DEval

DEUTSCHES
EVALUIERUNGSINSTITUT
DER ENTWICKLUNGS-
ZUSAMMENARBEIT

Von der Meta-Evaluierung zur Effizienz zur Stärkung einer effizienzorientierten Steuerung in der Entwicklungszusammenarbeit

DeGEval-Jahrestagung 2026, Session D1

Freitag, 25. September 2026, 09:30-11:00

Dr. Karolin Süß, Dr. Martin Noltze

Einleitung

- Zunehmende **globale Herausforderungen**, Zweifel an der Legitimität der EZ und **Mittelkürzungen** in der internationalen Entwicklungsfinanzierung
- **BMZ-Reformagenda** zielt auf die Stärkung der wirkungs- und **effizienzorientierten Steuerung**
- **Evaluierung als wirkmächtiges Instrument** für eine evidenzbasierte Politikgestaltung
- **Gut aufgestellte deutsche EZ-Evaluierung**:
  - Arbeitsteiliges Evaluierungssystem, Leitlinien und Standards, einschlägige Kriterien und hoher Deckungsgrad im Vergleich deutscher Politikfelder
  - Hohe Ansprüche und Transparenz im internationalen Vergleich

Evaluierungskriterium Effizienz

Definition

Wirtschaftliche und zeitgerechte Produktion von Ergebnissen

Dimensionen

Produktionseffizienz

Allokationseffizienz

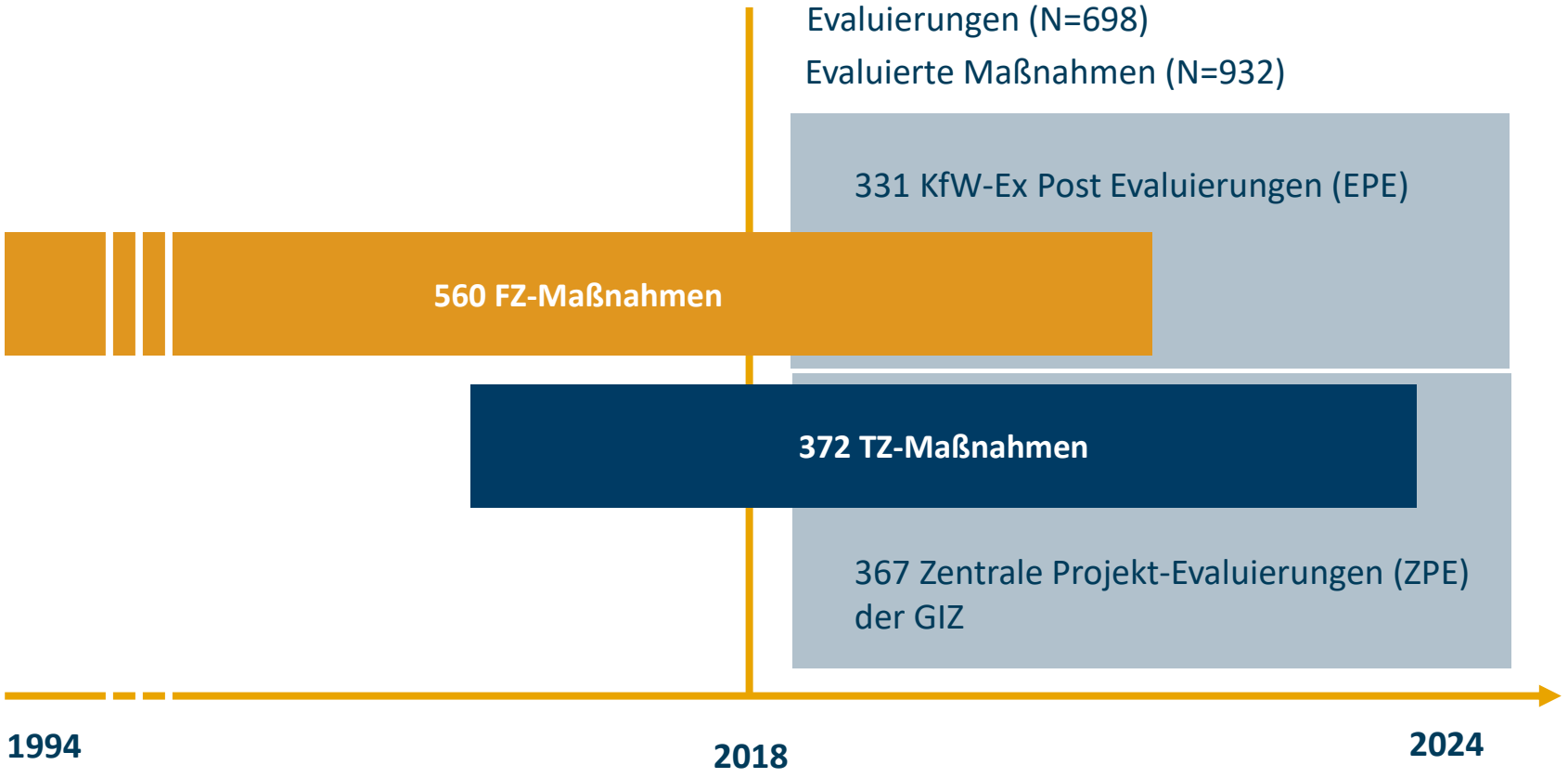
Prinzipien

Sparsamkeit

Ergiebigkeit

DEval

Evaluierungsgegenstand



Design

Vergleich Politikfelder

Evaluierungs-
synthese

Merkmale
Maßnahmen
Kontext
Evaluierung

Klassifizierung

QUAL
Inhaltsanalyse

Stichprobe:
KfW 46 EPE.
GIZ 51 ZPE

KI

Claude 4.0,
4.5 Sonnet

Meta-
Evaluierung

Internat.
Vergleich

Qualität

QUANT
Statistik

KI

BERTopic, sLDA,
Sentiment-BERT

Modellierung

Die Evaluierung findet eine hohe Produktionseffizienz der Maßnahme. Dabei wurden die Mittel sparsam eingesetzt. Gemeinsam mit dem Partner konnten Effizienzgewinne erzielt werden. Die finanzielle Stabilität im Sektor hat zu Zinsvergünstigungen und einer nachhaltigen Re-Finanzierung beigetragen.

TOP1 TOP2 TOP3

Themen &
Zusammenhänge

DEval

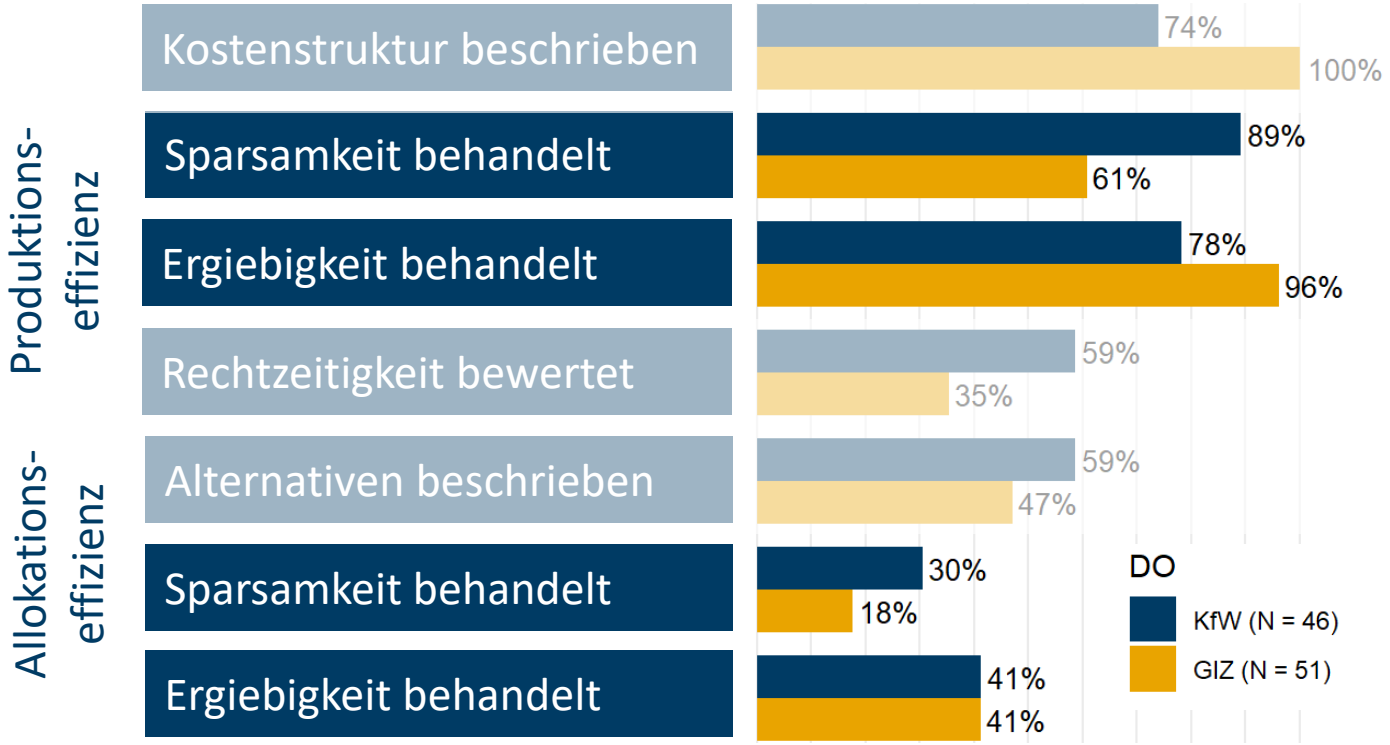
Meta-Evaluierung

- Qualität der Evaluierung

DEval

Anwendung Effizienzkriterium

Befund: Evaluierung der Produktionseffizienz und Ergiebigkeit gut aufgestellt

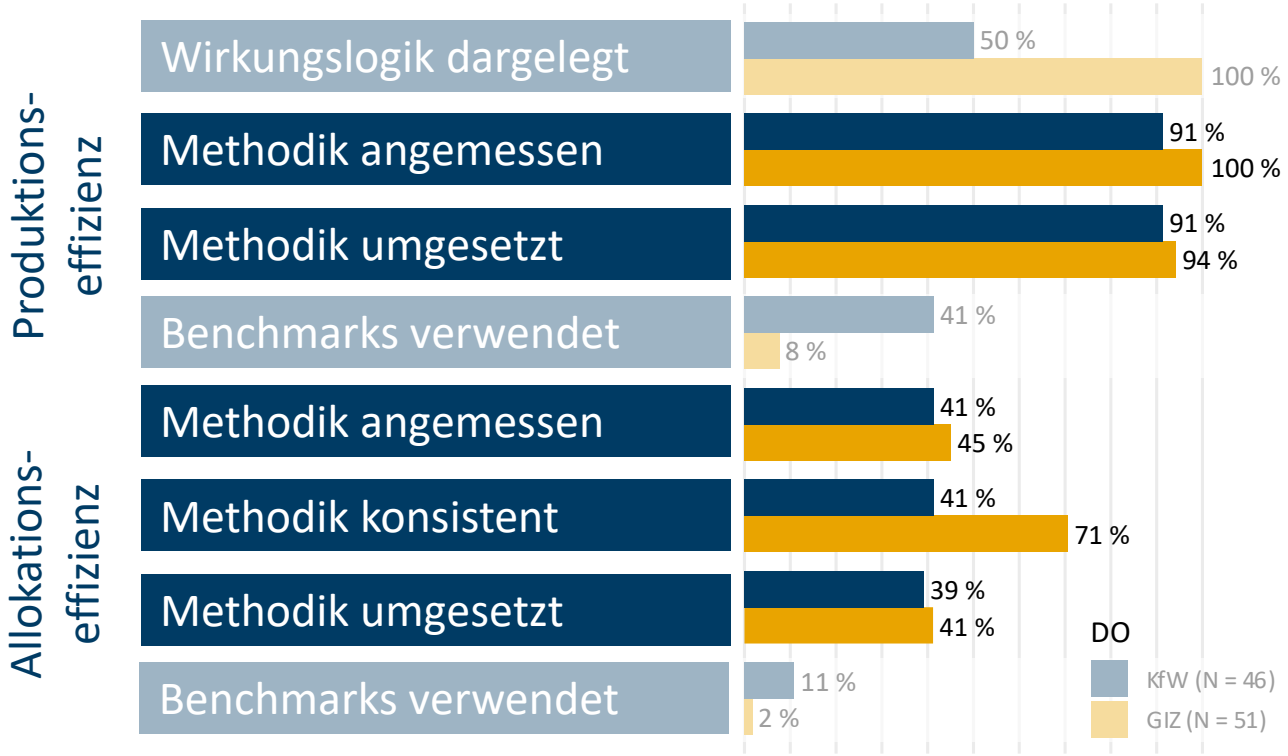


Das Prinzip der Sparsamkeit könnte mehr Aufmerksamkeit erfahren

DEval

Methodische Qualität

Befund: Methodische Qualität der Evaluierungen größtenteils angemessen



Die Evaluierungen könnten die Bewertung der Allokations-effizienz stärken

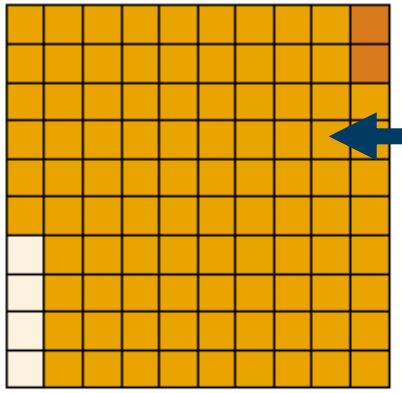
DEval

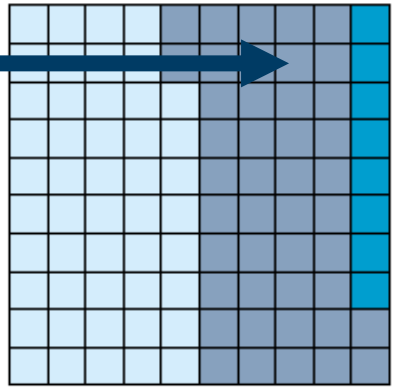
Angewandte Methoden

GIZ

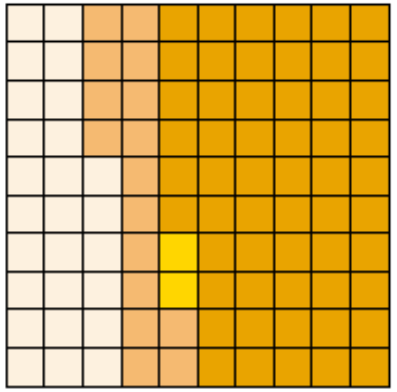
KfW

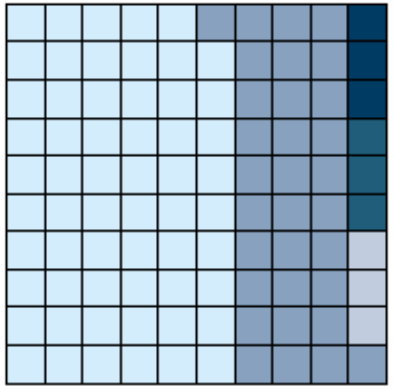
Produktions-effizienz





Allokations-effizienz



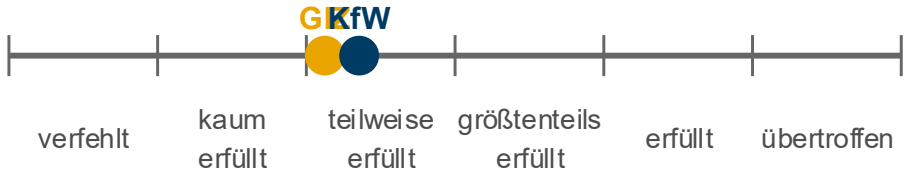


Methoden zur Evaluierung einzelner Maßnahmen
Benchmarking of unit costs
***Discounted Cash Flow Analysis* (DCF)**
Systematic Screening Approaches
Comparative ratings by stakeholders
Structured expert judgement (SEJ)
Stakeholder-driven approaches

Methoden zum Vergleich von Maßnahmen
Multi-Attribute Decision-Making (MADM)
Cost-Benefit Analysis (CBA)
Social Return on Investment (SROI)
Cost-Effectiveness Analysis (CEA)
Cost-Utility Analysis (CUA)
Production frontier methods

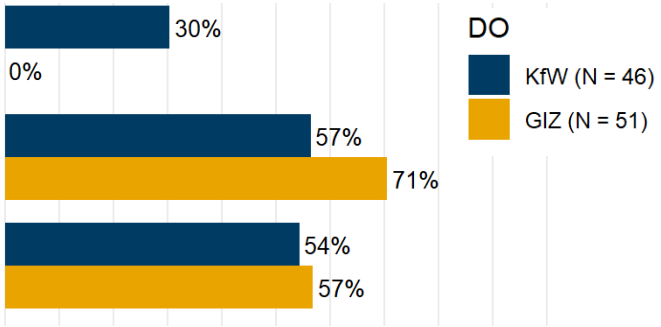
DEval

Evaluierungseffizienz



Befund: Evaluierungen arbeiten konzeptionell und methodisch effizient

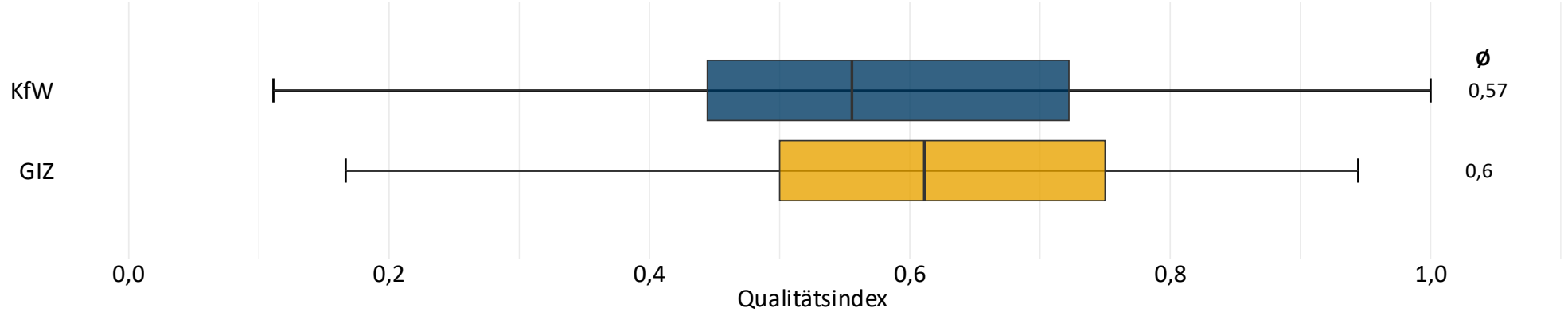
- Wirtschaftlichkeitsuntersuchungen
- Konsistenz der Ergebnisse
- Lernen



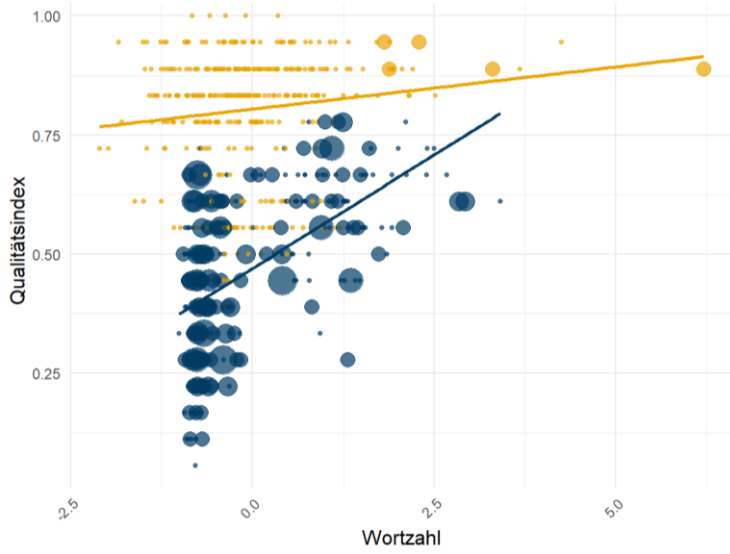
Die Evaluierungen könnten stärker auf vorangegangene Wirtschaftlichkeits-untersuchungen zurückgreifen

DEval

Evaluierungsqualität (insgesamt) und deren Einflussfaktoren



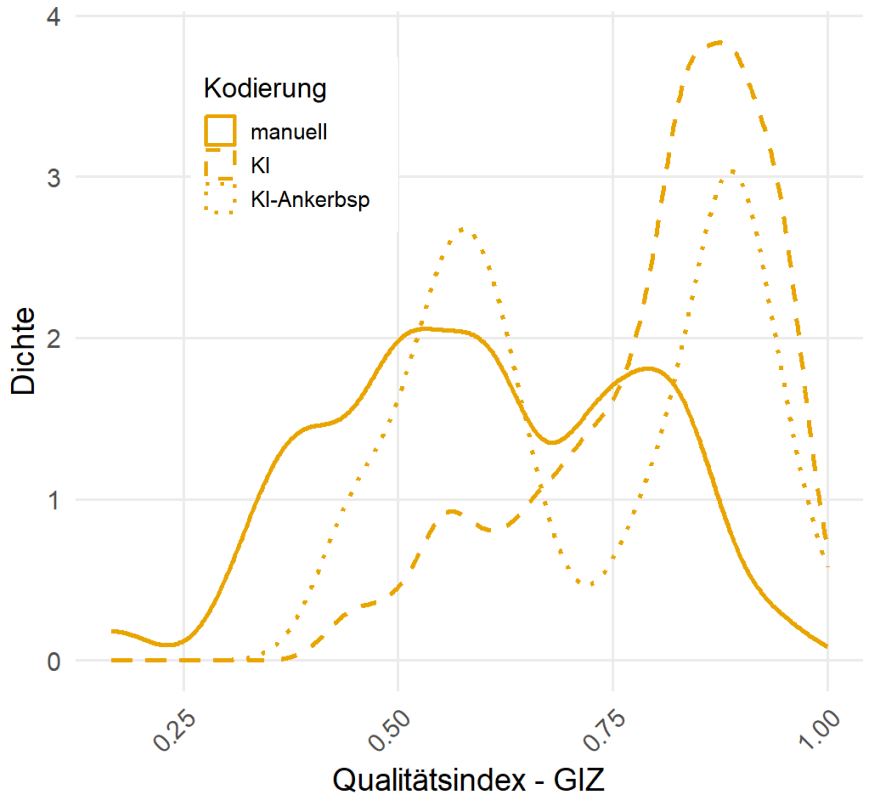
- Höhere Qualität bei großen Maßnahmen (KfW), geringere Qualität bei komplexen Maßnahmen (GIZ)
- Keine Zusammenhänge zwischen Qualität & Kontext
- Längere Berichte gehen mit einer höheren Qualität einher

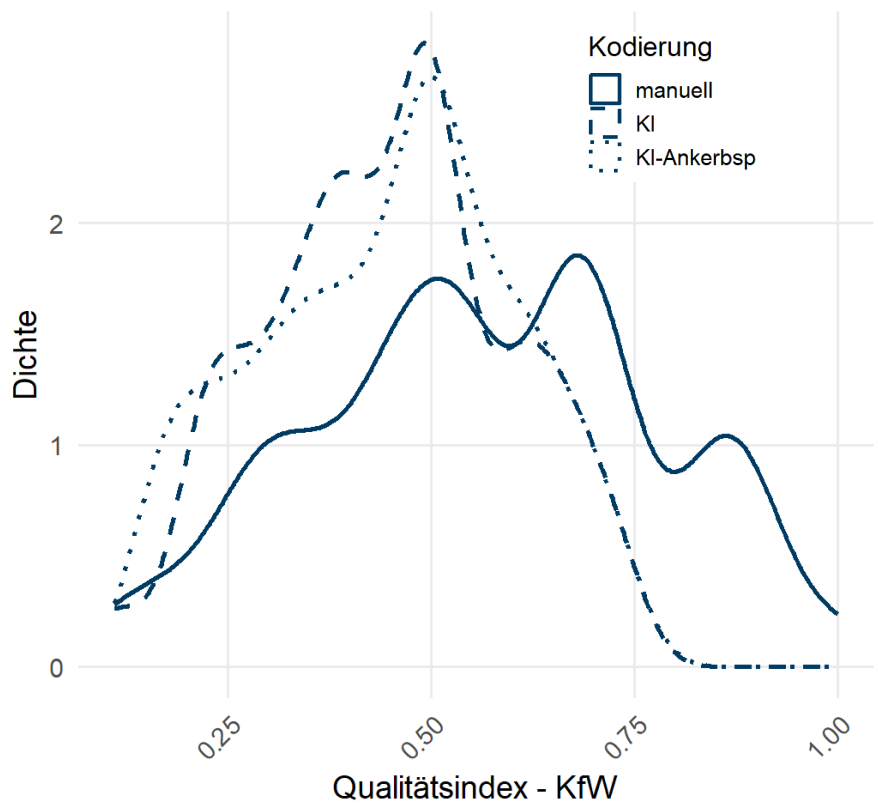


KI Ergebnisse basieren auf KI-Klassifizierung – Modell: Claude Sonnet 4.5 (Anthropic, 2025) und multivariater Statistik

DEval

KI-Analysen bestätigen die Ergebnisse der manuellen Qualitätsbewertung teilweise



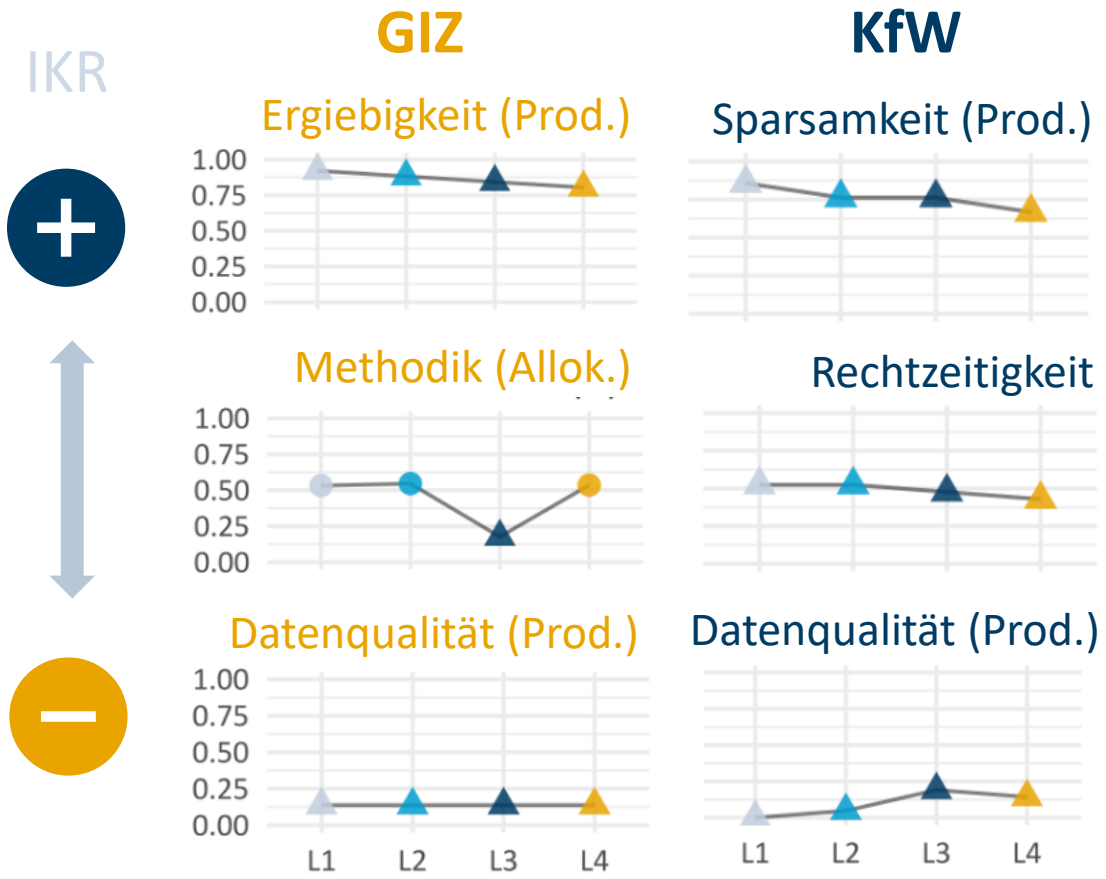


- Unterschiede zwischen Menschen und KI und
- Qualität der GIZ ZPE durch KI tendenziell besser bewertet
- Qualität der KfW EPE durch KI tendenziell schlechter bewertet

Ergebnisse basieren auf manueller Inhaltsanalyse & KI-Klassifizierung,
Modell: Claude Sonnet 4.5 (Anthropic, 2025)

DEval

Unterschiede in der Interkodierreliabilität zwischen einzelnen Kriterien



- KI-Klassifizierungen können die Ergebnisse der manuellen Kodierung nur für bestimmte Qualitätskriterien reproduzieren
- Erkannte Ursachen für Diskrepanzen
  - Layout (eingeschränkte Maschinenlesbarkeit)
  - Vertrauliche Berichtsteile (keine KI-Analysen)
  - Komplexität (z.B. bei gebündelten Maßnahmen)

Abbildung zeigt IKR zwischen manueller Kodierung und vier KI-Durchläufen
L1 = Claude Sonnet 4.0
L2 = Claude Sonnet 4.5
L3 = Claude Sonnet 4.5 (mit angepassten Prompts)
L4 = Claude Sonnet 4.5 (mit angepassten Prompts + Ankerbeispielen)

Lauf L1 L2 L3 L4 Metrik (dyn) Kappa PABAK

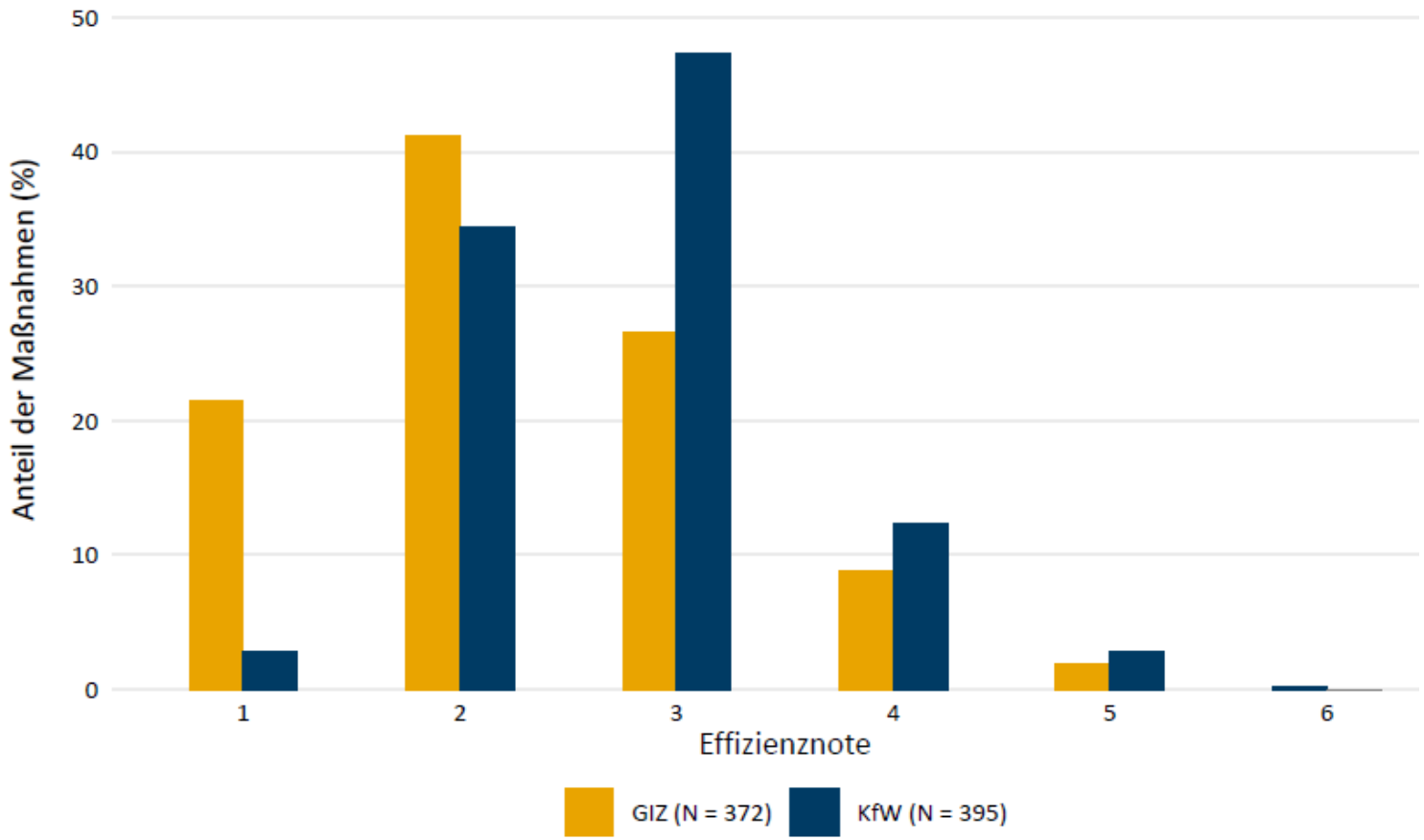
Ergebnisse basieren auf manueller Inhaltsanalyse & KI-Klassifizierung,
Modell: Claude Sonnet 4.0 & 4.5 (Anthropic, 2025)

DEval

Evaluierungssynthese

- Ergebnisse der Evaluierung und Einflussfaktoren

Verteilung der Effizienznote



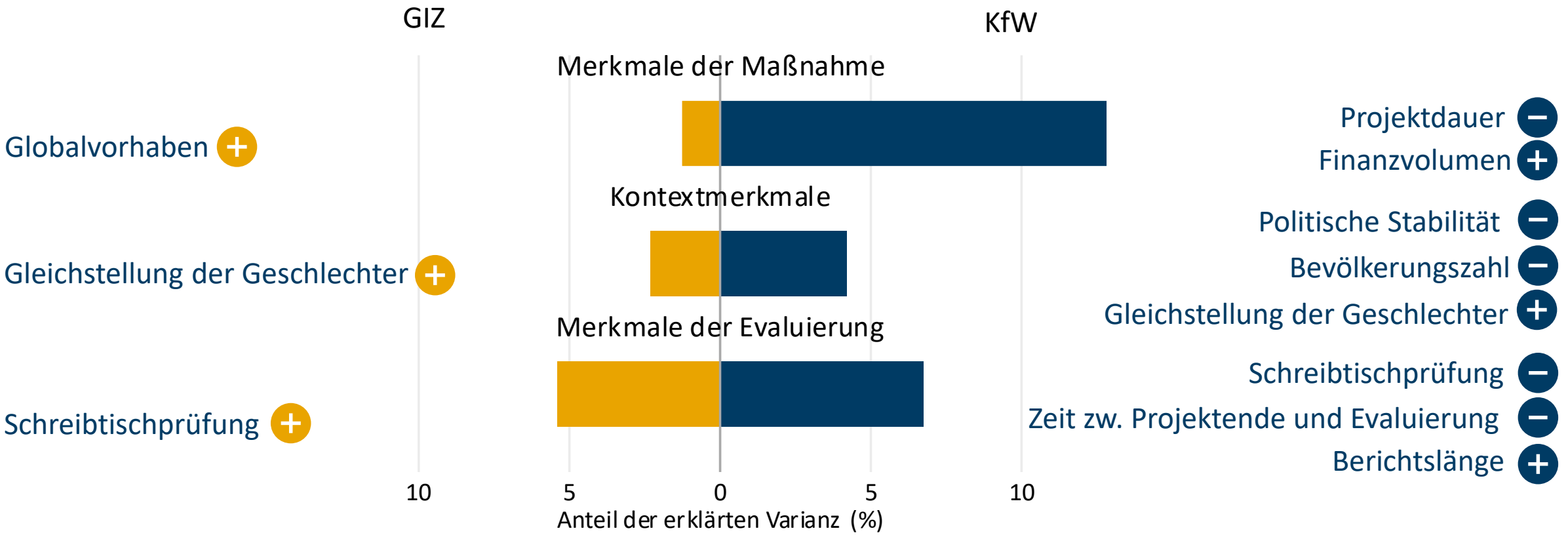
- Maßnahmenwerden als überwiegend effizient bewertet
- Maßnahmen der technischen Zusammenarbeit werden systematisch besser bewertet

Anmerkung: Dargestellt sind die relativen Häufigkeiten der vergebenen Effizienznoten nach Durchführungsorganisation. Die Abbildung enthält alle evaluierten Maßnahmen der Grundgesamtheit (N = 932).

DEval

Effizienzbewertung und Einflussfaktoren

Insgesamt geringe Nachvollziehbarkeit der Bewertung und nicht-intendierte Effekte



Fazit

- Empfehlungen der Meta-Evaluierung und Evaluierungssynthese
- Weitere Implikationen

Empfehlungen der Meta-Evaluierung und Evaluierungssynthese

1. Sparsamkeit und Ergiebigkeit gleichrangig behandeln
2. Bewertung auf Vergleichen basieren
3. Evaluierung Allokationseffizienz stärken
4. Wirtschaftlichkeitsuntersuchungen nutzen
5. Unabhängigkeit von Evaluierungsformat und Bewertung
6. Auseinandersetzung mit Effizienzansprüchen
7. Nachvollziehbarkeit und Replizierbarkeit der Bewertung erhöhen
8. Aufbau Evidenzarchitektur unterstützen

Weitere Implikation für die (EZ-)Evaluierung zur Stärkung der effizienzorientierten Steuerung und Durchführung von Maßnahmen

- Evaluierungen (noch) **stärker auf die Bewertung der Effizienz ausrichten**
- **Effizienzorientierte Evaluierungsprogramme** (über Projektevaluierungen hinaus) bilden und **nachfrageorientierte Produkte** (Efficiency Gaps Maps, Synthesen, Benchmark-Datenbanken, ChatBots) entwickeln
- **Breite des Methodenkatalogs** ausnutzen
  - Evaluierung einzelner Maßnahmen
  - Evaluierungen zum Vergleich mehrerer Maßnahmen
- **Lernpartnerschaften** (mit Wissenschaft und EZ-Organisationen) bilden und **Innovationen (Digitalisierung, KI)** vorantreiben
- **Effizienz der Evaluierung stärken**

Vielen Dank für die Aufmerksamkeit!

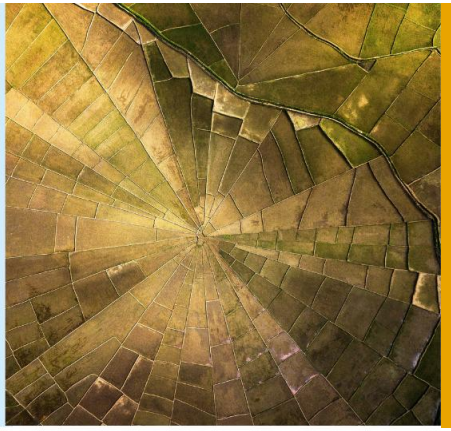
www.deval.org

➡️ Zur Anmeldung geht es hier entlang: https://lnkd.in/gAz6-T58

DEval

Effizienz in der deutschen Entwicklungszusammenarbeit

Meta-Evaluierung und Evaluierungssynthese



online 28.09.2026 14:00-15:30 Uhr

Paneldiskussion mit:
Uwe Gehlen (Referatsleiter Ressortforschung und Evaluierung, BMZ)
Prof. Dr. Jochen Kluve (Leitung Evaluierung, KfW)
Christian Schulze-Koch (Gruppenleitung Stabsstelle Evaluierung, GIZ)

DEval

Beispiel einer Effizienz-Geschichte, KfW: Finanzierung von Infrastrukturmaßnahmen

Themenmodellierung *(offen)*

Themenmodellierung *(Sentiment)*

Themenmodellierung *(zielgerichtet. sLDA)*

LASSO *(Regressionsanalysen)*

Literatur *(Analyse)*

OLS *(Regressionsanalyse)*

- **Vergabe- und Bauverzug** und damit Kosten als relevantes Thema in den Effizienzkapiteln
- **Verzögerte Umsetzung** insbesondere in CORONA-Zeiten negativ besetzt
- **Straßenbaukosten und -standards** gehen mit einer besseren Effizienzbewertung einher
- Die **finanzielle Ausstattung von Maßnahmen** zeigt sich als potenziell positiver Einflussfaktor
- Vor allem **großvolumige Infrastrukturmaßnahmen** gehen mit einer erhöhten Effizienz einher
- **Je höher das Finanzvolumen, desto positiver die Effizienzbewertung**

Ergebnisse basieren auf KI-Themenmodellierung und multivariater Statistik

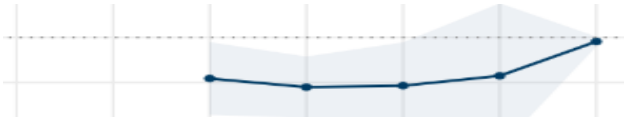
DEval

Effizienz-Geschichte, GIZ: Ressourceneinsatz für Beratungsmaßnahmen

Themenmodellierung *(offen)*

- **Personalkosten** als relevantes Thema in den Effizienzkapiteln

Themenmodellierung *(Sentiment)*



- Zunehmend positive Tonalität im Hinblick auf **Personalkosten und Budgetverteilung**

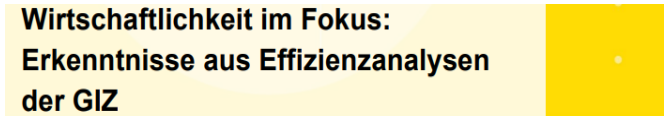
Themenmodellierung *(zielgerichtet. sLDA)*

- **Personal- und Reisekosten** gehen mit keiner besseren/schlechteren Bewertung einher

LASSO *(Regressionsanalysen)*

- **Kosten einer Maßnahme** zeigen sich nicht als potenzieller Einflussfaktor in der Bewertung

Literatur *(Analyse)*



- **Personalressourcen** gehören zu den potenziellen Einflussfaktoren der Effizienzbewertung…

OLS *(Regressionsanalyse)*

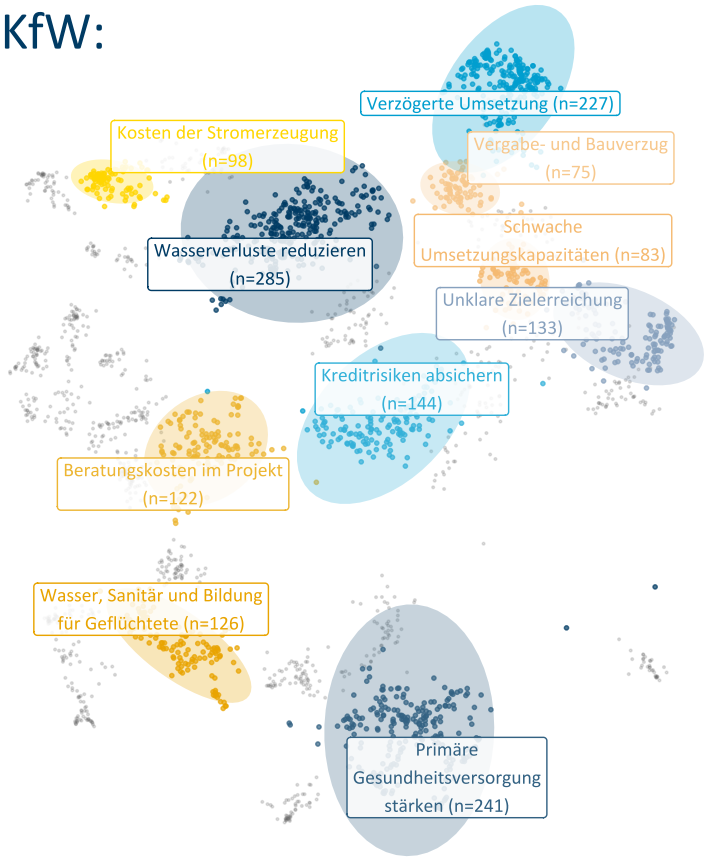
- … aber gehen **nicht eindeutig** mit einer besseren/ schlechteren Effizienzbewertung einher

Ergebnisse basieren auf KI-Themenmodellierung und multivariater Statistik

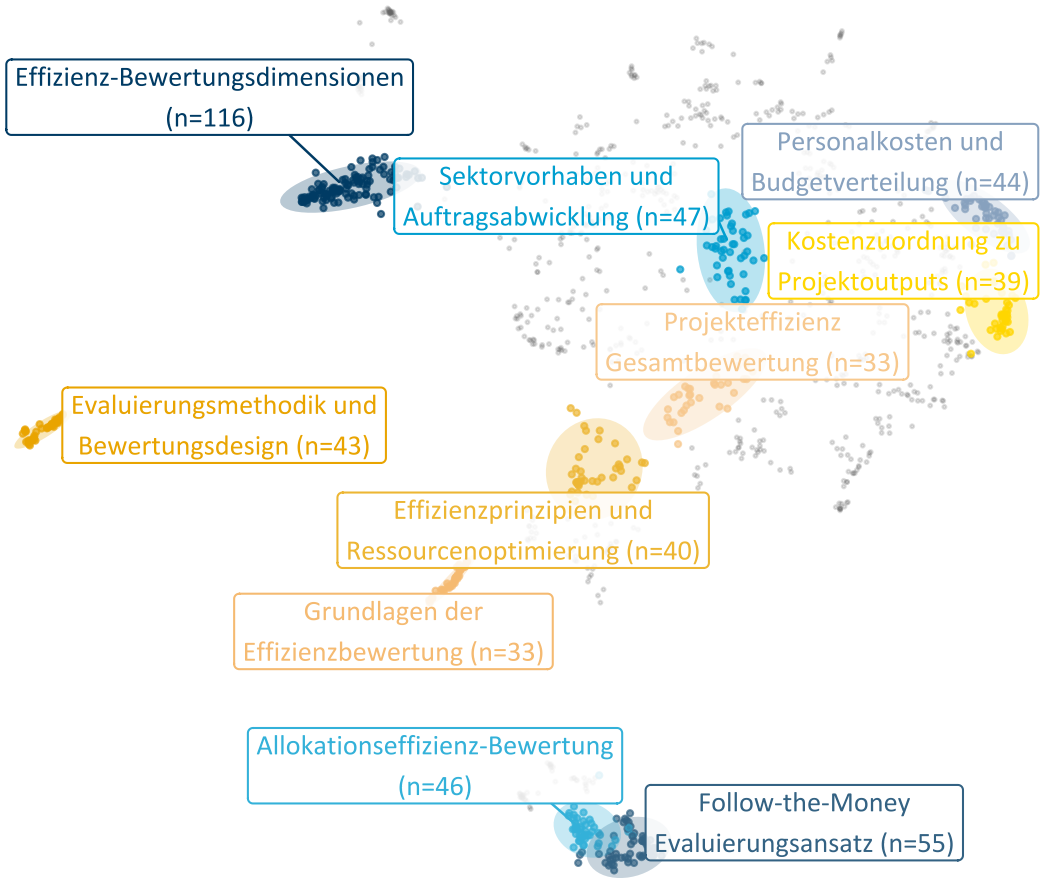
DEval

Themenvielfalt im Effizienzkapitel der Evaluierungsberichte

KfW:



GIZ:



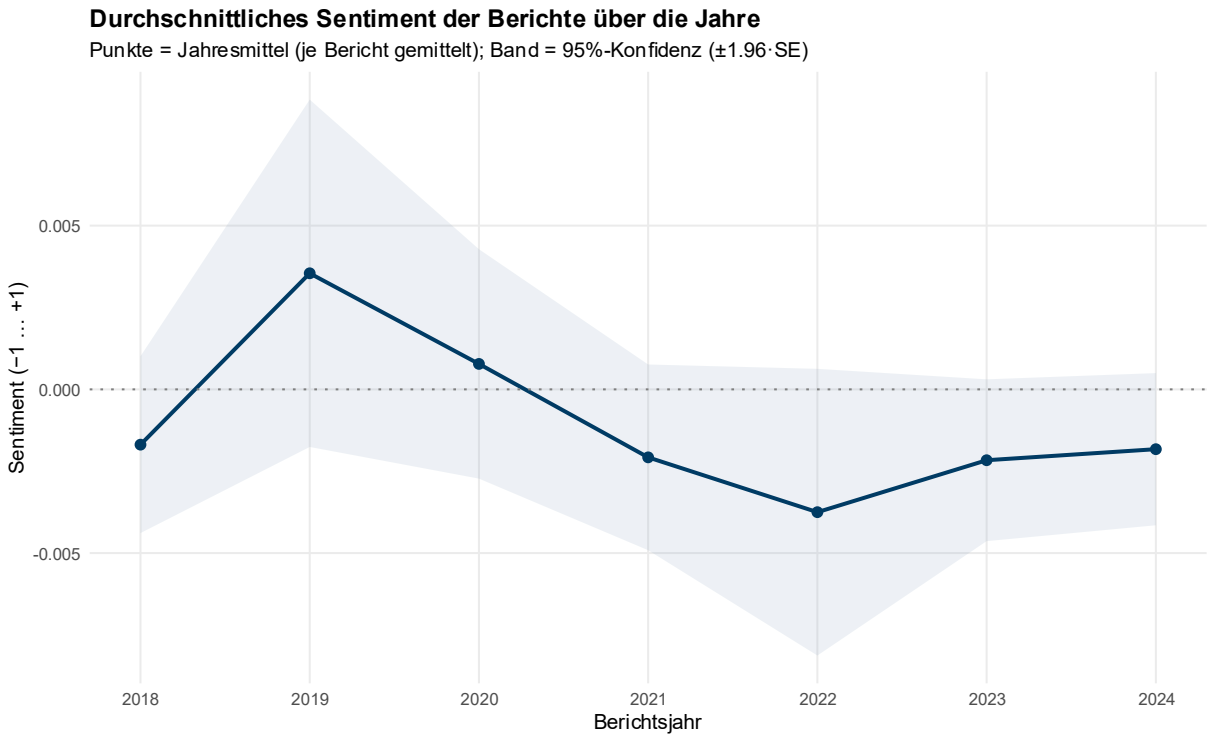
KI

Ergebnisse basieren auf offener Themenmodellierung,
Modell: BERTopic (Grootendorst, 2022)

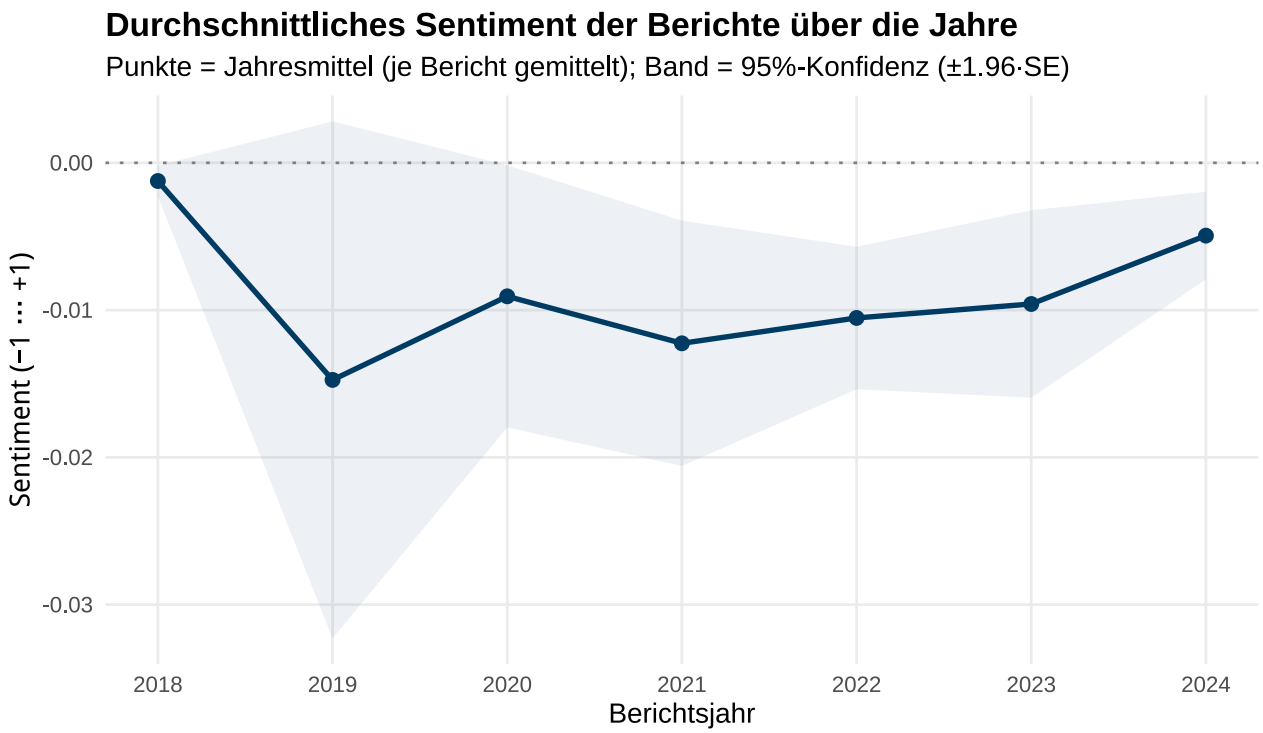
DEval

Insgesamt sehr nüchterne Berichterstattung

KfW:



GIZ:

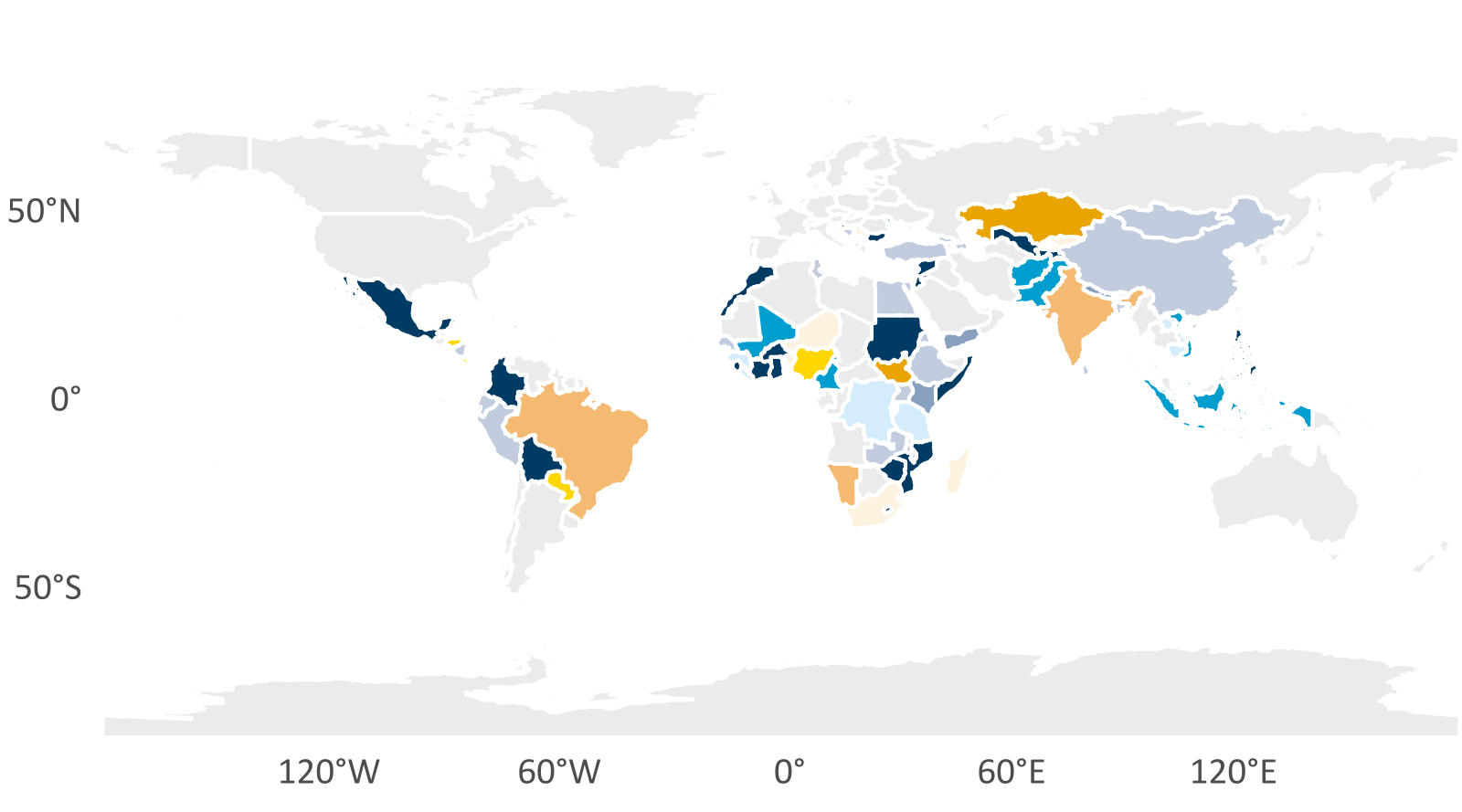


KI

Ergebnisse basieren auf Sentimentanalyse,
Modell: German Sentiment BERT (Guhr el al. 2020)

DEval

Länderspezifische Themenverteilung (am Beispiel der KfW)



Top-Thema

- Verzögerte Umsetzung
- Wasserverluste reduzieren
- Schwache Umsetzungskapazitäten
- Wirkungen kaum nachweisbar
- Kostensteigerungen im Projekt
- Sonstiges
- Beratungskosten im Projekt
- Kreditrisiken absichern
- Projektkosten versus Planung
- Vergabe- und Bauverzug

KI

Ergebnisse basieren auf offener Themenmodellierung,
Modell: BERTopic (Grootendorst, 2022)

DEval

Hintergrund: Qualitätsindex
18 Qualitätskriterien

Nr.	Kriterium	Nr.	Kriterium
Q-1.1	Kostenstruktur	Q-2.3	Umsetzung (Produktionseffizienz)
Q-1.2	Sparsamkeit (Produktionseffizienz)	Q-2.4	Methodik (Allokationseffizienz)
Q-1.3	Ergiebigkeit (Produktionseffizienz)	Q-2.5	Konsistenz Methodik (Allokationseffizienz)
Q-1.4	Rechtzeitigkeit	Q-2.6	Umsetzung (Allokationseffizienz)
Q-1.5	Alternativen	Q-3.1	Datenqualität (Produktionseffizienz)
Q-1.6	Sparsamkeit (Allokationseffizienz)	Q-3.2	Datenqualität (Allokationseffizienz)
Q-1.7	Ergiebigkeit (Allokationseffizienz)	Q-4.1	Anschlussfähigkeit Wirtschaftlichkeitsunters.
Q-2.1	Wirkungslogik	Q-4.2	Konsistenz Ergebnisse
Q-2.2	Methodik (Produktionseffizienz)	Q-4.3	Lernen

Stichprobe (für die manuelle Klassifizierung der Evaluierungsqualität)

Manuelle Inhaltsanalyse anhand aussagekräftiger Stichprobe

max. 10% Abweichung (95%-Signifikanzniveau) vom wahren Mittelwert Grundgesamtheit

Automatisierte Sprachverarbeitung der Grundgesamtheit

$$n \geq \frac{S^2}{\frac{e^2}{z^2_{1-\frac{\alpha}{2}}} + \frac{S^2}{N}}$$

S^2 bildet die erwartete Varianz des Outcomes in der Grundgesamtheit N. $e = 0.1$ gibt den maximalen Fehler in der Stichprobe an und z den kritischen Wert für das 95%-Signifikanzniveau

Durchschnittliche Varianz:

$S^2_{KfW} = 0{,}140$

$S^2_{GIZ} = 0{,}154$

Minimale Stichprobengröße:

KfW 46 Evaluierungen (davon 38 erfolgreich)

GIZ 51 Evaluierungen (davon 45 erfolgreich)

Annäherung über durchschnittliche Varianz der Qualitätsbewertung nach Noltze et al. (2018, „Meta_Nachhaltigkeit“)

zurück zum Hauptteil

DEval

Auswahl potenzieller Einflussfaktoren für die Regressionsanalysen

$$N = \beta_0 + \sum \beta M + \sum \beta K + \sum \beta E + \beta \overline{\mathrm{DAC}} + \epsilon$$

Merkmale der Maßnahme ($\sum \beta M$)	Merkmale Kontext ($\sum \beta$K)	Merkmale Evaluierung ($\sum \beta$E)
• Dauer • Finanzierung • Komplexität • Sektor • Region	• Wirtschaftliche Entwicklung • Politisches System	• Zeitpunkt • Komplexität • Kosten • Qualitätsindex • Erfolgsbewertung ohne Effizienz

Herangehensweisen

Deduktiv: Evaluierungen und Studien zu den Einflussfaktoren auf die Erfolgsbewertung internationaler Entwicklungsmaßnahmen (Assefa et al., 2014; Bulman et al., 2015; Denizer et al., 2013; Dollar und Levin, 2005; Isham et al., 1995; Kilby, 2013; Noltze et al., 2018), insbesondere zur deutschen finanziellen (Eilers et al., 2025) und technischen Zusammenarbeit (Syspons, 2023)

Induktiv: Ergebnisse der Themenmodellierung mit Zielvariable (sLDA) und Lasso-Regressionen

unter Berücksichtigung von Korrelationen zwischen potenziellen Einflussfaktoren.

zurück zum Hauptteil

DEval