar. Die Vielfältigkeit, Dynamik und Komplexität der verschiedenen Ökosysteme erschweren häufig die Entwicklung einheitlicher Indikatoren und Messmethoden: Dank des Einsatzes moderner Satellitendaten gelingt es inzwischen aber immer besser, auch längere Wirkungszeiträume zu erfassen und zu bewerten (*Madagaskar*, *Ecuador*, *Côte d'Ivoire*, *Brasilien* und *CCAD*).



Marktfrauen verkaufen landwirtschaftliche Erträge auf einem Stadtmarkt in Zimba, Sambia.

Der Spitzensektor: Gründe für das überdurchschnittlich erfolgreiche Abschneiden des Landwirtschaftssektors

Das überdurchschnittliche Abschneiden des Landwirtschaftssektors ist insbesondere drei Vorhaben zu verdanken, welche in der Evaluierung jeweils die Höchstbewertung 1 „sehr erfolgreich“ erzielten (zwei Vorhaben in Indien und eines in Ecuador). Alle drei Vorhaben erzeugten hohe landwirtschaftliche Produktivitäts- und Einkommenssteigerungen. In Peru und Indien reduzierte dies wiederum die zuvor durch Armut und Not erzwungene Wanderarbeit in nahegelegene Städte. In Indien waren die Vorhaben so erfolgreich, dass sie landesweit repliziert wurden.

Zu den zentralen Erfolgsfaktoren der Vorhaben gehörten die enge Ausrichtung der Projekte an den Bedarfen und Fähigkeiten der geförderten landwirtschaftlichen Familienbetriebe, eine hohe Eigenmotivation der Beteiligten, und ein holistischer Ansatz: Die Vorhaben förderten nicht nur Bewässerungsinfrastruktur und deren Instandhaltung, sondern auch die Produktion und Vermarktung der landwirtschaftlichen Produkte. In Ecuador spielte die Nähe des Vorhabens zu regionalen Wirtschaftszentren eine begünstigende Rolle, da die landwirtschaftlichen Erzeugnisse dort vertrieben werden können. Solche externen Faktoren können

Evaluierungsergebnisse gelegentlich jedoch auch negativ beeinflussen: Ein Vorhaben zur Modernisierung von Bewässerungsperimetern in Tunesien konnte aufgrund der sich drastisch verschlechterten Wasserverfügbarkeit kaum Wirkungen entfalten und wurde mit 4 („eher nicht erfolgreich“) bewertet.

Generell weisen die Erkenntnisse der Evaluierungen auf zwei zentrale Spannungsfelder im Sektor hin. Zum einen können Konflikte zwischen Effizienz und Nachhaltigkeit entstehen, wenn die Einführung landwirtschaftlich ressourcenschonender Technologien mit einem Wartungsaufwand verbunden ist, der von der Zielgruppe nicht ohne dauerhafte technische Begleitung geleistet werden kann. Zum anderen sind die Folgen des Klimawandels ambivalent: Bewässerungsansätze und integriertes Wasserressourcenschutzmanagement können bei der Anpassung an die Folgen des Klimawandels nachweislich helfen (Peru, Ecuador, Indien und Niger). Im Gegensatz hierzu kann die Zunahme von Extremwetterereignissen die Wirksamkeit und Nachhaltigkeit solcher Vorhaben jedoch auch beeinträchtigen (Tunesien).

Herausforderungen in konfliktgeprägten Kontexten: Ergebnisse des Sektors UN und Fragilität

Evaluierungsergebnisse von UN-Vorhaben, die in hoch fragilen Kontexten implementiert werden, deuten einerseits auf die strukturellen Herausforderungen in diesem Sektor hin und untermauern zugleich, wie relevant die Durchführung von Vorhaben in schwierigen Kontexten über UN-Organisationen ist. So werden zwei Drittel der evaluierten Vorhaben trotz der schwierigen Rahmenbedingungen insgesamt als noch erfolgreich bewertet.

Die strukturelle Herausforderung solcher Vorhaben besteht oftmals darin, langfristige Wirkungen zu entfalten, obwohl die Vorhaben primär an akuten Bedarfen der betroffenen Menschen ausgerichtet sind. Aufgrund der fragilen und volatilen Lebensumstände in extrem konfliktgeprägten Interventionskontexten brauchen die Betroffenen dringend schnellwirksame Maßnahmen, welche von Anfang an nicht – oder nur teilweise – auf dauerhafte Wirkungen

ausgelegt sind (u.a. Äthiopien, Uganda, Syrien und Südsudan). Hierzu zählen beispielsweise Cash-for-Work Programme, bei denen Menschen temporär angestellt werden, beispielsweise im lokalen Straßenbau, um ein kurzfristiges Einkommen erzielen zu können. Auch wenn übergeordnete Wirkungen oftmals weniger ausgeprägt festgestellt werden, so zeigen die Evaluierungsergebnisse doch ganz klar auf, dass UN-Vorhaben wichtige kurzfristige Wirkungen erfolgreich entfalten (z.B. Syrien).

Besonders positiv zu vermerken ist aus den Evaluierungen, dass die Durchführung der Vorhaben über UN-Organisationen eine wichtige Voraussetzung war, um auch in schwierigen Interventionskontexten, Maßnahmen – effizient und effektiv – durchführen zu können (Äthiopien, Uganda, Syrien, Südsudan, Sudan und Jordanien).



Die jungen Sozialaktivistinnen informieren Mütter von Neugeborenen über Immunisierung, die Bedeutung von Muttermilch und Ernährung in einer Unterkunft für Binnenvertriebene in Dollow, Somalia.

Die Erfolgsquote 2023–2024

85 Prozent

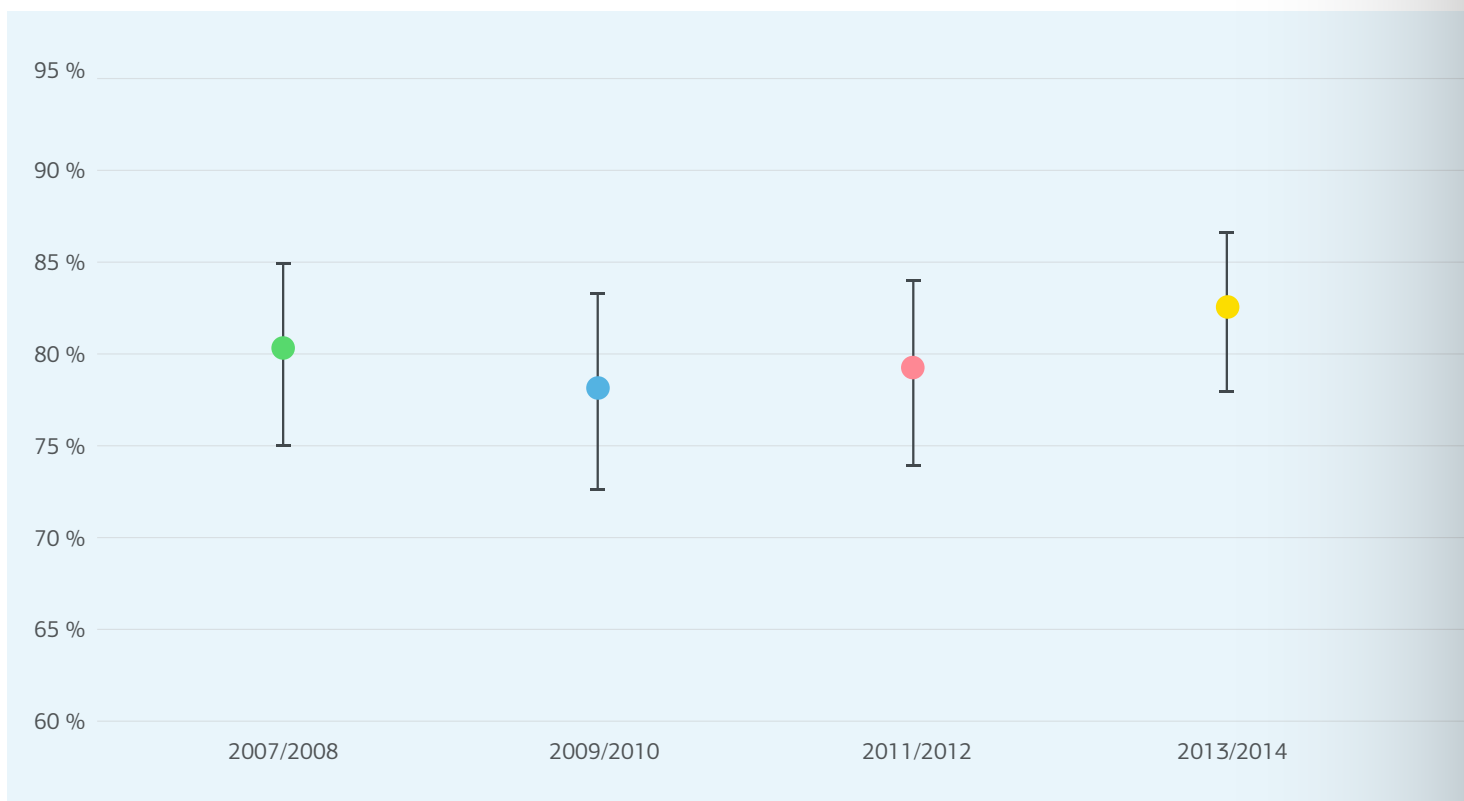
Besondere Aufmerksamkeit gilt immer der Erfolgsquote, denn sie legt auf einen Blick dar, wie die FZ-Vorhaben in einem zweijährigen Evaluierungszeitraum im Durchschnitt insgesamt abgeschnitten haben. Das Besondere: Die Erfolgsquote gilt – mit einer gewissen Schätzgenauigkeit – für alle von der Finanziellen Zusammenarbeit (FZ) umgesetzten Vorhaben einer Zweijahresperiode bzw. Grundgesamtheit (siehe Methodik, S. 44).

Der Bewertungsdurchschnitt der 65 bereits evaluierten FZ-Vorhaben der Stichprobe 2023/2024 liegt bei 2,71. Die Verteilung der Bewertungen zeigt, dass knapp die Hälfte (46 %) der 65 Vorhaben die Bewertung 2 (erfolgreich) und mehr als ein Drittel (38 %) das Gesamtergebnis 3 (eingeschränkt erfolgreich) erreichten. Insgesamt neun Vorhaben wurden mit 4 (eher nicht erfolgreich), und ein Vorhaben mit 5 als nicht mehr erfolgreich eingestuft.

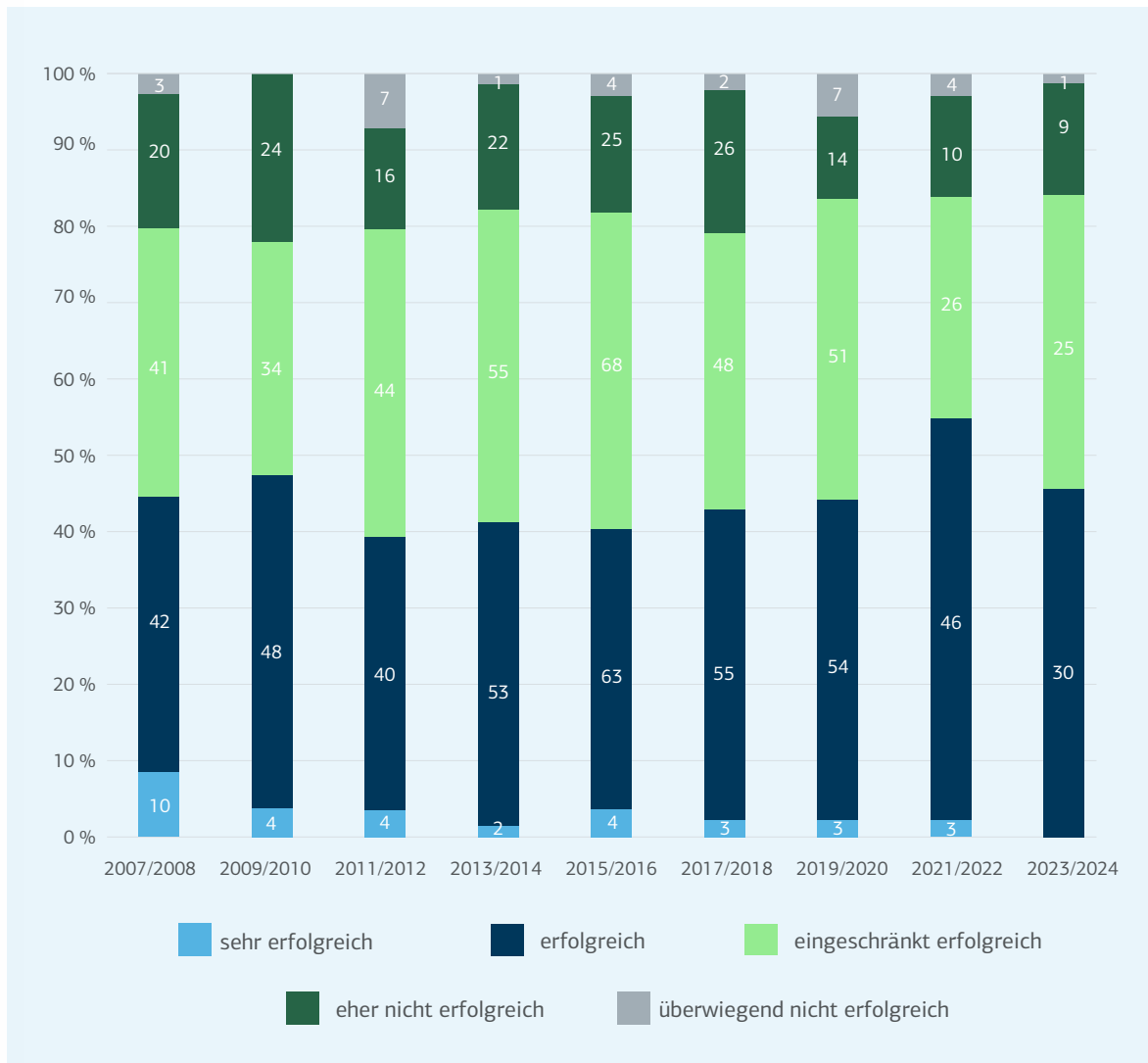
Die Erfolgsquote liegt bei 85 %. Das bedeutet, dass 85 % der Projekte mit mindestens 3 (eingeschränkt erfolgreich) bewertet wurden. Damit liegt die aktuelle Erfolgsquote über dem langjährigen Durchschnitt der Erfolgsquoten von 2007–2024 von 81 % und reiht sich ein in den leicht positiven Trend über die vergangenen 18 Jahre.

Dank der repräsentativen Stichprobenziehung kann diese Erfolgsquote von 85 % auch auf nicht-evaluierte Vorhaben und somit auf alle 268 FZ-Vorhaben des Evaluierungszeitraums übertragen werden. Hier gibt es jedoch eine gewisse statistische Unsicherheit, welche durch das Konfidenzintervall ausgedrückt wird: Die Erfolgsquote aller (evaluierten und nicht evaluierten) 268 Vorhaben des Berichtszeitraums 2023–2024 liegt mit 95-prozentiger Sicherheit zwischen 77 % und 92 %.

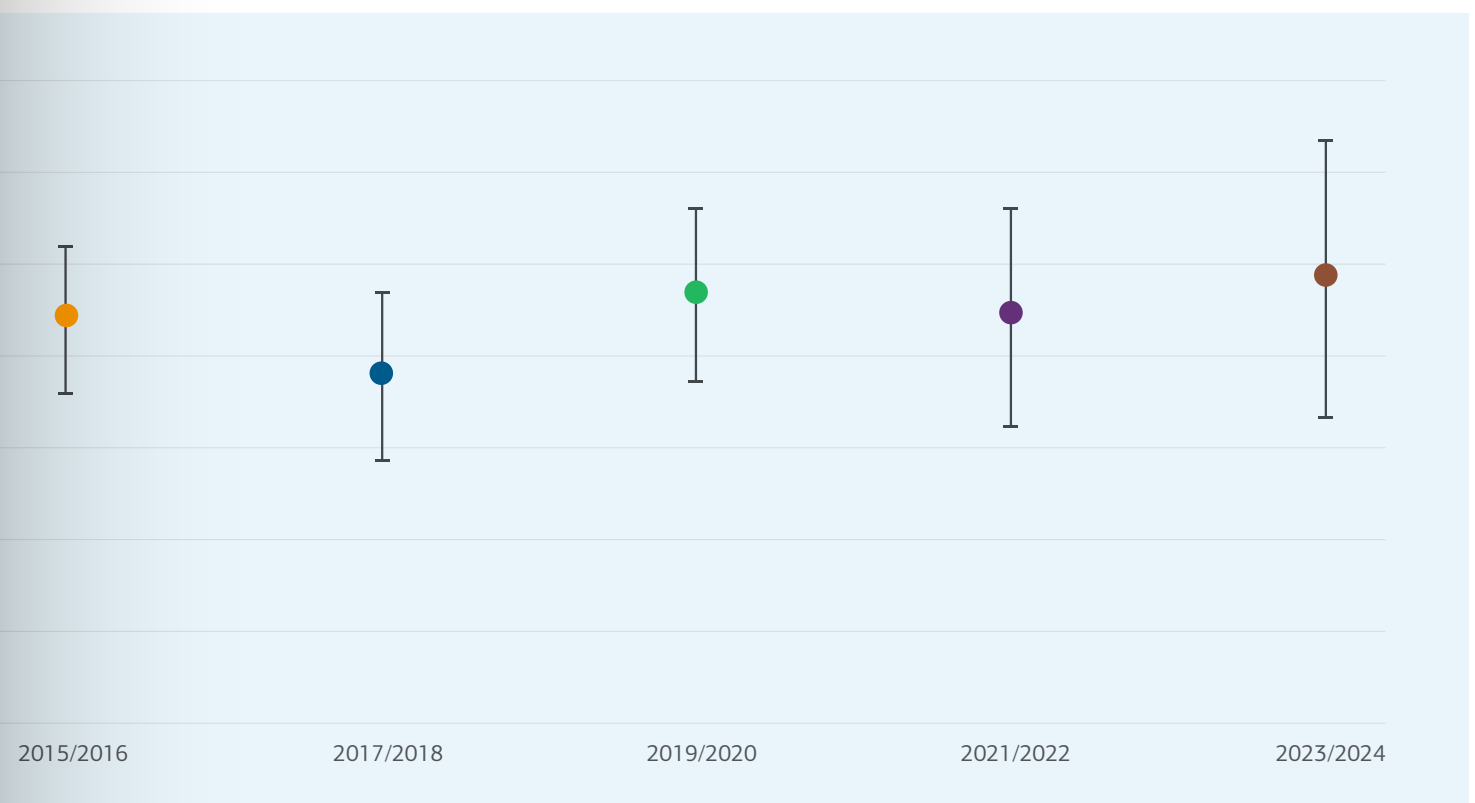
Das Konfidenzintervall für 2023–2024 ist noch vergleichsweise breit, da bislang 50 % der Vorhaben aus der Stichprobe 2023–2024 evaluiert wurden. Die Schätzgenauigkeit wird folglich mit jeder weiteren durchgeführten Evaluierung steigen, bis die Stichprobe vollständig evaluiert ist. In diesem Berichtszeitraum konnten 36 Evaluierungen aus der vorherigen Stichprobe für die Grundgesamtheit 2021–2022 abgeschlossen werden, wodurch sich die Breite des Konfidenzintervalls um 2,5 Prozentpunkte verringerte (die Schätzgenauigkeit also stieg). Damit liegt die Erfolgsquote des Berichtszeitraums 2021/22 bei 82 % (84 % im Evaluierungsbericht 2021–2022).



Erfolgsquote der FZ-Vorhaben: Anteil der mit mindestens Gesamtnote 3 bewerteten Vorhaben und 95%-Konfidenzintervall, nach zweijähriger Stichprobe von 2007/2008 - 2023/2024.



Anteil der Einzelbewertungen, nach zweijähriger Stichprobe von 2007/2008 - 2023/2024.



Drei Sektoren im Fokus

Lessons Learned aus mehreren Jahren Ex-post-Evaluierungen



Lessons Learned aus Bildungsvorhaben

Ein Blick in die Lessons Learned aus Ex-post-Evaluierungen der letzten fünf Jahre zeigt zunächst, dass sich der Fokus der Vorhaben im Bildungsbereich von der Grundbildung hin zur Berufsbildung verschoben hat. Dennoch gibt es weiterhin umfassende Programme zur Sicherung der Grundbildung in fragilen Kontexten, oft in Zusammenarbeit mit UN-Organisationen.

In Vietnam, einem aufstrebenden Schwellenland, hat sich erneut gezeigt, wie wichtig in der Berufsbildung die enge Zusammenarbeit mit der Wirtschaft ist. Das als erfolgreich bewertete Vorhaben unterstützte nicht nur den Übergang der Absolventinnen und Absolventen in den Job, sondern sorgte auch dafür, dass die Ausbildung und Ausstattung der Bildungseinrichtungen den Bedürfnissen der Industrie entsprechen. Hierzu dienten Investitionen in moderne Hardware und die

Unterstützung bei der Entwicklung von Curricula und Lehrkräftefortbildungen.

In fragilen Kontexten stehen Berufsbildungsansätze vor besonderen Herausforderungen. Wenn die Aussichten auf dem Arbeitsmarkt aufgrund geringer Arbeitsnachfrage und niedrigen Löhnen schlecht sind, ist der Wert von Bildung nicht immer unmittelbar ersichtlich. Dies kann zu Frustration bei jungen Menschen führen. Die Evaluierungen haben vor diesem Hintergrund gezeigt, dass dennoch positive Wirkungen auf die Resilienz und Beschäftigungsfähigkeit erzielt werden können.

Das Engagement auf Schulebene, die Sicherstellung der Finanzierung laufender Schulkosten und die Institutionalisierung von Wartung und Instandhaltung von Schulinfrastruktur sollten von Anfang an zentraler Bestandteil jeder Unterstützung sein. Gleichzeitig ist es wichtig, die Standarddesigns von

Bildungsvorhaben an lokale Gegebenheiten anzupassen. In abgelegenen Gebieten ohne Strom- und Wasseranschluss können beispielsweise Elektroinstallationen und Wasserklosetts nicht genutzt werden. Im Senegal hat sich gezeigt, dass der Bau von Außenmauern um Schulen den Vandalismus reduziert und vor der Nutzung durch unbefugte Dritte schützt.

Generell zeigt sich: Mit Engagement und kohärenter Zusammenarbeit kann die Bildungslandschaft nachhaltig verbessert werden. Ein zentraler Erfolgsfaktor bleibt hierbei auch das Engagement der Schulleitung sowie der Lehrkräfte und Eltern. In Tadschikistan wurden die Gemeinden erfolgreich in die Schulorganisation einbezogen, was trotz begrenzter finanzieller Mittel zu einer besseren Unterstützung bei Betrieb und Wartung von Schulinfrastruktur führte.



214 Schülerinnen und Schüler besuchen eine neue, von der KfW Entwicklungsbank finanzierte Schule (Klasse 1-11) im Danghara Distrikt, Tadschikista.



Lessons Learned aus Finanzsektorvorhaben

Die Ex-post-Evaluierungen von Vorhaben des Finanzsektors der letzten Jahre belegen zwei wichtige Trends: Erstens verschiebt sich der Fokus von Vorhaben mit Umweltschwerpunkt, wie zum Beispiel Umweltkreditlinien, hin zu Vorhaben, die sich auf die Verbesserung der Energieeffizienz konzentrieren (z.B. Energieeffizienzkreditlinien). Zweitens gibt es immer mehr Vorhaben, die als Fonds strukturiert sind.

Die Evaluierungen zeigen: Ein entscheidender Erfolgsfaktor für Fondsvorhaben ist eine erfahrene Fondsmanagerin oder ein erfahrener Fondsmanager. Diese sollten über umfangreiche Marktkenntnisse, Erfahrungen und ein gutes Netzwerk von qualifizierten Fachleuten in der Region verfügen. Außerdem ist es wichtig, dass das Fondsmanagement ein internes Impact- und Nachhaltigkeitsteam hat und zudem genügend

Zeit, um Kundinnen und Kunden zu unterstützen. Das zeigte sich zum Beispiel bei der Evaluierung des Programms Mikrofinanzinitiative für Subsahara-Afrika, wo die gute Fachkenntnis und das Netzwerk des Fondsmanagers entscheidend für die hohe Qualität des Fondsportfolios waren.

Vor besonderen Herausforderungen stehen Projekte in sehr armen oder eventuell konfliktbeladenen Gebieten. Besonders erfolgreich waren hingegen Vorhaben, in denen die Finanzierung an die individuellen Bedürfnisse der Kreditnehmenden angepasst wurde.

Oder es konnten flexible Rückzahlungsfristen vereinbart werden, in denen die Kreditnehmenden keine Zahlungen leisten müssen. Dies ist z.B. bei Agrarkrediten wichtig, bei denen Rückzahlungen beispielsweise an den Erntezeiten ausgerichtet werden.

So konnten Kreditnehmende verschiedene Arten von Sicherheiten anbieten, wie zum Beispiel Autos, Gold oder auch Gruppensicherheiten von Frauenkooperativen.

Die Evaluierungen zeigen außerdem: In Notsituationen, z.B. nach einer Naturkatastrophe, sollten bereits vorhandene Strukturen für die Durchführung von Vorhaben genutzt werden. Das kann – neben weiteren Vorteilen – dabei helfen, die benötigte Unterstützung schnell bereitzustellen. Ein erfolgreiches Beispiel dafür ist ein Vorhaben nach einer Flut in Bosnien-Herzegowina, bei dem dank der Nutzung existierender Strukturen in nur wenigen Wochen über 2.000 Darlehen an private Haushalte sowie an kleine und mittlere Unternehmen ausbezahlt werden konnten.



Unterlagen und Sparbuch einer Spargruppe in Afar, Äthiopien. Die FZ-unterstützten Darlehen, welche über die Bank of Ethiopia vergeben wurden, nutzt die Frauengruppe für die Zucht von Kühen, Ziegen und die Fleischproduktion.



Lessons Learned aus Energiesektorvorhaben

Der Energiesektor ist – gemessen am Finanzierungsvolumen – der wichtigste Sektor in der KfW Entwicklungsbank. Er besteht aus den Bereichen Kreditlinien, erneuerbare Energien, Stromübertragung sowie Energieeffizienz. Aus Ex-post-Evaluierungen der letzten 10 Jahre lassen sich sowohl grundsätzliche wie auch subsektor-spezifische Lessons Learned ableiten.

Um Partnerländern Investitionen in erneuerbare Energien zu ermöglichen, stellte die FZ in den letzten Jahren zunehmend Kreditlinien bereit. Kreditlinien sind ein wichtiges Instrument, um die Nutzung innovativer (Energie-)Technologien zu steigern und gleichzeitig zur Entwicklung nationaler Finanzmärkte beizutragen. Da insbesondere nationale Kapitalmärkte oft nur Kredite mit kurzen Laufzeiten und hohen Zinssätzen bereitstellen, können attraktivere (zinsgünstige, langfristige) Kredite die Investitionen in erneuerbare Energien erfolgreich steigern. Grundvoraussetzung für eine solche erfolgreiche Umsetzung: Länder und Finanzintermediäre, die bereits Erfahrungen mit der

Umsetzung von Finanzprodukten im Energiesektor haben. Eine weitere Lessons Learned: Es ist zentral, beim Ausbau der Energiequellen zu diversifizieren, also unterschiedliche Technologien zu nutzen, um den Herausforderungen des Klimawandels zu begegnen. So gibt es zum Beispiel eine hohe Komplementarität zwischen den Energieträgern Wind, Sonne und Wasser: Es ist daher langfristig sinnvoll, diese in einem Land parallel einzubeziehen.

Die Evaluierungsergebnisse im Bereich erneuerbare Energien verdeutlichen, dass Vorhaben mit privatwirtschaftlicher Beteiligung in Ländern ohne staatliche Subventionen oder garantierten Einspeisetarifen in der Regel erst bei größeren Finanzvolumina zu Stande kommen: Für private Firmen lohnt sich die Beteiligung beispielsweise an einem Bauprojekt sonst nicht. Garantierte Einspeisetarife können dabei helfen Investitionsrisiken zu reduzieren. Evaluierungen im Bereich der Stromübertragung zeigen außerdem, dass eine Erweiterung der Stromnetze nur dann effektiv ist, wenn gleichzeitig die Erzeugungs-

und Übertragungskapazitäten entsprechend mit ausgebaut werden, und andersherum.

Aus den Evaluierungen lernen wir außerdem, dass die Einbeziehung von Wohnungseigentümern und Mietern entscheidend ist für den Erfolg von Sanierungsprojekten zur Steigerung der Energieeffizienzen von Gebäuden. Wenn Menschen ausreichend beteiligt werden, weichen die dank des Vorhabens erreichten CO₂-Einsparungen weniger von den (vorab modellierten) geplanten CO₂-Einsparungen ab. Die Wirkungen treten also nur dann im gewünschten Ausmaß ein, weil Verhaltensannahmen bei Planung und tatsächliches Verhalten besser übereinstimmen. Des Weiteren zeigen die Ex-post-Evaluierungen, dass Bewohnerinnen und Bewohner effizienter Energie nutzen, wenn die Energie-rechnungen nach Konsum (verbrauchsabhängig) berechnet werden. Für den Wohnungsmarkt zeigen die Evaluierungen: Die Existenz eindeutiger Energieeffizienz-Kriterien und nationaler Baustandards steigert die Attraktivität von Gebäuden für Käuferinnen und Käufer.



Das Stromkraftwerk von Puerto Villamil, Galapagosinseln, Ecuador. Die Anlage wird etwa zur Hälfte (800KW) mit Solarenergie und zur anderen Hälfte mit Dieselgeneratoren betrieben.



Umspannstation in Gaita, Tansania. Die Landwirtin arbeitet in Teilzeit auch als Wachfrau für das Werk. Sie ist Mutter von zwei Kindern. Der Strom fließt durch drei Regionen: Von den Kraftwerken Kidatu und Mtera über die Buly-Umspannstation in Richtung Nyakanazi.

Evaluierungsbericht 2023-2024: Lernen



Aus Erfahrung lernen

Das Lernen aus abgeschlossenen Vorhaben zu fördern, ist die Motivation unserer Arbeit bei FZ Evaluierung.





Künstliche Intelligenz

Chancen und Risiken für die Bereitstellung von Evidenz

Generative Künstliche Intelligenz (KI) sorgt anhaltend für Schlagzeilen und hat seit der öffentlichen Einführung von ChatGPT 3.5 im November 2022 außergewöhnliche Fortschritte gemacht. Der bemerkenswerte Anstieg der KI-Fähigkeiten und -Nutzung löste erhebliche Investitionen in KI-Technologien aus, was wiederum die Raffinesse und Anwendbarkeit von KI über Sektoren hinweg weiter verbesserte. Mit dem Versprechen, Produktivität und Effizienz zu steigern, transformiert KI den Arbeitsplatz, indem sie

Arbeitsrollen und -bedarfe neu definiert. Diese Transformation erstreckt sich natürlich auch auf den Bereich der Evaluation: Die KI kann Evaluatoren befähigen, schnellere, bessere oder wirkungsvollere Projektevaluationen durchzuführen, aber auch ihre Rolle und Aufgaben neu definieren. Bei der Untersuchung der Chancen und Risiken von KI für die Generierung von Evidenz ergeben sich jedoch einige Fragestellungen.



Im Gespräch



Dr. Estelle Raimondo ist die Leiterin der Gruppe für Methoden in der unabhängigen Evaluierungsgruppe (Independent Evaluation Group, IEG) der Weltbank. Bei der Weltbank berät und schult sie Teams in einer Vielzahl von Evaluierungsmethoden, fördert Innovationen und integriert Data Science und Künstliche Intelligenz in die Evaluierungspraxis. Mit über 15 Jahren Erfahrung in der Evaluierung und Evaluierungsforschung wurden ihre Arbeiten in internationalen, renommierten Fachzeitschriften und Blogs veröffentlicht. Sie ist die Autorin des Buches „Dealing with Complexity in Development Evaluation“.



Prof. Dr. Jochen Kluge ist Abteilungsleiter der unabhängigen Evaluierungsabteilung (FZ E) der KfW Entwicklungsbank und Professor für Volkswirtschaftslehre an der Humboldt-Universität zu Berlin. Seit 25 Jahren sammelt er Erfahrungen mit Impact Evaluierungen und der Analyse von Wirtschaftspolitik und hat dabei mit zahlreichen internationalen Entwicklungsorganisationen, Regierungen und der EU zusammengearbeitet. In der KfW Entwicklungsbank liegt ihm die Entwicklung innovativer und nützlicher Wissensprodukte sehr am Herzen. So entstanden in den vergangenen Jahren das „KfW Development Impact Lab“ und die IDEaL-App (Interactive Database Evaluation and Learning).



Estelle Raimondo und Jochen Kluge im Gespräch in Washington, DC.

Estelle, als Verantwortliche für Methoden der IEG verfügst Du über ein umfassendes Verständnis der Evaluierungsmethoden. Welche bedeutenden Trends hast Du in den letzten 5 bis 10 Jahren beobachtet?



„Bevor ich größere Trends herausarbeite, möchte ich betonen, dass ich aus der Perspektive einer Person spreche, die im Bereich der Evaluierung in internationalen Organisationen tätig ist. Insgesamt identifiziere ich drei bedeutende Trends:

Der erste Trend liegt darin, die **Bandbreite der Methoden und Ansätze, die wir in der Evaluierung zur Erfassung von Wirkungen verwenden, zu erweitern**. Eine Veröffentlichung von Elliot Stern und anderen von 2012 hat entscheidend dazu beigetragen, das Wissen aus verschiedenen Disziplinen, Bereichen und Paradigmen aufzugreifen, um unser Instrumentarium erheblich zu erweitern. Zum Beispiel haben qualitative vergleichende Analysen, process Tracing und Kontributionsanalysen eine Art Wiedergeburt erfahren. Ich denke, wir sind jetzt sehr offen für viele verschiedene Ansätze und versuchen wirklich zu verstehen, was für den Kontext und das gewünschte Komplexitätsniveau angemessen ist. Tom Aston und Marina Apgar haben den Begriff „Bricolage“ geprägt, um diese Idee zu erfassen: Praktisch und pragmatisch mit den Werkzeugen umzugehen und sie nach Bedarf so zu kombinieren, dass sie der Komplexität der Fragestellung gerecht werden.

Zweitens habe ich das Gefühl, dass wir viel besser in der Lage sind, **mit Komplexität aus einer empirischen Perspektive umzugehen**. Wir haben viel Arbeit geleistet, um Komplexitätstheorie und systemisches Denken in die Evaluierung zu integrieren und auch um das richtige empirische Werkzeugset zu entwickeln. Wo ich jedoch ein Problem sehe, ist, dass es auf der Nutzerseite nicht viel Interesse dafür gibt. Wir stehen vor einem neuen Paradox: Während es eine Zeit lang die Intuition gab, dass wir mit unseren konventionellen Methoden nicht ausreichend mit der Komplexität von

Entwicklungsfragen umgehen konnten, sind wir jetzt viel besser gerüstet, um dies zu tun. Auf der Empfängerseite gibt es jedoch kein Interesse, sodass wir wie ein **Pendel eine Tendenz zurück zu einfacheren Ansätzen** wie Scorecards, ergebnisorientierten Managementansätzen (RBM), Indikator-Frameworks und randomisierten Kontrollstudien erleben, welche jedoch dazu neigen, eine komplexe Realität zu stark zu vereinfachen.

Drittens gibt es natürlich die **Datenrevolution**. In den 2010er Jahren wurde sie als Big Data bezeichnet. Jetzt denken wir mehr in Richtung künstlicher Intelligenz, aber der gemeinsame Nenner besteht darin, dass wir in der Lage sind, Daten zu nutzen, auf die wir zuvor im Bereich der Evaluierung nicht zugreifen konnten. In der IEG haben wir uns vor allem auf die Analyse von Bildern und Texten konzentriert.

Im Bereich der Bilder beginnt die Evaluierung endlich aufzuholen durch die Nutzung von geospatialer Analyse. Im Bereich der Texte haben wir nun einige Jahre Erfahrung mit maschinellem Lernen und machen Fortschritte beim Experimentieren mit generativer KI. Letztendlich ist die Evaluierung im Kern ein interdisziplinäres Feld, und die Nutzung von Methoden und Ansätzen aus verschiedenen Disziplinen ist das, was uns voranbringt.“

Jochen, folgt auch die Evaluierungsabteilung der KfW Entwicklungsbank einem solchen Bricolage-Ansatz?



„Das resoniert sehr gut mit unserem Denken und unseren Erfahrungen. Unsere Arbeit folgt vollständig dem Bricolage-Ansatz, den Estelle erwähnt hat: In den letzten fünf Jahren haben wir aktiv das Instrumentarium zur Wirkungsmessung mit einer pragmatischen und praktischen Denkweise genutzt; in unserem Team sagen wir gerne „form follows function“.

Um die Früchte der von Estelle erwähnten Datenrevolution zu ernten, haben wir 2023 das Development Impact Lab gegründet, das neben Wirkungsbewertungen alle Aktivitäten im Bereich Data Science bündelt und verschiedene Datenquellen nutzt, einschließlich Satellitenbildern und hochfrequenten Verkehrsdaten.

Das begrenzte Interesse an komplexen Evaluierungsergebnissen ist auch etwas, mit dem wir uns intensiv auseinandersetzen. In den letzten 5 bis 10 Jahren haben wir daher unseren Fokus von „Evaluierung für Rechenschaftspflicht“ auf die **Bereitstellung praktischer und anwendbarer Werkzeuge für unsere Nutzer erweitert** oder besser gesagt, neu kalibriert. Auf diese Weise hoffen wir, das Interesse zu steigern. Ich bin gespannt, welche Potenziale die KI in Zukunft zur Anregung des Evidenzhungers bietet.“

Estelle, KI als ein Werkzeug zur Verarbeitung und Bereitstellung von Evidenz: Siehst du das auch so?



„Absolut. Die Zusammenfassung von Einzeldokumenten ist ein klarer Wettbewerbsvorteil von generativer KI und großen Sprachmodellen; unser Team testet gerade deren Leistung bei der Zusammenfassung oder Synthese von vielen Dokumenten. Darüber hinaus können diese Arten von Modellen sehr hilfreich als „Schreibcoach“ sein – sie ermöglichen es Evaluatorinnen und Evaluatoren, kostengünstiger für verschiedene Zielgruppen zu schreiben, insbesondere im englischsprachigen Raum.

Aber ich würde gerne noch einen weiteren Gedanken hinzufügen. Es gibt einige erste Hinweise darauf, wie KI unser kritisches Denken beeinflussen kann, insbesondere wenn wir in eine Überabhängigkeit von KI verfallen. Dies kann unser kritisches Denken und unsere Fähigkeit beeinträchtigen, mit Komplexität umzugehen, sie zu verarbeiten und zu verstehen. Daher denke ich, dass es Potenzial gibt, aber auch Herausforderungen oder sogar Gefahren.“

Jochen, bevor wir noch ausführlicher über KI sprechen, lass uns ein gemeinsames Verständnis finden: Was ist KI und was ist der Unterschied zwischen der „neuen“ KI im Vergleich zur Vergangenheit?



„Nun, wenn man KI in einem Wörterbuch nachschlägt, wird sie als ein Computersystem definiert, das Aufgaben ausführt oder imitiert, die typischerweise menschliche Intelligenz erfordern. Diese Unterscheidung zwischen der Ausführung und der Imitation von Aufgaben ist genau der Unterschied zwischen diskriminativer und generativer KI: Vor 2022 war nur diskriminative

KI verfügbar, nämlich maschinelles Lernen oder überwachte Lernmodelle, die in der Lage waren, Entscheidungen zu reproduzieren, für welche sie trainiert worden waren. Mit der Einführung von ChatGPT erleben wir einen Wandel hin zur generativen KI, was die Schaffung neuer Inhalte bedeutet.“

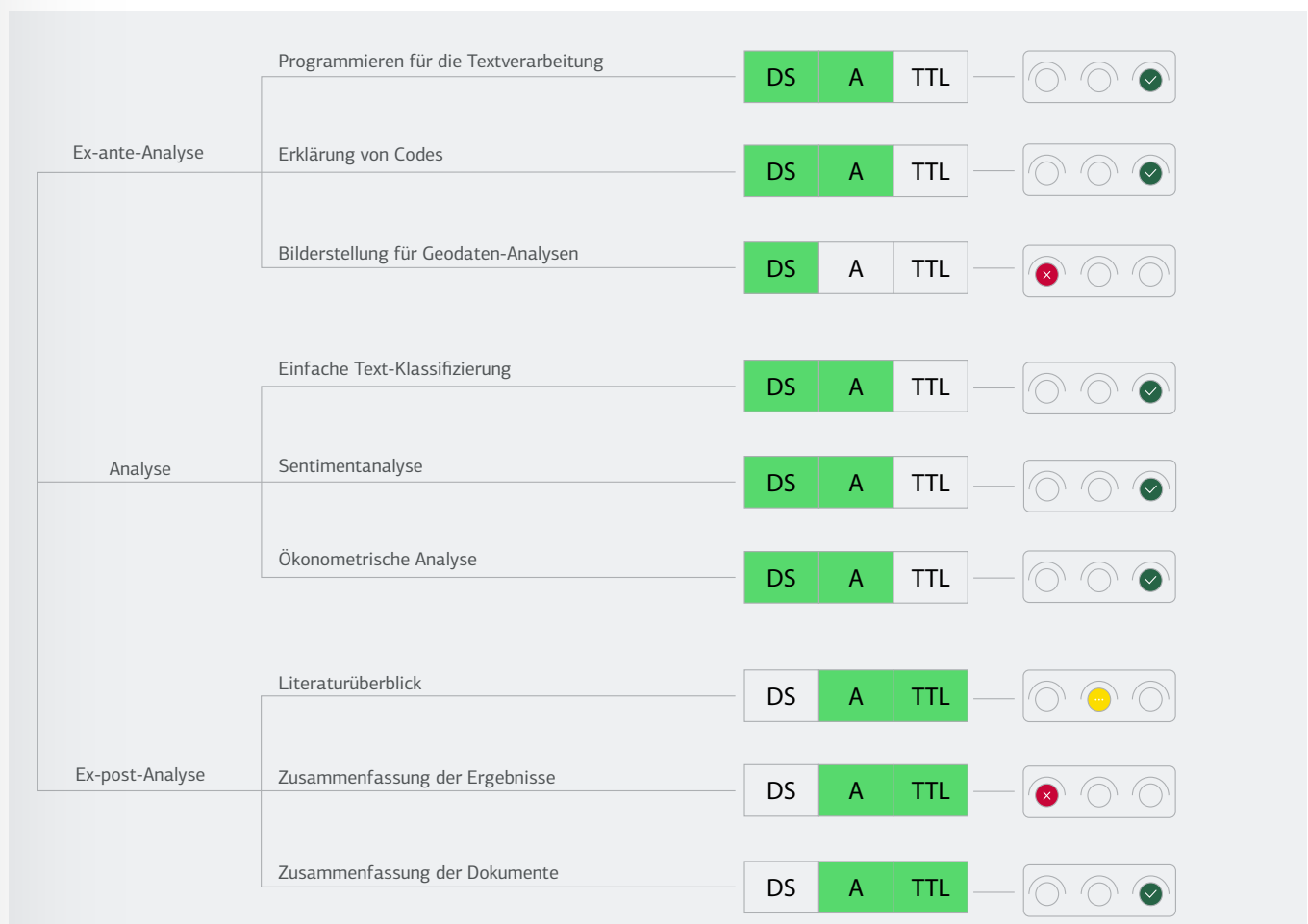
Estelle, bei der letzten Konferenz der European Evaluation Society hast Du eine Reihe von Experimenten zu den Anwendungen von KI für verschiedene Evaluierungsaufgaben erwähnt. Könntest Du unseren Leserinnen und Lesern davon berichten?



„Gerne. Wir haben im Frühjahr 2023 unsere erste **Reihe von Experimenten durchgeführt, bei denen wir sowohl die Unternehmensversion von ChatGPT der Weltbank als auch andere Open-Source-Modelle getestet haben.** Wir wollten klare Experimente zu zehn Routineaufgaben durchführen, die normalerweise unsere Data Scientists, Analysten und Evaluatoren in der IEG erledigen.

Wir haben verschiedene Anwendungen getestet, von der Erstellung und Erklärung von Codes über die Durchführung ökonomischer Analysen, das Zusammenfassen von Dokumenten bis hin zu Literaturübersichten. Anschließend haben wir die Arbeitsergebnisse, die durch konventionelle KI oder menschliche Intelligenz erzeugt wurden, mit denen verglichen, die von generativer KI produziert wurden, um die Genauigkeit zu bewerten und auf Fehler und Halluzinationen zu überprüfen. Zu diesem Zeitpunkt haben wir unsere wichtigsten Empfehlungen in einem einfachen Diagramm festgehalten (siehe folgende Grafik).

Seitdem hat sich viel verändert. Es gibt viele weitere generative KI-Modelle, und deren Leistung hat sich bei den meisten Aufgaben enorm verbessert. Wir haben aus dieser ersten Reihe von Experimenten viel gelernt und unsere Nutzung von generativer KI verbessert. Zum Beispiel wissen wir jetzt, dass wir die komplexen Aufgaben für Literaturübersichten und Synthesen viel besser aufschlüsseln müssen, als wir es in der ersten Reihe von Experimenten getan haben, um bessere Ergebnisse von der KI zu erhalten. Wir wissen auch, wie wir besser prompten können, um Halluzinationen zu vermeiden. Wir integrieren jetzt routinemäßige Experimente in unseren Evaluierungs-Workflow und planen, diesen Ansatz in absehbarer Zeit fortzusetzen, während wir stets sicherstellen, dass wir gute Praktiken des „Mensch-im-Prozess“ anwenden – und ich würde sogar sagen „Mensch-am-Steuer“.



Empfohlene Anwendungen von künstlicher Intelligenz für die Evaluierungspraxis.

Quelle: IEG, Weltbank. Hinweis: A = Analyst, DS = Data Scientist, TTL = Teamleiter.

Eure Experimente zeigen, dass viele der getesteten Aufgaben durch KI ersetzt werden können. Es wird oft angenommen, dass KI zu Arbeitsplatzverlusten und großen Veränderungen in den Arbeitsaufgaben führen wird. Würdest Du das auch aus Euren experimentellen Ergebnissen für den Bereich der Evaluierung ableiten?



„Es ist wichtig klarzustellen, dass wir zwar einige erfolgreiche KI-Experimente durchgeführt haben, aber nicht unsere Analysten ersetzt haben. Wenn wir jedoch eine vielversprechende Anwendung von KI finden, versuchen wir, die Experimente in einer Handvoll Evaluierungskontexte zu replizieren und teilen dann unsere Ergebnisse mit unseren Kollegen und Kolleginnen, um Analysten und Evaluatoren zu ermutigen, KI bei ähnlichen Aufgaben in Betracht zu ziehen. In Bezug auf Eure Frage glaube ich, dass es einen **Wandel der erforderlichen Fähigkeiten geben wird, aber dies wird schrittweise passieren**. Derzeit liegt der Fokus darauf, eine **grundlegende KI-Kompetenz für alle zu entwickeln, da wir gebildete Nutzer benötigen und keine blinden Anwender**. Für diejenigen, die analytische Aufgaben durchführen – also mit Daten arbeiten und Codes schreiben – wird es Veränderungen in der Arbeitsteilung geben. Zum Beispiel haben wir in diesem Jahr fünf Analytinnen und Analysten eingestellt, und ihre Stellenbeschreibungen betonen jetzt viel stärker die Fähigkeiten im Bereich Data Science als vorher.

Was Arbeitsplatzverluste generell betrifft, möchte ich nicht spekulieren. Brillante Kolleginnen und Kollegen bei der Weltbank haben daran gearbeitet, was generative KI für die Zukunft der Arbeitsplätze bedeuten könnte. Was die Evaluatoren betrifft, hoffe ich, dass wir lernen, KI für mechanische und routinemäßige Aufgaben (z.B. Klassifizierung, Zusammenfassung, Dokumentenüberprüfungen usw.) zu nutzen, um so mehr Zeit, Fachwissen und menschliche Fähigkeiten für Analysen, den Austausch mit Interessengruppen, den Aufbau von Theorien und komplexe Aufgaben aufwenden zu können.

Das gesagt, müssen wir unbedingt lernen, wie wir KI effektiv nutzen und steuern können, während wir ihre Risiken, einschließlich ihrer Umweltauswirkungen, mindern. Es wird weiterhin viele Arbeitsplätze für Menschen geben; wir müssen nur die Herausforderung annehmen und die Rolle der KI besser verstehen und definieren. Persönlich glaube ich, dass wir in dieser Hinsicht noch Verbesserungspotenzial haben.“



Estelle Raimondo setzt sich bei der Weltbank für die Integration von Data Science und Künstlicher Intelligenz in der Evaluierungspraxis ein.



Jochen Kluge und Estelle Raimondo im angeregten Austausch über die Rolle von künstlicher Intelligenz in der Evaluierungspraxis.

Jochen, wir haben wichtige Einblicke in die Potenziale, Chancen und Risiken bei der Anwendung von KI in der Evaluierung gehört. Was kannst Du uns über KI und Evaluierung bei der KfW Entwicklungsbank erzählen? Gibt es spannende Anwendungsfälle oder Beobachtungen?



„Zunächst möchte ich betonen, dass die systematische Experimentierung, die Estelle und ihr Team durchgeführt haben, für uns äußerst nützlich ist und unsere Arbeit mit KI anleitet. Wir machen gute Fortschritte bei der Nutzung von generativer KI in unserem Alltag, z.B. indem wir eigenes ChatKfW für Evaluierungsbedürfnisse nutzen und dabei die institutionellen Datenschutzverpflichtungen einhalten. Neben den täglichen Anwendungen integrieren wir zunehmend strukturierte KI-Anwendungen. **Eine Idee ist beispielsweise, KI für die Transkription und Übersetzung von Interviews in lokalen Sprachen zu nutzen, was es uns ermöglicht, Beiträge zu berücksichtigen, die zuvor möglicherweise untergegangen sind.** Ein weiteres Beispiel ist der Versuch, **eine KI-basierte Plattform für standardisierte Ex-post-Evaluierungen aufzubauen:** Wir möchten KI von Anfang bis Ende nutzen, d. h. von der Formulierung von Indikatoren und

Wirkungslogiken zu Beginn bis hin zur benutzerfreundlichen Kommunikation der Evaluierungsergebnisse.

Neben diesen positiven Entwicklungen ist meine Hauptsorge, dass wir, während wir KI-Tools zur Steigerung der Effizienz unserer täglichen Evaluierungsarbeit nutzen, **ein potenzielles Risiko der zu starken Abhängigkeit von der KI haben, was das Vertrauen in unsere Evaluierungsergebnisse gefährden könnte.** In dem Moment, in dem die von der KI bereitgestellten Informationen falsch sind und die menschliche Aufsicht unzureichend ist, könnten die Nutzer der Evaluierung beginnen, an der Gültigkeit unserer Ergebnisse zu zweifeln. **Dies könnte das Fundament unserer Rolle als Evaluatoren untergraben,** das darin besteht, wahrheitsgemäße Daten zu vermitteln und 100 % Rechenschaft abzulegen.“

Somit würde das Vertrauen geschwächt, das unserer Arbeit zugrunde liegt. Was denkst Du, Estelle?



„Ich verstehe Deine Bedenken, aber meine Einstellung hat sich ein wenig weiterentwickelt, und ich betrachte dieses Thema jetzt aus zwei Perspektiven. Zunächst gibt es die **Annahme, dass menschliches Urteilsvermögen immer korrekt ist**. Für bestimmte Aufgaben, wie z.B. Projektbewertungen, die einen Einzelfall betreffen, ist menschliche Intelligenz in der Tat unerlässlich. In meinem Bereich erfordern jedoch einige Aufgaben monatelanges manuelles Codieren und Klassifizieren von mehreren Analysten, die einem systematischen Template folgen und relativ einfache und sich wiederholende Entscheidungsregeln anwenden. In diesen Fällen besteht das Risiko menschlicher Fehler aufgrund der repetitiven Natur der Arbeit. In einigen unserer Experimente haben wir festgestellt, dass Maschinen in der Lage sind, Menschen bei dieser Art von Aufgaben zu übertreffen.

Zweitens müssen wir unsere Toleranz für Fehler berücksichtigen. Für einige Aufgaben benötigen wir 100 % Genauigkeit, da die Ergebnisse von großen Publikationen in der Öffentlichkeit aufgegriffen werden. In solchen Fällen ist menschliche Aufsicht entscheidend. Umgekehrt können wir für Ergebnisse, die früher im Evaluierungsproduktionsprozess stattfinden, einige Ungenauigkeiten akzeptieren. Zum Beispiel können wir bei der

Durchführung von Literaturübersichten auf Fehler bei Ein- oder Ausschlusskriterien in unserer Suche stoßen, aber **wenn die Gesamtergebnisse mit dem vergleichbar sind, was ein Mensch produzieren würde, könnten wir die Effizienz über die absolute Genauigkeit setzen**.

Letztendlich sehe ich zwei Bereiche, in denen ich bereit bin, nicht 100 % genau zu sein. Erstens wenn wir KI nutzen, um Aufgaben zu erledigen, die zuvor nicht machbar waren; dann können wir kalkulierte Risiken eingehen. Ob es darum geht, Bilder als Datenquellen zu nutzen oder große Textsegmente zu bearbeiten – das könnten wir ohne KI nicht tun, daher habe ich in diesem Bereich eine höhere Risikobereitschaft. Zweitens gibt es bestimmte Aufgaben, die weniger erfüllend, aber notwendig sind. Beispiele könnten manuelles Codieren, Datenklassifizierung usw. sein. In diesen Fällen ziehe ich es vor, Effizienz zu erhöhen, damit Analysten, Evaluatoren und Data Scientists ihre Zeit auf sinnvollere Arbeiten verteilen können. Das könnte bedeuten, dass wir die Arbeit anders ausrichten, um mehr Interviews zu führen, mehr Zeit und Energie für sinnvollen Austausch mit einer breiteren Palette von Interessengruppen aufzuwenden und somit tiefergehende Analysen durchzuführen.“

Danke, Estelle und Jochen, für das Gespräch.

Das Interview wurde durchgeführt von Dr. Jörg Langbein und Dr. Luciane Lenz.



Estelle Raimondo und Jochen Kluve besuchen eine Ausstellung in der Weltbank in Washington, DC.

Ein persönlicher Blickwinkel

Was ist die bedeutendste Auswirkung von KI auf Dein tägliches Leben?

„Meine eigene Rolle und Verantwortung haben sich durch KI erheblich weiterentwickelt. Ich habe das Privileg, mit wunderbaren Data Scientists zu arbeiten und unsere Daten- und Technologieinitiativen zu leiten.“



„Die Existenz von generativer KI zwingt uns dazu, die Art und Weise, wie wir arbeiten, neu zu überdenken, wie wir KI effektiv integrieren und nutzen können, wie wir Fallen vermeiden und wie wir die Zukunft gestalten wollen. Das ist inspirierend.“



Welche KI-basierte Anwendung nutzt Du am häufigsten?

„Copilot benutze ich viel zur Unterstützung bei alltäglichen Aufgaben.“



„ChatKfW.“



Was sind Deine größte Angst und Hoffnung in Bezug auf KI?

„Angst: Dass wir nicht in der Lage sind, KI zu steuern. Hoffnung: Dass wir es schaffen, KI zum Wohle aller zu meistern.“



„Angst: Dass wir und unsere Kinder bald nicht mehr unterscheiden können, was von KI erstellt wurde und was nicht. Hoffnung: Dass KI und menschliche Intelligenz auf eine vorteilhafte Weise interagieren.“



Kooperation in der deutschen Entwicklungszusammenarbeit

Eine gemeinsame Evaluierung

Ein Großteil der Projekte der deutschen Entwicklungszusammenarbeit werden im Rahmen von Regierungsverhandlungen zwischen dem Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) und der Regierung im jeweiligen Partnerland ins Leben gerufen. **Die tatsächliche Umsetzung obliegt dann dem Partnerland gemeinsam mit einer der staatlichen Durchführungsorganisationen (DO):** Neben der KfW Entwicklungsbank zählen hierzu die Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), die Physikalisch-Technische Bundesanstalt (PTB) und die Bundesanstalt für Geowissenschaften und Rohstoffe (BGR).

Die DOs begegnen sich also häufig – ausgestattet mit unterschiedlichen Kompetenzen – vor Ort in den Partnerländern und arbeiten teilweise in ähnlichen Sektoren oder mit denselben Partnern, um die Ziele der deutschen Entwicklungszusammenarbeit umzusetzen. Ganz im Sinne von „alleine sind wir stark, zusammen sind wir stärker“ **ist eine angemessene Kooperation mit Fokus auf die jeweiligen Kompetenzen also wesentlich, um ein Zusammenwirken verschiedener Projekte zu ermöglichen, um Synergien auszuschöpfen und im Ausland mit „einer Stimme“ aufzutreten.** Insbesondere die Kooperation zwischen Vorhaben der GIZ, der sog. „Technischen Zusammenarbeit (TZ)“, und Vorhaben der KfW Entwicklungsbank, der sog. „Finanziellen Zusammenarbeit“ (FZ), ist in vielen Partnerländern Realität und wird vom BMZ mitunter strategisch aufgesetzt.

Gemeinsam mit der Stabstelle Evaluierung der GIZ hat die Evaluierungsabteilung der KfW Entwicklungsbank diese Kooperation untersucht, um das gemeinsame institutionelle Lernen zu fördern. Die beiden unabhängigen Evaluierungseinheiten widmeten sich den Fragen, welche Kooperationsformen gibt es in den Partnerländern in der Praxis? Welche Faktoren der Zusammenarbeit sind relevant für Erfolge? Welchen Einfluss hat die Kooperation auf die entwicklungspolitische Wirkung der Projekte, also auf Bereiche wie Armutsbekämpfung, Umweltschutz, Bildungs-, Gesundheits-, oder Infrastrukturförderung? An den Interviews, Fokusgruppensitzungen und einer Online-Befragung nahmen über 1.000 Mitarbeitende der GIZ und der KfW Entwicklungsbank teil.

Die Ergebnisse der Evaluierung zeigen, dass Kooperation zwischen den beiden DOs für viele Mitarbeitende zum Arbeitsalltag gehört: Rund zwei Drittel der GIZ-Befragten und 90 % der KfW-Befragten berichteten von Kooperationserfahrungen. Dabei wird auf Grund unterschiedlichster Kontexte und Konstellationen nicht nur viel, sondern auch auf vielfältige Weise kooperiert: Die Bandbreite reicht von festen Austauschformaten, gemeinsamen Planungsworkshops und Fortschrittskontrollen bis hin zu Ko-Produktionen von Studien und gemeinsamer Öffentlichkeitsarbeit.



Ein Eindruck aus der Entwicklungszusammenarbeit vor Ort in Cotonou, Benin.

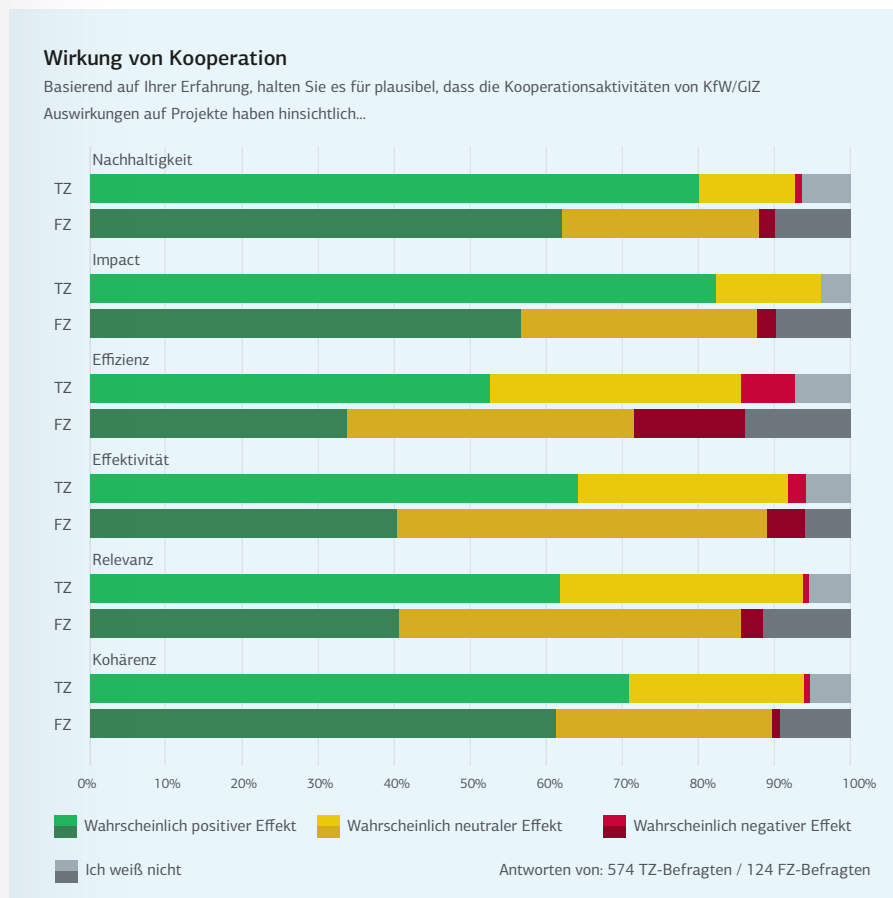
Kooperation lohnt sich also – jedoch nicht in allen Vorhabens-Konstellationen gleichermaßen: Sie hat dort den größten Mehrwert, wo es Potentiale gibt für Synergien, gegenseitige Risikominderung oder die Auflösung von Widersprüchen und Konkurrenz. Synergien entstehen zum Beispiel im Sektor Berufsbildung, wenn die GIZ die Lehrplanentwicklung und Trainings für Auszubildende unterstützt während die KfW Entwicklungsbank die für die Ausbildung erforderlichen Räumlichkeiten und technischen Geräte finanziert. In einem anderen Beispiel führte die Kapazitätsstärkung von kleinunternehmerischen Kreditnehmern durch die GIZ zu höheren Rückzahlungsraten von durch die KfW Entwicklungsbank finanzierten Kreditlinien. Dies minderte die finanziellen Risiken der FZ.

Insgesamt zeigt die Evaluierung, dass eine enge Kooperation positive Auswirkungen auf die Projekte der deutschen Entwicklungszusammenarbeit haben kann: Insbesondere auf die Effektivität, Impact und Nachhaltigkeit von Vorhaben. Gleichzeitig trägt die Kooperation zur Kohärenz der deutschen Entwicklungszusammenarbeit bei.

Die Evaluierung identifiziert darüber hinaus Erfolgsfaktoren sowie gute Rahmenbedingungen für erfolgreiche Kooperation. Dazu gehört zum Beispiel das Wissen über die Arbeitsweise der jeweils

anderen Organisation, das aktive Mitdenken von Kooperationen, wenn Projekte von einem an den nächsten Projektmanagerin oder Projektmanager übergeben werden, und die aktive Förderung von Kooperation durch Führungskräfte.

Für die zukünftige Gestaltung der Zusammenarbeit bietet die Evaluierung wertvolle Ansatzpunkte, welche auch schon in Unternehmensentscheidungen aufgegriffen wurden. So haben neue Mitarbeiterinnen und Mitarbeiter inzwischen Gelegenheit, zu Beginn ihrer Tätigkeit die jeweils andere Organisation kennen zu lernen und sich zu vernetzen. Außerdem läuft ein Austausch über konkrete „good practices“ in der Kooperation und es gibt Initiativen, die gemeinsame Öffentlichkeitsarbeit in den Partnerländern auszubauen. Auch die erfolgreiche Zusammenarbeit der beiden Evaluierungseinheiten für diese Evaluierung ist ein Beispiel der zielführenden und Mehrwert schaffenden Kooperation zwischen GIZ und KfW.



Die deutsche Entwicklungszusammenarbeit: GIZ und KfW Entwicklungsbank im gemeinsamen Büro in Yaoundé, Kamerun.

Die Wirkung von Kooperation auf Projekte der Entwicklungszusammenarbeit.

Evaluierungsbericht 2023-2024: Methodik



Von der Stichprobe zur Erfolgsquote

FZ Evaluierung zieht jährlich eine repräsentative Stichprobe, die für die Berechnung der Erfolgsquote genutzt wird.





Ex-post-Evaluierung

Von der Grundgesamtheit bis zur Erfolgsquote

Die Evaluierungsabteilung der KfW Entwicklungsbank zieht jährlich eine nach Sektoren geschichtete, repräsentative Zufallsstichprobe aus der Grundgesamtheit der abgeschlossenen FZ-Vorhaben. Für alle Vorhaben in der Stichprobe muss eine Ex-post-Evaluierung durchgeführt werden. Dieses Verfahren sichert die statistische Repräsentativität und Aussagekraft der Evaluierungen für das FZ-Portfolio und wahrt die Unabhängigkeit bei der Auswahl der zu evaluierenden Vorhaben. Im Zweijahreszeitraum 2023-2024 wurden aus einer Grundgesamtheit von 268 abgeschlossenen Vorhaben 129 in die Stichprobe gezogen. Die Grundgesamtheit entsprach in der Größenordnung in etwa dem langjährigen Durchschnitt.

Nicht alle Evaluierungen können direkt im Jahr der Stichprobenziehung durchgeführt werden: Aus der Stichprobe von 2023-2024 wurden bereits 65 Vorhaben evaluiert. Zusätzlich wurden 49 Vorhaben aus den Stichproben der Vorjahre evaluiert.

Des Weiteren evaluiert FZ Evaluierung zusätzlich Vorhaben anhand einer „Zubündelung“, wenn diese Vorhaben stimmig mit einer durchzuführenden Evaluierung der Stichprobe kombinierbar sind – also auch, wenn diese nicht selbst Teil einer Stichprobe sind. Kriterien hierfür wären zum Beispiel räumliche und zeitliche Nähe zu Vorhaben aus der Stichprobe. Im Evaluierungszeitraum 2023-2024 wurden 26 solcher „zugebundelter“ Vorhaben evaluiert.

Zusammenfassend: Insgesamt wurden in den Jahren 2023-2024 140 Vorhaben der Finanziellen Zusammenarbeit evaluiert. Hiervon stammten 65 aus der Stichprobe 2023-2024 und 49 aus vorherigen Stichproben, während 26 zugebündelt wurden.

Aus der Grundgesamtheit 2023-2024 von 268 Vorhaben wurden also 65 bereits evaluiert (blauer Balken), 64 werden noch evaluiert (grüner Balken) und 139 wurden nicht für eine Evaluierung gezogen (grauer Balken).

Die zu evaluierenden Vorhaben werden von FZ Evaluierung anhand der sechs OECD DAC-Kriterien bewertet: Relevanz, Kohärenz, Effektivität, Effizienz, Impact und Nachhaltigkeit.

Die Bewertung der Kriterien richtet sich nach den Evaluierungsleitlinien des Bundesministeriums für wirtschaftliche Entwicklung und Zusammenarbeit (BMZ). Konkret bedeutet das: Jedes Kriterium wird separat betrachtet und jeweils in der Spanne von eins (sehr erfolgreich) bis sechs (gänzlich erfolglos) bewertet. Darauf basierend wird die Gesamtbewertung des Vorhabens ermittelt, welche in der Regel das arithmetische Mittel der sechs Teilnoten darstellt. Dies ist allerdings nicht immer der Fall: Besonders schwerwiegende Mängel bei einzelnen Teilkriterien sollen nicht durch positive Ergebnisse bei anderen Teilkriterien ausgeglichen werden. Deshalb bewirken Bewertungen schwächer als „3 = eingeschränkt erfolgreich“ in den Kriterien Effektivität, Impact und

Von der Grundgesamtheit...



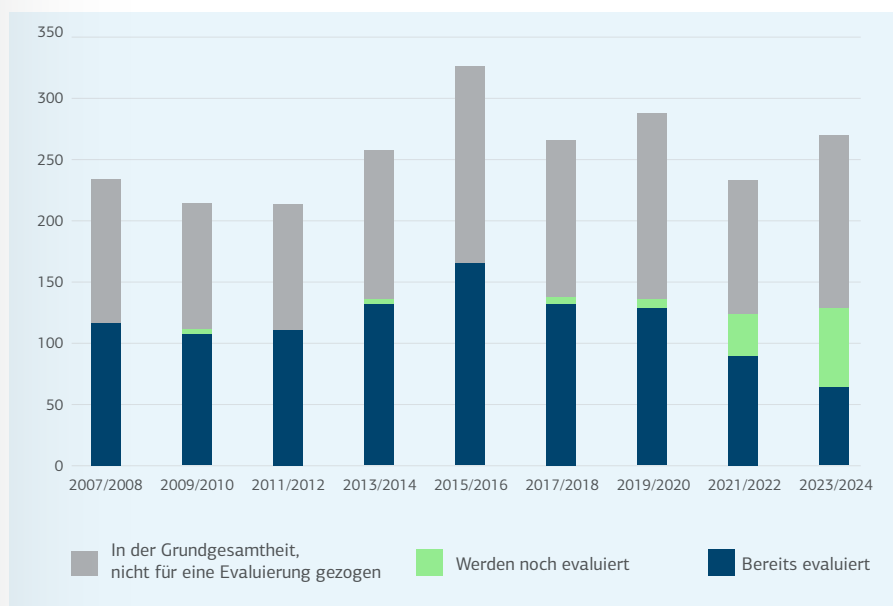
Von der Grundgesamtheit bis zur Erfolgsquote: Ein Prozess in neun Schritten.

Nachhaltigkeit, dass das Projekt insgesamt nicht als erfolgreich bewertet werden kann. Die grundlegende Idee: Vorhaben, die nicht (nachhaltig) wirken, können nicht „erfolgreich“ sein.

Die übergreifende Erfolgsquote wird dann basierend auf allen Einzelevaluierungen geschätzt: Die Quote ist der statistische Anteil der erfolgreichen FZ-Vorhaben – also der Anteil der Vorhaben, die in der Gesamtbewertung eine 3 oder besser erhielten. Zur Schätzung der Erfolgsquote werden nur Vorhaben herangezogen, die als Teil der Stichprobe desselbigen Jahreszeitraums, hier 2023-2024, gezogen wurden (unverzerrte Schätzung). Für den Zeitraum 2023-2024 gehen also nur die 65 bereits evaluierten Vorhaben ein, da sie Teil der Stichprobe 2023-2024 sind.

Die Erfolgsquote für den Zeitraum 2023-2024 liegt bei 85 %, also etwas über dem Durchschnitt der Erfolgsquoten von 2007-2024 (81 %), kann sich infolge der noch zu erfolgenden Evaluierungen der Stichprobe aber noch verändern.

Bei den Erfolgsquoten gilt: Sie sind Schätzungen und daher mit statistischer Unsicherheit behaftet. Das Konfidenzintervall gibt Auskunft über die Genauigkeit der Schätzung. Die Erfolgsquoten der vergangenen Zweijahresperioden seit Einführung des Stichprobenverfahrens 2007/2008 unterscheiden sich zwar leicht voneinander (81 % und 85 %). Dieser Unterschied ist jedoch nicht statistisch signifikant, d.h. die Werte liegen im gleichen Konfidenzintervall und sind somit nicht sicher unterschiedlich voneinander, da der geringe Unterschied zwischen ihnen dem Zufall geschuldet sein kann.



Anzahl aller abgeschlossenen FZ-Vorhaben (Grundgesamtheit), in Stichprobe und bereits evaluiert, nach zweijähriger Stichprobe von 2007/2008 - 2023/2024

...bis zur Erfolgsquote



FZ Evaluierung stellt sich vor

Wir sind eine unabhängige Evaluierungsabteilung, die anerkannte und moderne Methoden in drei Säulen der Evaluierungsarbeit einsetzt: Ex-post-Evaluierungen, KfW Development Impact Lab und Institutionelles Lernen.

Vision

Wir stärken die Wirksamkeit der Finanziellen Zusammenarbeit (FZ) im Sinne der Agenda 2030 und ihrer nachhaltigen Entwicklungsziele (SDGs), indem wir Vorhaben der FZ mit unseren evaluativen Kenntnissen konstruktiv begleiten, sie kritisch bewerten und Lessons Learned für zukünftige Vorhaben ableiten.

Mission Statement

Wir kombinieren interne und externe Evaluierungskompetenz, um Wirkungen abgeschlossener Vorhaben der FZ systematisch zu bewerten und laufende Vorhaben der FZ durch begleitende Wirkungsmessung zu unterstützen. Wir stärken das Institutionelle Lernen, indem wir die Ergebnisse unserer Evaluierungsarbeit in

zielgruppengerechten und effizient nutzbaren Produkten aufbereiten. Wir streben nach fachlicher Exzellenz, fortlaufender Weiterentwicklung und Innovation.

Über uns

Als agiles Team aus Sektorexpertinnen und -experten mit wissenschaftlicher und/oder operativer Erfahrung verstehen wir uns innerhalb der KfW Entwicklungsbank als „Knowledge Hub“ zum Thema Evaluierung und Wirkungsmessung. Wir kooperieren mit dem Bundesministerium für wirtschaftliche Zusammenarbeit (BMZ) und vernetzen uns mit anderen nationalen wie internationalen Institutionen der Evaluierung in der Entwicklungszusammenarbeit. Die Einbeziehung lokaler Partnerinnen und Partner sowie ihr Erkenntnisinteresse sind uns wichtig. Wir berücksichtigen in unserer Arbeit relevante Standards der Evaluierungscommunity (DeGEval, OECD-DAC, Global Evidence Commitment) sowie der empirischen Forschung, insbesondere hinsichtlich des Schutzes von Persönlichkeitsrechten und der Anonymität persönlicher Daten sowie eines „ethical code of conduct“.

Drei Säulen der Evaluierungsarbeit



KfW Development Impact Lab

- Wir begleiten ausgewählte laufende Vorhaben der FZ evaluativ, indem wir maßgeschneiderte Designs zur fortlaufenden Wirkungsmessung entwickeln und umsetzen.
- Wir beantworten inhaltliche Fragestellungen, die von besonderer Relevanz für die FZ sind.
- Wir verwenden moderne Methoden der Rigorosen Impact Evaluierung (RIE) und wählen aus der RIE-Toolbox das für das jeweilige Vorhaben passende Design (form follows function).
- Wir streben die innovative Sammlung und Nutzung von Daten an, insbesondere Satellitendaten.
- Wir kooperieren fallspezifisch mit passenden externen Partnern, insbesondere Forschungseinrichtungen und anderen Entwicklungsbanken.



Ex-post-Evaluierungen

- Wir evaluieren jährlich eine nach Sektoren geschichtete, repräsentative Stichprobe abgeschlossener Vorhaben der FZ.
- Wir führen Ex-post-Evaluierungen (EPE) selbst durch und arbeiten mit externen Gutachtern sowie Abordnungen aus den operativen Bereichen der KfW Entwicklungsbank zusammen.
- Ex-post-Evaluierungen werden seit 2000 systematisch durch FZ Evaluierung durchgeführt. Sie dienen der Rechenschaftspflicht und bilden den Erfahrungsschatz für Institutionelles Lernen.
- Die Bewertung erfolgt systematisch nach OECD-DAC Kriterien.
- Ex-post-Evaluierungen sind die tragende Säule unserer Evaluierungsarbeit.



Institutionelles Lernen

- Wir orientieren uns am Erkenntnisinteresse der Zielgruppe und informieren aktuell und in modernen Formaten.
- Wir bieten thematische Lernformate an, veranstalten Fortbildungen zu Evaluierungsmethoden und bringen unser Wissen in Peer Discussions und Sector Retreats ein.
- Operative Kolleginnen und Kollegen führen gemeinsam mit uns aktiv Ex-post-Evaluierungen durch (Abordnungen).
- Über Newsletter und Publikationsreihen informieren wir über neue Ergebnisse.
- Digital und interaktiv stellen wir die Inhalte aller seit 2007 durchgeführten Ex-post-Evaluierungen in einer IDEaL-App (Interactive Database Evaluation and Learning) bereit.

**KfW Bankengruppe**

Geschäftsbereich KfW Entwicklungsbank
Palmengartenstraße 5 – 9
60325 Frankfurt am Main

Tel. +49 69 7431-0

Fax. +49 69 7431-1234

FZ-Evaluierung@kfw.de

www.kfw.de

Gestaltung und Umsetzung

new office GmbH, Frankfurt am Main

Stand: Mai 2025

Bildnachweis

Titel: Jonas Wresch; S. 1 Jonas Wresch; S. 2 Thomas Meyer; S. 2 + 3 new office; S. 4 + 5 Jonas Wresch; S. 6 + 7 Jonas Wresch; S. 8 + 9 new office, Svenja Klingspor; S. 9 Jens Drillisch, Sebastian Manthey; S. 10 Yahia Maraka; S. 12 + 13 Nicole Karnaus, Jonas Wresch; S. 14 new office; S. 15 Jean Joseph Manampisoa RATOVOSON, Ony Rakotoarisoa, Ingrid Dallmann; S. 16 Kati Krähnert; S. 17 Nguyen Minh Phuong, Mahesh Joshi, Sven Lorenz; S. 18 + 19 Jonas Wresch; S. 20 + 21 new office, Svenja Klingspor, André Hensel; S. 22 + 23 Svenja Klingspor, Jonas Wresch; S. 24 + 25 Jonas Wresch, Mark Naftalin; S. 26 + 27 Svenja Klingspor, Noten, die FZ; S. 28 + 29 Jonas Wresch; S. 30 + 31 Jonas Wresch; S. 32 + 33 Jonas Wresch; S. 34 Estelle Raimondo, Jochen Kluge; S. 35 + 38 + 39 + 40 Melvin Wong; S. 42 Jonas Wresch; S. 43 Franziska Grimm, S. 44 + 45 Jonas Wresch; S. 46 + 47 new office; S. 50 Jonas Wresch

