

Wolfgang Meyer, Reinhard Stockmann

Institutionalisierung der Evaluation in den politischen Systemen Europas. Eine vergleichende Analyse

Zusammenfassung

Der Beitrag präsentiert die ersten vergleichenden Ergebnisse eines weltweiten Forschungsprojekts zur Institutionalisierung der Evaluation. Mehr als 30 Autorinnen und Autoren haben sich mit der Situation in 16 Ländern Europas beschäftigt und diese anhand eines einheitlichen Analyseleitfadens untersucht. Bei der Betrachtung der Integration von Evaluation innerhalb des politischen Systems standen drei Aspekte im Vordergrund: Die formalen Verankerungen in der Legislative (z. B. in Form von Gesetzen und Verordnungen) und in der Exekutive (z. B. der Evaluationspraxis in verschiedenen Politikfeldern, Ministerien und Behörden) sowie die Nutzung der Evaluationsergebnisse für evidence-based policies.

Es handelt sich um die erste umfassende systematische Bestandsaufnahme der Institutionalisierung von Evaluation in den politischen Systemen Europas. Herausgearbeitet werden die Gemeinsamkeiten und Unterschiede sowie die unterschiedlichen Einflussfaktoren auf die nationalen Institutionalisierungsprozesse. In die Betrachtung gehen z. B. Spezifika nationaler politischer Systeme, die Einflüsse transnationaler Organisationen (insbesondere der Europäischen Union), die sektoralen Besonderheiten und ihre Leitfunktionen sowie die Rolle der Zivilgesellschaft mit ein.

Schlagworte: Institutionalisierung der Evaluation, Politik, Europa, systematischer Vergleich

Abstract

Institutionalization of evaluation in European political systems. A comparative analysis

This paper presents the first comparative results of a global research project on institutionalization of evaluation. More than 30 authors investigated the situation in 16 European countries by using a consistent analytical framework. While observing the integration of evaluation into the political system, three key elements are in focus: the formal anchoring within legislation (e. g. in form of acts, laws and decrees) and within the executive (e. g. the practice of evaluation in various policy fields, ministries and public authorities) as well as the use of evaluation results for evidence-based policies.

This is the first systematic and comprehensive stock-taking of the institutionalization of evaluation within the European political systems. It carves out the similarities and differences (especially for the drivers) of the national institutionalization processes. Among others, the specifics of national political systems, the influence of transnational organizations (notably the European Union), the leading function of particular sectors and the role of civil societies will be mentioned.

Keywords: Institutionalization of Evaluation, Politics, Europe, systematic comparison

1 Einführung

Dieser Beitrag präsentiert zentrale Befunde zur Institutionalisierung der Evaluation in Europa, welche im Mai 2020 umfassend in einer Buchpublikation vorgestellt werden (Stockmann, Meyer & Taube, 2020a). Buch und Beitrag sind Teil eines globalen Forschungsprojektes Evaluation-GLOBE des Centrums für Evaluation (CEval) an der Universität des Saarlandes mit dem Ziel, den ersten weltweiten systematischen Überblick zu diesem Thema in einer vierbändigen Buchserie zu veröffentlichen. Der zweite Band über Nord- und Südamerika befindet sich ebenfalls bereits in der Endredaktion und soll Anfang 2021 erscheinen (Stockmann & Meyer, 2021). Der vorliegende Artikel zur Lage in Europa konzentriert sich ausschließlich auf die Darstellung der vergleichenden Ergebnisse zum politischen System und verzichtet zugunsten von ersten Erklärungsversuchen auf die im Buch zusätzlich dargestellte Implementierung in der Zivilgesellschaft sowie im System der Professionen. Er stellt damit zugleich den ersten und vorsichtigen Schritt von der reinen Deskription zu einer weitergehenden Analyse dar.

Das Gesamtprojekt ist sowohl hinsichtlich seiner ambitionierten geographischen Ausrichtung als auch hinsichtlich seines Publikationsumfangs bisher einmalig. Am ehesten vergleichbar ist es mit dem 2002 erschienenen „International Atlas of Evaluation“ (Furubo, Rist & Sandahl, 2002), der ebenfalls auf Basis eines einheitlichen Analysekonzepts und mit Hilfe von Expertenratings eine Beurteilung des Stands der Evaluation in einer Vielzahl von Ländern – vorrangig in Nordamerika und Europa – vorgenommen hat. Das GLOBE-Projekt orientiert sich jedoch nur bedingt an diesem Vorbild und bietet keine vergleichbaren, aktualisierten Befunde (siehe hierzu die Studie von Jacob, Speer & Furubo, 2015). Auf die theoretischen wie methodischen Grundlagen des GLOBEs insbesondere auch in Abgrenzung zum *International Atlas* und einigen weiteren Projekten zur Darstellung der Institutionalisierung von Evaluation wird im zweiten Abschnitt dieses Beitrags näher eingegangen.

Die Fokussierung auf die Institutionalisierung im politischen System begründet sich nicht nur durch das Konzept dieses Schwerpunkthefts und die inhaltliche Ausrichtung der Zeitschrift, sondern auch durch dessen besondere Rolle für die Herausbildung, Implementierung und Nutzung von Evaluation in Europa. Während sich in der Privatwirtschaft andere, in ihrer Zielrichtung und Wirkungsweise durchaus vergleichbare Konzepte herausgebildet haben (Stockmann, 2008), ist die Evaluation sehr eindeutig als „Kind der Politik“ zu verorten und kommt primär bei politischen Programmen und Projekten zur Anwendung (hierzu z. B. Pattyn, Voorst, Mastenbrock & Dunlop, 2018; Vedung, 2017). Hieraus ergeben sich einige Spezifika hinsichtlich der Institutionalisierung, die im Rahmen der Vorstellung der Ergebnisse im dritten Abschnitt näher herausgearbeitet werden.

Besonders bemerkenswert ist allerdings die Tatsache, dass weder die unterschiedlichen Charakteristika der politischen Systeme noch die historische Entwicklung die abweichenden Muster der Institutionalisierungsprozesse in den verschiedenen Ländern erklären können. Einige der zentralen Einflussfaktoren und deren Gewichtung werden in der vergleichenden Analyse im vierten Abschnitt vorgestellt.

Schließlich bleibt zum Schluss noch ein Ausblick auf die potentielle weitere Entwicklung der Evaluation in Europa, die in den letzten Jahrzehnten angesichts der exponentiellen Verbreitung in höchst unterschiedlichen Politikfeldern zweifellos als eine

beeindruckende Erfolgsgeschichte zu beschreiben ist, nun aber an einen kritischen Punkt kommt und eventuell zum Opfer des eigenen Erfolgs werden könnte. Der fünfte und letzte Abschnitt widmet sich diesen kritischen Aspekten und den notwendigen nächsten Schritten, die für eine weiterführende Verankerung von Evaluation in den politischen Systemen Europas von entscheidender Bedeutung sein werden.

2 Der GLOBE Europe – Konzept und Leitfragen

Ziel des GLOBE-Projektes ist ein systematischer, weltweit ländervergleichender Überblick zum Stand der Institutionalisierung von Evaluation. Die Entscheidung für die Nationalstaaten als Einheit des Vergleichs begründet sich durch ihre zentrale Rolle für die gesellschaftliche Steuerung mittels des nationalen politischen Systems (Hay, Lister & March, 2006). Dem entsprechend sind Gesetze oder andere Formen rechtlicher Grundlagen zur Evaluation Aufgabe des Nationalstaats, selbst wenn es ergänzende internationale, regionale oder sektorale Regelungen oder Steuerungsmodelle gibt (die dann allerdings durch nationale Regulationen abgesichert sein müssen, siehe hierzu die Beispiele in Ansell & Torfing, 2016). Auch hinsichtlich der Nutzung von Evaluationen kommt dem Nationalstaat eine Führungsrolle zu, da viele öffentliche Aufgaben von nationalen Ministerien und Behörden wahrgenommen werden. Hier unterscheiden sich die Staaten bezüglich der Verteilung von Zuständigkeiten und der Organisation dieser öffentlichen Aufgabenerfüllung erheblich voneinander – und dies gilt im besonderen Maße für Europa mit seinen historisch gewachsenen Staatsformen (Hroch, 2005) und z. T. grundsätzlich voneinander abweichenden politischen Systemen (Ismayr, 2009; 2010).

In den Europaband wurden insgesamt 16 Länderkapitel und ein Beitrag über die Europäische Union (EU) aufgenommen (*Abbildung 1*). Da in der Buchpublikation aus Platzgründen nicht alle europäischen Länder zu berücksichtigen waren, musste eine Auswahl vorgenommen werden. Im ersten Schritt sind vor allem kleinere Länder (z. B. Andorra, Luxemburg, Liechtenstein, Malta und San Marino) aufgrund ihrer geringen Größe und der Wahrscheinlichkeit, hier keinen eigenständigen nationalen Evaluationsmarkt vorzufinden, ausgeschlossen worden. Im Falle der baltischen Staaten wurde versucht, einen vergleichenden Beitrag zu initiieren, was aber an den großen Unterschieden hinsichtlich der Institutionalisierung der Evaluation zwischen den drei Ländern scheiterte. Mit Lettland wurde deshalb das Land mit der fortgeschrittensten Evaluationskultur für das Buch ausgewählt. Die Balkanregion sowie Belarus und die Ukraine wurden nach Vorgesprächen mit einigen nationalen Experten aufgrund des geringen Entwicklungsstands der Evaluation nicht aufgenommen. In vier für die Entwicklung der Evaluation in Europa wichtigen Ländern – Norwegen, Österreich, Schweden und Ungarn – konnten keine Autorinnen und Autoren gefunden werden, die in dem vorgegebenen Zeitraum einen Beitrag entsprechend der Vorgaben ausarbeiten konnten. Die vorliegenden Befunde für Europa sind somit nicht vollständig, umfassen aber die bedeutendsten Länder. Von den knapp 600 Mio. Bewohnern¹ Europas (Russland und die Türkei wurden hier nicht mitgezählt, die beiden Staaten werden nach Rücksprache in den später erscheinenden Australasia-Band aufgenommen) leben ca. 80% in den 16 erfassten Nationen.

Abbildung 1: Ausgewählte Länder Europas



Quelle: Eigene Darstellung.

Inhaltlich ist das „casing“ nicht so einfach, wie es auf den ersten Blick erscheint. Die Entscheidung für Nationalstaaten als Untersuchungseinheiten impliziert z. B., dass die sektorübergreifenden Gemeinsamkeiten innerhalb eines Staates größer sind als die länderübergreifende Homogenität innerhalb eines Handlungsfelds. Im Falle der Evaluation lässt sich diese Annahme aber durchaus bezweifeln (siehe z. B. die Ergebnisse der Studie zu den großen Unterschieden zwischen öffentlichem und privatem Sektor in Kanada von Cousins, Goh, Elliott, Aubrey & Gilbert, 2014). Als Folge wird der Steuerungseinfluss des Nationalstaats auf die Institutionalisierung von Evaluation innerhalb eines Landes tendenziell über-, die Bedeutung sektorspezifischer internationaler Organisationen, multilateraler Vereinbarungen und transnationaler Vernetzung dagegen eher unterschätzt.

Generell stellt die Fallauswahl in der international vergleichenden Forschung immer eine besondere Hürde dar, die auf verschiedene Weise genommen werden kann und jeweils mit unterschiedlichen Schwierigkeiten und Einschränkungen verbunden ist (Goerres, Siewert & Wagemann, 2019, pp. 83-93). Durch die hier getroffenen, eher pragmatisch orientierten Auswahlentscheidungen ergibt sich ein Selektionsbias in Richtung

Überbewertung der Institutionalisierung von Evaluation in Europa (insbesondere durch den Ausschluss von Ländern, in denen keine oder nur eine sehr rudimentäre Institutionalisierung vermutet wurden) und trotz dem bevölkerungsbezogen hohen Abdeckungsgrad kann weder statistisch noch inhaltlich von Repräsentativität gesprochen werden.

Das Ziel der Studie ist allerdings nicht die Repräsentation Europas, sondern die deskriptive Darstellung der zentralen Unterschiede zwischen Nationalstaaten innerhalb Europas. Es geht darum herauszufinden, ob und inwieweit die Institutionalisierung der Evaluation gleichmäßig verläuft und welche unterschiedlichen Entwicklungspfade festzustellen sind. Die Beschreibung historischer Abweichungen und nicht der Test elaborierter Theorien der Institutionalisierung stehen (als Überblick Peters, 2019) im Vordergrund. Gleichwohl orientiert sich die Auswahl der Indikatoren an neo-institutionalistischen Überlegungen und der Fokus wird auf drei verschiedene Elemente gesetzt (*Abbildung 2*):

- Erstens geht es um die Frage des legalen Rahmens und der vorhandenen Regulationen, die den Einsatz, die Form und die Nutzung von Evaluationen steuern.
- Zweitens steht die regelmäßige Einbindung des Instruments der Evaluation in die praktische Politik sowohl generell als auch in den verschiedenen Teilbereichen im Fokus.
- Drittens schließlich beschäftigen sich die Länderstudien mit der Nutzung von Evaluationen und ihren Ergebnissen bei der Planung, (Weiter-)entwicklung, Steuerung und Umsetzung von Politik durch die verschiedenen Beteiligten im politischen System.

Abbildung 2: Leitfragen zur Institutionalisierung von Evaluation im politischen System

National Regulations on Evaluation	Evaluation Practice (Evaluation Policies)	Use of Evaluation within the Political System
Are there national laws or regulations about evaluation or use of evaluation? Are there sectoral laws or regulations about evaluation or use of evaluation (e.g. a law about school evaluation or evaluation in the higher education system as example for laws in the educational sector)?	Is it possible to speak of a frequent rhythm of evaluations, for instance for every new legislation or for every national program? Or is it rather non-specific?	Which sectors are "good performer", which are "bad performer" regarding use of evaluation and evaluation findings? Please describe up to 3 sectors that can be considered as leading and 3 sectors that are lagging behind in the field of evaluation's use.
Are there policies or strategies about evaluation or use of evaluation, either national or sectoral? Are there administrative regulations about evaluation or use of evaluation in different policy fields (instructions, guidelines, etc.)?	Does evaluation take place in all sectors/ policy fields of a country? And within one sector, is evaluation applied for measures funded in different ways or maybe only the ones that received funding by the European Union?	On which aspect do most evaluations focus in these sectors (e.g. Planning and Steering, Accountability and Legitimacy, Enlightenment)? Which professional groups use evaluation and evaluation findings regularly (e.g. political decision makers, program or project manager, administrative staff)?
Is evaluation and use of evaluation findings embedded in parliamentary structures? Do parliamentarians in your country deal with evaluation findings for their own political work? Do parliamentarians in your country demand evaluations for their own political work?	How would you describe the relation between internal and external evaluations? Which form is carried out more often and for what purposes? How would you describe the relation between process and impact/ outcome evaluations? Which form is used more often and for what purposes? Do independent internal departments exist, in ministries or elsewhere?	How is the quality of evaluations guaranteed (e.g. regular conduction of meta-evaluations analyses, competence requirements for evaluators, quality requirements for evaluations)?

Quelle: Stockmann, Meyer & Taube, 2020a: Appendix (von den Autoren gekürzte Fassung).

Die eigentliche Datenerhebung erfolgte durch Expertinnen und Experten vor Ort, die angehalten waren, zur Abdeckung der Breite ihrer Informationen zusätzlich Interviews mit Ressourcenpersonen durchzuführen. Die Auswahl der Expertinnen und Experten erfolgte über die Netzwerke der lokalen Evaluationsgesellschaften (VOPEs = Voluntary Organisation of Professional Evaluation) sowie entsprechende Aufrufe bei der Europäischen Evaluationsgesellschaft (EES = European Evaluation Society) und während der Kick-Off-Veranstaltung bei der EES-Tagung in Maastricht, Niederlande, im Herbst 2016. Insgesamt konnten so 30 namhafte Expertinnen und Experten als Autorinnen und Autoren für den Band gewonnen werden.

Zur Gewährleistung der Vergleichbarkeit der Beiträge wurde ein für alle Autoren verpflichtendes Analyseraster sowie ein gemeinsam entwickeltes Glossar zur Verfügung gestellt. Die eingereichten Rohfassungen sind von den Herausgebern hinsichtlich der Vollständigkeit und Plausibilität der Aussagen geprüft und mit Bitte um Vervollständigung bzw. Schärfung der Argumentation zurückgegeben worden. Dadurch konnten natürlich nicht sämtliche Lücken geschlossen, unterschiedliche Gewichtungen oder Interpretationen jedoch weitgehend vermieden werden.

Bezüglich der Datenqualität lässt sich sagen, dass einige der erfassten Aspekte (z. B. die Gesetze und Regulationen zu Evaluationen) gut dokumentiert und nachvollziehbar belegt sind, andere jedoch auf Einschätzungen beruhen (z. B. die Verbreitung der Evaluationspraxis in den verschiedenen Evaluationsefeldern). Die Herausgeber haben sich durch die beschriebene Vorgehensweise bemüht, solche Einschätzungen soweit möglich zu homogenisieren. Die methodischen Schwierigkeiten sind aber primär durch den Stand der Institutionalisierung selbst bedingt: Nur dort, wo sich dauerhaft implementierte Institutionen herausgebildet haben, ist eine entsprechende Erfassung einfach und empirisch eindeutig machbar. Noch dominiert aber ein gewisser „Wildwuchs“, d. h. ein wenig geregelter und nicht dauerhaft festgelegter Einsatz der Evaluation im politischen System.

Erschwerend kommt hinzu, dass der Begriff Evaluation in den letzten Jahren zunehmend inflationär genutzt wird und in den verschiedenen europäischen Sprachen sich zusätzlich unterschiedliche Interpretationen bzw. Konnotationen herausgebildet haben. Eine Abgrenzung der Begrifflichkeiten ist dementsprechend nicht in allen Fällen trennscharf und eindeutig möglich, auch wenn mittels des Glossars zwischen den Beteiligten an dem Europaband ein allgemein geteiltes Verständnis gewährleistet werden konnte. Trotz aller Bemühungen konnte so letztlich nicht vollständig ausgeschlossen werden, dass in manchen Ländern Dinge als Evaluationen bezeichnet werden, die in anderen mit abweichenden Begriffen oder Konzepten erfasst sind. Auch diese Schwierigkeiten sind aber weniger der Methodik denn der Dynamik eines fortschreitenden Verbreitungs- und Institutionalisierungsprozesses geschuldet (zur Komplexität solcher Prozesse siehe vor allem die Ergebnisse der Organisationsforschung z. B. Barley & Tolbert, 1997; Lawrence, Winn & Deveraux, 2001; Yu, 2013).

Im Vergleich zu Jan-Eric Furubo, Ray Rist & Rolf Sandahl (2002) und auch den eigenen Vorarbeiten zur zukünftigen Entwicklung der Evaluation (Stockmann & Meyer, 2017a) ist sowohl die Länderauswahl als auch der Datenerhebungsprozess deutlich systematischer und kontrollierter erfolgt. Zwar beziehen auch hier sich die Befunde auf Experteneinschätzungen in ausgewählten Ländern, die Expertinnen und Experten waren aber gehalten, nicht nur ihre eigenen Einschätzungen widerzugeben, sondern die Befunde empirisch abzusichern und möglichst umfassend zu belegen. Im Unterschied

zu der eher willkürlichen Länder- und Themenauswahl bei Furubo, Rist und Sandahl (2002), die wichtige Aspekte nicht behandelt und andere kaum nachvollziehbar unterschiedlich gewichtet haben, werden die Fragen aus einem theoretisch fundierten Raster abgeleitet. Insgesamt wurde ein deutlich höherer Aufwand zur Gewährleistung eines einheitlichen Bewertungsmaßstabs und des wissenschaftlich überprüfbaren Belegs der getroffenen Einschätzungen betrieben.

3 Institutionalisierung der Evaluation im politischen System

Die zur Untersuchung der Institutionalisierung der Evaluation in den politischen Systemen Europas verwendeten Leitfragen beziehen sich auf drei verschiedene Aspekte, welche sowohl die Legislative als auch die Exekutive betreffen. Kollektives gesellschaftliches Handeln wird idealtypischer Weise durch die auf nationaler Ebene von der Legislative festgelegten (und von der Judikative kontrollierten) Regeln und eine strikt regelkonforme Sachbearbeitung durch Regierung und staatliche Verwaltung in der Exekutive gewährleistet. Kollektive Verhaltensänderungen werden dabei primär durch die Bildung neuer bzw. die Veränderung bestehender Institutionen erreicht. Diese bestehen nicht nur aus der Verabschiedung neuer Regeln in Form von Gesetzen oder Verordnungen, sondern auch aus deren Interpretation (z. B. mittels der Ausgestaltung durch die Rechtsprechung oder die praxisorientierte Umsetzung durch zuständige Verwaltungen) sowie der Herausbildung von Kontroll- und Vermittlungsmechanismen zur Überwachung regelkonformen Verhaltens und zur Sanktion abweichenden Verhaltens.

Der erste Themenblock befasst sich mit der formalen Einbindung von Evaluation durch Regulationen in der Legislative und den darauf basierenden Kontrollen des Regierungshandelns. Solche nationalen Regeln sind für die Sicherstellung eines kohärenten Verständnisses und Einsatzes des Instruments der Evaluation von größter Bedeutung. Dies hat sich auf supranationaler Ebene durch die Initiativen und Aktivitäten zentraler multinationaler Organisationen (EU, Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD), Vereinte Nationen (UN), Weltbank) gezeigt (Meyer, 2019a).

Aus globaler Perspektive ist zunächst festzuhalten, dass es eine unüberschaubare Vielfalt von Regelwerken in Form von Standards, Richtlinien, Selbstverpflichtungen, Durchführungsempfehlungen etc. zu Evaluationen mit einem mehr oder weniger hohen Verpflichtungsgrad innerhalb von nationalen und transnationalen Organisationen, Netzwerken, Politikfeldern, Verwaltungen usw. gibt (Meyer, 2019b). Für die nationale Ebene hat Barbara Rosenstein (2013) 115 Staaten untersucht und in immerhin 20 Staaten eine Nationale Evaluationspolitik gefunden, die auf bestimmten formalen Regeln beruht.

In Europa gibt es lediglich drei Länder (Frankreich, die Niederlande und die Schweiz), die Evaluation auf höchster nationaler Ebene als allgemeingültiges Instrument gesetzlich verankert haben. Vier weitere Staaten (Deutschland, Finnland, Großbritannien und Lettland) sehen Evaluationen in verschiedenen Verordnungen und Dekreten als allgemein verbindlich vor (in Deutschland z. B. in den Bundes- und Landshaushaltsordnungen sowie der gemeinsamen Geschäftsordnung der Bundesministerien, siehe ausführlich in Stockmann & Meyer, 2017b). Weitaus häufiger sind aber spezielle Regelungen in einzelnen Politikfeldern verbreitet, die zumeist durch die zuständigen

Ressorts (Ministerien) oder die ihnen nachgeordneten Behörden erstellt und umgesetzt werden. Hier bleibt es den zuständigen Stellen weitgehend selbst überlassen, wie sie Evaluation verstehen, in welcher Form und für welche Zwecke sie eingesetzt wird und welche Qualitätsanforderungen sowohl an die Evaluation als auch an die geprüften Maßnahmen gestellt werden. Dies bedingt den Aufbau von Verwaltungseinheiten zur Beauftragung, Überwachung und Durchführung von Evaluationen, die es mit der Ausnahme von Portugal und Rumänien in allen untersuchten europäischen Staaten gibt. Selten sind dies allerdings auf Evaluationen spezialisierte Einheiten, zumeist wird diese Funktion mit anderen Aufgaben (z. B. Accounting, Controlling oder Monitoring) verknüpft und die Evaluation spielt in diesem Kontext eher eine untergeordnete Rolle. Eigenständige staatliche Evaluationsinstitute wie das Deutsche Evaluierungsinstitut DEval im Bereich der deutschen Entwicklungszusammenarbeit gibt es nur äußerst selten.

Zentrales Organ der Legislative ist das Parlament und zur wichtigsten Aufgabe gehört die Kontrolle der Exekutive (insbesondere der Regierung). Diese Kontrollfunktion wird zumeist durch die parteiübergreifende Arbeit der Abgeordneten in Ausschüssen ausgeübt, die neben der Rechtmäßigkeit auch die Effektivität des Regierungshandelns überwachen sollen. Darüber hinaus gibt es unabhängige Organe wie z. B. Rechnungshöfe und *Audit Offices*, welche das Regierungshandeln und die administrative Umsetzung von Gesetzen eigenständig oder im Auftrag der Legislative prüfen können.

Aus diesem Grund liegt es nahe, dass die Parlamentarier und Parlamentarierinnen sowie speziell die parlamentarischen Ausschüsse ein besonderes Interesse an Evaluationen und an einer Nutzung für diesen Zweck haben sollten. Dieses Interesse manifestiert sich z. B. in der Tatsache, dass sich 2014 Parlamentarierinnen und Parlamentarier weltweit in einem „Global Parliamentarians Forum for Evaluation“ (<https://globalparliamentarianforum.wordpress.com/>) zusammengeschlossen haben. U. a. wird hier der institutionelle Aufbau von Monitoring und Evaluationssystemen und die Entwicklung nationaler Evaluationspolitiken unterstützt. Bemerkenswert ist allerdings, dass Europa in diesem Forum mit Ausnahme der Schweiz keine Rolle spielt und es vor allem Parlamentarier und Parlamentarierinnen aus dem globalen Süden sind, die sich hier engagieren und die Verbreitung der Evaluation zu befördern suchen.

Die Ergebnisse des GLOBES bestätigen dies: Außer der Schweiz gibt es kein Land, in dem das Parlament Evaluationsergebnisse systematisch zur Kontrolle des Regierungshandelns einsetzt oder gar selbst Evaluationen beauftragt. Evaluationsbefunde sind nur in drei weiteren Ländern (Dänemark, Deutschland und den Niederlanden) in der parlamentarischen Arbeit und in den Debatten von Bedeutung. Einige weitere Länder verfügen über einen wissenschaftlichen Dienst, der sich allerdings höchst selten mit Evaluationen beschäftigt (Finnland, Irland, Italien und Polen). Deutlich mehr Einfluss auf die Etablierung der Evaluation haben allerdings die zumeist eigenständig agierenden Rechnungshöfe / *National Audit Offices*, die in einer Reihe von Ländern (Belgien, Dänemark, Finnland, Lettland, Niederlande, Portugal, Schweiz und Tschechien) nicht nur wie generell üblich Performance-Audits, sondern auch Policy-Evaluationen zur Wirkungskontrolle durchführen.

Diese Heterogenität und Diversität der legislativen Verankerung spiegelt sich in einer ebenso unübersichtlich wie vielfältigen Anwendungspraxis auf Seiten der Exekutive wider. Der Einsatz hängt dabei stark von den einzelnen Politikfeldern ab: In einigen Ländern – insbesondere in Mitteleuropa – wird die Evaluation fast ausschließlich dort genutzt, wo sie von den Evaluationsrichtlinien der EU verlangt wird. Auf der

anderen Seite ist nur in den Niederlanden der Einsatz von Evaluation in praktisch allen Politikfeldern üblich. Ansonsten gibt es einige Bereiche, in denen Evaluation überall eingesetzt wird (Beschäftigung, Bildung, Entwicklungszusammenarbeit, Forschung und Entwicklung sowie Gesundheit) und andere (Außenpolitik, Innenpolitik, Justiz, Kultur und Landwirtschaft), in denen dies nur selten geschieht.

Dies führt allerdings innerhalb der einzelnen Sektoren nicht zu einer gleichartigen Nutzung in verschiedenen Ländern, im Gegenteil kennzeichnet Europa eine extreme Heterogenität hinsichtlich der Art, des Umfangs, der Häufigkeit und der Intensität der Anwendung von Evaluationen. Generell ist ein regelmäßiger, gleichförmiger und standardisierter Einsatz gegenüber dem punktuellen und den jeweiligen Anforderungen entsprechender, individueller Gebrauch die Ausnahme. Selbst in den Sektoren, in denen es auf internationaler Ebene Vereinbarungen und Vorgaben zum Einsatz von Evaluationen gibt, unterscheidet sich die Praxis auf nationaler Ebene sehr stark. Die Intensität des Evaluationseinsatzes nimmt dabei mit der formalen und legalen Verankerung auf nationaler Ebene zu, wobei allerdings unklar bleibt, ob die Verankerung den höheren Einsatz bedingt oder nicht eher umgekehrt die wachsende Anwendung die Notwendigkeit einer legalen Verankerung befördert (hierzu ausführlicher Stockmann & Meyer, 2020).

Eine häufige Anwendung bedeutet jedoch nicht unbedingt, dass die Evaluationsergebnisse auch für die Weiterentwicklung von Programmen und Projekten sowie im weiteren Sinne für die Politikgestaltung genutzt werden (zu den Besonderheiten der Nutzung von Evaluationen im politischen System siehe Bjørnholt & Larsen, 2014; zu den Einflussfaktoren Vanlandingham, 2011). Letztlich sind der Gebrauch und die Nutzung von Evaluation zur Politikgestaltung ein politischer Prozess, dem in den einzelnen Ländern verschiedene Barrieren und Widerstände entgegenstehen. Häufig fehlt es an systematischen, fest institutionalisierten Wegen zur Verbreitung und Weiterverarbeitung der Evaluationsergebnisse (z. B. in Belgien und Frankreich). An manchen Stellen gibt es eine generelle Abwehrhaltung gegen Evaluationen, die zumeist auf Vorurteilen oder Ängsten basieren (z. B. in Italien). Insbesondere in den Ländern mit einem weitentwickelten formalen Rahmen (Niederlande und Schweiz) wird die zunehmende Routinisierung und bürokratische „Abarbeitung“ der „Evaluationspflicht“ als ein Problem gesehen. Hinzu kommt in manchen Ländern (z. B. Dänemark und Finnland) die fortschreitende Vermengung mit anderen Konzepten und Managementpraktiken, welche die zentrale Bedeutung der Evaluation zur Faktengenerierung in Frage stellt und die distanzierte, neutrale und faktenorientierte Perspektive aufgibt. Dies führt dazu, dass sich die Evaluierungspraxis zunehmend von ihrer wissenschaftlichen Basis entfernt und von den aktuellen Trends in der universitären Forschung entkoppelt wird. Umgekehrt verliert der akademische Bereich teilweise das Interesse an Evaluation und die bereits erreichte, aber vielfach noch sehr schwach verankerte Professionalisierung in Forschung und Lehre wird in Frage gestellt (hierzu ausführlicher Meyer & Stockmann, 2020).

Zusammenfassend lässt sich somit sagen, dass es zur systematischen Nutzung von Evaluationen und ihrer Ergebnisse im politischen System formalisierte und fest institutionalisierte Bearbeitungswege geben muss, die dann aber zu einer bürokratischen Verarbeitung im Sinne eines Pflichtprogramms führen und der Evaluationen impliziten Intention einer kreativen, lernorientierten und diskursiven Weiterentwicklung der Programme und Projekte entgegenstehen. Offensichtlich ist es noch nirgendwo gelungen,

eine systematische Be- und Verarbeitung von Evaluation innerhalb der bestehenden Politikpraxis zu institutionalisieren, die den kontinuierlichen Einsatz und gleichzeitig die angemessene Nutzung zufriedenstellend gewährleisten kann. Dies bedeutet jedoch nicht unbedingt, dass Evaluationen prinzipiell nicht oder nur ungenügend genutzt werden (siehe dazu die Fallbeispiele von Petra Kaps und Regine Paul in diesem Heft).

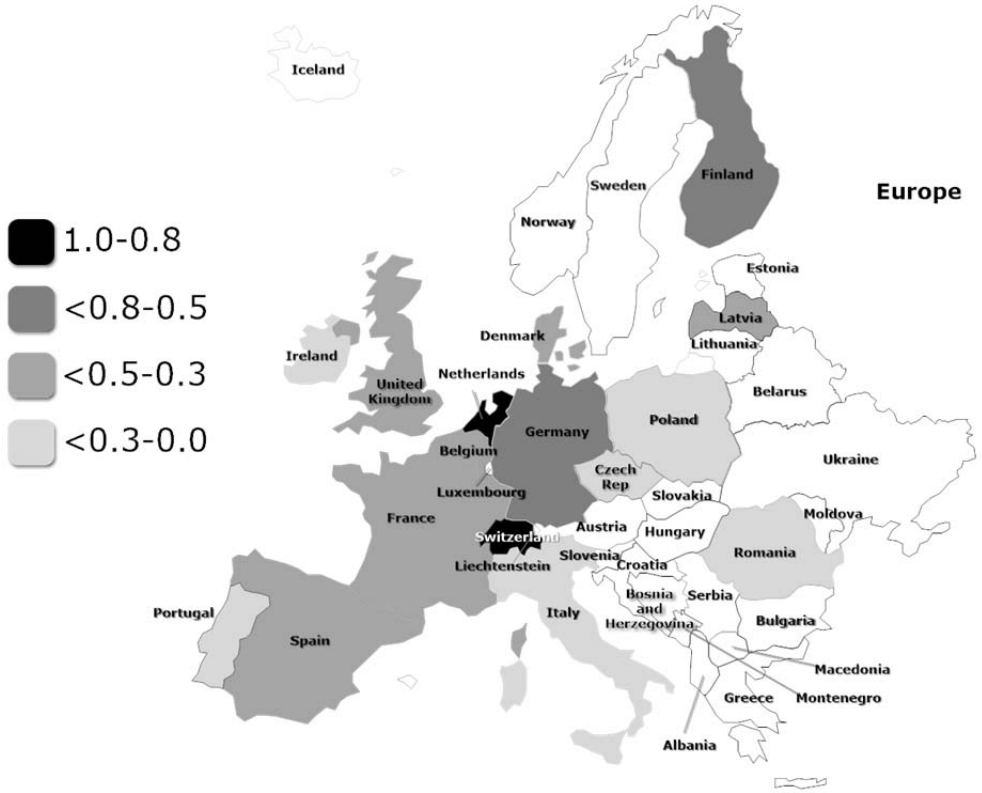
Die Institutionalisierung von Evaluation im politischen System stellt sich also aus mehreren Gründen als Problem dar. Erstens müssen bei der Nutzung von Evaluation als Steuerungsinstrument der Exekutive Ängste, Vorurteile und Widerstände der Akteure überwunden werden, die staatliche Kontrollen und den Verlust eigenständiger Handlungsspielräume fürchten. Zweitens entstehen u. a. hierdurch bedingt innerhalb der verschiedenen Politikfelder und den jeweils zuständigen Ressorts voneinander abweichende Praktiken des Einsatzes von Evaluation, die in spezifische Regelungen münden – häufig zunächst in Form unverbindlicher Empfehlungen und dem Teilen von Erfahrungswerten. Drittens führt diese Regulierung auf sektoraler Ebene, die zumeist durch transnationale Organisationen (insbesondere die EU) zur Homogenisierung der europäischen Politik und zur Absicherung eigener Investitionen gefördert wird, zu einer sehr heterogenen Evaluationspraxis auf nationaler Ebene. Bis auf wenige Ausnahmen ist es deshalb den europäischen Nationalstaaten bisher nicht gelungen, eine einheitliche nationale legale Grundlage für Evaluationen zu entwickeln und diese in ihre jeweilige politische Ordnung als festen Bestandteil der Politikgestaltung aller Ressorts über die Legislative einzubinden.

Die Institutionalisierung der Evaluation im politischen System folgt aus diesem Grund in Europa keineswegs einem gleichmäßigen Verlaufsmuster. Entsprechend der aktuellen Bestandsaufnahme des GLOBE lassen sich grob vier verschiedene Ausmaße der Institutionalisierung identifizieren (*Abbildung 3*), die jedoch nicht als unterschiedliche Stadien eines einheitlichen Entwicklungsprozesses zu bezeichnen sind.

An der Spitze befinden sich die Niederlande und die Schweiz, denen es als einzige Länder (zudem vergleichsweise früh) gelungen ist, auf nationaler Ebene eine legale Grundlage zu schaffen und Evaluation als Standardinstrument zu implementieren. In beiden Staaten werden dadurch Evaluationen regelmäßig eingesetzt und überdurchschnittlich häufig als Entscheidungsgrundlage genutzt. Das hohe Ausmaß an Institutionalisierung bedeutet jedoch nicht, dass alle Widerstände überwunden werden konnten und eine einheitliche, allgemein akzeptierte nationale „Evaluationskultur“ entstanden ist.

Mit Deutschland und Finnland folgen zwei weitere Staaten, die sich etwas vom Durchschnitt der europäischen Länder absetzen konnten. Hier gibt es zwar keinen allgemeingültigen legalen Rahmen und die Evaluationspraxis konzentriert sich stärker auf bestimmte Politikfelder, im Vergleich zur dritten Gruppe sind aber Evaluationen in deutlich mehr Bereichen gut verankert und werden regelmäßig eingesetzt. Die dritte Gruppe umfasst mit Belgien, Dänemark, Frankreich, Großbritannien und Spanien die meisten westlichen Staaten zu denen bemerkenswerterweise Lettland als einziges der mitteleuropäischen Länder aufgeschlossen hat. Kaum Institutionen und nur wenig Evaluationseinsatz gibt es in Irland, Italien, Polen, Portugal, Rumänien und Tschechien. Es ist natürlich eine interessante Frage, wie sich diese Unterschiede erklären lassen und welche Ursachen hierfür verantwortlich zu machen sind.

Abbildung 3: Stand der Institutionalisierung von Evaluation in Europa 2019



Quelle: Eigene Berechnung; Skala von 0 „Evaluation ist nicht institutionalisiert“ bis 1 „Evaluation ist voll und ganz institutionalisiert“ (Mittelwert von 7 Indikatoren, jeweils 0-1 codiert).

4 Einflussfaktoren und ihre Bedeutung

Die beschriebenen Unterschiede und Gemeinsamkeiten hinsichtlich der Institutionalisierung von Evaluation im politischen System sind eine Momentaufnahme und könnten sowohl durch die spezifischen nationalen Entwicklungen als auch durch abweichende Rahmenbedingungen und daraus resultierende Unterschiede zwischen den Ländern bedingt sein. Auf einige dieser potentiellen Einflussfaktoren und ihre Bedeutung zur Erklärung der Varianz zwischen den europäischen Ländern soll im Folgenden näher eingegangen werden.

Als erstes lassen sich die Länder Europas hinsichtlich der *Entwicklung einer nationalen Evaluationstradition* differenzieren. Einige westeuropäische Länder (speziell Deutschland, Frankreich, Großbritannien, die Niederlande, Norwegen, Schweden und die Schweiz) können bereits auf eine lange Evaluationsgeschichte zurückblicken, die bis in die 1960er Jahre zurückreicht. Erste Institutionalisierungsansätze gab es dort bereits in den 1980er Jahren (als Überblick siehe Derlien, 1990; für Deutschland Hellstern & Wollmann, 1984). In anderen Ländern konnte sich Evaluation dagegen erst relativ spät als Steuerungsinstrument etablieren. Dies gilt im Westen für Spanien und

Portugal, deren Demokratisierungsprozesse nach Ende der Diktaturen in den 1980er Jahren abgeschlossen waren und die neuen politischen Strukturen sich vor allem nach dem Beitritt zur Europäischen Union 1986 stabilisieren konnten (siehe dazu Barrios, 2009; Fonseca, 2009). Der Aufbau und die Entwicklung der Evaluation ist hier eng mit dem Beitritt zur Europäischen Union verbunden.

Ähnliches gilt für die mittelosteuropäischen Länder (Lettland, Polen, Rumänien und Tschechien), deren Demokratisierungsprozesse mit der Erosion des Ostblocks in den 1980er Jahren begannen und erst Ende der 1990er Jahre beendet werden konnten. Auch hier prägte der Beitritt zur Europäischen Union 2004 die Institutionalisierung der Evaluation wesentlich mit. Dies begründet sich vor allem durch die Kohäsionspolitik der Europäischen Kommission und die Verknüpfung von Regional- und Infrastrukturförderung mit dem Aufbau eines Evaluationssystems zur Wirkungskontrolle dieser Maßnahmen (vgl. zur „Good Governance“-Politik der EU in Mitteleuropa Börzel, Pamuk & Stahn, 2008; Börzel, 2009, zur Verankerung der Evaluation in der EU Bloemeyer & Eußner, 2020).

Diese *historisch bedingten Unterschiede* prägen vor allem die Art und Weise, wie Evaluationen in den Ländern institutionalisiert worden sind, sowie die Politikfelder, in denen Evaluation betrieben wird. In den Mitteleuropäischen Ländern führte der hohe Bedarf an europäischen Beihilfen zur Strukturförderung zu einer Implementierung von Evaluationsstellen auf nationaler Ebene, die es in dieser Form in den westlichen Pionierstaaten nicht oder nur ansatzweise gegeben hat. Die Evaluationspraxis ist hier deutlich stärker von den europäischen Fördermitteln und damit von deren Richtlinien zur Mittelvergabe abhängig, die eine Implementierung von Evaluationen vorsehen. In anderen Politikfeldern kann sie sich nur allmählich etablieren.

Neben der Europäischen Union gibt es noch weitere *internationale Organisationen und Netzwerke*, welche die nationale Entwicklung der Evaluationspraxis stark geprägt haben. Dies gilt vor allem für die OECD, die insbesondere im Bildungsbereich und bei der Entwicklungszusammenarbeit Einfluss erlangt hat. In der Entwicklungszusammenarbeit z. B. kam es bereits mit der Gründung der OECD 1961 zur Institutionalisierung eines Development Assistance Committees (DAC) zur Koordination der Geberländer, in dessen Rahmen ab den 1980er Jahren auch eine permanente Arbeitsgruppe zur Entwicklung der Evaluation entwicklungspolitischer Programme und Projekte gebildet wurde (DAC Network on Development Evaluation, <http://www.oecd.org/dac/evaluation/>). Dies hat zu einer kontinuierlichen Weiterentwicklung und Institutionalisierung der Evaluation auf nationaler Ebene geführt (u. a. bedingt durch die regelmäßigen *Peer Reviews* der nationalen Evaluationspraxis, siehe u. a. OECD, 2016). Ähnliche Reviews durch die OECD gibt es auch in anderen Politikbereichen (z. B. Nusche, Braun, Halász & Santiago, 2014) und darüber hinaus wurden eine Vielzahl von Handlungsempfehlungen, *Guidelines* und Praxisberichte zu Anwendung und Einsatz von Evaluationen erstellt (als Beispiel hier die Richtlinien für Evaluationen in Krisengebieten OECD, 2012).

Im Unterschied zur EU ist der Verpflichtungsgrad der OECD und anderer internationaler Organisationen geringer, d. h. die Teilnahme und die Einhaltung von Regeln und Vorgaben ist freiwillig. Der Einfluss ist aber in bestimmten Politikfeldern trotzdem hoch und führt zumindest punktuell und mit zeitlichem Abstand zu Veränderungen auf nationaler Ebene. Die Entwicklung der Evaluation wurde und wird durch diese internationale Zusammenarbeit wesentlich geprägt, wobei hierdurch die sektorale Differenzierung auf nationaler Ebene angesichts einer fehlenden allgemeinen Evaluationspolicy

internationaler Organisationen befördert wird. Hinzu kommt noch der Fakt, dass nicht alle Nationalstaaten in Europa in allen Politikfeldern auf internationaler Ebene in gleicher Weise engagiert und in den supranationalen Austausch eingebunden sind. Insgesamt führt der internationale Einfluss deshalb nicht zu einer generellen Stärkung der Evaluation auf nationaler Ebene und beeinflusst die Institutionalisierung im nationalen politischen System lediglich indirekt, ein direkter Einfluss (weder positiv noch negativ) lässt sich mit den GLOBE-Daten nicht belegen. Dies betrifft sowohl die legale Verankerung als auch die Anwendung und Nutzung in der Exekutive.

Ein wichtiger Unterschied zwischen den Ländern Europas ist deren Größe sowohl hinsichtlich der Fläche als auch hinsichtlich der Einwohnerzahl. Besonders die *Bevölkerungsgröße* ist für die Herausbildung eines eigenständigen nationalen Evaluationsmarkts und die Organisation eines Austauschs zwischen den in Evaluationen eingebundenen Personen im Land wichtig. Dies hat unmittelbar Auswirkungen auf die Professionalisierung und die Etablierung professioneller Aus- und Weiterbildungsstrukturen – hier tun sich kleinere Länder deutlich schwerer als die größeren.

Aus Sicht der Anbieter von Evaluationsdienstleistungen bedarf es einer kritischen Masse von Aufträgen, damit sich eine Beratungsfirma auf diese Art von Dienstleistungen spezialisieren kann. Wenn diese im nationalen Kontext nicht vorhanden ist, muss das Unternehmen sich entweder auf dem internationalen Markt etablieren (was angesichts des offenen europäischen Marktes und den Regulierungen zum Schutz des Marktzugangs europäischer Firmen durch die EU vergleichsweise einfach möglich ist) oder ein breiteres Angebot an Dienstleistungen neben der Evaluation auf dem nationalen Markt bereitstellen. Allerdings besteht dabei die Gefahr, dass Anbieter aus anderen (größeren) Staaten als Konkurrenten aufgrund ihrer umfangreicheren Expertise in Evaluation bessere Marktchancen haben.

Die Größe des Nationalstaats hat nicht nur Auswirkungen auf die Angebotsseite, sondern auch auf die Nachfrage, weil bedingt durch das Steueraufkommen und die *Wirtschaftsleistung* in kleinen Ländern in der Regel weniger Programme und Projekte durchgeführt werden bzw. durchgeführt werden können. Bei gleichen Rahmenbedingungen wird es dementsprechend in kleinen Staaten weniger Evaluationen als in größeren geben. Allerdings existieren in Europa kleine Länder mit extrem hoher Wirtschaftskraft (z. B. Luxemburg, die Schweiz oder Norwegen), während einige größere Länder (z. B. Polen, Rumänien oder die Ukraine) über ein deutlich niedrigeres Wohlstandsniveau verfügen.

Bei den 16 untersuchten Ländern schwankte 2019 die Bevölkerungsgröße zwischen 83 Mio. (Deutschland) und knapp 2 Mio. (Lettland), das Bruttoinlandsprodukt zwischen 3,4 Billionen EUR (Deutschland) und etwas über 30 Mrd. EUR (Lettland) (EUROSTAT Online Data TPS00001 und TEC00001). Wie zu erwarten haben beide Größen einen Einfluss auf den Professionalisierungsgrad der Evaluation auf nationaler Ebene, d. h. die Etablierung eigenständiger Studiengänge und Evaluationsgesellschaften, von Fachzeitschriften oder anderen Kommunikationsmedien sowie von Qualitätssicherungsmechanismen wie z. B. Standards, gelingt mit zunehmender Bevölkerungsgröße und Wirtschaftskraft besser. Interessanterweise gilt das jedoch nicht für die Institutionalisierung im politischen System: Hier ergibt sich kein nennenswerter Einfluss von Bevölkerungsgröße und Wirtschaftskraft.²

Die europäischen Nationalstaaten unterscheiden sich aber nicht nur durch ihre Größe und Wirtschaftskraft, sondern auch bezüglich der historisch gewachsenen und in

der Vergangenheit vielfachen Verwerfungen ausgesetzten *politischen Systeme*. Hinsichtlich der Herrschaftsform dominieren mittlerweile in Europa parlamentarische Regierungssysteme (Deutschland, Irland, Italien, Finnland, Lettland und Tschechien), von denen allerdings einige als parlamentarische Monarchien (Belgien, Dänemark, Großbritannien, Niederlande und Spanien) oder als semi-präsidentielle Republiken (Frankreich, Polen, Portugal und Rumänien) verfasst sind. Eine Besonderheit stellt die Schweiz dar, die sich von den anderen Republiken durch eine deutlich stärkere Konsensorientierung in Richtung Konkordanzdemokratie mit ausgeprägten basisdemokratischen Elementen (z. B. Volksinitiativen und Referenden) unterscheidet. Im Vergleich zu den stärker zentralistisch organisierten Monarchien und semi-präsidentiellen Republiken ist die Schweiz genauso wie Belgien und Deutschland als Föderalstaat organisiert, d. h. die Teilregionen (in der Schweiz Kantone) verfügen über eigenständige Entscheidungskompetenzen in einer Reihe politischer Fragen.

Weitere wichtige Unterschiede ergeben sich durch das Wahlsystem und die daraus resultierende Verteilung der Parlamentssitze, die Organisation der Gewaltenteilung (insbesondere die Abgrenzung zwischen Legislative und Exekutive) sowie die Rolle und Funktion des Staatsoberhauptes (insbesondere in Relation zum Parlament und zum Regierungschef). All dies hat Auswirkungen auf die formalen Grundlagen, Politikstil und informellen Gepflogenheiten innerhalb der Exekutive, speziell des Regierungshandelns und der administrativen Praxis in der öffentlichen Verwaltung (allgemeinen zu den Unterschieden in der Qualität der Regierungsführung in Europa Berg-Schlosser, 2006; zu den Einflussfaktoren und Veränderungen politischer Steuerung des Staats Mayntz, 2004).

Insgesamt lässt sich allerdings feststellen, dass die Art des politischen Systems und seine prägenden Elemente erstaunlich wenig direkten Einfluss auf die Institutionalisierung von Evaluation haben. Es gibt z. B. keine Hinweise, dass die Institutionalisierung in zentralstaatlichen einfacher und schneller als in föderalen Systemen erfolgt. Im Gegenteil war es gerade die Schweiz, deren Konkordanzdemokratie aufwendige konsensorientierte Entscheidungsprozesse erforderlich macht, welche als bisher einziger europäischer Staat frühzeitig und umfassend Evaluation im politischen System institutionalisieren und kontinuierlich weiterentwickeln konnte (hierzu die ausführlichen Analysen von Sager, Widmer & Balthasar, 2017, siehe auch den Beitrag von Thomas Widmer in diesem Heft). Die anderen Bundesstaaten (Belgien und Deutschland) belegen allerdings, dass nicht die föderale Ordnung per se die Institutionalisierung befördert – einige zentralstaatliche Systeme (Finnland und die Niederlande) haben deutlich stärker Evaluation als politisches Instrument integriert.

Die Befunde des GLOBEs legen also nahe, dass die Institutionalisierung von Evaluation im politischen System weniger durch die strukturellen Rahmenbedingungen und Besonderheiten als durch Entwicklungsprozesse und Reformen in der Regierungspraxis geprägt wird. Besonders in den Pionierländern (Deutschland, Frankreich, Großbritannien und Schweden) sind dabei die *Konzepte des New Public Management und der evidence-based policy* wegweisend gewesen und haben zu einer breiten Nutzung von Evaluationen geführt (allgemein zur Modernisierung des Staates Naschold & Bogumil, 2000, zur Rolle der Evaluation Wollmann, 2003, kritisch Pawson, 2006). In der Tat ist die Korrelation zwischen der Institutionalisierung von Evaluation und den beiden zentralen World Governance Indikatoren zur regulatorischen Qualität und zur Effektivität des Regierens sehr hoch (Stockmann, Meyer & Taube, 2020b, p. 515; zu

den World Governance Indikatoren der Weltbank siehe Kaufmann, Kraay & Mastruzzi, 2010). Es besteht also offensichtlich ein Zusammenhang zwischen guter Regierungsführung und Implementierung von Evaluation, wobei vermutlich die Evaluationspraxis mehr das Ergebnis guter Regierungsführung ist als umgekehrt. Die Institutionalisierung in Form von festen, verbindlichen und allgemeingültigen Regeln ist dagegen offenbar eher die Folge denn die Voraussetzung für eine Nutzung des Instruments der Evaluation im politischen System.

5 Schlussfolgerungen und Ausblick

Insgesamt lassen sich gegenwärtig fünf Trends erkennen, die in den Ländern in unterschiedlichem Maße wirksam und für die Institutionalisierung von Evaluation bedeutsam sind.

Erstens prägen in den meisten Ländern immer noch eine *fortschreitende Expansion und Verbreitung* des Instruments der Evaluation die nationale Entwicklung. Dies bedeutet, dass weniger eine Verfestigung und Institutionalisierung denn die Übernahme und Anpassung in neuen Anwendungsgebieten im Vordergrund stehen.

Zweitens gibt es insbesondere in den Politikfeldern, in denen sich die Nutzung von Evaluationen als Standardinstrument etablieren konnte, einen *Trend zur Professionalisierung* im Sinne einer kontinuierlichen Qualitätsentwicklung der Evaluationen als auch einer Diversifikation des Instrumenteneinsatzes. Hier steht die Herausbildung von Institutionen und Regulationen zur Nutzung von Evaluationen im Vordergrund, auch wenn es bisher nur wenige Länder und wenige Handlungsfelder gibt, in denen die Evaluation bereits routinemäßig und klaren Vorgaben folgend eingesetzt wird (dazu für Deutschland die Beiträge in Hense, Böttcher, Kalman & Meyer, 2019; zum weltweiten Professionalisierungstrend siehe Meyer, 2015).

Diesen positiven Entwicklungstrends stehen aber auch negative zur Seite, die in Zukunft mehr Gewicht erlangen könnten. So ist drittens insbesondere in den Pionierländern mit dem fortgeschrittensten Stand der Evaluation eine gewisse Stagnation oder zumindest *nachlassende Dynamik* zu beobachten, die eher mit einer Deprofessionalisierung denn mit fortschreitender Weiterentwicklung einhergeht. Kritische Stimmen beklagen den übertriebenen Einsatz des Instruments in Form einer „Evaluitis“ (Frey, 2007; 2008) und den bescheidenen Nutzwert von Evaluationen (als Überblick zu den Befunden hierzu Johnson, Greenseid, Toal, King, Lawrenz & Volkov, 2009). Diese Tendenz könnte für die Evaluation trotz der in diesen Ländern bestehenden Institutionen gefährlich werden, weil die meisten Institutionen noch vergleichsweise schwach verankert sind (z. B. der gesetzliche Rahmen für die Nutzung fehlt) und deshalb relativ einfach und schnell zurückgenommen werden können (wie z. B. bei der Einrichtung eines nationalen Evaluationsinstituts in Schweden geschehen).

In ähnliche Richtung weist auch der vierte Trend einer *Vermischung und Verfremdung* der Evaluation mit anderen Konzepten wie z. B. dem Auditing oder Controlling, die einer eigenständigen Profilbildung entgegenwirken und ggf. zu einer Auflösung führen können. Ansätze einer solchen Entwicklung lassen sich entsprechend der Befunde des GLOBES in Dänemark, Frankreich und der Schweiz finden.

Schließlich ist noch der generelle Trend weg von faktenbasierter und konsensorientierter Politik hin zu einem seine Wahrheiten selbst produzierenden *Populismus* her-

vorzuheben. Diese allgemeine Entwicklung in der Politik ist nicht allein auf Europa beschränkt, hat aber viele Länder erfasst und zu fundamentalen Umbrüchen im Parteiensystem geführt (Greven, 2016; Bugaric & Kuhelj, 2018; Biorcio, 2019). Bis jetzt sind die Auswirkungen auf die Institutionalisierung von Evaluation noch gering, es ist jedoch zu erwarten, dass ein weiteres Wachstum populistischen Gedankenguts einen ähnlichen Effekt wie die Verbreitung des Konzepts der *evidence-based policy* in den 1980er Jahren haben könnte – nur mit umgekehrtem Vorzeichen, d. h. eine ideologiebasierte Politik wird problemlos auf Fakten produzierende Verfahren wie die Evaluation verzichten können. Einher geht diese Entwicklung zudem mit einer Erosion internationaler Institutionen und einer Renationalisierung, die in Europa bereits zu einer politischen Krise der Europäischen Union geführt hat (Dinan, Nugent & Paterson, 2017). Wie die Befunde belegten, hat die Europäische Union in vielen Ländern eine bedeutende Rolle für die Etablierung und Institutionalisierung gespielt und eine Schwächung der EU könnte sich dementsprechend auch auf nationaler Ebene negativ auswirken.

Insgesamt ist aber nicht zu erwarten, dass die Evaluation als Steuerungsinstrument schnell aus den politischen Systemen Europas verschwinden wird. Vor allem vier Dinge stehen dem entgegen:

- Erstens die ungebrochene Tendenz *transnationaler Organisationen* (insbesondere der EU), Evaluationen von ihren Mitgliedsnationen als Gegenleistung für Förderungsmaßnahmen einzufordern und dafür die Evaluation sowie deren Weiterentwicklung zu fördern.
- Zweitens der mittlerweile in einer Reihe von Sektoren erreichte *Status des Einsatzes von Evaluationen* und dessen Nutzung für die Weiterentwicklung ihrer Instrumente. Dies ist nicht ausschließlich dem Druck der transnationalen Organisationen, sondern auch einer gewissen sektoralen Eigendynamik geschuldet (mit spezifischen nationalen Formungen und Entwicklungstrends).
- Drittens der in manchen Ländern bereits erreichte *Grad der Institutionalisierung* auf nationaler Ebene, nicht nur in Form von Gesetzen und Verordnungen, sondern auch hinsichtlich des regelmäßigen Einsatzes von Evaluation und der geregelten Nutzung von Ergebnissen.
- Und schließlich viertens die *Verfestigung der output- und outcome-orientierten Steuerung* in den politischen Systemen als Folge des *New Public Management*, die kaum zurückgenommen und wieder durch eine reine Inputsteuerung ersetzt werden wird. Als hierfür notwendiges Steuerungsinstrument ist die Evaluation mit dieser Entwicklung verbunden und nicht von ihr trennbar.

Dies bedeutet jedoch auf der anderen Seite nicht, dass Evaluation „fest im Sattel sitzt“ und wie in der Vergangenheit expandieren wird. Dem stehen nicht nur einige allgemeine Entwicklungstrends, sondern auch ein ungelöstes Problem der Institutionalisierung im politischen System entgegen. Die beschriebene „top-down“-Steuerung moderner Gesellschaften mittels bürokratischer Abarbeitung ist nicht, wie von verschiedenen Autoren (z. B. Beck, 1993) erwartet, in eine prinzipielle Krise geraten und ein Wandel von „Government“ zu „Governance“ ist nur partiell eingetreten. Eine kompromissorientierte, partizipative Politikgestaltung konnte sich in Europa nicht verbreiten, in manchen Ländern sind sogar deutlich gegenläufige Tendenzen erkennbar.

Das Beispiel der Schweiz zeigt aber auch, dass sich Evaluation in einem solchen Umfeld deutlich besser entfalten kann als in einer bürokratischen regelgesteuerten und

allein über politische Institution gelenkten Gesellschaft. Dies liegt an einem bisher nicht aufgelösten *Grundsatzkonflikt* zwischen dem Selbstverständnis der Evaluation als Instrument der Aufklärung mit der Aufgabe, zusammenzuführen und die Zivilgesellschaft zu „empowern“, und dem Wunsch der Exekutive, mittels Evaluation das von ihr initiierte gesellschaftliche Handeln hinsichtlich seiner Wirkungen zu überwachen und durch zielorientierte Maßnahmen zu lenken. Diese Ambivalenz prägt gegenwärtig die Institutionalisierung der Evaluation in Europa.

Anmerkungen

- 1 Aus Gründen der besseren Lesbarkeit wird im Folgenden ausschließlich die männliche Sprachform verwendet. Selbstverständlich sind alle Geschlechter gleichermaßen gemeint.
- 2 Getestet wurden neben den absoluten Zahlen auch die Bevölkerungsdichte, das Pro-Kopf-Einkommen, der Human-Development-Index, die Staatsinvestitionen und eine Reihe anderer gängiger demographischer und ökonomischer Kennzahlen. Keiner dieser Indikatoren kann zur Erklärung der beschriebenen Unterschiede hinsichtlich des Institutionalisierungsgrads im politischen System beitragen.

Literatur

- Ansell, Christopher & Torfing, Jacob (Eds.) (2016). *Handbook on Theories of Governance*. Cheltenham: Edward Elgar.
- Barley, Stephen R. & Tolbert, Pamela S. (1997). Institutionalization and Structuration: Studying the Links between Action and Institution. *Organizational Studies*, 18 (1), 93-117. <https://doi.org/10.1177%2F017084069701800106>.
- Barrios, Harold (2009). Das politische System Spaniens. In Wolfgang Ismayr (Hrsg.), *Die politischen Systeme Westeuropas. 4. aktualisierte und überarbeitete Auflage* (S. 713-764). Wiesbaden: VS Verlag.
- Beck, Ulrich (1993). *Die Erfindung des Politischen. Zu einer Theorie reflexiver Modernisierung*. Frankfurt: Suhrkamp.
- Berg-Schlosser, Dirk (2006). The quality of democracies in Europe as measured by current indicators of democratization and good governance. *Journal of Communist Studies and Transition Politics*, 20 (1), 28-55. <https://doi.org/10.1080/13523270410001687091>.
- Biorcio, Roberto (2019). Populism as a challenge for party systems: A comparison between Italy and Spain. In Gregor Fitz, Jürgen Mackert & Brian S. Turner (Eds.), *Populism and the Crisis of Democracy. Vol. 2: Politics, Social Movements and Extremism* (pp. 137-151). London/New York: Routledge.
- Bjørnholt, Bente & Larsen, Flemming (2014). The politics of performance measurement: 'Evaluation use as mediator for politics'. *Evaluation*, 20 (4), 400-411. <https://doi.org/10.1177/1356389014551485>.
- Blomeyer, Roland & Eußner, Ansgar (2020). European Union. In Reinhard Stockmann, Wolfgang Meyer & Lena Taube (Eds.), *The Institutionalization of Evaluation in Europe: Evaluation Globe. Vol. I.* (pp. 463-479). Cham: Springer Nature.
- Börzel, Tanja A., Pamuk, Yasemin & Stahn, Andreas (2008). *Good Governance in the European Union*. Berlin: FU Berlin Berliner Arbeitspapier zur Europäischen Integration 7. Verfügbar unter https://refubium.fu-berlin.de/bitstream/handle/fub188/19026/2008-7_Boerzel_Pamuk_Stahn.pdf?sequence=1 [06.11.19].
- Börzel, Tanja A. (2009). *Transformative Power Europe? The EU Promotion of Good Governance in Areas of Limited Statehood*. Paper prepared for the ERD Workshop "Transforming Political Structures: Security, Institutions, and Regional Integration Mechanisms", Florence, April 16-17, 2009. Verfügbar unter

- <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.551.2884&rep=rep1&type=pdf> [06.11.2019].
- Bugaric, Bojan & Kuhelj, Alenka (2018). Varieties of Populism in Europe: Is the Rule of Law in Danger? *Hague Journal on the Rule of Law*, 10 (1), 21-33.
<https://doi.org/10.1007/s40803-018-0075-4>.
- Cousins, J. Bradley, Goh, Swee C., Elliott, Catherine, Aubry, Tim & Gilbert, Nathalie (2014). Government and voluntary sector differences in organizational capacity to do and use evaluation. *Evaluation and Program Planning*, 44, 1-13. <https://doi.org/10.1016/j.evalprogplan.2013.12.001>.
- Derlien, Hans-Ulrich (1990). Genesis and Structure of Evaluation Efforts in Comparative Perspective. In Ray C. Rist (Ed.), *Program Evaluation and the Management of Government. Patterns and Prospects across Eight Nations* (pp. 147-176). New Brunswick/London: Transaction Publishing.
- Dinan, Desmond, Nugent, Neill & Paterson, William E. (Eds.) (2017). *The European Union in Crisis*. Palgrave MacMillan: London.
- EUROSTAT (o. J.). *Europäische Statistik Online Datenbank*. Verfügbar unter: <https://ec.europa.eu/eurostat/de/data/database> [22.04.2020].
- Fonseca, Sara Claro da (2009). Das politische System Portugals. In Wolfgang Ismayr (Hrsg.), *Die politischen Systeme Westeuropas. 4. aktualisierte und überarbeitete Auflage* (S. 765-818). Wiesbaden: VS Verlag.
- Frey, Bruno S. (2007). Evaluierungen, Evaluierungen ... Evaluitis. *Perspektiven der Wirtschaftspolitik*, 8 (3), 207-220. <https://doi.org/10.1111/j.1468-2516.2007.00248>.
- Frey, Bruno S. (2008). Evaluitis — eine neue Krankheit. In Hildegard Matthies & Dagmar Simon (Hrsg.), *Wissenschaft unter Beobachtung* (S. 125-140). VS Verlag.
- Furubo, Jan-Eric, Rist, Ray & Sandahl, Rolf (2002). *International Atlas of Evaluation*. New Brunswick, NJ: Transatlantic Publishers.
- Goerres, Achim, Siewert, Markus B. & Wagemann, Claudius (2019). Internationally Comparative Research Designs in the Social Sciences: Fundamental Issues, Case Selection Logics, and Research Limitation. In Hans-Jürgen Andreß, Detlef Fetchenhauer & Heiner Meulemann (Eds.), *Cross-National Comparative Research* (pp. 75-97). Wiesbaden: VS Verlag, Sonderheft 59 der Kölner Zeitschrift für Soziologie und Sozialpsychologie.
- Grevén, Thomas (2016). *The Rise of Right-wing Populism in Europe and the United States. A Comparative Perspective*. Berlin: Friedrich-Ebert-Stiftung.
- Hay, Colin, Lister, Michael & Marsh, David (Eds.) (2006). *The State. Theories and Issues*. London: Red Globe Press, MacMillan International.
- Hellstern, Gerd-Michael & Wollmann, Hellmut (1984). Evaluierung und Evaluierungsforschung – ein Entwicklungsbericht. In dies. (Hrsg.). *Handbuch zur Evaluierungsforschung. Bd. 1* (S. 17-90), Opladen: Westdeutscher Verlag.
- Hense, Jan, Böttcher, Wolfgang, Kalman, Michael & Meyer, Wolfgang (Hrsg.) (2019). *Evaluation: Standards in unterschiedlichen Handlungsfeldern, Einheitliche Qualitätsansprüche trotz heterogener Praxis?* Münster: Waxmann.
- Hroch, Miroslav (2005). *Das Europa der Nationen. Die moderne Nationsbildung im europäischen Vergleich*. Göttingen: Vandenhoeck & Ruprecht.
- Ismayr, Wolfgang (Hrsg.) (2009). *Die politischen Systeme Westeuropas. 4. aktualisierte und überarbeitete Auflage*. Wiesbaden: VS Verlag.
- Ismayr, Wolfgang (Hrsg.) (2010). *Die politischen Systeme Osteuropas. 3. aktualisierte und erweiterte Auflage*. Wiesbaden: VS Verlag.
- Jacob, Steve, Speer, Sandra & Furubo, Jan-Eric (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21 (1), 6-31.
<https://doi.org/10.1177/1356389014564248>.
- Johnson, Kelli, Greenseid, Lija O., Toal, Stacie A., King, Jean A., Lawrenz, Frances & Volkov, Boris (2009). Research on Evaluation Use. A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, 30 (3), 377-410. <https://doi.org/10.1177/1098214009341660>.
- Kaps, Petra (2020). Formen der Evaluation experimenteller Politikgestaltung. *dms – der moderne staat*, 13 (1), 81-100.

- Kaufmann, Daniel, Kraay, Aart & Mastruzzi, Massimo (2010). *The Worldwide Governance Indicators Methodology and Analytical Issues*. Washington: World Bank Policy Research Working Paper 5430.
- Lawrence, Thomas B., Winn, Monika I. & Jennings, P. Deveraux (2001). The Temporal Dynamics of Institutionalization. *Academy of Management Review*, 26 (4), 624-644.
<https://doi.org/10.5465/amr.2001.5393901>.
- Mayntz, Renate (2004). Governance im modernen Staat. In Arthur Benz (Hrsg.). *Governance – Regieren in komplexen Regelsystemen* (S. 65-76). Wiesbaden: VS Verlag.
- Meyer, Wolfgang (2015). Professionalisierung von Evaluation: Ein globaler Blick. In: *Zeitschrift für Evaluation*, 14 (2), 215-246.
- Meyer, Wolfgang (2019a). Professionalisierung zwischen Kohärenz, Koexistenz und Konkurrenz. Das Fallbeispiel der Evaluation. In Nicole Burzan (Hrsg.). *Komplexe Dynamiken globaler und lokaler Entwicklungen. Verhandlungen des 39. Kongresses der Deutschen Gesellschaft für Soziologie in Göttingen 2018*. Verfügbar unter [http://publikationen.sozioologie.de/index.php/kongressband_2018/issue/view/33_\[08.11.2019\]](http://publikationen.sozioologie.de/index.php/kongressband_2018/issue/view/33_[08.11.2019]).
- Meyer, Wolfgang (2019b). Die DeGEval-Standards im weltweiten Vergleich. In Jan Hense, Wolfgang Böttcher, Michael Kalman & Wolfgang Meyer (Hrsg.), *Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis?* (S. 277-294) Münster: Waxmann.
- Meyer, Wolfgang & Stockmann, Reinhard (2020). Evaluation in Europe 2019. A comparative analysis. (Manuskript eingereicht zur Publikation in *Evaluation*, im Begutachtungsverfahren).
- Naschold, Frieder & Bogumil, Jörg (2000). *Modernisierung des Staates. New Public Management in deutscher und internationaler Perspektive* (2. Aufl.). Wiesbaden: VS Verlag.
- Nusche, Deborah, Braun, Henry, Halász, Gábor & Santiago, Paulo (2014). *OECD Reviews of Evaluation and Assessment in Evaluation: Netherlands 2014*. Paris: OECD.
- Pattyn, Valérie, Voorst, Stijn van, Mastenbroek, Ellen & Dunlop, Claire A. (2018). Policy Evaluation in Europe. In Edoardo Ongaro & Sandra van Thiel (Eds.), *The Palgrave Handbook of Public Administration and Management in Europe* (pp. 577-594), London: Palgrave MacMillan.
- Paul, Regine (2020). Analyse and rule? A conceptual framework for explaining the variable appeals of ex-ante evaluation in policymaking. *dms – der moderne staat*, 13 (1), 124-142.
- Pawson, Ray (2006). *Evidence-based Policy. A realist perspective*. London et al: Sage.
- Peters, B. Guy (2019). *Institutional Theory in Political Science. The New Institutionalism* (4th edition). Cheltenham, UK / Northampton, MA, USA: Edward Elgar.
- OECD (2016). *Evaluation Systems in Development Co-operation. 2016 Review*. Paris: OECD.
- OECD (2012). *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility. Improving Learning for Results*. Paris: OECD.
- Rosenstein, Barbara (2013). *Status of National Evaluation Policies. Global Mapping Report*. Washington: EvalPartners.
- Sager, Fritz, Widmer, Thomas & Balthasar, Andreas (2017). *Evaluation im politischen System der Schweiz. Entwicklung, Bedeutung und Wechselwirkungen*. Basel: NZZ Libro.
- Stockmann, Reinhard (2008). *Evaluation and Quality Development: Principles of Impact-based Quality Management*. Frankfurt am Main: Peter Lang.
- Stockmann, Reinhard & Meyer, Wolfgang (Hrsg.) (2017a). *Die Zukunft der Evaluation. Trends, Herausforderungen, Perspektiven*. Münster: Waxmann.
- Stockmann, Reinhard & Meyer, Wolfgang (2017b). Evaluation in Deutschland. Woher sie kommt, wo sie steht, wohin sie geht. *Zeitschrift für Evaluation*, 16 (2), 57-110.
- Stockmann, Reinhard & Meyer, Wolfgang (2020). Die Institutionalisierung der Evaluation in Europa. *Zeitschrift für Evaluation* 1/2020 (im Erscheinen).
- Stockmann, Reinhard & Meyer, Wolfgang (eds.) (2021). *The Institutionalization of Evaluation in the Americas: Evaluation Globe. Vol. II*. Cham: Springer Nature (in Vorbereitung).
- Stockmann, Reinhard, Meyer, Wolfgang & Taube, Lena (Eds.) (2020a). *The Institutionalization of Evaluation in Europe: Evaluation Globe. Vol. I*. Cham: Springer Nature.

- Stockmann, Reinhard, Meyer, Wolfgang & Taube, Lena (2020b). The Institutionalization of Evaluation in Europe: A Synthesis. In: dies. (Eds.), *The Institutionalization of Evaluation in Europe: Evaluation Globe. Vol. I.* (pp. 483-522). Cham: Springer Nature.
- Vanlandingham, Gary R. (2011). Escaping the Dusty Shelf: Legislative Evaluation Offices' Efforts to Promote Utilization. *American Journal of Evaluation*, 32 (1), 85-97.
- Vedung, Evert (2017). *Public Policy and Program Evaluation*. Abingdon/New York: Routledge.
- Widmer, Thomas (2020). Wechselwirkungen von Politik und Evaluation: Befunde aus der Schweiz. *dms – der moderne staat*, 13 (1), 44-60.
- Wollmann, Hellmut (Ed.) (2003). *Evaluation in Public-Sector Reform. Concepts and Practice in International Perspective*. Cheltenham/Northampton: Edward Elgar.
- Yu, Kyoung-Hee (2013). Institutionalization in the Context of Institutional Pluralism: Politics as a Generative Process. *Organizational Studies*, 34 (1), 105-131.

Anschriften der Autoren:

Adj. Prof. Dr. Wolfgang Meyer, Centrum für Evaluation, Universität des Saarlandes, Campus C 5.3, 66123 Saarbrücken, E-Mail: w.meyer@mx.uni-saarland.de.

Prof. Dr. Reinhard Stockmann, Centrum für Evaluation, Universität des Saarlandes, Campus C 5.3, 66123 Saarbrücken, E-Mail: r.stockmann@ceval.de.