

# zeppelin universität

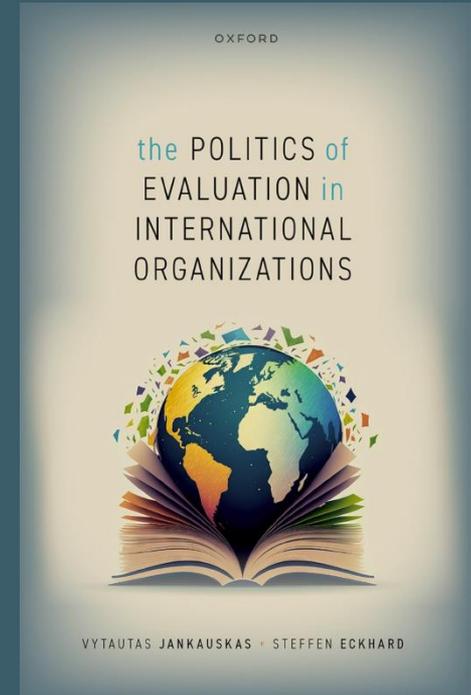
zwischen

Wirtschaft Kultur Politik

# Die Rolle von Evaluation bei der evidenzbasierten Politikgestaltung: Das Beispiel internationaler Organisationen

Prof. Dr. Steffen Eckhard

02.10.2024



AUSGANGSPUNKT

# Kein systematisches Evaluationssystem in Deutschland

## Ex ante

- I Gesetzesfolgenabschätzungen seit 2000
- I EU: Integrated Impact Assessment seit 2002
- I Normenkontrollrat
- I Programm “Bürokratieabbau und bessere Rechtsetzung” der B’Reg (seit 2012)

## Ex post

- I Gutachten/ Gegengutachten
- I Auftragsforschung
- I Wissenschaftliche Beiräte
- I Vereinzelte Evaluationen

## AUSGANGSPUNKT

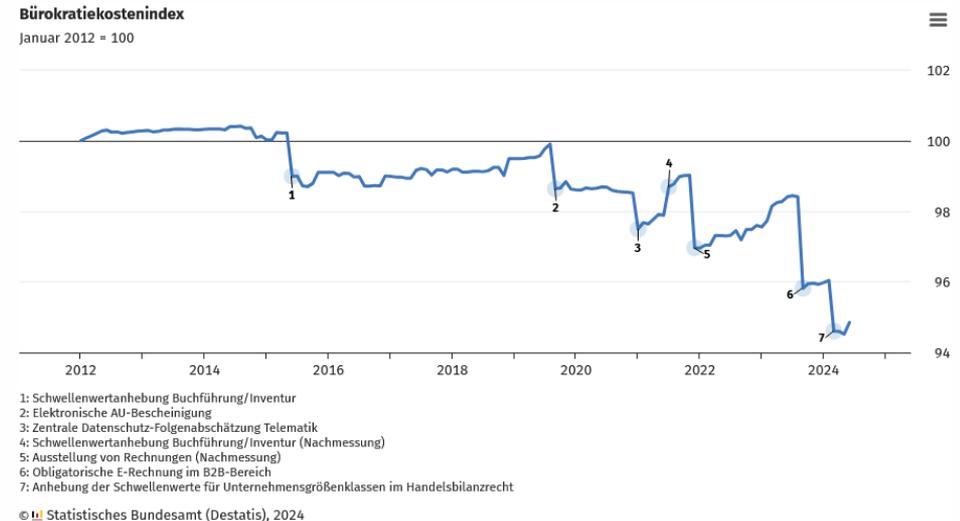
# Kein systematisches Evaluationssystem in Deutschland

## Ex ante

- | Gesetzesfolgenabschätzungen seit 2000
- | EU: Integrated Impact Assessment seit 2002
- | Normenkontrollrat
- | Programm “Bürokratieabbau und bessere Rechtsetzung” der B'Reg (seit 2012)

## Ex post

- | Gutachten/ Gegengutachten
- | Auftragsforschung
- | Wissenschaftliche Beiräte
- | Einzelne Evaluationen



## AUSGANGSPUNKT

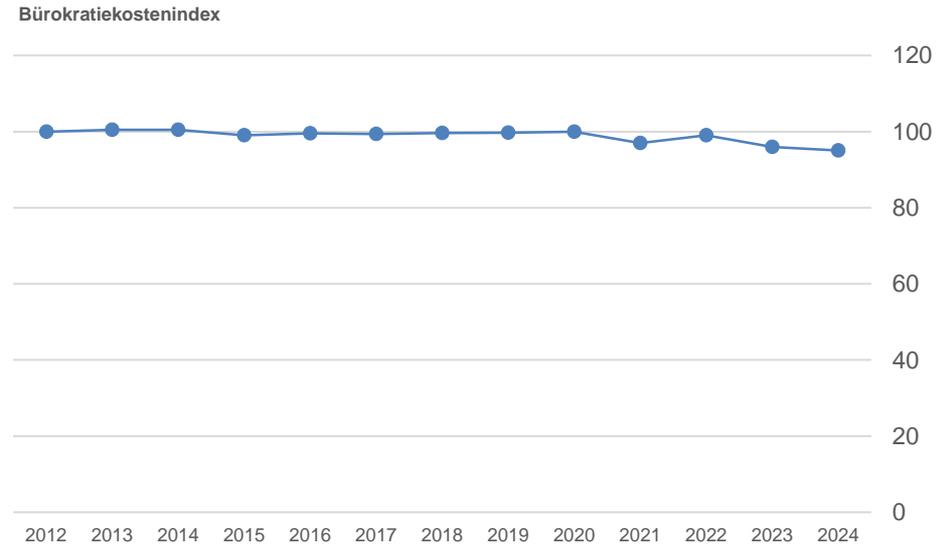
# Kein systematisches Evaluationssystem in Deutschland

## Ex ante

- I Gesetzesfolgenabschätzungen seit 2000
- I EU: Integrated Impact Assessment seit 2002
- I Normenkontrollrat
- I Programm “Bürokratieabbau und bessere Rechtsetzung” der B’Reg (seit 2012)

## Ex post

- I Gutachten/ Gegengutachten
- I Auftragsforschung
- I Wissenschaftliche Beiräte
- I Vereinzelt Evaluationen



<https://www.destatis.de/DE/Themen/Staat/Buerokratiekosten/Erfuellungsaufwand/buerokratiekostenindex.html>

AUSGANGSPUNKT

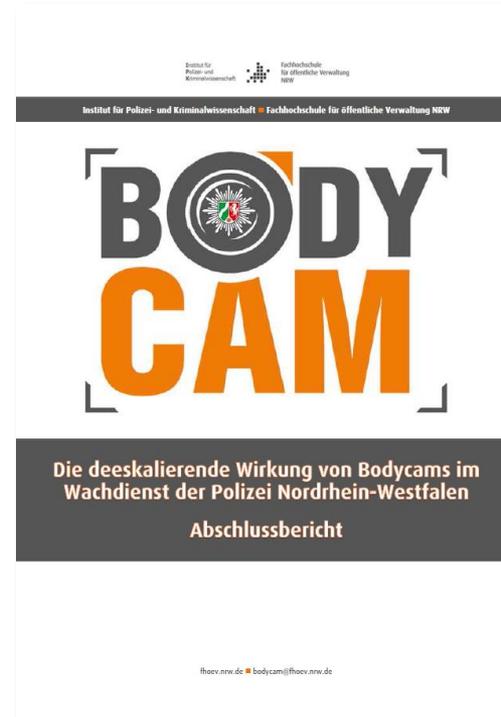
# Kein systematisches Evaluationssystem in Deutschland

## Ex ante

- | Gesetzesfolgenabschätzungen seit 2000
- | EU: Integrated Impact Assessment seit 2002
- | Normenkontrollrat
- | Programm “Bürokratieabbau und bessere Rechtsetzung” der B’Reg (seit 2012)

## Ex post

- | Gutachten/ Gegengutachten
- | Auftragsforschung
- | Wissenschaftliche Beiräte
- | Vereinzelte Evaluationen



## AUSGANGSPUNKT

# Kein systematisches Evaluationssystem in Deutschland

## Ex ante

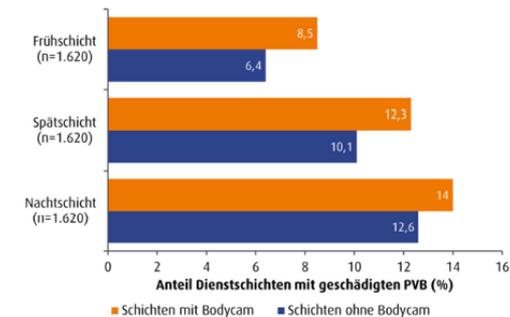
- | Gesetzesfolgenabschätzungen seit 2000
- | EU: Integrated Impact Assessment seit 2002
- | Normenkontrollrat
- | Programm “Bürokratieabbau und bessere Rechtsetzung” der B’Reg (seit 2012)

## Ex post

- | Gutachten/ Gegengutachten
- | Auftragsforschung
- | Wissenschaftliche Beiräte
- | Vereinzelte Evaluationen

Abbildung 4.1

Anteil geschädigter PVB in Schichten mit und ohne Bodycam und nach Schichtzeiten



Anmerkung: Datenbasis ist die Schichtebene

Entgegen der Erwartung liegt der Anteil geschädigter Polizeibeamt:innen in Schichten mit Bodycam über dem Anteil in Schichten ohne Bodycam (57)

AUSGANGSPUNKT

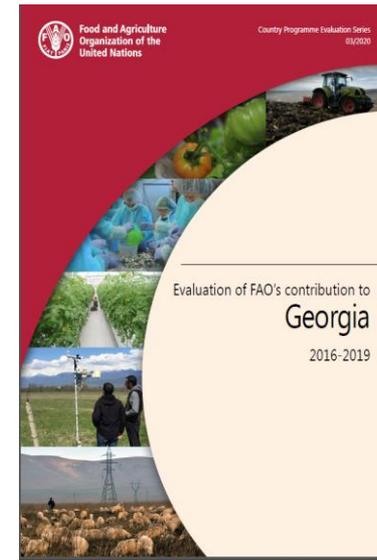
# These

**Ein unabhängiges und systematisches Evaluationssystem könnte in Deutschland einen wichtigen Beitrag zum Bürokratieabbau leisten**

## HINTERGRUND

# Evaluationen in internationalen Organisationen

- I “eine möglichst *systematische* und *unparteiische* *Bewertung* einer Tätigkeit, eines Projekts, eines Programms, einer Strategie, einer Politik, eines Themas, eines Sektors, eines operativen Bereichs oder einer institutionellen Leistung” (UNEG 2016).



Beispiel eines  
Evaluierungsberichtes

## HINTERGRUND

# Evaluationen in internationalen Organisationen

- I “eine möglichst *systematische* und *unparteiische* *Bewertung* einer Tätigkeit, eines Projekts, eines Programms, einer Strategie, einer Politik, eines Themas, eines Sektors, eines operativen Bereichs oder einer institutionellen Leistung” (UNEG 2016).
- I Standardisiert; OECD DAC Kriterien



**UNEG Code of Conduct for  
Evaluation in the UN System**

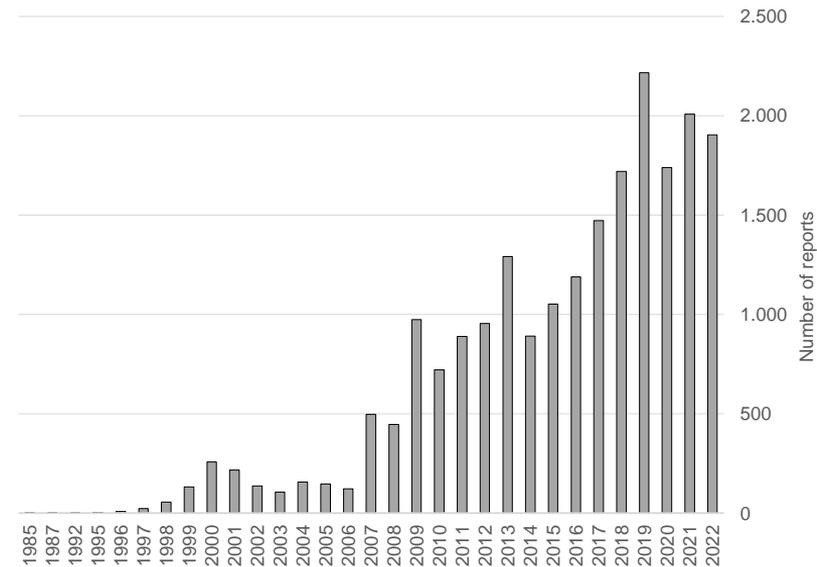
UNEG, March 2008



## HINTERGRUND

# Evaluationen in internationalen Organisationen

- I “eine möglichst *systematische* und *unparteiische Bewertung* einer Tätigkeit, eines Projekts, eines Programms, einer Strategie, einer Politik, eines Themas, eines Sektors, eines operativen Bereichs oder einer institutionellen Leistung” (UNEG 2016).
- I Standardisiert; OECD DAC Kriterien
- I Jährlich werden mehr als 1.500 Evaluationsberichte



Source: [UN Evaluation Group Database, 2023](#)

# Ein Goldstandard für institutionelle Bewertung und Lernen?

## Funktionale Perspektive:

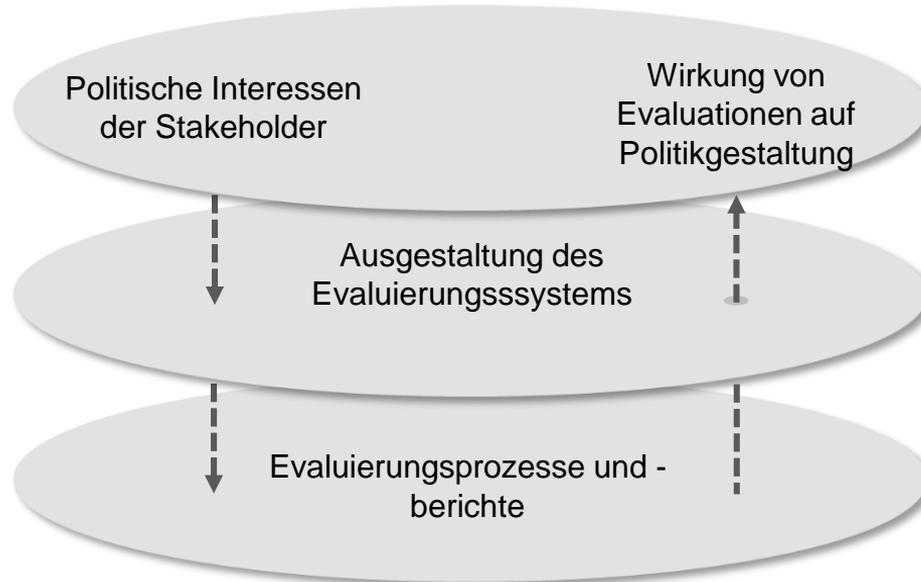
- I Evaluierung dient der Transparenz, der Rechenschaftspflicht und dem Lernen; ermöglicht eine evidenzbasierte Politikgestaltung.
- I Evaluation als „der ‚Goldstandard‘ der institutionellen Bewertung auf internationaler Ebene“ (Lall, 2017:245; Honig, 2019).

## Politische Perspektive:

- I „Evaluation und Organisation sind ... widersprüchliche Begriffe“ (Wildavsky 1972).
- I Evaluatoren werden ständig unter Druck gesetzt, ihre Ergebnisse falsch darzustellen, indem sie sie „positiver oder weniger negativ aussehen lassen“ (Morris und Clark, 2012: 57; Azzam 2010).
- I anekdotische Belege für IOs (Raimondo 2018; Eckhard & Jankauskas 2019, 2020)

THEORETISCHES ARGUMENT

# Stakeholder-Interessen und institutionelle Gestaltung

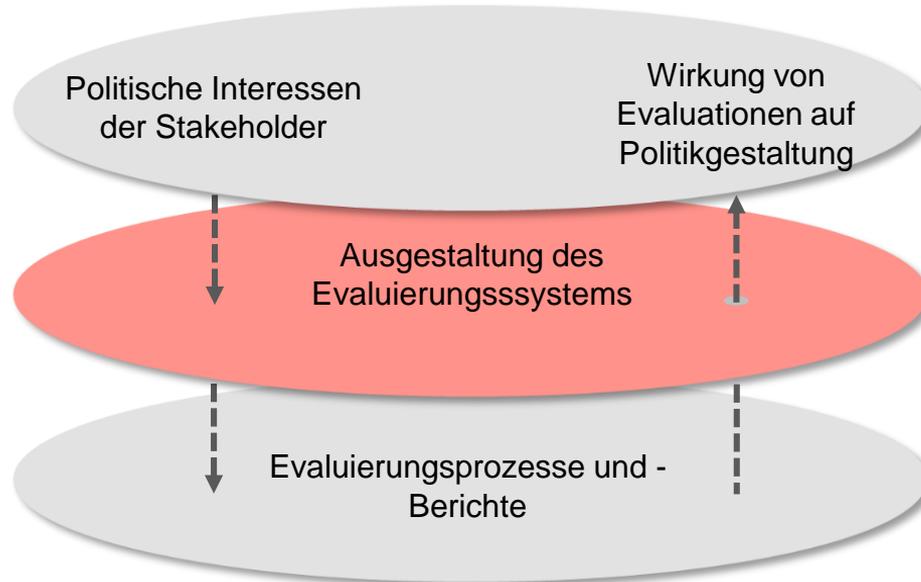


## Forschungs-Design

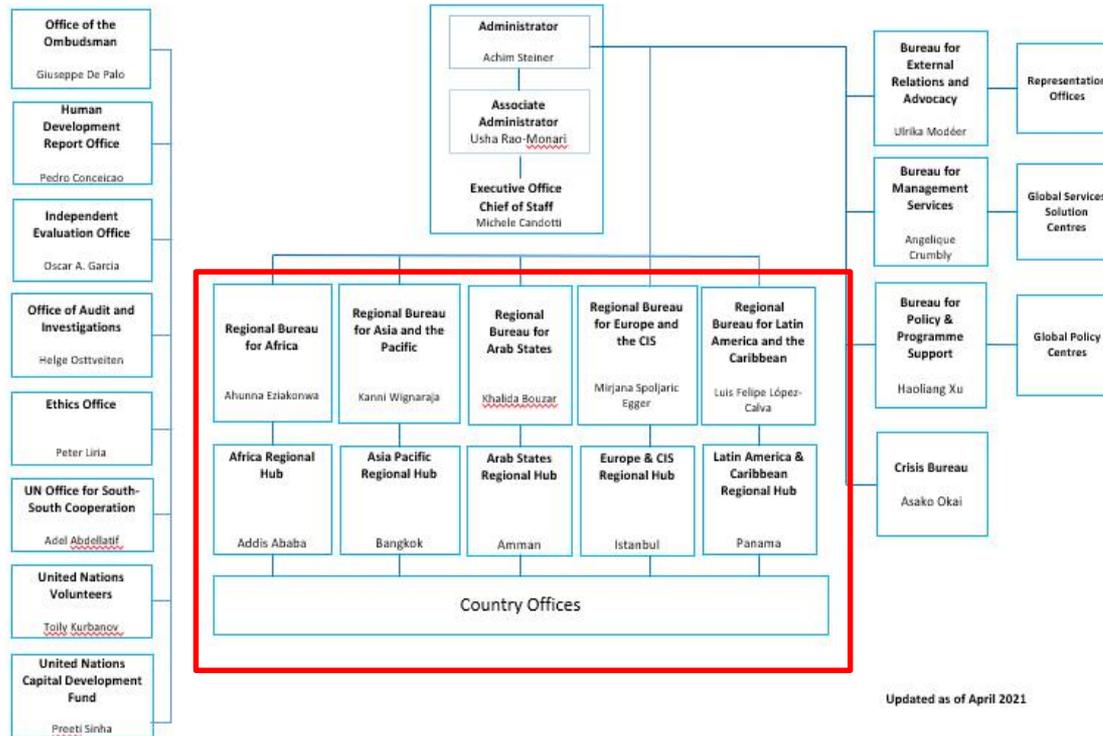
- I empirischer Fokus auf Evaluierungseinheiten von IOs des UN-Systems (N=21)
- I 120 halbstrukturierte Experteninterviews mit Vertretern von Mitgliedsstaaten, IO-Beamten und Mitarbeitern von Evaluierungseinheiten
- I qualitative Analyse von Evaluierungsrichtlinien
- I qualitative und quantitative Analyse von Evaluierungsberichten (N=1028)

BEFUNDE

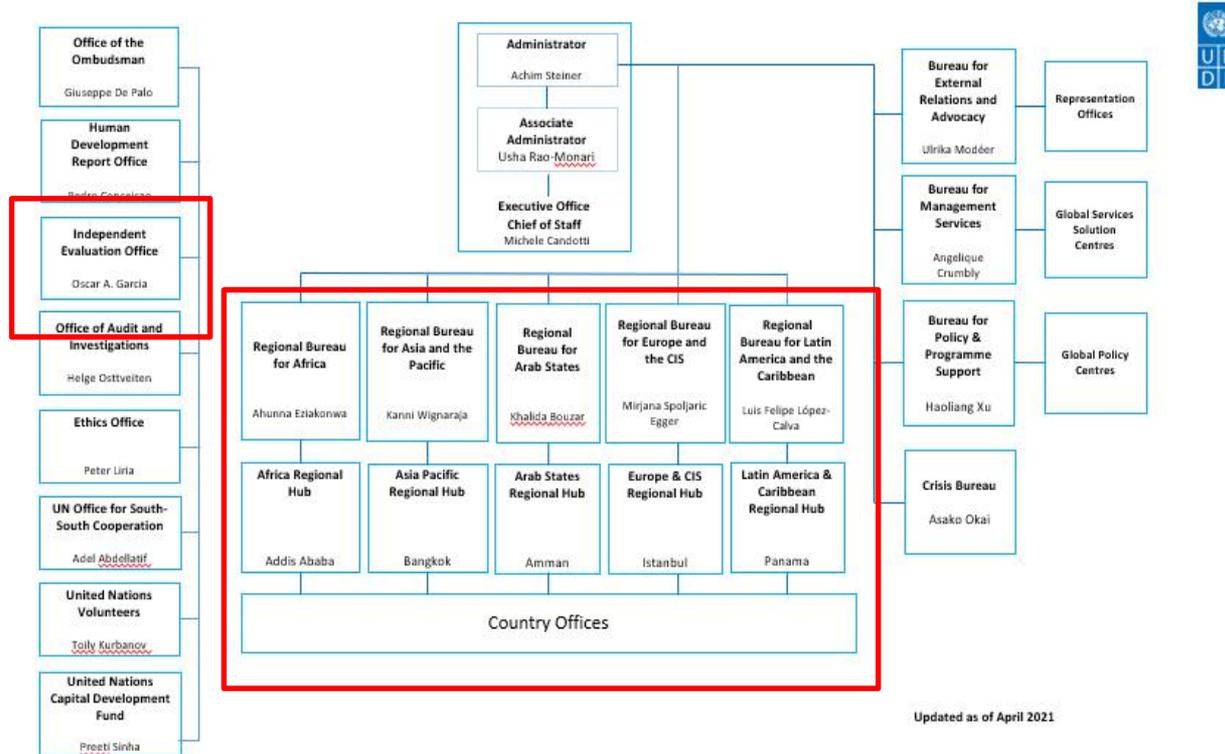
# Ausgestaltung Evaluierungssystem



# Ausgestaltung Evaluierungssystem



# Ausgestaltung Evaluierungssystem



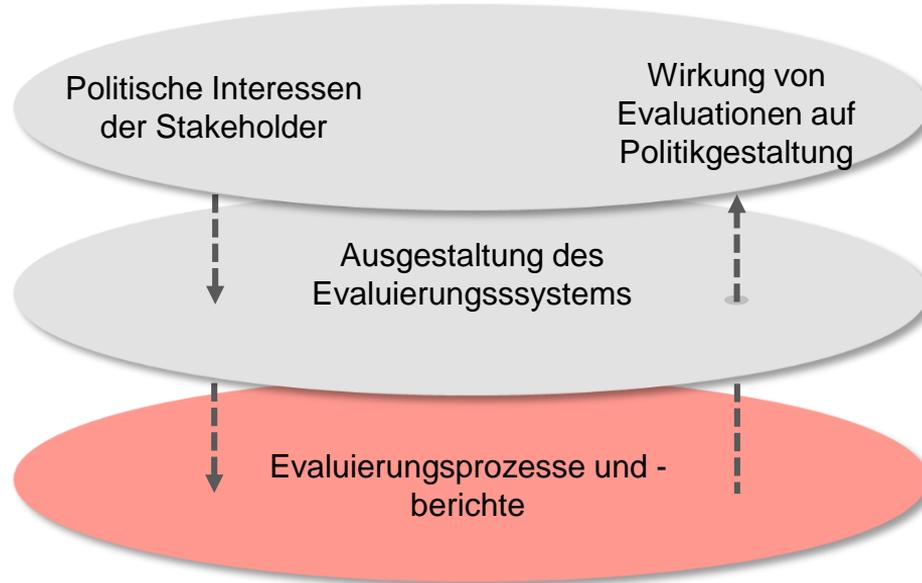
BEFUNDE

# Ausgestaltung Evaluierungssystem



## BEFUNDE

# Einfluss auf Evaluierungsberichte



## Hypothese

- I *Evaluierungsberichte, die von **dezentralen operativen Einheiten** in Auftrag gegeben wurden, enthalten systematisch **mehr positive Bewertungen** als solche, die von zentralen Evaluierungseinheiten in Auftrag gegeben wurden.*

## EXKURS

# Extrahieren von Leistungsbewertungen aus Evaluierungsberichten

I Evaluation erfolgt in Berichtsform

I Bewertungen auf **Satzebene**:

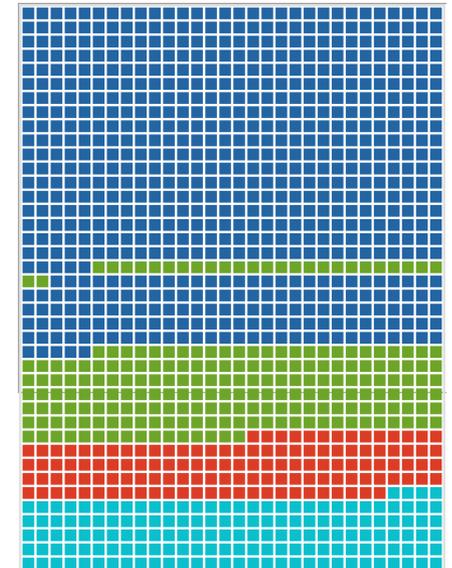
**Positive Bewertung:** *“Das Projekt erwies sich als erfolgreich bei der Bildung von Partnerschaften und dem Aufbau von Synergien mit anderen Organisationen ...” (ILO 2018)*

**Negative Bewertung:** *“Die FAO hat ihr anerkanntes Wissen über den Agrarsektor nicht systematisch genutzt, um strategische Partnerschaften mit wichtigen Akteuren aufzubauen ...” (FAO, 2019)*

**Deskriptiver Text:** *“Das Hauptziel der Bewertung ist die Beurteilung der Leistung des Weltwasserbewertungsprogramms ...” (UNESCO, 2015)*

**Empfehlungen:** *“Bis Ende 2019 sollte das UNHCR die Mehrjahres-/Multipartner-Pilotprojekte (MYMP) mit Lebensunterhaltsplänen ausweiten” (UNHCR, 2018)*

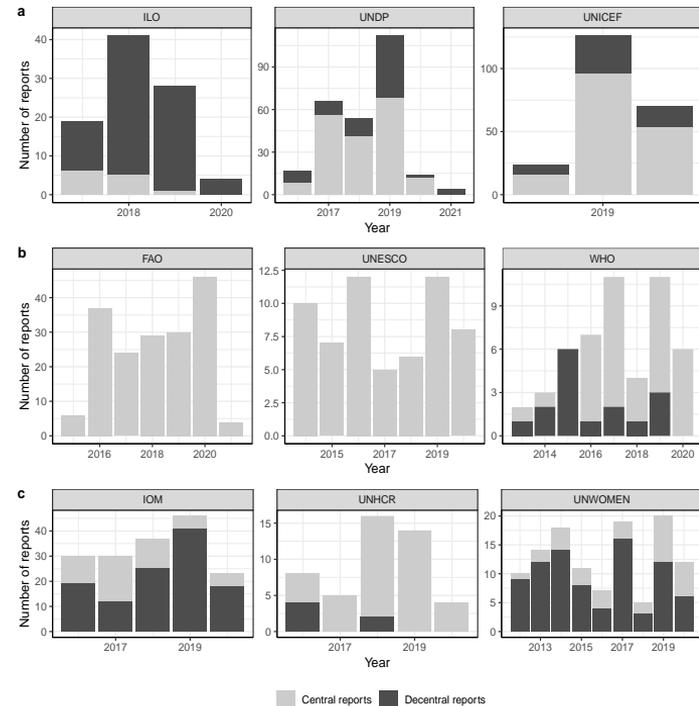
I Berechnung des relativen Anteils der **positiven** gegenüber den **negativen** Sätzen



## EXKURS

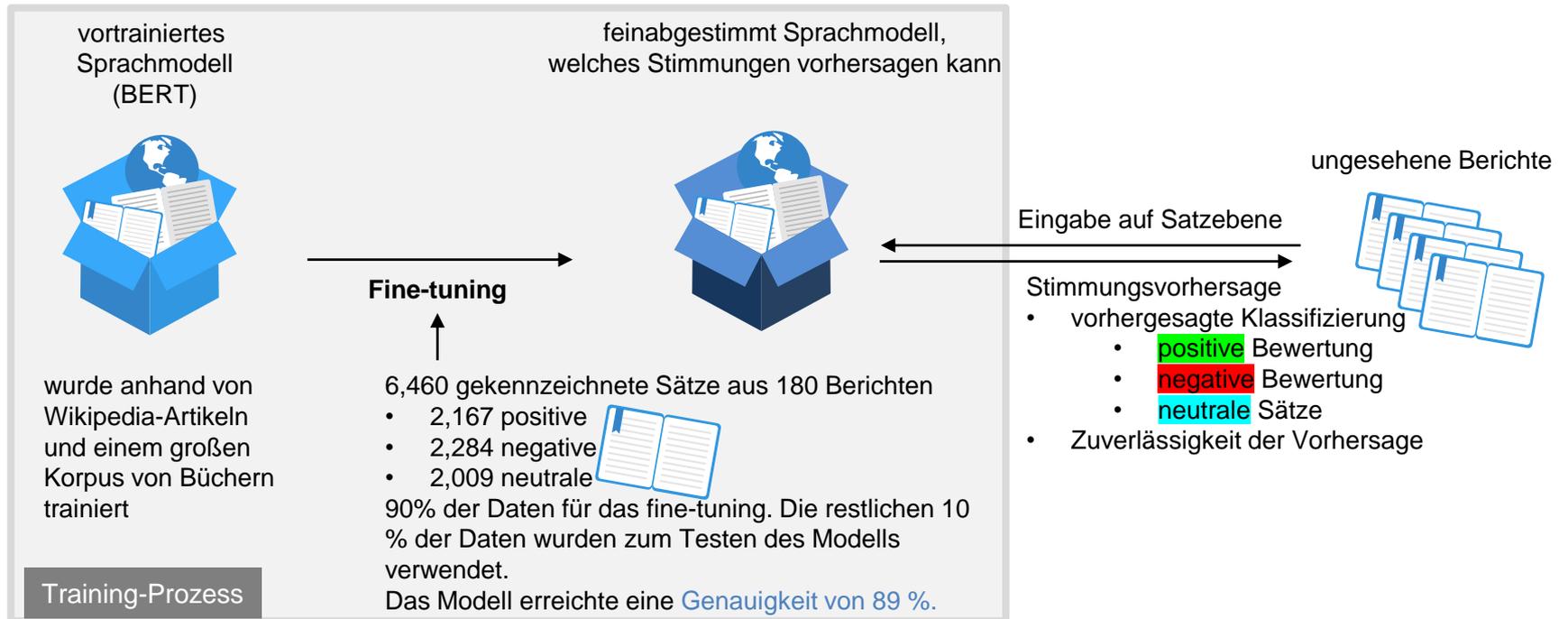
# Extrahieren von Leistungsbewertungen aus Evaluierungsberichten

- I Auswahl von **9 IOs** mit 1,082 Evaluierungsberichten: ILO, UNDP, UNICEF, IOM, UNHCR, UN WOMEN, WHO, UNESCO, FAO
- I Variationen in Berichte, die von der **zentralen** Auswertungsstelle oder **dezentralen** operativen Einheiten in Auftrag gegeben werden
- I mehr als eine Millionen Sätze, die klassifiziert werden müssen



## EXKURS

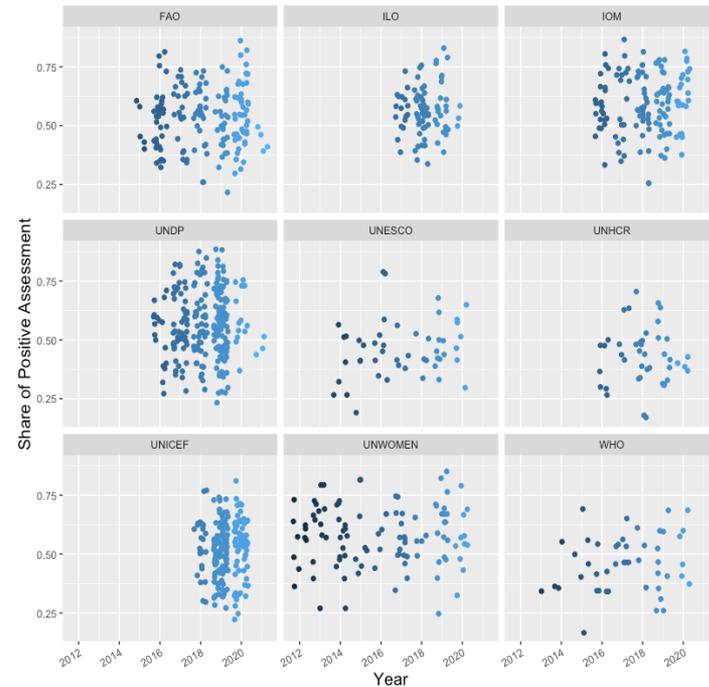
# Extrahieren von Leistungsbewertungen aus Evaluierungsberichten



## BEFUNDE

# Einfluss auf Evaluierungsberichte

I Anteil an positiven Bewertungen für 1.082 Berichte



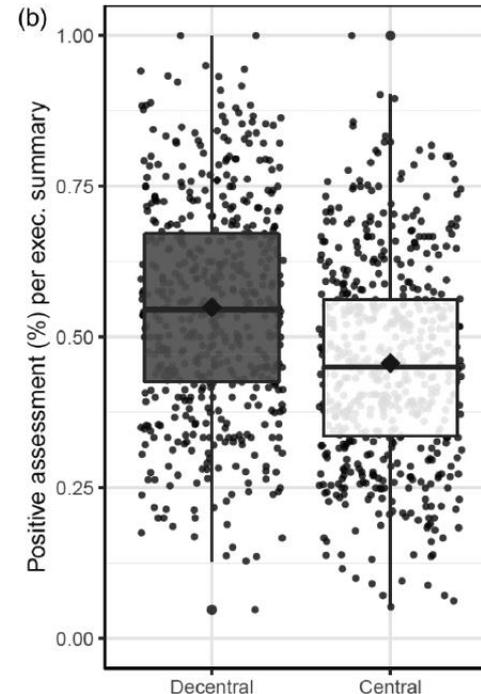
## BEFUNDE

## Einfluss auf Evaluierungsberichte

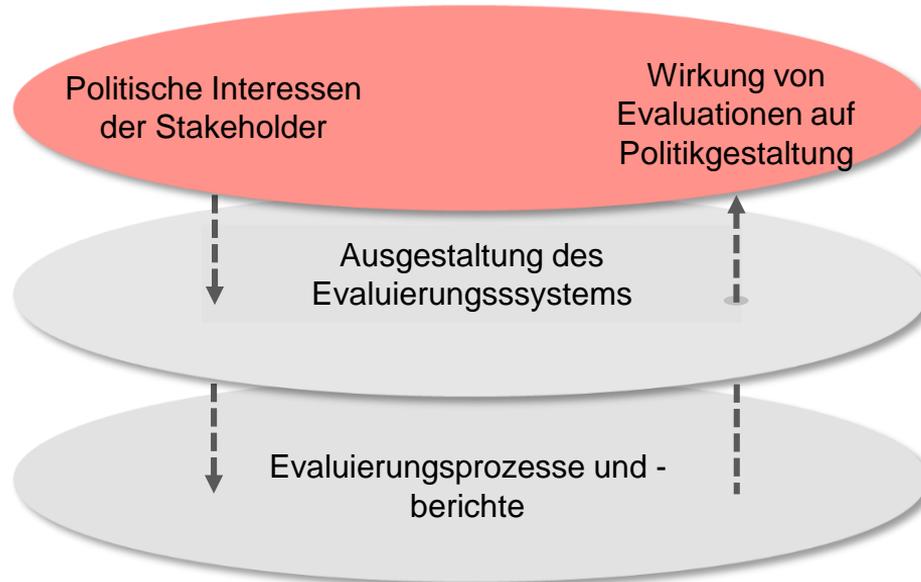
- I Anteil an positiven Bewertungen für 1.082 Berichte
- I Zentrale Berichte 49% positiv
- I Dezentrale Berichte 58% positiv

= Dezentrale Berichte haben durchschnittlich 6 positive Sätze mehr im „Executive Summary“

→ **Zentrale Evaluations-Einheiten liefern ausgewogenere Berichte**



# Politische Wahrnehmung zentraler Evaluationseinheiten



## BEFUNDE

## Politische Wahrnehmung zentraler Evaluationseinheiten

### Vorstandsmitglied einer IO:

- I *“Wir nutzen die Empfehlungen des unabhängigen Evaluierungsbüros, um zu sagen: ‘Wir haben eindeutig Anhaltspunkte dafür, dass dieses Thema von dieser Organisation umfassender oder kohärenter angegangen werden sollte’ [. . .] Es gibt also diese Art von **Kontrolle und Gegenkontrolle durch die Strukturen des Verwaltungsrats**, um zu sehen, wie sie vorankommen.”*

### Führungsebene einer IO-Verwaltung

- I *“Wenn wir über potenziellen Widerstand oder zögerliches Engagement sprechen, dann denke ich, dass der Blick auf die Fakten und die Bereitstellung von Beweisen eines der **wichtigsten Instrumente ist, um die Akteure tatsächlich zu mobilisieren.**”*

# Zusammenfassung

These: **Ein unabhängiges Evaluationssystem könnte in Deutschland einen wichtigen Beitrag zum Bürokratieabbau leisten**

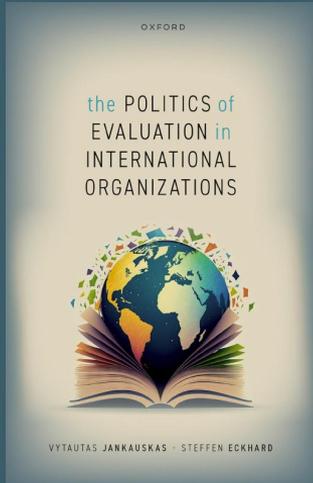
- I Evaluationen finden in einem politisierten Umfeld statt
- I Die institutionelle Ausgestaltung von Evaluationssystemen beeinflusst die Ergebnisse von Evaluationen
- I Unabhängige Evaluationseinheiten liefern ausgewogenere Evaluations Berichte
- I Chance für mehr Evidenz im politischen Prozess und Bürokratieabbau in Deutschland



Vielen Dank für die Aufmerksamkeit!

zu.de

# Vielen Dank für die Aufmerksamkeit!



SPRINGER LINK

Find a journal Publish with us Track your research Search

Home > The Review of International Organizations > Article

**The performance of international organizations: a new measure and dataset based on computational text analysis of evaluation reports**

Open Access | Published: 04 May 2023  
Volume 18, pages 193–176, (2023) | Cite this article

Download PDF You have full access to this open access article

Steffen Eckhard , Vytautas Jankauskas , Elena Leuschner , Jan Burton, Tilmann Keri & Rita Sevastianova

3583 Accesses 4 Citations 31 Altmetric Explore all metrics

**Abstract**

International organizations (IOs) of the United Nations (UN) system publish around 750 evaluation reports per year, offering insights on their performance across project, program, institutional, and thematic activities. So far, it was not feasible to extract quantitative performance measures from these text-based reports. Using deep learning, this article presents a novel text-based performance metric: We classify individual sentences as containing a negative, positive, or neutral assessment of the evaluated IO activity and then compute the share of positive sentences per report. Content validation yields that the measure adequately reflects the underlying concept of performance; convergent validation finds high correlation with human-provided performance scores by the World Bank; and construct validation shows that our measure has theoretically expected results. Based on this, we present a novel dataset with performance measures for 1,082 evaluated activities implemented by nine UN system IOs and discuss avenues for further research.

WILEY Online Library Search

PUBLIC ADMINISTRATION REVIEW

RESEARCH ARTICLE

**Institutional design and biases in evaluation reports by international organizations**

Steffen Eckhard , Vytautas Jankauskas , Elena Leuschner

First published: 14 July 2023 | <https://doi.org/10.1111/puar.13705>

SECTIONS

**Abstract**

Governments spend hundreds of millions on evaluations to assess the performance of public organizations. In this article, we scrutinize whether variation in the institutional design of evaluation systems leads to biases in evaluation findings. Biases may emerge because influence over evaluation processes could enable the bureaucracy to present its work in a more positive way. We study evaluation reports published by nine international organizations (IOs) of the United Nations system. We use deep learning to measure the share of positive assessments at the sentence level per evaluation report as a proxy for the positivity of evaluation results. Analyzing 1082 evaluation reports, we find that reports commissioned by operative units, as compared to central evaluation units, systematically contain more positive assessments. Theoretically, this link between institutional design choices and evaluation outcomes may explain why policymakers perceive similar tools for evidence-based policymaking as functional in some organizations, and politicized in others.

Kontakt:

Steffen.eckhard@zu.de

