

Nicola Döring

---

## 11.1 Einleitung

Mit „Evaluationsforschung“, „wissenschaftlicher Evaluation“ oder kurz „Evaluation“ ist die Bewertung eines Gegenstandes mit Hilfe sozialwissenschaftlicher Methoden durch Evaluationsfachleute gemeint (Widmer/De Rocchi 2012: 11). Die Evaluationsforschung ist ein boomendes interdisziplinäres Forschungsfeld, in dem sowohl qualitative als auch quantitative Methoden der empirischen Sozialforschung zum Einsatz kommen (Döring 2022). In zahlreichen Bereichen des öffentlichen Lebens werden Evaluationsmaßnahmen inzwischen rechtlich vorgeschrieben. So verlangt beispielsweise das Hochschulgesetz vieler Bundesländer die Evaluation von Lehrveranstaltungen und Studiengängen. Der Trend zu allgegenwärtigen Evaluationen stößt dabei auch auf Kritik („Evalutitis“). Mit eigenen Fachgesellschaften und Fachzeitschriften hat sich die Evaluationsforschung wissenschaftlich fest etabliert. Im deutschsprachigen Raum sind die *Gesellschaft für Evaluation (DeGEval; www.degeval.de)* sowie die *Zeitschrift für Evaluation (ZfEv)* einschlägig. Das vorliegende Handbuchkapitel grenzt zunächst die *Evaluations- von der Grundlagenforschung* ab, stellt zentrale *Begriffe der Evaluationsforschung* vor und unterscheidet wichtige *Typen von Evaluationsstudien*. Anschließend präsentiert es die für gute Evaluationsstudien bindenden *Evaluationsstandards*, erläutert verschiedene *Ansätze der Evaluationsforschung* und geht abschließend auf die einzelnen *Phasen einer Evaluationsstudie* ein.

---

## 11.2 Abgrenzung der Evaluations- von der Grundlagenforschung

Die sozialwissenschaftliche Methodenausbildung setzt üblicherweise bei der *akademischen Grundlagenforschung* (Przyborski/Wohlrab-Sah, Stein, Kelle, Kapitel 7, 8 und 9 in diesem Band) an: Ein selbst gewähltes und als wissenschaftlich relevant er-

achtetes Forschungsproblem wird mit dem Ziel des Erkenntnisgewinns eigenverantwortlich bearbeitet und gemäß wissenschaftlicher Gütekriterien der Fachöffentlichkeit präsentiert. Nehmen wir ein *Beispiel aus der Forschung zur sexuellen Arbeit*, die u. a. in Psychologie, Soziologie, Kriminologie, Sozialarbeit, Medizin empirisch untersucht wird (Döring, 2014). Vor dem Hintergrund der Beobachtung, dass sich Prostitutionswerbung, Kontaktanbahnung und sexuelle Dienstleistungen in den letzten Jahren stark in das Internet verlagert haben und hier auch Männer als Anbieter sichtbarer werden, könnte man der Forschungsfrage nachgehen, welche Merkmale und Motive Männer haben, die ihre sexuellen Dienste über Internet-Plattformen vermarkten und die man auf diesem Wege auch für Forschungsinterviews erreichen kann.

### 11.2.1 Evaluationsforschung als Anwendungs- und Auftragsforschung

Im Unterschied zu der primär wissenschaftlichen Kriterien verpflichteten Grundlagenforschung handelt es sich bei der Evaluationsforschung um *praxisnahe Anwendungs- und meist auch Auftragsforschung*. Themen werden also nicht selbst gewählt, sondern von den Auftraggebenden an die Evaluierenden herangetragen. Und die Ergebnisse müssen nicht nur wissenschaftlich überzeugen, sondern auch nützlich für die Praxis sein. Praxisbezogene Forschungsprobleme sind häufig umfassender und schwerer greifbar (da zahlreiche Betroffenen Gruppen und Kontextfaktoren eine Rolle spielen) als grundlagenwissenschaftliche Themen, die sich bei Bedarf stark eingrenzen lassen. So kann man bei einer grundlagenwissenschaftlichen Studie selbst entscheiden, ob man – wie im obigen Beispiel – lediglich männliche oder auch weibliche und Transgender-Personen untersucht und ob man sich auf den Status Quo bezieht oder Veränderungen über die Zeit erfasst usw.

Bei einer Evaluationsstudie dagegen geben die Auftraggebenden bestimmte Zielsetzungen und Rahmenbedingungen des Forschungsprozesses vor. Oft werden hier relativ umfangreiche Analysen und praktisch direkt anwendbare Resultate erwartet. So gab die deutsche Bundesregierung im Jahr 2007 – fünf Jahre nach Einführung des Prostitutionsgesetzes (ProstG) – eine Evaluationsstudie in Auftrag, die beurteilen sollte, ob das Gesetz die Lebens- und Arbeitsbedingungen von Prostituierten in Deutschland tatsächlich wie intendiert verbessert hatte (SOFFI K, 2007). Gemäß dem 2002 in Kraft getretenen ProstG gilt Prostitution in Deutschland nicht länger als sittenwidrig; dementsprechend haben Prostituierte ein Anrecht auf ihre Entlohnung und können am Kranken-, Sozialversicherungs- und Rentensystem teilnehmen. In der Begriffslogik quantitativer Sozialforschung ist hier also die Gesetzesänderung (Einführung des ProstG) die unabhängige Variable, deren Auswirkungen auf die Lebens- und Arbeitsbedingungen von Prostituierten als abhängige Variablen zu betrachten sind.

Eine umfassende Beantwortung der genannten Evaluationsfrage erfordert die Berücksichtigung unterschiedlicher Formen der Prostitution bzw. sexuellen Arbeit, un-

terschiedlicher Gruppen von Dienstleistenden sowie ihrer Angehörigen. Neben den Sichtweisen der unmittelbar Betroffenen sind auch Informationen von weiteren Anspruchsgruppen zu beachten (z. B. Polizei, Staatsanwaltschaften, Gerichte, Fachberatungsstellen, Sozialarbeit). Auch ist zu klären, was im Einzelnen als „Verbesserung“ der Lebens- und Arbeitsbedingungen gelten und empirisch erfasst werden kann (z. B. Gefühl der Entstigmatisierung und gesellschaftlichen Inklusion; Anzahl der offiziellen Anstellungsverhältnisse von Prostituierten; Anzahl und Art der neu abgeschlossenen Krankenversicherungsverträge; Häufigkeit von Problemen mit Kunden, die den vertraglich vereinbarten Preis nicht bezahlen). Gefordert ist in jedem Fall eine *Veränderungsmessung*, denn es geht um den Vergleich der Situation vor und nach der Einführung des neuen Gesetzes. Dabei muss jede festgestellte Veränderung der Situation der Prostituierten nach 2002 *kausal* eindeutig auf die Gesetzesänderung zurückführbar sein, damit sie zur Bewertung des Gesetzes dienen kann (andere Einflussfaktoren müssen also beispielsweise durch Berücksichtigung einer Kontrollgruppe ausgeschlossen werden). Die Erwartungen an eine aussagekräftige Evaluationsstudie sind typischerweise sehr hoch, können aber oft nur partiell eingelöst werden. Denn für eine kausal möglichst eindeutig interpretierbare Veränderungsmessung sind prospektive Längsschnittdesigns notwendig (Pötter/Prein und Schupp, Kapitel 52 und 85 in diesem Band). Diese sind sehr teuer und aufwändig – und eben definitionsgemäß im Nachhinein nicht zu realisieren: Ein im Jahr 2007 beauftragtes Evaluationsteam kann die Situation vor 2002 nur begrenzt empirisch rekonstruieren.

### 11.2.2 Besondere Anforderungen an Evaluationsforschende

Evaluationsstudien sind also oft wesentlich komplexer als Grundlagenstudien und meist nur in *interdisziplinärer Teamarbeit* zu bewältigen. Nicht zuletzt ist zu beachten, dass Evaluationsforschung als wissenschaftliche Auftragsforschung im Praxisfeld, von deren Ergebnissen weitreichende Entscheidungen abhängen können (z. B. Weiterführung einer Bildungsmaßnahme, Schließung einer sozialen Einrichtung, Änderung eines Gesetzes), besondere soziale Verantwortung verlangt und zahlreiche äußere und innere Konflikte mit sich bringen kann. So wurde u. a. mit Bezug auf die Evaluationsbefunde zum PostG im Jahr 2016 ein neues Gesetz, das Prostituiertenschutzgesetz (ProstSchG), verabschiedet, dessen Wirksamkeit wiederum rechtzeitig und umfassend evaluiert werden muss (Döring 2018). Die Forschenden sind einerseits *Dienstleistende für die Auftraggebenden* und sollen diesen mit ihren Evaluationsergebnissen bei Entscheidungen helfen. Andererseits sind sie *wissenschaftlicher Unabhängigkeit* verpflichtet und dürfen keine beliebigen Wunschergebnisse produzieren (etwa um von einem zufriedengestellten Auftraggeber dann ein Anschlussprojekt zu erhalten). Sie müssen eine Datenerhebung nach wissenschaftlichen Standards in die Wege leiten und umsetzen (auch wenn das im Praxisfeld Schwierigkeiten bereitet). Letztlich müssen sie auch unliebsame Ergebnisse in die Auswertung ein-

beziehen, sie in geeigneter Weise nach innen und außen kommunizieren und fair interpretieren.

Indem die Evaluationsforschung sich der wissenschaftlichen Bewertung praxisrelevanter und teilweise auch politisch sehr brisanter Untersuchungsgegenstände widmet, stellt sie an die Forschenden eine Reihe zusätzlicher Anforderungen: Wer das methodische Handwerkszeug der qualitativen und quantitativen Grundlagenforschung beherrscht, ist damit noch lange nicht in der Lage, eine gute Evaluationsstudie erfolgreich zu realisieren. Dafür sind vor allem spezifische *soziale und persönliche Kompetenzen* notwendig: Es geht beispielsweise darum, die Sichtweisen unterschiedlicher (mitunter zerstrittener) Betroffenengruppen überparteiisch einzubeziehen, überzogene Erwartungen von Auftraggebern zugunsten einer realistischen Evaluationsfragestellung zu korrigieren, dem bisweilen massiven sozialen oder politischen Druck standzuhalten, bestimmte Wunschergebnisse zu produzieren oder Negativbefunde zu verschweigen sowie mehr oder minder subtile Manipulations- und Vereinnahmungsversuche unterschiedlicher Interessengruppen zu erkennen und geeignet gegenzusteuern. Die DeGEval (2008) hat u. a. deswegen konkrete *Anforderungen an Evaluierende* formuliert und daraus auch *Empfehlungen für die Ausbildung* abgeleitet. Wesentlicher Baustein der Ausbildung ist die praktische Evaluationstätigkeit in allen Phasen des Evaluationsprozesses unter Anleitung langjährig erfahrener Evaluationsfachleute. Denn gerade die für Evaluationsforschung (im Unterschied zur Grundlagenforschung) essentiellen *sozial-kommunikativen* und *ethischen Problemlösungsstrategien*, die zudem passgenau auf den jeweiligen Evaluationskontext und dessen soziale Normen und rechtliche Rahmenbedingungen zuzuschneiden sind (z. B. Schule, Betrieb, Partei, Museum, Bordell), lassen sich allein theoretisch kaum vermitteln.

---

## 11.3 Zentrale Begriffe der Evaluationsforschung

Die Merkmale der Evaluationsforschung werden im Folgenden anhand zentraler Evaluationsbegriffe wie *Evaluationsgegenstand*, *Anspruchsgruppen*, *Evaluationskriterien*, *Evaluationsfunktionen* und *Evaluationsnutzung* verdeutlicht.

### 11.3.1 Evaluationsgegenstand

Während man in der Grundlagenforschung vom Untersuchungsgegenstand bzw. Forschungsobjekt (research object) spricht, ist in der Evaluationsforschung analog vom *Evaluationsgegenstand* bzw. *Evaluationsobjekt* oder auch vom *Evaluandum* (evaluation object, evaluand) die Rede. Das Spektrum der Evaluationsgegenstände ist dabei sehr breit und kann kaum trennscharf und erschöpfend dargestellt werden. Evaluieren werden Bildungseinrichtungen und Lehrveranstaltungen ebenso wie medizinische

und psychotherapeutische Behandlungsmethoden, Gesetze und Verordnungen, Bauvorhaben und Entwicklungshilfeprojekte, Methoden der Unternehmens- und Mitarbeiterführung, technologische Innovationen oder politische Kampagnen. Besonders typisch in der Evaluationsforschung ist die Bewertung von größer angelegten Programmen (z. B. Weiterbildungs-, Frauenförder-, Integrations-Programme), die auf bestimmte individuelle und kollektive Veränderungen abzielen, man spricht deswegen auch von *Programm-Evaluation* (program evaluation).

### 11.3.2 Anspruchsgruppen

Da die Evaluationsforschung auf eine wissenschaftlich fundierte Bewertung des jeweiligen Evaluationsgegenstandes hinausläuft, welche oft als Entscheidungsgrundlage dient und weitreichende praktische Konsequenzen haben kann, ist es wichtig, die Sichtweisen der *Betroffenen* (v. a. der Zielgruppen einer Maßnahme) und der *Beteiligten* (v. a. der Durchführenden einer Maßnahme) bzw. allgemein der zentralen *Anspruchsgruppen* (stakeholder) zu berücksichtigen. Wesentliche Informationen über den Evaluationsgegenstand sind oft nur über die Anspruchsgruppen zu erhalten. Zudem besteht ethisch der Anspruch, die Interessen der Betroffenen und Beteiligten zu berücksichtigen und nicht zu übergehen. Im Rahmen der Evaluationsforschung gilt es, für die jeweilige Evaluationsfrage die relevanten Stakeholder zu identifizieren und deren Anliegen angemessen in den Forschungsprozess einzubeziehen. Ein gravierender Mangel einer Evaluationsstudie besteht darin, wichtige Stakeholder zu übersehen. Bei der angesprochenen Evaluationsstudie zum Prostitutionsgesetz wären u. a. die Krankenkassen als Stakeholder einzubeziehen, etwa um zu eruieren, ob und unter welchen Konditionen sie Prostituierten unter dieser Berufsbezeichnung Versicherungsverträge anbieten (würden).

### 11.3.3 Evaluationskriterien

Ziel der Evaluationsforschung ist die wissenschaftlich fundierte Bewertung des Evaluationsgegenstandes. Dazu müssen klare und messbare Bewertungskriterien definiert, entsprechende Daten erhoben und interpretiert werden. Die *Auswahl und Festlegung der Evaluationskriterien* bestimmt maßgeblich das spätere Evaluationsergebnis. Kriterien sind unbrauchbar, wenn sie sich nicht empirisch erfassen lassen oder wenn sie inhaltlich nicht zur Lösung des Evaluationsproblems beitragen. Es ist zu beachten, dass neben der Definition der Evaluationskriterien auch noch die jeweiligen Zielwerte bzw. *Bewertungsmaßstäbe* festzulegen sind: Welche Ausprägungen des Evaluationskriteriums sollen als positives, ausreichendes oder negatives Ergebnis eingeordnet werden? Solche Maßstäbe können beispielsweise als *Mindest-, Regel- oder Maximalstandards* gefasst werden. Sie können empirisch relativ zu einem Aus-

gangs- oder anderen Vergleichszustand formuliert sein (z. B. Symptomreduktion im Vorher-Nachher-Vergleich um 10 %) oder auch als absolute Werte bzw. Idealnormen festgelegt werden (z. B. Symptomfreiheit aller Patienten als Behandlungsziel; siehe zu Vergleichskriterien Gollwitzer/Jäger 2014).

Wenn also bei der Evaluation des ProstG untersucht werden sollte, ob sich die Lebens- und Arbeitsbedingungen von Prostituierten verbessert haben, so mussten zunächst die relevanten Kriterien abgeleitet werden. Da Prostituierte bekanntermaßen stark unter gesellschaftlicher Ausgrenzung und Stigmatisierung leiden, könnte ein Evaluationskriterium die Erfahrung von Stigmatisierung sein. Da Stigmatisierung sehr belastend ist, könnte bereits ein geringer Effekt der Entstigmatisierung (z. B. von einem halben Punkt auf einer Stigmatisierungsskala) als Erfolg gewertet werden. Derartige Festlegungen von Evaluationskriterien und Bewertungsmaßstäben sind auf der Basis des aktuellen Forschungsstandes sowie mit Rückgriff auf Experten- und Betroffenenurteile detailliert zu begründen.

### 11.3.4 Evaluationsfunktionen

Die im Zuge einer Evaluation erarbeitete Bewertung des Evaluationsgegenstandes anhand unterschiedlicher Evaluationskriterien (und Bewertungsmaßstäbe) soll in der Praxis verschiedenen Funktionen, Zwecken bzw. Zielen dienen. Die *Evaluationsfunktionen* (bzw. *Evaluationszwecke* oder *Evaluationsziele*) werden in der Literatur zwar unterschiedlich gegliedert (Stockmann 2002: 3 ff., Widmer/De Rocchi 2012: 27 f.), sie hängen aber eng zusammen. Es lassen sich pragmatisch fünf Funktionen abgrenzen:

- *Erkenntnisfunktion*: Evaluationsforschung trägt definitionsgemäß dazu bei, wissenschaftliche Erkenntnisse über die Eigenschaften und Wirkungen von Evaluationsgegenständen zu sammeln.
- *Lern- und Dialogfunktion*: Im Zuge der Evaluationsforschung vollziehen die Beteiligten Lernprozesse und es treten verschiedene Anspruchsgruppen miteinander in Dialog. Eine Evaluationsstudie kann die Funktion haben, einen solchen Dialog über den Evaluationsgegenstand zu fördern und fruchtbar zu gestalten.
- *Optimierungsfunktion*: Über die reine Wissenserweiterung und damit verbundene Lern- und Dialogprozesse hinaus soll Evaluation oft Erkenntnisse liefern, die zur zielgerichteten Verbesserung des Evaluationsgegenstandes beitragen können.
- *Entscheidungsfunktion*: Evaluationsergebnisse bilden nicht selten die Basis für Wahlentscheidungen zwischen verschiedenen Evaluationsgegenständen (soll Maßnahme A oder B umgesetzt werden?) oder auch für Entscheidungen über die Implementierung oder Weiterführung einer Maßnahme (lohnt es sich, die Maßnahme fortzusetzen oder sollte sie abgebrochen werden?). Damit eine Evaluationsstudie die Entscheidungsfunktion erfüllt, muss das Evaluationsteam die Evaluationsergebnisse zu klaren Praxis-Empfehlungen verdichten. Die Entscheidung

selbst wird indessen nicht vom Evaluationsteam getroffen. Es liefert im Zuge der unabhängigen wissenschaftlichen Bewertung des Evaluationsgegenstandes allenfalls eine Entscheidungshilfe.

- *Legitimationsfunktion*: Sowohl die Durchführung von Evaluationsforschung als auch ihre Befunde sollen dazu beitragen, die Entwicklung und Durchführung einer Intervention zu kontrollieren und nach außen zu legitimieren. Vor allem bei öffentlich geförderten Programmen und Maßnahmen dient regelmäßige Evaluation dazu, über die sachgerechte Verwendung öffentlicher Gelder Rechenschaft abzulegen.

### 11.3.5 Evaluationsnutzung

Verantwortung im Hinblick auf die Evaluationsnutzung (evaluation use; Patton/Campbell-Patton 2021) zu übernehmen bedeutet für die Evaluierenden zunächst, ihre Ergebnisse in *Zwischen- und Abschlussberichten* verständlich darzustellen und sich vor eindeutigen Bewertungen und daraus abgeleiteten Praxisempfehlungen nicht zu scheuen. Anders als in der Grundlagenforschung, in der zu Recht vor überinterpretierten Ergebnissen zu warnen ist, kann die Evaluationsforschung ihrer Ratgeberpflicht nur nachkommen, wenn sie ihre Ergebnisse zu praxisbezogenen Empfehlungen verdichtet. So ist in den vorliegenden Evaluationsberichten zum ProstG (SOFFI K 2007, Renzikowski 2007, Kavemann/Steffan 2013) die Einschätzung zu finden, dass Entkriminalisierung der richtige Weg ist, um die Arbeits- und Lebensbedingungen von freiwilligen Prostituierten zu verbessern, ohne die Verfolgung von Zwangsprostitution zu erschweren. Zudem wird in den Berichten diskutiert, welche zusätzlichen rechtlichen Änderungen (z. B. im Steuer-, Gewerbe- und Baurecht) für eine Gleichstellung notwendig wären und welche sozialen Interventionen (z. B. Ausstiegshilfen) sinnvoll sind.

Evaluationsergebnisse und darauf aufbauende Praxisempfehlungen in Berichten darzustellen, reicht für eine erfolgreiche Nutzung der Evaluationsergebnisse im Sinne der genannten Optimierungs- und Entscheidungsfunktion aber meist nicht aus. Denn oft kann nicht davon ausgegangen werden, dass Evaluationsberichte auch wirklich gelesen, verstanden und akzeptiert werden und dann auch noch zur Ableitung und Umsetzung der richtigen Konsequenzen führen. Die Förderung der Evaluationsnutzung umfasst deswegen weitere Aktivitäten wie *Beratung, Coaching, Workshops, Weiterbildungskurse* usw., in denen Evaluierende in der Ratgeberrolle Bedeutung und praktische Konsequenzen der Evaluationsergebnisse mit den intendierten Evaluationsnutzenden (also denjenigen Personen und Organisationen, die aus den Evaluationsergebnissen praktische Konsequenzen ziehen sollen) behandeln.



## 11.4 Wichtige Typen von Evaluationsstudien

Um Evaluationsstudien griffig zu kennzeichnen, werden sie als *formative* oder *summative Evaluationen*, als *interne* oder *externe Evaluationen*, als *Konzept-, Prozess- oder Ergebnisevaluationen* bezeichnet. Damit sind unterschiedliche Aspekte einer Evaluationsstudie angesprochen. Größere Evaluationsprojekte kombinieren nicht selten mehrere dieser Typen miteinander, um umfassendere Erkenntnisse zu generieren.

### 11.4.1 Formative und summative Evaluation

Eine *formative Evaluation* liefert Ergebnisse, die der schrittweisen Verbesserung des Evaluationsgegenstandes dienen sollen (oft wird mit qualitativen Methoden gearbeitet: z. B. offene Befragung zu Verbesserungsvorschlägen). Eine *summative Evaluation* verfolgt das Ziel einer abschließenden Beurteilung (hier sind quantitative Methoden einschlägig: z. B. Bewertung auf einer Schulnotenskala). Im Hinblick auf die Evaluationsfunktionen erfüllt eine formative Evaluation eher eine Lern- und Dialog- sowie Optimierungsfunktion, eine summative Evaluation dagegen eine Entscheidungs- und/oder Legitimationsfunktion. Sowohl formative als auch summative Evaluation erfüllen immer auch eine Erkenntnisfunktion.

### 11.4.2 Interne und externe Evaluation

Bei der *internen Evaluation* wird die Evaluationsforschung von Personen durchgeführt, die derselben Organisation angehören (z. B. die Lehrveranstaltungen und Lehrenden einer Universität werden unter Leitung des universitätseigenen Didaktik-Zentrums evaluiert). Ein Sonderfall der internen Evaluation ist die *Selbstevaluation*, hier organisieren diejenigen, die für den Evaluationsgegenstand verantwortlich sind, auch selbst die Evaluation (z. B. eine Dozentin nutzt für ihre eigenen Lehrveranstaltungen ein spezielles Evaluationsverfahren). Bei der *externen Evaluation* werden Außenstehende tätig (z. B. ein Studiengang einer Hochschule wird durch ein Expertengremium einer anderen Hochschule evaluiert). Die interne Evaluation hat den Vorteil, dass den Beteiligten vor Ort meist mehr detaillierte Informationen über den Evaluationsgegenstand zugänglich sind, die externe Evaluatoren kaum erlangen können. Sie hat gleichzeitig den Nachteil, dass sie eher als parteiisch eingeschätzt wird, die externe Evaluation dagegen als neutraler.



### 11.4.3 Konzept-, Prozess- und Ergebnisevaluation

Ein Evaluationsgegenstand kann in unterschiedlichen Phasen seines Lebenszyklus evaluiert werden. Eine *Konzeptevaluation* bewertet das Konzept einer Maßnahme bereits im Vorfeld einer möglichen Umsetzung. Die *Prozessevaluation* bewertet die Durchführung einer laufenden Maßnahme. Die *Ergebnisevaluation* schließlich konzentriert sich auf die Bewertung der Effekte der Maßnahme. Prozess- und Ergebnisevaluation spielen in der Praxis eine so große Rolle, dass für sie umfassendere Evaluationsansätze entwickelt wurden (siehe Abschn. 11.6.1 für Modelle der Ergebnisevaluation und 11.6.2 für Modelle der Prozessevaluation).

---

## 11.5 Evaluationsstandards

Woran ist eine gute Evaluationsstudie zu erkennen und von einer durchschnittlichen oder gar mangelhaften Studie abzugrenzen? Gute Evaluationsstudien müssen nach aktuellem nationalem und internationalem Diskussionsstand in der Wissenschaftsgemeinschaft vier grundlegende Eigenschaften aufweisen (DeGEval 2016: S. 34, 39, 41, 44):

- *Nützlichkeit* (utility): „Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an abgestimmten und geklärten Evaluationszwecken sowie soweit möglich am Informationsbedarf der vorgesehenen Nutzerinnen und Nutzer ausrichtet.“
- *Durchführbarkeit* (feasibility): „Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird.“
- *Fairness* (propriety): „Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit allen betroffenen Personen und Gruppen umgegangen wird.“
- *Genauigkeit* (accuracy): „Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige und nachvollziehbare Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt. In diesem Zusammenhang kommt wissenschaftlichen Gütekriterien eine besondere Bedeutung zu.“

Den vier Qualitätsdimensionen sind insgesamt 25 *Einzelstandards* zugeordnet. Wer selbst eine Evaluationsstudie durchführt, sollte bereits in der Planungsphase prüfen, ob und wie das avisierte Vorgehen die Evaluationsstandards erfüllt, im Zweifelsfall ist das Studienkonzept anzupassen.

Wie jede wissenschaftliche Studie ist auch eine Evaluationsstudie in ihrem Ablauf sorgfältig zu dokumentieren (z. B. Anfertigung und Archivierung von Protokollen

aller Stakeholder-Meetings), so dass das Vorgehen im Nachhinein auch für Außenstehende nachvollziehbar und nachprüfbar ist. Zuweilen ist es gewünscht oder erforderlich, eine Evaluationsstudie ihrerseits zu evaluieren (z. B. wenn Auftraggebende Zweifel an der Qualität der Studie anmelden und der Meinung sind, sie sei den vertraglich vereinbarten Preis nicht wert). Die systematische Bewertung einer Evaluationsstudie im Hinblick auf die Erfüllung der Evaluationsstandards wird als *Meta-Evaluation* bezeichnet und typischerweise als externe Evaluation durchgeführt: Fachkollegen bekommen alle Unterlagen zu der betreffenden Evaluationsstudie, halten ggf. Rücksprache mit den Beteiligten und/oder sammeln anderweitig weitere Daten, um die Qualität der Evaluationsstudie zu beurteilen. Eine Meta-Evaluation muss ihrerseits Evaluationsstandards genügen.

---

## 11.6 Verschiedene Ansätze der Evaluationsforschung

Um den vielfältigen praxisbezogenen Anforderungen gerecht zu werden, die an Evaluationsstudien gestellt werden, sind in den letzten Jahren eine Reihe von unterschiedlichen *Evaluationsmodellen* vorgeschlagen und erprobt worden. Die einzelnen Evaluationsmodelle unterscheiden sich in der Ausrichtung des methodischen Vorgehens und im Evaluationsverständnis. Grob lassen sich die einzelnen Modelle vier *Evaluationsansätzen* zuordnen (Döring 2022, Kap. 18). Es ist zu beachten, dass die verschiedenen Evaluationsansätze und -modelle wichtige Hilfestellungen zur Durchführung von Evaluationsstudien liefern, allerdings nicht als starre Vorschriften zu verstehen sind. Evaluationsforschende gehen meistens nicht streng modellkonform, sondern *eklektisch* vor, d. h. sie kombinieren Elemente aus unterschiedlichen Evaluationsansätzen miteinander, um das vorliegende Evaluationsproblem optimal bearbeiten zu können.

### 11.6.1 Der ergebnisorientierte Evaluationsansatz

Der ergebnisorientierte Evaluationsansatz konzentriert sich auf einige wenige ergebnisbezogene Merkmale des Evaluationsgegenstandes.

- Bei einer *Wirksamkeitsevaluation* geht es darum, ob und wie stark eine Maßnahme ihre intendierten Wirkungen tatsächlich erzeugt (*Effektivität*). Wichtig ist dabei, dass die gemessenen Wirkungen wirklich kausal auf das Evaluationsobjekt zurückgeführt werden können, was methodisch am eindeutigsten mit experimentellen oder quasi-experimentellen Kontrollgruppenstudien gelingt. So muss die Wirksamkeitsevaluation eines neuen Kopfschmerzmedikaments nachweisen, dass das Medikament besser wirkt als ein Placebo.

- Bei der *ökonomischen Evaluation* setzt man den Nutzen der Maßnahme mit den anfallenden Kosten in Relation (*Effizienz*). Die ökonomische Evaluation steht vor der Aufgabe, sowohl den Nutzen des Evaluationsgegenstandes als auch seine Kosten möglichst exakt zu quantifizieren, am besten in Geldwerteinheiten.

### 11.6.2 Der systemische Evaluationsansatz

In Abgrenzung zu Evaluationen, die sich auf einzelne Merkmale einer Maßnahme – vor allem Ergebnismerkmale der Effektivität und Effizienz – konzentrieren, sind unter dem systemischen Ansatz hier solche Evaluationsmodelle angesprochen, die eine Maßnahme *ganzheitlicher* als die ergebnisorientierten Modelle betrachten: Der systemische Ansatz untersucht den Evaluationsgegenstand also nicht nur hinsichtlich seiner Wirkungen, sondern auch im Hinblick auf Prozesse seiner Konzeption, Umsetzung und Nutzung sowie im Hinblick auf die Bedingungen des Umfeldes.

Das prominenteste Evaluationsmodell im systemischen Ansatz ist das bereits in den 1960er Jahren entwickelte *CIPP-Evaluations-Modell von Daniel Stufflebeam* (Stufflebeam et al. 1971), das aus den vier Komponenten Context, Input, Process und Product Evaluation besteht. Zur Durchführung von Evaluationen nach dem CIPP-Modell liegt eine umfassende Evaluations-Checkliste vor ([www.wmich.edu/evaluation/checklists](http://www.wmich.edu/evaluation/checklists)). Eine konkrete Evaluationsstudie im CIPP-Modell kann dabei stets nur eine kleine Auswahl aus dem großen Spektrum möglicher Evaluationsfragen bearbeiten:

- Context
  - Welche Personengruppen haben Bedarf an einer geplanten (oder bereits realisierten) Intervention?
  - Welche vergleichbaren Maßnahmen existieren bereits?
  - Wie lassen sich Mittel zur kurz- und langfristigen Finanzierung der Intervention beschaffen?
- Input
  - Wie gut entspricht das Konzept der Intervention den Bedürfnissen und Besonderheiten der Zielgruppen?
  - Wie berücksichtigt das Konzept mögliche Widerstände gegen seine Umsetzung?
  - Wie viel Personal steht zur Umsetzung der Maßnahme zu welchen Zeitpunkten zur Verfügung und welche Qualifikationen bringt es mit?
- Process
  - In welchen Situationen bzw. an welchen Orten wird die Intervention wie intensiv genutzt?
  - Welche Probleme treten bei der Umsetzung der Intervention auf?
  - Wie kann die Intervention verbessert werden?

- Product
  - Wie groß ist die Nutzergruppe der Intervention? In welchem Verhältnis steht die aktuelle Nutzergruppe zur Zielgruppe? (*Impact*) Spezieller: Ist die Intervention Menschen mit körperlichen Beeinträchtigungen zugänglich (*Barrierefreiheit*)? Ist die Intervention für Menschen mit unterschiedlichem kulturellen Hintergrund gleichermaßen nutzbar (*Internationalisierung*)?
  - Erreicht die Intervention die intendierten Wirkungen? In welchem Umfang werden die intendierten Wirkungen erzielt? (*Effektivität*) Spezieller: Gibt es differenzielle Wirkungen in unterschiedlichen Gruppen? Profitieren Frauen und Männer gleichermaßen von der Intervention, oder wird eine Gruppe benachteiligt (*Gender Mainstreaming*)?
  - In welchem Verhältnis stehen Kosten und Nutzen der Intervention? Erreicht die Intervention erwünschte Ziele kostengünstiger als andere Interventionen? (*Effizienz*)
  - Welche Vorkehrungen werden getroffen, um die Intervention an anderen Orten/für andere Zielgruppen anzubieten? (*Portabilität*)
  - Welche Vorkehrungen werden getroffen, um die Intervention langfristig oder dauerhaft zu erhalten? Welchen Status hat die Intervention z. B. zwei Jahre nach Abschluss des ersten Evaluationsprojektes? (*Nachhaltigkeit*)

### 11.6.3 Der theorieorientierte Evaluationsansatz

Bei der so genannten theorieorientierten Evaluation wird über die Frage der reinen Kausalität (die in der wirksamkeitsorientierten Evaluation untersucht wird) hinausgegangen: Man versucht, die Wirkungsmechanismen zu rekonstruieren: Welche Interventionskomponenten funktionieren für welche Zielgruppe unter welchen Umständen *wie* und *warum*? Es geht also darum, die Wirkprozesse des Evaluationsgegenstandes theoretisch zu modellieren (Programm-Theorie) und dann empirisch zu prüfen. Die *theoriegeleitete Evaluation* (theory-driven evaluation; Chen 1990, 2015) ist ein bekanntes Evaluationsmodell innerhalb dieses Ansatzes. Es besteht aus einem theoretischen *Aktionsmodell* (das genau beschreibt und erklärt, wie die Maßnahme funktioniert) und einem *Veränderungsmodell* (das genau beschreibt und erklärt, wie und wodurch die Maßnahme die gewünschten Wirkungen erzeugt).

### 11.6.4 Der aktorsorientierte Evaluationsansatz

Der aktorsorientierte Evaluationsansatz (Widmer/De Rocchi 2012) schließlich ist dadurch gekennzeichnet, dass seine Forschungsaktivitäten in den Dienst einzelner Stakeholder-Gruppen gestellt werden (klientenorientierte Modelle) oder alle Stakeholder gleichermaßen aktiv einbeziehen sollen, bis hin zur Steuerung der Evaluation durch die Evaluierten selbst (stakeholderorientierte Modelle).

- *Klientenorientierte Evaluationsmodelle*: Die Evaluierenden stellen hier ihre Dienste primär auf eine ausgewählte Stakeholder-Gruppe ab. Seien es die Auftraggebenden (managementorientierte Evaluation), seien es sozial unterprivilegierte Bevölkerungsgruppen, deren Situation durch eine Maßnahme verbessert werden soll (anwaltschaftliche Evaluation) oder seien es die Maßnahmenverantwortlichen, die Evaluationsergebnisse zur Veränderung nutzen sollen (nutzungsfokussierte Evaluation; Patton/Campbell-Patton 2021).
- *Stakeholderorientierte Evaluationsmodelle*: Während sich die klientenorientierten Evaluationsmodelle jeweils primär in den Dienst einzelner Stakeholdergruppen stellen, verfolgen stakeholderorientierte Modelle das Ziel, alle Beteiligten und Betroffenen gleichermaßen so aktiv wie möglich in den Evaluationsforschungsprozess zu involvieren. Die *Partizipative Evaluation* (participatory evaluation) bezieht die Beteiligten und Betroffenen nicht nur als Informanten, sondern auch als *Forschungspartner* in das Evaluationsprojekt mit ein, etwa durch deren Mitwirkung an der Ausformulierung von Evaluationsfragen und/oder der Planung des Untersuchungsdesigns. Die *Empowerment-Evaluation* (empowerment evaluation) geht noch einen Schritt weiter als die Partizipative Evaluation: Die Stakeholder werden nicht nur zu Forschungspartnern der Evaluierenden, sondern sollen die Evaluation gänzlich selbst in die Hand nehmen, die Evaluierenden agieren hier von Anfang an nur beratend. Stakeholderorientierte Evaluationsmodelle sind weniger gebräuchlich als klientenorientierte Modelle. Denn zum einen besteht oft das Problem, dass nicht genügend Maßnahmen-Betroffene zu der zeitintensiven und anstrengenden Arbeit an der Evaluationsforschung bereit oder in der Lage sind. Zum anderen wird die wissenschaftliche Qualität von Evaluationsstudien in Frage gestellt, deren Leitung (wie im Falle der Empowerment-Evaluation) ausdrücklich nicht in den Händen von Evaluationsfachleuten liegt.

---

## 11.7 Phasen einer Evaluationsstudie

Generell folgt eine wissenschaftliche Evaluationsstudie einerseits den typischen *Phasen des empirischen Forschungsprozesses* (mit entsprechenden Varianten, je nachdem, ob ein quantitativer, qualitativer oder Mixed-Methods-Forschungsprozess ansteht) und kann andererseits aus Projektmanagement-Perspektive in *Projektphasen* ein-

geteilt werden. Im Folgenden wird ein Fünf-Phasen-Modell vorgestellt, wobei jeweils die Besonderheiten der Evaluationsforschung herausgestellt werden (Döring 2022, Kap. 18).

### 11.7.1 Initiierungsphase der Evaluation

In der Initiierungsphase geht es aus Sicht der Evaluierenden darum, *potenzielle Evaluationsprojekte zu akquirieren*, indem Evaluationsmöglichkeiten identifiziert (z. B. über öffentliche Ausschreibungen von Evaluationsprojekten oder über Kontakte zu den Auftraggebern) und deren Evaluationsziele kritisch beleuchtet werden (z. B. hinsichtlich Aufwand, eigener Expertise oder politischer und ethischer Brisanz). Da Evaluation meist Auftragsforschung ist, müssen sich Evaluierende darum bemühen, entsprechende Aufträge zu erhalten. Umgekehrt geht es für die Auftraggeber darum, für ihr jeweiliges Evaluationsanliegen ein geeignetes Evaluationsteam zu finden.

### 11.7.2 Konzeptionsphase der Evaluation

Wurden mögliche Evaluationsprojekte identifiziert, deren Zielsetzungen machbar und verantwortbar erscheinen, so müssen im nächsten Schritt die Evaluationsbedingungen genauer exploriert und das *Evaluationskonzept* bzw. *Projektangebot* ausgearbeitet werden. Dabei ist es notwendig sicherzustellen, dass die gängigen Evaluationsstandards von Anfang an und über den weiteren Evaluationsprozess hinweg eingehalten, das Projektmanagement vorbereitet (Zusammensetzung des Evaluationsteams, Zeit- und Kostenplan) und verbindliche Vereinbarungen zwischen Evaluierenden und Auftraggebern getroffen werden. Erhält ein Projektangebot den Zuschlag oder nimmt man einen Projektauftrag an, so folgt der Abschluss des *Evaluationsvertrags* (evaluation contract), der die Rechte und Pflichten von Auftraggebern und Auftragnehmern fixiert. Insbesondere bei umfangreichen Projekten sind die rechtlichen Bedingungen genau zu prüfen, etwa die Zahlungsmodalitäten, Möglichkeiten des Vertragsrücktritts oder Konsequenzen bei Nichteinhaltung einzelner vertraglicher Regelungen (für Checklisten zur Vorbereitung von Evaluationsverträgen siehe Kuster et al. 2011; Stufflebeam/Coryn 2014).

### 11.7.3 Planungsphase der Evaluation

Nachdem der Evaluationsvertrag abgeschlossen wurde, stehen die *Detailplanung* und konkrete Vorbereitung der Evaluationsstudie an. Dies umfasst die Formulierung von detaillierten Evaluationsfragen bzw. -hypothesen, die Untersuchungsplanung inklu-

sive Untersuchungsdesign, Stichprobenplänen und Instrumenten sowie die organisatorische Projektplanung (z. B. Rekrutierung des Evaluationsteams). Im Sinne der Evaluationsstandards der Nützlichkeit und Durchführbarkeit ist in der Planungsphase wiederum ein enger Kontakt zu den Stakeholdern notwendig, ebenso eine enge Rückkopplung mit den Abläufen im Praxisfeld (z. B. Identifikation geeigneter Zeitpunkte und Orte für Datenerhebungen). Andernfalls besteht die Gefahr, dass „am grünen Tisch“ ein Studienentwurf ausgearbeitet wird, der forschungspraktisch nicht umsetzbar ist und/oder am Widerstand derjenigen Stakeholder scheitert, die nicht von Anfang an genügend einbezogen wurden.

#### 11.7.4 Realisierungsphase der Evaluation

Auf der Basis der Ergebnisse der Planungsphase werden in der Realisierungs- oder Durchführungsphase die für die Evaluationsziele notwendigen *Evaluationsdaten erhoben, analysiert und interpretiert*, zudem wird der operative Projektablauf durch ein Projektcontrolling begleitet. Je nach Umfang und Dauer einer Intervention sowie der aktuellen Phase im Interventionslebenszyklus werden Erhebungs-, Analyse- und Interpretationsschritte mehrfach nacheinander oder auch parallel stattfinden, wobei es im Sinne eines *iterativen Vorgehens* oft Rücksprünge in die Planungsphase gibt.

Aus dem Satz der Evaluationsstandards sind bei der Datenerhebung vor allem Standards der Durchführbarkeit und Fairness relevant, während bei der Datenauswertung und Dateninterpretation der Schwerpunkt auf Standards der Genauigkeit liegt (DeGEval 2016).

#### 11.7.5 Abschlussphase der Evaluation

Die in der Realisierungsphase der Evaluation gewonnenen Ergebnisse gilt es in der Abschlussphase an unterschiedliche Stakeholder der Intervention zu *kommunizieren* und ihre *praktische Umsetzung* durch die intendierten Evaluationsnutzenden vorzubereiten. Auch eine wissenschaftliche Bewertung der Evaluation in Form von systematischer *Meta-Evaluation* kann sinnvoll oder notwendig sein (z. B. wenn Auftraggebende oder andere Stakeholder die Qualität der Studie anzweifeln). Schließlich ist es wünschenswert, Evaluationsstudien zu *publizieren*, um sie der Fachcommunity vorzustellen und kritisch diskutieren zu können. Dabei geht es zum einen um kumulativen Erkenntnisgewinn über den jeweiligen Evaluationsgegenstand und zum anderen um die methodische Weiterentwicklung des Feldes der Evaluationsforschung.



## Literatur

- Chen, Huey-tsyh (1990): *Theory-Driven Evaluations*. London: Sage
- Chen, Huey-tsyh (2015): *Practical Program Evaluation: Theory-Driven Evaluation and the Integrated Evaluation Perspective* (2. Auflage). London: Sage
- DeGEval (2008): *Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile an Evaluatorinnen und Evaluatoren* (2. Aufl.). Mainz: DeGEval
- DeGEval (2016): *Standards für Evaluation. Erste Revision 2016* (1. Aufl.). Mainz: DeGEval
- Döring, Nicola (2014): *Prostitution in Deutschland: Eckdaten und Veränderungen durch das Internet*. *Zeitschrift für Sexualforschung* 2: 99–137
- Döring, Nicola (2018): *Das neue Prostituiertenschutzgesetz: Wie ist es aus fachlichen Perspektiven zu beurteilen?* *Zeitschrift für Sexualforschung* 1: 44–56
- Döring, Nicola (2022): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (6. vollst. überarb., aktual. und erw. Aufl.). Heidelberg: Springer
- Gollwitzer, Mario/Jäger, Reinhold S. (2014): *Evaluation kompakt* (2. Auflage). Weinheim: Beltz
- Kavemann, Barbara/Steffan, Elfriede (2013): *Zehn Jahre Prostitutionsgesetz und die Kontroverse um die Auswirkungen*. *Aus Politik und Zeitgeschichte* 9: 9–15
- Kuster, Jürg/Huber, Eugen/Lippmann, Robert/Schmid, Alphons/Schneider, Emil/Witschi, Urs/Wüst, Roger (2011): *Handbuch Projektmanagement* (3. erw. Aufl.). Berlin: Springer
- Patton, Michael Q./Campbell-Patton, Charmagne E. (2021): *Utilization-Focused Evaluation: The New Century Text* (5. Auflage). Thousand Oaks: Sage
- Renzikowski, Joachim (2007): *Reglementierung von Prostitution. Ziele und Probleme: Eine kritische Betrachtung des Prostitutionsgesetzes*. Berlin: Bundesministerium für Familie, Senioren, Frauen und Jugend
- SOFFI K (Sozialwissenschaftliches Frauenforschungsinstitut, Freiburg Kontaktstelle praxisorientierte Forschung der Evangelischen Fachhochschule Freiburg) (2007): *Untersuchung „Auswirkung des Prostitutionsgesetzes“*. Berlin: Bundesministerium für Familie, Senioren, Frauen und Jugend
- Stockmann, Reinhard (2002): *Was ist eine gute Evaluation? Ceval-Arbeitspapiere* (Vol. 9). Saarbrücken: Centrum für Evaluation
- Stufflebeam, Daniel L./Coryn, Chris L. S. (2014): *Evaluation Theory, Models, and Applications* (2. Auflage). San Francisco: Wiley
- Stufflebeam, Daniel L./Foley, Walter J./Gephart, W. J./Guba, Egon G./Hammond, Robert L./Marriman, Howard O. et al. (1971): *Educational Evaluation and Decision Making*. Itasca: Peacock
- Widmer, Thomas/De Rocchi, Thomas (2012): *Evaluation. Grundlagen, Ansätze und Anwendungen*. Zürich: Rüegger

**Nicola Döring** ist Professorin für Medienpsychologie und Medienkonzeption an der Technischen Universität Ilmenau. *Ausgewählte Publikationen:* Forschungsmethoden und Evaluation für die Sozial- und Humanwissenschaften, 6. Auflage. Heidelberg: Springer (2022); Zur Operationalisierung von Geschlecht im Fragebogen: Probleme und Lösungsansätze aus Sicht von Mess-, Umfrage-, Gender- und Queer-Theorie, in: Gender – Zeitschrift für Geschlecht, Kultur und Gesellschaft 2 (2013); Mobile Market Research/Mobile Marktforschung. Köln: Herbert von Halem (zusammen mit Emanuel Maxl und Astrid Walisch, 2009). *Webseite:* [www.nicola-doering.de](http://www.nicola-doering.de). *Kontaktadresse:* [nicola.doering@tu-ilmenau.de](mailto:nicola.doering@tu-ilmenau.de).