

KOMWEID-Impulse

Kurzeinführung in die Grundlagen der Evaluationsforschung und -praxis

Benjamin Ditzel
Hochschule für Angewandte Wissenschaften Hamburg,
Projekt KOMWEID

KOMWEID-Impulse, Jahrgang 2022, Nr. 11, November 2022,
www.haw-hamburg.de/hochschule/qualitaet-in-der-lehre/komweid/impulse

Abstract

Diese Handreichung richtet sich als ein erster konzeptioneller Einstieg an alle, die sich mit der Evaluation ihrer Handlungspraxis – ob Lehrveranstaltungen, Studiengänge, Unterstützungsangebote oder Projekte – auseinandersetzen möchten. Es werden wichtige Begriffe der Evaluationspraxis und -forschung vorgestellt. Ziel ist es, ein Verständnis dafür zu entwickeln, was Evaluation bedeutet, was sie von Alltagsformen der Bewertung abgrenzt und welche grundsätzlichen Möglichkeiten und Ansätze der Evaluation existieren.

Inhaltsübersicht

1	Was bedeutet Evaluation? – Merkmale von Evaluation	1
2	Was wird evaluiert? – Objekte & Gegenstandsbereiche der Evaluation.....	3
3	Warum wird evaluiert? – Funktionslogiken von Evaluation.....	5
4	Wann wird evaluiert? – prospektive, formative und summative Evaluation.....	8
5	Wer evaluiert? – Selbst- und Fremdevaluation	9
6	Wie wird evaluiert? – Ansätze der Evaluation.....	10
7	Kurzzusammenfassung.....	15
8	Aufgaben für das Selbststudium.....	16
9	Weiterführende Materialien und Literatur	16

Im Zentrum des Anliegens einer wissenschaftsgeleiteten Wirkungsreflexion steht die Frage danach, wie sich die Wirkung, Wirkungsweise und Wirksamkeit einer spezifischen Handlungspraxis als Ausgangspunkt für deren Reflexion und Weiterentwicklung analysieren lassen. Den primären Bezugspunkt für eine solche Betrachtung liefert die Evaluations- und Wirkungsforschung. Deren Erkenntnisinteresse richtet sich darauf, „*welche Wirkungen durch gezielte Interventionen unter gegebenen Rahmenbedingungen*“ erreicht werden (Stockmann/Meyer 2014, S. 21).

Diese Handreichung dient als eine erste Grundlage, sich mit der Frage auseinanderzusetzen, was Evaluation bedeutet. Denn Evaluation ist ein weites Feld; es existieren sehr unterschiedliche Möglichkeiten der Fokussierung und der methodischen Ausgestaltung. Ziel dieser Handreichung ist es, ausgehend von einer Bezugnahme auf die evaluationstheoretische und -praktische Literatur eine Sprachfähigkeit hinsichtlich grundlegender Konzepte und Begriffe der Evaluation zu entwickeln. Es geht darum zu klären, was Evaluation bedeutet (und was nicht), wie sie grundsätzlich ausgestaltet werden kann, worauf sie sich beziehen und welche Aspekte sie in den Blick nehmen kann. Dies wird entlang von allgemein relevanten Fragestellungen bzw. Dimensionen erläutert. Wie diese Fragen und Dimensionen jeweils konkret ausbuchstabiert werden können, wird schrittweise in einer eigenen Handreichung zu den Verfahrensschritten erläutert.

1

Was bedeutet Evaluation? – Merkmale von Evaluation

Um Evaluation von einem alltagsprachlichen Gebrauch des Begriffs abzugrenzen, ist es üblich, drei Bestimmungsmomente zur Beschreibung dessen heranzuziehen, was Evaluation in einem wissenschaftlichen Sinne ausmacht.

Evaluation als Bewertungspraxis

Zunächst sollen mit Evaluation Daten generiert werden, die eine Aussage darüber erlauben, „*ob und unter welchen Bedingungen*“ eine Intervention, ein Projekt oder ein Programm erfolgreich ist (Altfeld et al. 2015, S. 57). Es geht damit explizit darum, Merkmale, Strukturen und/oder Effekte/Wirkungen einer Intervention zu bewerten (Meyer 2017, S. 519), also zu evaluieren.

Wichtig ist dabei, dass diese Bewertung einer normativen Komponente bedarf, die es überhaupt erst ermöglicht, zu einem Werturteil zu gelangen. Zwar wird in der evaluationstheoretischen Literatur vielfach behauptet, dass mit einer Evaluation objektive und wertneutrale Aussagen über ein Evaluationsobjekt generiert werden (z. B. Kromrey 2001, S. 114). Doch wird dabei eher darauf rekurriert, dass die Werturteile auf eine systematische, weil durch wissenschaftliche Methoden abgesicherte Methode erfolgen. In Abgrenzung zu Alltagsurteilen handelt es sich somit um wissenschaftlich fundierte Bewertungen.

Und doch zeichnet sich diese Evaluationsmethodik in Form eines Vergleichs beobachteter Zustände mit zuvor festgelegten Bewertungskriterien (Kardorff/Schönberger 2010, S. 268; Wolbring 2016, S. 58) durch Selektionen aus einer Vielzahl möglicher Beobachtungen, Zeitpunkte und Referenzpunkte aus. Die mit einer Evaluation herausarbeitenden Werturteile sind somit nicht nur abhängig von der Ausprägung der betrachteten Merkmale eines Evaluationsobjekts, sondern auch von den zugrundeliegenden Referenzpunkten der Bewertung sowie von den Methoden der Datenerhebung und -auswertung. Darauf weisen vor allem qualitativ-interpretative Evaluationsansätze hin (Bohnsack/Nentwig-Gesemann 2020; Kardorff 2012).

Evaluation durch Forschung – Forschungsbezug

Das im Zuge von Evaluation zu entwickelnde Werturteil wird wissenschaftlich begründet hergeleitet. Es handelt sich bei Evaluation also um eine Bewertungspraxis auf der Grundlage (sozial-)wissenschaftlicher Forschungsmethoden. Dies soll dazu dienen, „belastbare Evidenz zu den Wirkungen einer Intervention“ vorzulegen und diese einzuordnen (Wolbring 2014, S. 245). In diesem Zusammenhang wird gerne auch der nicht ganz unproblematische Begriff der ‚Evidenzbasierung‘ verwendet (Bellmann/Müller 2011).¹

Hinsichtlich des Forschungszugangs kann prinzipiell das komplette Spektrum quantitativer und qualitativer Methoden der Datenerhebung und -auswertung zum Einsatz kommen (Kromrey 2000, S. 20). Üblicherweise werden in der Evaluationspraxis schwerpunktmäßig quantitative Methoden eingesetzt. Qualitative Methoden haben in den vergangenen Jahren zwar an Bedeutung gewonnen (Flick 2009; Kardorff 2012), spielen aber noch immer eine ‚Nebenrolle‘, indem sie eher als Vorstudie, Ergänzung oder zur Illustration (Kardorff/Schönberger 2010, S. 374) und nur selten als primärer Forschungszugang genutzt werden.

Mit dem Konzept einer wissenschaftsgeleiteten Wirkungsreflexion werden qualitative Methoden bevorzugt. Dieser Fokus auf qualitative Methoden dient einerseits dazu, unterschiedliche Perspektiven und insbesondere die Perspektive der Adressantinnen und Adressaten von zu evaluierenden Programmen zu berücksichtigen und andererseits dazu, im Sinne einer ‚dokumentarischen Evaluationsforschung‘ (Bohnsack 2020) den Blick auf implizite Wirkannahmen der Handelnden zu richten und nicht nur bei der Rekonstruktion expliziter Bewertungen stehen zu bleiben.

Evaluation für die Praxis – Anwendungsbezug

Die mit Evaluation einhergehende Bewertung erfolgt auf der Grundlage wissenschaftlicher Methoden. Die Verwendung der damit produzierten Daten erfolgt mit Blick auf handlungspraktische Verwertungsinteressen. Es geht also weniger darum, zu einem wissenschaftlichen Erkenntnisgewinn beizutragen, sondern vielmehr darum, die Evaluationsergebnisse zur Rechenschaftslegung gegenüber externen Stakeholdern, zur Rationalisierung von Entscheidungen oder zum Identifizieren von Verbesserungspotenzialen zu nutzen. Auf den Anwendungsbezug wird weiter unten mit der Differenzierung unterschiedlicher Funktionslogiken von Evaluation, mit denen unterschiedliche Anwendungszwecke einhergehen, noch näher eingegangen (vgl. Abschnitt 3).

¹ Problematisch ist daran vor allem ein allzu fester Glaube an die ‚Objektivität‘ der über Evaluation bereitgestellten Daten. Denn aus einer sozialkonstruktivistischen Perspektive existiert die Welt nicht jenseits sozialer Konstruktionen, die infolge von Deutungs- und Aushandlungsprozessen entstehen (Berger/Luckmann 1969). Wie weiter oben bereits ausgeführt, ist ein über Evaluation generiertes Werturteil unter anderem abhängig davon, zu welchem Zeitpunkt, aus welcher Perspektive und entlang welcher Referenzpunkte der Bewertung es zustande kommt.

2

Was wird evaluiert? – Objekte & Gegenstandsbereiche der Evaluation

Aufmerksamkeit richtet sich auf Interventionen, Programme, Projekte

Objekte oder Gegenstände der Betrachtung von Evaluation sind ‚Interventionen‘, die mit dem Ziel untersucht werden, ein wertendes Urteil zu generieren. Grundsätzlich kann sich die Aufmerksamkeit dabei auf sehr unterschiedliche Dinge richten: auf Projekte, Maßnahmen, Strategien oder allgemein gesprochen auf ‚Programme‘. Im Kontext von Studium und Lehre kann sich Evaluation z. B. auf die Lehre, auf Studiengänge, Studienbedingungen, aber auch auf Unterstützungsangebote für Studierende wie für Lehrende richten. Allgemein kann Evaluation sehr unterschiedliche Aspekte der Handlungspraxis in den Blick nehmen.

Gegenstandsbereiche von Evaluation

Aus der Evaluationsperspektive werden Interventionen/Programme auf eine sehr spezifische Art und Weise in den Blick genommen. Insbesondere bei einer quantitativ-objektivistischen Evaluation wird ein Programm als eine mehr oder weniger klar definierbare Entität aufgefasst, die dazu dient, Ziele zu erfüllen; entsprechend wird auch von *Programmzielen* gesprochen. Das Erkenntnisinteresse einer Evaluation bezieht sich darauf, das Programm, seine Ziele, seine Ausgestaltung und vor allem seine Wirkungen zu analysieren. Auch können der Bedarf für ein Programm sowie die Rahmenbedingungen bzw. der Kontext, in dem ein Programm geplant oder umgesetzt wird, Gegenstand der Evaluation sein.

Eine Heuristik zur Differenzierung relevanter Gegenstandsbereiche eines Evaluationsobjekts lässt sich aus dem Diskurs zur Qualitätsentwicklung und zum Qualitätsmanagement ableiten. In Anschluss an *Donabedian (1980)* lässt sich zwischen dem Interaktions- bzw. Dienstleistungsprozess (Prozessqualität), den daraus entstehenden Ergebnissen (Ergebnisqualität) sowie den rahmenden Strukturen (Strukturqualität) differenzieren. Auf diese Differenzierung wird insbesondere im Diskurs zum Anwendungsbereich der Dienstleistungsqualität Bezug genommen. Für den Kontext von QM an Hochschulen ergänzt *Schmidt (2010)* die drei Qualitätsdimensionen durch Zielqualität. Darüber hinaus lässt sich für Interventionen, Programme bzw. Projekte der Blick darauf richten, inwieweit mit diesen auf einen konkreten Anlass oder Bedarf reagiert wird.

In diesem Sinne lassen sich die fünf in Tabelle 1 aufgeführten Gegenstandsbereiche von Evaluation – Bedarf, Konzept, Prozess, Strukturen/Rahmenbedingungen sowie Ergebnis/Wirkungen – unterscheiden. Diese Gegenstandsbereiche korrespondieren mit den unterschiedlichen Phasen bzw. Reifegraden eines Programms oder Projekts. Daraus leiten sich nicht nur unterschiedliche Formen der Evaluation ab. Je nach Gegenstandsbereich ergeben sich darüber hinaus bereits erste Hinweise darauf, entlang welcher Referenzrahmen eine Bewertung der Intervention erfolgt.

Tabelle 1: Unterschiedliche Gegenstandsbereiche der Evaluation eines Programms

	Form der Evaluation	Bewertungsmaßstab	zeitlicher Horizont
Bedarf	Bedarfsanalyse	Bedienen eines Bedarfs	primär ex-ante
Ziele, Konzept	Konzeptanalyse	Angemessenheit der Ziele, Plausibilität des Konzepts	primär ex-ante
Ausgestaltung der Implementierung/Durchführung	Prozessanalyse	Prozessqualität	prozessbegleitend
Rahmenbedingungen	Kontextanalyse	Strukturqualität	ex-ante, prozessbegleitend, ex-post
Resultate	Ergebnisanalyse	Ergebnisqualität, Zielerreichung	primär ex-post

Der Unterschied zwischen unterschiedlichen Gegenstandsbereichen und einer darauf fokussierten Evaluation soll am Beispiel eines Projekts veranschaulicht werden:

- Eine Bedarfsanalyse bezieht sich primär darauf, den Bedarf für ein Projekt z. B. ausgehend von einer Analyse potenzieller Adressatinnen und Adressaten und deren Anforderungen oder basierend auf einer Marktanalyse zu ermitteln und/oder zu evaluieren, inwieweit ein geplantes Projekt den ermittelten Bedarf bedienen kann.
- Eine Konzeptanalyse richtet sich darauf, die Angemessenheit der Ziele eines Projekts (mit Blick auf einen spezifischen Bedarf) und/oder die Plausibilität der geplanten Maßnahmen (im Hinblick auf die Zielerreichung) sowie der zugrundeliegenden Wirkannahmen zu evaluieren. Im Vordergrund stehen die Passung der Ziele zum Bedarf, die Passung der Maßnahmen zu den Zielen sowie die Umsetzbarkeit des Konzepts.
- Bei einer Prozessanalyse steht im Vordergrund, den Prozess der Implementierung oder Durchführung eines Projekts in den Blick zu nehmen und den Fokus vor allem darauf zu richten, wie gut die prozesshafte Ausgestaltung des Projekts funktioniert und weiterentwickelt werden kann.
- Eine Kontextanalyse bezieht sich darauf, die für ein Projekt relevanten Rahmenbedingungen zu analysieren und ihren positiven (Erfolgsfaktoren) oder negativen (Hürden, Stolpersteine) Beitrag zum Erfolg eines Projektes zu evaluieren.
- Mit einer Ergebnisanalyse wird der Blick in erster Linie auf Resultate eines Projekts gerichtet und die Frage, inwiefern diese zur Zielerreichung beitragen. Im Vordergrund steht hier also die Passung der Resultate (Ergebnisse/Effekte) zu den zuvor definierten Zielen.

Eine Evaluation kann mit dem in Tabelle 1 dargestellten spezifischen Fokus durchgeführt werden. Das Erkenntnisinteresse richtet sich dann in erster Linie auf eine Evaluation des Bedarfs, des Konzepts, des Prozesses, der Rahmenbedingungen oder des Ergebnisses bzw. der Wirkungen. Gleichwohl können bei einer Evaluation unterschiedliche Perspektiven integriert werden. So spielen die Programmtheorien (Konzept) insbesondere für quantitativ-objektivistische Evaluationsansätze einen wichtigen Ausgangspunkt für die Operationalisierung relevanter Indikatoren. Auch die Rahmenbedingungen spielen eine Rolle bei einer Wirkungsevaluation im Sinne endogener Variablen.

Die Frage nach dem Gegenstandsbereich der Evaluation ist durch den zeitlichen Horizont eng verbunden mit dem Zeitpunkt der Evaluation und einer damit einhergehenden Differenzierung zwischen prospektiver, formativer und summativer Evaluation (vgl. Abschnitt 4).

3 Warum wird evaluiert? – Funktionslogiken von Evaluation

Hinsichtlich der Frage, welche Ziele einem Evaluationsvorhaben zugrunde liegen, wird in der evaluationstheoretischen Literatur üblicherweise zwischen unterschiedlichen Funktionen bzw. Paradigmen von Evaluation differenziert (*Chelimsky 1997; Kromrey 2001; Stockmann 2002*). Um hervorzuheben, dass es sich dabei nicht nur um unterschiedliche Ziele handelt, die einer Evaluation bzw. Wirkungsreflexion zugrunde liegen können, schlägt *Ditzel (2022)* den Begriff der Funktionslogiken vor. Denn ausgehend von den Zielen folgt Evaluation unterschiedlichen Rationalitäten, die sich anhand des jeweils primären Bezugssystems sowie der zugrundeliegenden Informationsinteressen charakterisieren lassen (vgl. Abbildung 1) und mit denen jeweils unterschiedliche Implikationen zum Erhebungsdesign und zu den zugrundeliegenden Bewertungskriterien einhergehen. Folgende Ziele, Anwendungszwecke bzw. Funktionslogiken lassen sich idealtypisch differenzieren:

Legitimationslogik –
legitimationsorientiert

- Die *Legitimationslogik* richtet sich in einer Außenperspektive auf Anforderungen externer (oder seltener auch interner) Stakeholder. Es geht darum, diesen Rechenschaft hinsichtlich der Erfüllung ihrer Anforderungen abzulegen; es wird auch von Accountability gesprochen. Eine summativ (überprüfend) angelegte Evaluation (vgl. Abschnitt 4) dient dazu, eine Leistungs- bzw. Qualitätsbewertung bezogen auf die Anforderungen der Stakeholder vorzunehmen und damit Daten für ein Reporting und Monitoring bereitzustellen. Damit soll Konformität gegenüber den Anforderungen nachgewiesen und Legitimität für die zu evaluierende Intervention hergestellt werden. Ziel ist es, zu überprüfen und zu dokumentieren, mit welchem Aufwand, welche Leistungen erbracht und welche Wirkungen erzielt werden. Mitunter richten sich Anforderungen der Rechenschaftslegung eher darauf, dass evaluiert wird, ohne näher zu definieren, mit welchem Ziel und anhand welcher Kriterien evaluiert werden soll. Im Vordergrund steht nicht selten, die grundsätzliche Konformität mit den zugrundeliegenden Zielen und Anforderungen festzustellen.

Kontrolllogik –
entscheidungsorientiert

- In der *Kontrolllogik* dient Evaluation in einer hochschulinternen Betrachtungsperspektive dazu, Daten für eine ‚evidenzbasierte‘ Entscheidungsfindung zu Steuerungszwecken bereitzustellen. Evaluation kommt demnach dann zum Einsatz, wenn Entscheidungen zu treffen sind und dazu eine Datengrundlage geschaffen werden soll. Es geht also darum, Entscheidungsprozesse zu rationalisieren. Auch hier handelt es sich im Sinne einer summativen (überprüfenden) Evaluation um eine Erfolgskontrolle bzw. Leistungsbewertung. Hierbei steht die Beurteilung der Qualität, Wirksamkeit oder Effizienz einer Intervention im Vordergrund. Den Bewertungsmaßstab für eine entscheidungsorientierte Evaluation bilden zuvor definierte programmbezogene Ziele und Kriterien.

Entwicklungslogik –
entwicklungsorientiert

- In einer auf Lern- und Veränderungsprozesse gerichteten *Entwicklungslogik* liegt der Fokus darauf, Handlungspotentiale zur (Weiter-)Entwicklung eines Programms aufzudecken. Eine solche entwicklungs- bzw. lernorientierte Evaluation erfolgt nicht anhand zuvor definierter Bewertungskriterien und mit dem Ziel, ein abschließendes Urteil zur Qualität, Wirksamkeit oder Effizienz zu fällen, sondern im Sinne einer formativen Evaluation prozessbegleitend (vgl. Abschnitt 4). In einem iterativen Prozess können Fragestellungen und Erhebungsmethoden angepasst und geschärft werden. Das Erheben von Daten dient dazu, einen Dialog zu ermöglichen und Lernprozesse zu fördern.

Forschungslogik –
wissensorientiert

- In der *Forschungslogik* dient Evaluation primär dazu, neues Wissen zu generieren. Es lässt sich auch von Evaluation als einer (angewandten) ‚Wirkungsforschung‘ (Kromrey 2001, S. 114) sprechen. Im Sinne eines wissenschaftlichen Erkenntnisinteresses geht es darum, zu einem tieferen Verständnis des Evaluationsobjekts – seiner Eigenschaften, Wirkungen, Wirkungsweise sowie relevanter Einflussfaktoren – beizutragen. Wie für Forschung üblich, dienen wissenschaftliche Methoden und Gütekriterien dazu, belastbare Aussagen zu generieren. Evaluation als Forschung zu begreifen, geht über diese wissenschaftliche Fundierung der Methoden deutlich hinaus. Die Ausgestaltung des Evaluationsvorhabens (als Forschungsprozess) orientiert sich insgesamt an einer wissenschaftlichen Logik. Was als relevante Frage angesehen wird und welche Methoden zum Einsatz kommen, entscheidet sich primär aus wissenschaftlicher Perspektive und weniger aus Perspektive der lokalen Handlungspraxis (Entwicklungslogik), eines Steuerungsanliegens (Kontrolllogik) oder der Stakeholder (Legitimationslogik).

In Abbildung 1 werden die vier Funktionslogiken noch einmal zusammengefasst. Im Sinne einer Heuristik dienen sie der idealtypischen Beschreibung unterschiedlicher Anwendungsformen von Evaluation. Sie lassen sich nicht immer trennscharf auseinanderhalten; auch können sie sich mitunter überlappen bzw. ein Evaluationsvorhaben kann unterschiedliche Ziele gleichzeitig verfolgen.

Die Funktionslogiken von Evaluation als Reflexionsfolie heranzuziehen kann jedoch helfen, ehrlich und selbstkritisch über die Ziele eines Evaluationsvorhabens nachzudenken und die Dinge klar beim Namen zu nennen. Denn damit gehen weitreichende Implikationen nicht nur bezogen auf die Ausgestaltung der Datenerhebung und -auswertung, sondern auch bezogen darauf einher, wie die Beteiligten und Betroffenen mit den erhobenen Daten umgehen und sich dazu verhalten werden.

Funktionslogiken	Bezugssysteme	Informationsinteressen	Ziele von Evaluation
Legitimation externe Accountability	interne/externe Stakeholder	Leistungsbewertung ... bezogen auf definierte Anforderungen als Grundlage für ein Reporting/Monitoring → summative Evaluation ⇒ quantitativer Evaluationsansatz üblich	Erfüllen von Anforderungen nachweisen → z.B. Verwendung von Fördergeldern gegenüber Mittelgebern legitimieren
Kontrolle evidenzbasierte Entscheidungsfindung	Management		Entscheidungen rationalisieren → z.B. Datengrundlage für Entscheidungen schaffen
Entwicklung organisationales Lernen	Handlungspraxis	Erkenntnisinteresse ... im Hinblick auf (praktische) Handlungs- bzw. Lernpotenziale → formative Evaluation ⇒ qualitativer Evaluationsansatz notwendig	organisationales Lernen fördern → z.B. Ideen zur Weiterentwicklung von Lehr-/Lernformen entwickeln
Forschung Erkenntnisgewinn, Verstehen	Wissenschaft		neues Wissen generieren → z.B. Erkenntnisse über die Interaktion zwischen Lehrenden und Lernenden generieren ⇒ qualitativer Evaluationsansatz sinnvoll

Abbildung 1: Ziele bzw. Funktionslogiken von Evaluation, Quelle: in Anlehnung an Ditzel (2020, S. 232) und Ditzel (2022, S. 63)

Die Sichtweisen auf Evaluation ausgehend von den vier zuvor beschriebenen Funktionslogiken sollen am Beispiel des Qualitätsmanagements an Hochschulen veranschaulicht werden:

- Paradebeispiel für eine legitimationsbezogene Evaluation sind peer-review-basierte Verfahren der Programmakkreditierung oder der internen Akkreditierung. Sie sind in der Regel auf Aspekte der Qualitätssicherung ausgerichtet und überprüfen die Konformität von Studiengängen hinsichtlich formaler Vorgaben wie Mindestgrößen von Modulen, Vollständigkeit und Ausgestaltung der Modulbeschreibungen etc. In der QM-Forschung wird gezeigt, dass die handelnden Akteure bei legitimationsbezogenen Evaluationsansätzen ein Interesse haben, Konformität mit den zugrundeliegenden Anforderungen zu signalisieren, im Zweifelsfall auch durch Strategien des ‚window dressing‘.
- Entscheidungsorientierte Formen der Evaluation finden sich im QM dann, wenn Daten erhoben und ausgewertet werden, um z. B. Entscheidungen zur Einrichtung oder Schließung von Studiengängen von den Evaluationsergebnissen abhängig zu machen. Akkreditierungsverfahren werden für solche Entscheidungen genutzt, aber z. B. auch statistische Analysen von Daten aus der Prüfungsverwaltung im Sinne eines Monitoring.
- Ein Beispiel für entwicklungsorientierte Evaluation im QM ist die Lehrveranstaltungsevaluation, wenn sie als Feedbackinstrument für Lehrende eingesetzt wird und nicht als Kontrollinstrument z. B. für Entscheidungen zur Personalauswahl oder -entwicklung. Im Gegensatz zur Legitimations- und Kontrolllogik haben Lehrende in der Entwicklungslogik ein Interesse daran, dass Verbesserungspotenziale z. B. bezogen auf einen Studiengang aufgedeckt werden, ohne dafür durch Sanktionen wie Auflagen etc. ‚bestraft‘ zu werden.
- Wissensorientierte Evaluation kommt im Kontext des QM beispielsweise dann zum Einsatz, wenn Untersuchungen im Sinne eines Institutional Research durchgeführt werden. Ein Beispiel dafür kann sein, dass der Einfluss von Unterstützungsmaßnahmen in der Studieneingangsphase auf den Studienerfolg der Studierenden untersucht wird.

 Literatur zur Vertiefung

- Chelimsky, E. (1997):* Thoughts for a New Evaluation Society. In: *Evaluation* 3, 1997, 1, S. 97–109.
- Ditzel, B. (2022):* Paradigmen und Paradoxien des Qualitätsmanagements an Hochschulen. Ein theoretisch informierter Blick hinter die Kulissen formaler Implementierung. In: Reinbacher, P. (Hrsg.): *Qualität und Qualitätsmanagement im Universitäts- und Hochschulbetrieb: Paradoxien, Probleme, Perspektiven*. Weinheim: Beltz Juventa, 57–108.
- Kromrey, H. (2001):* Evaluation - ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. In: *Sozialwissenschaften und Berufspraxis* 24, 2001, 2, S. 105–132.
- Stockmann, R.:* *Qualitätsmanagement und Evaluation - Konkurrierende oder sich ergänzende Konzepte*, 2002. Saarbrücken.

4

Wann wird evaluiert? – prospektive, formative und summative Evaluation

In der evaluationstheoretischen und -praktischen Literatur wird üblicherweise zwischen ‚formativer‘ und ‚summativer‘ Evaluation differenziert (Kromrey 2001, S. 118; Stockmann 2006, S. 17ff.). Gollwitzer und Jäger (2009, S. 15) ergänzen diese Differenzierung durch ‚prospektive‘ Evaluation. Evaluationsvorhaben lassen sich dabei dahingehend differenzieren, wann die Evaluation eines Programms erfolgt, auf welche Phase bzw. welchen Reifegrad eines Programms sich die Evaluation bezieht und wozu die Evaluationsergebnisse genutzt werden sollen. Letztere Frage steht in engem Zusammenhang mit der in Abschnitt 3 behandelten Frage nach dem Verwendungszweck der Evaluationsergebnisse.

Prospektive Evaluation (Bedarfs- und Konzeptanalyse)

- Als prospektiv lässt sich eine Evaluation bezeichnen, wenn sie auf der Grundlage erster Ideen oder ausgearbeiteter Konzepte erfolgt, ohne dass diese bereits implementiert wurden. Die Evaluation erfolgt damit *ex-ante*. Dabei lassen sich zwei Formen der Evaluation unterscheiden: Eine *Bedarfsanalyse* richtet sich darauf, den Bedarf für eine geplante Intervention zu evaluieren und mögliche Anforderungen aus Sicht von Auftraggeber*innen, Adressat*innen oder anderen Stakeholdern zu ermitteln. Eine *Konzeptanalyse* richtet die Aufmerksamkeit hingegen auf die Plausibilität der zugrundeliegenden Annahmen bzw. Programmtheorien sowie die Durchführbarkeit der geplanten Intervention angesichts gegebenen Ressourcen, vorhandener Qualifikationen und zur Verfügung stehender Zeit. Auch lassen sich *ex-ante* potenzielle Wirkungen abschätzen. Ziel einer prospektiven Evaluation ist es, Daten für Entscheidungen zur Implementierung sowie zur Ausgestaltung einer Intervention bereitzustellen.

Formative Evaluation (Prozessanalyse)

- Als formativ wird Evaluation bezeichnet, wenn sie sich *prozessbegleitend* auf die Implementierung oder Durchführung einer Intervention bezieht. Die Evaluation ist integraler Bestandteil der Intervention selbst. Die Evaluationsergebnisse werden rückgekoppelt, um die Ausgestaltung einer Intervention bei Bedarf im Sinne der Entwicklungslogik anpassen zu können. Formative Evaluation hat überwiegend beratenden Charakter. Es wird auch von ‚Begleitforschung‘ (Kromrey 2006, S. 235), ‚Prozessevaluation‘ (Tietgens 1986, S. 24) oder ‚Verbesserungsevaluation‘ (Widmer 2006, S. 88) gesprochen.

Summative Evaluation (Ergebnisanalyse)

- Als summativ wird Evaluation bezeichnet, wenn sie der Intervention zeitlich nachgelagert stattfindet und einen Gesamtüberblick über die Effekte einer Intervention liefern soll. In einer *ex-post*-Betrachtung wird ein zusammenfassendes Urteil über die Intervention und ihre Qualität, Wirksamkeit oder Effizienz herausgearbeitet. Das Idealmodell einer summativen Evaluation stellt das (quasi-)experimentelle Design dar. Dabei findet die Evaluation zwar in der Regel prozessbegleitend statt, indem Daten zu unterschiedlichen Zeitpunkten erhoben werden. Das Urteil wird allerdings erst zu einem späteren Zeitpunkt gefällt, wenn Änderungen beobachtbar sind, die sich auf die Intervention zurückführen lassen. Eine summative Evaluation dient damit überwiegend der Überprüfung bzw. Leistungsbewertung im Sinne der Legitimations- und Kontrolllogik. Es wird auch von ‚Wirkungsforschung‘ (Kromrey 2006, S. 235), ‚Ergebnisevaluation‘ (Tietgens 1986, S. 24) oder ‚Bilanzevaluation‘ (Widmer 2006, S. 88) gesprochen.

Bei der Frage, ob es sich um eine prospektive, formative oder summative Evaluation handelt, spielt neben den zugrundeliegenden Zielen (vgl. Abschnitt 3) insbesondere eine Rolle, an welchem Punkt sich eine Intervention im Planungs-, Konzeptions- und Implementierungsprozess befindet. Eine prospektive Evaluation bietet sich an, wenn eine Intervention geplant, aber noch nicht umgesetzt ist. Eine

formative Evaluation erfolgt begleitend zum Implementierungsprozess oder zur (dauerhaften) Durchführung. Für zeitlich begrenzte Interventionen wie Projekte entscheidet der Zeitpunkt der Evaluation darüber, ob sie formativ oder summativ erfolgen kann. Denn eine summative Evaluation bezieht sich in einer retrospektiven Betrachtung auf eine – im Sinne der relevanten Bewertungskriterien – möglichst umfassende Betrachtung des Programms. Das macht es erforderlich, dass ein Programm bereits mindestens einmal erfolgreich durchlaufen wurde. Für Projekte bedeutet es, dass die Evaluation in der Regel erst nachgelagert nach Projektende stattfindet bzw. dass erst dann abschließende Ergebnisse vorliegen können.

Daraus folgt, dass sich eine projektbegleitende Evaluation im summativen Sinne eher auf Einzelaspekte beziehen aber nicht das gesamte Projekt umfassen kann bzw. dass eine projektbegleitende Evaluation per se nicht im Sinne einer Leistungsbewertung (summativ), sondern im Sinne einer Identifikation von Lernpotenzialen (formativ) erfolgen kann. Für eine Analyse kurzfristiger und längerfristiger Wirkungen im summativen Sinne braucht es eine Evaluation, welche das zu evaluierende Projekt überdauert, also erst nach Ende des Projektes zu einem Urteil gelangen kann. Das stellt eine Herausforderung für eine projektbegleitende Evaluation dar.

5

Wer evaluiert? – Selbst- und Fremdevaluation

Bezüglich der Frage, von wem eine Evaluation methodisch durchgeführt wird, wird in der evaluationstheoretischen Literatur üblicherweise zwischen interner und externer Evaluation differenziert (*Stockmann/Meyer 2014, 88ff.*). Dabei geht es darum, ob die Evaluierenden aus der eigenen oder aus einer fremden Organisation kommen. Da der Unterschied zwischen interner und externer Evaluation insbesondere dann, wenn die intern mit der Evaluation betrauten Personen über eine entsprechende Expertise verfügen, verschwimmen, scheint es zielführender zwischen Selbst- und Fremdevaluation zu unterscheiden:

- *Selbstevaluation* meint eine Evaluation, welche durch die das Programm Verantwortenden oder daran Beteiligten selbst durchgeführt wird. Die Evaluation erfolgt aus einer Innenperspektive. Das hat den Vorteil, dass die Evaluierenden in der Regel über ein detailliertes Sach- und Kontextwissen verfügen und somit einen fundierten Blick auf das zu evaluierende Programm richten und die Erkenntnisse fachlich einordnen können. Gleichzeitig geht damit der Nachteil einher, dass es den Evaluierenden schwerer fällt, eine kritische Distanz zum Evaluationsobjekt aufzubauen, als es bei einer Fremdevaluation üblicherweise der Fall ist. Auch kann es bei einer Evaluation durch die Programmverantwortlichen oder -durchführenden selbst sein, dass sich die Befragten – je nach Erhebungsmethode – mit kritischen Rückmeldungen zurückhalten.
- *Fremdevaluation* meint hingegen eine Evaluation durch unabhängige Personen, die nicht am Programm beteiligt oder davon betroffen sind. Dabei muss es sich nicht notwendigerweise um organisationsexterne Personen handeln. Der Vorteil einer Fremdevaluation liegt in einer stärkeren Distanzierung und Unabhängigkeit sowie einer externen Legitimation und Reputation, sofern eine entsprechende methodische Expertise vorhanden ist.



Die Unterschiede zwischen Selbst- und Fremdevaluation sowie die damit einhergehenden Implikationen sollen am Beispiel der Evaluation einer Organisationseinheit veranschaulicht werden:

- Eine Selbstevaluation durch die zu evaluierende Organisationseinheit selbst könnte beispielsweise darin bestehen, dass die Mitarbeitenden, die Kundinnen und Kunden sowie kooperierende Organisationseinheiten anhand standardisierter Fragebögen befragt werden. Auch qualitative Interviews wären als Erhebungsmethode denkbar, stünden aber vor der Herausforderung, eine offene und kritische Atmosphäre zu schaffen, wenn Mitarbeitende der zu evaluierenden Organisationseinheit selbst die Interviews führen und auswerten.
- Bei einer Fremdevaluation könnte z. B. eine völlig andere Erhebungsmethode gewählt werden. Im Sinne eines Peer Reviews könnte die Organisationseinheit durch Personen begutachtet werden, die an anderen Hochschulen in einem ähnlichen Tätigkeitsfeld arbeiten.

6

Wie wird evaluiert? – Ansätze der Evaluation

Auf die Frage danach, wie evaluiert werden soll/kann, lassen sich vielfältige Antworten finden; in der Literatur zur Evaluationspraxis und -forschung findet sich zahlreiche Evaluationsansätze mit unterschiedlichen Schwerpunktssetzungen und unterschiedlichen methodischen Zugängen. In diesem Abschnitt sollen einige Aspekte vorgestellt werden, die als Orientierungspunkte bei der Frage nach der Ausgestaltung eines Evaluationsvorhabens herangezogen werden können. Es kann sich dabei jedoch nur um eine stark verkürzte Darstellung handeln, die notwendigerweise eine Vielzahl von Aspekten und Zugängen ausspart.

 weiterführende Informationen

Eine ausführlichere Auseinandersetzung mit quantitativen und qualitativen Forschungszugängen und möglichen Evaluationsansätzen erfolgt im KOMWEID-Impuls *„Übersicht über unterschiedliche Formen und Formate der Wirkungsreflexion“*.

Phasen der Evaluation

Ein in der Evaluationspraxis gängiger Ansatz ist, die im Zuge eines Evaluationsvorhabens notwendigen Phasen und die dabei zu klärenden Fragen zu definieren und zu strukturieren. Damit soll eine systematische Vorgehensweise sichergestellt werden. Auch wenn die Phasen sequenziell dargestellt werden, sind vielfältige Iterationsschleifen denkbar und sinnvoll.

Die Auseinandersetzung mit einem Evaluationsvorhaben beginnt zunächst damit, zu präzisieren, worauf sich die Aufmerksamkeit der Evaluation richten und wie sie ausgestaltet werden soll.

Präzisieren des Evaluationsobjekts

- In einem ersten Schritt werden das *Evaluationsobjekt* und seine *Ziele* beschrieben. Erst danach kann das Evaluationsvorhaben als eine Analyse des Evaluationsobjekts präzisiert werden.

Präzisieren des Evaluationsverfahrens

- Im zweiten Schritt werden die in die Evaluation *involvierten Personen* einschließlich ihrer Interessen und Rollen identifiziert.
- Im dritten Schritt werden die *Ziele und Fragestellungen des Evaluationsvorhabens* definiert. Davon hängt ab, wie das Evaluationsdesign auszugestalten ist.
- Im vierten Schritt werden die der Evaluation zugrundeliegenden *Bewertungskriterien* definiert als Ausgangs- und Referenzpunkt für die evaluative Analyse.
- Im fünften Schritt wird ein im Hinblick auf das Evaluationsobjekt sowie die Ziele und Fragestellungen angemessenes *Evaluationsdesign* entwickelt und entsprechende Methoden ausgewählt.

Datenerhebung und -analyse

- Danach beginnen das Erheben, Auswerten und Interpretieren der Daten.
- Konstitutiv für Evaluation ist schließlich, dass das Dokumentieren, Veröffentlichen und Nutzen der Evaluationsergebnisse von Anfang an mitgedacht wird.

weiterführende Informationen

Im KOMWEID-Impuls ‚*Verfahrensschritte der Wirkungsreflexion bzw. Evaluation*‘ werden die einzelnen Schritte vorgestellt, die es für die Konzeption, Durchführung und Auswertung eines Evaluationsvorhabens braucht.

Forschungsmethoden: quantitativ / qualitativ

Die Ausgestaltung eines Evaluationsvorhabens, insbesondere hinsichtlich der zugrunde zu legenden Bewertungskriterien und des Evaluationsdesigns richtet sich zunächst danach, welche Methoden der Datenerhebung und -auswertung – ausgehend vom Evaluationsobjekt, den Zielen der Evaluation und den zu untersuchenden Fragestellungen zum Einsatz kommen sollen. Dabei lässt sich zunächst zwischen quantitativen und qualitativen Forschungszugängen sowie deren Kombination im Sinne von Mixed Methods differenzieren. Von Bedeutung ist dabei auch, wie sich die Methoden wissenschafts- und erkenntnistheoretisch einordnen lassen. Insgesamt steht ein breites Spektrum quantitativer und qualitativer Forschungsmethoden und damit einhergehender Evaluationsansätze zur Verfügung.

weiterführende Informationen

Im KOMWEID-Impuls ‚*Formen und Formate der Wirkungsreflexion*‘ werden unterschiedliche methodische Ansätze der Evaluation vorgestellt.

Eine wichtige Frage bei der Entwicklung des Evaluationsdesigns und der damit einhergehenden Wahl der Methoden besteht darin, wie sich das Erkenntnisinteresse des Evaluationsvorhabens ausgehend von der Identifikation des Evaluationsobjekts und der Wahl eines Gegenstandsbereichs (vgl. Abschnitt 2) sowie ausgehend von den Zielen der Evaluation und dem damit einhergehenden Verwendungszweck der Evaluationsergebnisse (vgl. Abschnitt 3) weiter fokussieren lässt. Drei mögliche Gestaltungsdimensionen zur Fokussierung des Erkenntnisinteresses einer Evaluation werden nachfolgend vorgestellt.

Analysefokus: Qualität, Wirkung, Effizienz

Erstens stellt sich die Frage, ob sich die Evaluation auf die Qualität, die Wirkung oder die Effizienz eines Programms bzw. Projekts bezieht. Mit der *Qualität* werden Merkmale des Evaluationsobjekts zu einem singulären Zeitpunkt betrachtet und mit definierten Vergleichshorizonten wie Zielen, Standards, Anforderungen oder anderen Entitäten verglichen, um daraus eine Qualitäts- oder Leistungsbewertung zu generieren. Mit der *Wirkung* geht es um die Frage, ob und welche Veränderungen sich beobachten lassen, die dem Programm/Projekt zuzuschreiben (zu attribuieren) sind. Anders als bei der Qualitätsbewertung braucht es dafür mindestens zwei Beobachtungszeitpunkte, um mögliche Veränderungen überhaupt beobachten zu können. Im Idealmodell eines (quasi-)experimentellen Erhebungsdesigns braucht es darüber hinaus eine Kontrollgruppe, um Kausalzusammenhänge zwischen Intervention und beobachteter Veränderung als Wirkung erkennen zu können. Mit der *Effizienz* werden ermittelte Wirkungen (Effektivität) schließlich in ein Verhältnis gesetzt mit dem Aufwand, der dafür erforderlich ist.

Tabelle 2: Stufen der Bewertung: Qualität, Wirkung, Effizienz

	Fokus der Evaluation	Gegenstandsbereiche
Qualitätsanalyse	Bewertung eines Merkmals/Zustands zu einem singulären Beobachtungszeitpunkt	<ul style="list-style-type: none"> • Konzept • Prozess • Rahmenbedingungen • Resultate
Wirkungsanalyse	(1) Effekte als Veränderungen: Bewertung eines Merkmals/Zustands zu unterschiedlichen Beobachtungszeitpunkten (2) Attribution als Ursache-Wirkungszusammenhang: – Vergleich mit Kontrollgruppe (quantitatives Paradigma) – Deutung/Zuschreibung (qualitatives Paradigma)	<ul style="list-style-type: none"> • Resultate: Ergebnisse/Leistungen, Wirkungen/Effekte • Wirksamkeit/Effektivität • Wirkungsweise (Prozess)
Effizienzanalyse	Verhältnis zwischen Aufwand (Kosten) und Nutzen (Effektivität)	<ul style="list-style-type: none"> • Effizienz/ Wirtschaftlichkeit

Die Unterschiede zwischen den drei Bewertungsstufen Qualität, Wirkung und Effizienz sollen am Beispiel medien- oder hochschuldidaktische Beratungsangebote veranschaulicht werden:

- Eine Evaluation im Hinblick auf die Qualität des Beratungsangebots kann sich darauf beziehen, das Beratungsangebot dahingehend zu evaluieren, inwiefern Lehrende, die daran teilgenommen haben, einzelne Aspekte bewerten. Hierbei stehen vermutlich Zufriedenheitsaspekte im Vordergrund. Es geht also darum, inwiefern das Beratungsangebot die Erwartungen oder Anforderungen der Zielgruppe erfüllen.
- Eine Wirkungsanalyse setzt sich mit den Effekten des Beratungsangebots auseinander. Dabei kann es beispielsweise darum gehen, Kompetenzbereich, um die es in dem Beratungsangebot geht, vor und nach der Beratung im Sinne eines Vorher/Nachher-Vergleichs zu bewerten. Oder in einem qualitativen Evaluationsdesign könnte danach gefragt werden, wie die Teilnehmenden die Beratung erlebt haben, wie mit ihren Fragen aufgehoben gefühlt haben oder was sich daraus für ihre Handlungspraxis ergeben hat.
- In einer Effizienzanalyse werden ermittelte Effekte in ein Verhältnis zum dafür erforderlichen Aufwand gestellt. In einer entsprechenden Analyse könnte folglich – aufbauen auf oder zusammen mit einer Wirkungsanalyse – der Frage nachgegangen werden, welcher Aufwand dafür seitens der Teilnehmenden und der Anbietenden erforderlich war. Eine solche Analyse bietet sich insbesondere im Vergleich unterschiedlicher Beratungsangebote an.

Analysetiefe einer Wirkungsanalyse

Zweitens lassen sich für den Spezialfall einer Wirkungsanalyse mit Blick auf die zum Einsatz kommenden Forschungsmethoden unterschiedliche Analyseebenen differenzieren. So unterscheiden *Balzer und Beywl (2018, S. 97)* zwischen einer subjektiven Wirkungseinschätzung und einem empirischen Wirkungsnachweis und rekurrieren damit auf die methodische Absicherung eines Evaluationsurteils. Insbesondere mit Blick auf qualitative Ansätze der Evaluation lässt sich dies durch eine Analyse bzw. Rekonstruktion der Wirkungsweise ergänzen.

- Wirkungseinschätzung: Hierbei erfolgt eine Bewertung der Qualität, Wirkung und/oder Wirksamkeit einer Intervention durch Befragung von Personen. Ergebnis ist nicht eine wissenschaftlich abgesicherte Aussage, sondern eine subjektive Zuschreibung. Dafür können unterschiedliche Formen quantitativer und qualitativer Befragungen zum Einsatz kommen. Eine Spezialform stellt das Peer-Review dar, bei dem die Einschätzung durch Peers als Expert*innen des zu evaluierenden Handlungsfelds erfolgt. Mit vordefinierten Frageitens oder Evaluationskriterien sollen zwar die Referenzpunkte der Bewertung transparent und überprüfbar gemacht werden. Die Einschätzung kommt jedoch in der Regel zustande, ohne die zugrundeliegenden Handlungstheorien näher zu explizieren.

- *Wirkungsnachweis*: Den Begriff des ‚Wirkungsnachweises‘ verwenden *Balzer und Beywl (2018, S. 97)* für Evaluationsurteile, die wissenschaftlich abgesichert im Sinne eines (quasi-)experimentellen Designs erfolgen. Dabei findet ein zweifacher Vergleich statt, einerseits im zeitlichen Verlauf, um Veränderungen zu erkennen und andererseits mit einer Kontrollgruppe, um identifizierte Veränderungen der untersuchten Intervention zuschreiben (attribuieren) zu können. In der Regel kommen dafür quantitative Forschungsmethoden zum Einsatz.
- *Wirkungsrekonstruktion*: Wie eine Intervention funktioniert bzw. wirkt, wie also die Wirkungen zustande kommen (oder weshalb sie ausbleiben), verbleibt in quantitativen Evaluationsdesigns in einer Black-Box. Hier setzen qualitative Ansätze an, die aus der Perspektive der Beteiligten und Betroffenen die *Wirkungsweise* und mögliche *Einflussfaktoren* analysieren, z. B. durch die Rekonstruktion von Deutungsprozessen (vgl. z.B. *Bohnsack 2020*). Zwar können Wirkmechanismen und Einflussfaktoren auch in quantitativen Evaluationsdesigns untersucht werden. Sie müssen dazu aber zuvor in einem komplexen Wirkmodell bereits als (potenziell) relevant berücksichtigt und operationalisiert worden sein.

Auch hier sei der Unterschied der drei Konstrukte ‚Wirkungseinschätzung‘, ‚Wirkungsnachweis‘ und Wirkungsrekonstruktion an einem konkreten Beispiel einer Lehrveranstaltung veranschaulicht:

- Werden die Studierenden z. B. im Rahmen der Lehrveranstaltungsevaluation über die üblichen Zufriedenheitsabfragen auch danach befragt, welchen Einfluss ein Gestaltungsmerkmal wie ein Planspiel auf ihre Selbstmotivation oder ihren Lernerfolg hatte, so handelt es sich dabei um eine subjektive Qualitäts- oder Wirkungseinschätzung durch die Studierenden.
- Für einen Wirkungsnachweis bedarf es einer Untersuchung mit mindestens zwei Beobachtungszeitpunkten (vor und nach der Veranstaltung) und einer Kontrollgruppe (in der kein Planspiel stattfand), um herauszufinden, ob sich die Motivation und die Kompetenzen der Studierenden im Laufe der Lehrveranstaltung verändern und ob das auf das Planspiel zurückzuführen ist.
- Eine qualitative Analyse könnte sich darauf fokussieren herauszuarbeiten, wie die Studierenden mit spezifischen didaktischen Interventionen wie einem Planspiel oder einem Projekt zurechtkommen und welchen Einfluss das auf ihren Lernprozess hat.

Stufen der Wirkung bzw. Wirksamkeit

Eine dritte Gestaltungsdimension bezieht sich auf die Stufen der Wirkung bzw. Wirksamkeit eines zu evaluierenden Programms bzw. Projekts. Dabei geht es um die Frage, was Wirkung genau bedeutet bzw. auf welcher Ebene sie beobachtet werden kann. In der Evaluations- und Wirkungsforschung wird üblicherweise zwischen erbrachten Leistungen bzw. Ergebnissen (Output), Zielerreichung (Outcome) und Wirkung (Impact) differenziert (*Stockmann/Meyer 2014, S. 70; Wachsmuth/Hense 2016, S. 81*). Beispielhaft werden diese Begriffe in der ‚Wirkungstreppe‘ von *Kurz und Kubek (2021, S. 5)* operationalisiert. In Tabelle 3 finden sich diese Stufen der Wirkung/Wirkungsweise in leicht abgewandelter Form.

Diese Differenzierung von Stufen bzw. Ebenen der Wirkung und Wirksamkeit lässt sich am Beispiel von medien- oder hochschuldidaktischen Weiterbildungen veranschaulichen:

- Auf der ersten Ebene lässt sich beobachten, inwiefern das Ereignis einer Weiterbildung überhaupt stattfindet, inwiefern oder wie viele Weiterbildungen also stattfinden. Hier deckt sich das Resultat weitgehend mit der Aktivität/Handlung, die Gegenstand der Betrachtung ist.
- Auf der zweiten Ebene wird beobachtet, inwiefern und in welchem Umfang Lehrende an dieser Weiterbildung teilnehmen. Das Ergebnis drückt sich folglich darin aus, dass die Lehrenden grundsätzlich erreicht werden und am Angebot partizipieren.
- Auf der dritten Ebene lässt sich beobachten, wie die Lehrenden auf die Weiterbildung reagieren und wie zufrieden sie mit ihr sind. Damit wird jedoch noch nichts darüber ausgesagt, was sie aus der Weiterbildung für ihre eigene Handlungspraxis an Erkenntnissen mitnehmen.

- Erst ab der vierten Ebene lässt sich von Effekten/Wirkung im engeren Sinne sprechen. Hier geht es um die Frage, inwiefern die Weiterbildung zu einer Veränderung kognitiver Strukturen beiträgt, z. B. beobachtbar in der Entwicklung von Kompetenzen, Denkweisen oder Rollenbildern.
- Auf der fünften Ebene zeigen sich Effekte darin, dass die Weiterbildung – vermittelt über eine Veränderung von Denkweisen – zu einer veränderten Handlungspraxis beiträgt. Bezogen auf das Beispiel könnte es also sein, dass sich Auswirkungen auf die Gestaltung der Lehre zeigen.
- Auf der sechsten Ebene schließlich kann danach gefragt werden, ob und in welcher Weise eine veränderte Lehrpraxis einen Einfluss auf das Lernen der Studierenden hat, also z. B. auf die Gestaltung des Lernprozesses, die Kompetenzentwicklung oder das berufliche Handeln.

Tabelle 3: Stufen der Wirkung bzw. Wirksamkeit, Quelle: in Anlehnung an Kurz und Kubek (2021, S. 5)

Stufe	Erläuterung	Einordnung	
(6) Impact	Handlungspraxis zeigt Wirkung.	Impact	Wirkung, Effekte
(5) Handeln	Handlungspraxis verändert sich.	Outcome	
(4) Denken	Kognitive Strukturen verändern sich.		
(3) Zufriedenheit	Maßnahme wird akzeptiert.	Output	Leistungen, Ergebnisse, Qualität
(2) Partizipation	Zielgruppe wird erreicht.		
(1) Ereignis	Ereignis findet statt.		

7

Kurzzusammenfassung

In den vorangegangenen Abschnitten wurde nicht nur der Frage nachgegangen, was Evaluation bedeutet, sondern es wurden bereits Themen angeschnitten, die einen Einfluss darauf haben, wie ein Evaluationsvorhaben auszugestalten ist:

Evaluationsobjekt und Gegenstandsbereich

- Erstens ist für die Ausgestaltung eines Evaluationsvorhabens von Bedeutung, worauf sich der Fokus richten soll. Die Bestimmung des Evaluationsobjekts stellt dabei den ersten Schritt dar. Wesentlich ist auch, auf welchen Gegenstandsbereich sich die Aufmerksamkeit richtet. So macht es für die Ausgestaltung des Evaluationsvorhabens einen Unterschied, ob der *Bedarf* für ein Programm, das *Konzept*, der *Prozess* der Implementierung/Durchführung, der *Kontext* oder die *Ergebnisse* zu analysieren sind (vgl. Abschnitt 1).

Funktionslogiken

- Zweitens ergeben sich Hinweise auf die Ausgestaltung eines Evaluationsvorhabens aus der Frage, welche Zwecke die Evaluation erfüllen soll, was mit den Ergebnissen passieren soll und welche Rationalitäten bzw. Funktionslogiken der Evaluation primär zugrunde liegen: ob sie also zur *Legitimation*, zur *Erfolgskontrolle*, als Grundlage für eine *Weiterentwicklung* oder zum Zwecke eines *wissenschaftlichen Erkenntnisgewinns* durchgeführt werden soll (vgl. Abschnitt 3).

prospektive, formative oder summative Evaluation

- Drittens hat einen Einfluss, wann eine Evaluation im Konzeptions-, Implementierungs- und Durchführungsprozess eines Projektes erfolgt und andererseits ob dabei ein *in die Zukunft gerichtetes*, ein *prozessbegleitendes* oder ein *abschließendes* Urteil gefällt werden soll. Abhängig davon lässt sich zwischen prospektiver, formativer und summativer Evaluation differenzieren (vgl. Abschnitt 4).

Selbst- oder Fremdevaluation

- Viertens hängt die Ausgestaltung des Evaluationsvorhabens davon ab, ob die Evaluation von internen, am zu evaluierenden Projekt beteiligten Personen evaluiert werden soll oder durch externe Personen. Bei einer *Selbstevaluation* kommen unter Umständen andere Erhebungs- und Auswertungsmethoden zum Einsatz als bei einer *Fremdevaluation* (vgl. Abschnitt 5).

Analysefokus: Qualität, Wirkung, Effizienz

- Fünftens macht es einen Unterschied, ob ein Evaluationsobjekt hinsichtlich der Qualität, Wirkung oder Effizienz evaluiert werden soll. Eine *Qualitätsanalyse* bezieht sich auf den Abgleich mit definierten Zielen, Standards oder Anforderungen. Eine *Wirkungsanalyse* fragt nach den – im Hinblick auf Ziele/Standards/Anforderungen – beobachtbaren Veränderungen, die sich auf das Evaluationsobjekt zurückführen lassen. Bei einer *Effizienzanalyse* kommt eine Betrachtung des Aufwands hinzu (vgl. Abschnitt 6).

Analysetiefe einer Wirkungsanalyse

- Sechstens lässt sich für den Spezialfall von Wirkungsanalysen ausgehend von den zum Einsatz kommenden Erhebungsmethoden nach der anzustrebenden Analysetiefe fragen. Eine *subjektive Wirkungseinschätzung* nimmt die Sichtweise der Beteiligten und Betroffenen z. B. über Befragungsinstrumente in den Blick, in der Regel jedoch ohne die zugrundeliegenden Handlungstheorien zu explizieren. Ein *Wirkungsnachweis* erfolgt in der Regel in Form eines (quasi-) experimentellen Forschungsdesigns. Dabei werden Veränderungen als Effekte ermittelt und nach ihrer Zuordnung zum Evaluationsobjekt als Ursache gefragt. Eine *Wirkungsrekonstruktion* nimmt die Wirkungsweise einer Intervention in den Blick und fragt danach, wie Wirkungen zustande kommen (vgl. Abschnitt 6).

Stufen der Wirkung bzw. Wirksamkeit

- Siebtens lässt sich – für den Fall von Wirkungsanalysen – fragen, auf welcher Stufe die Wirkung oder Wirksamkeit eines Evaluationsobjekts betrachtet werden soll, ob es darum geht, *beobachtbare Leistungen/Ergebnisse*, also den Output im Sinne des Stattfindens von Ereignissen, der Partizipation oder der Zufriedenheit

der Adressatinnen und Adressaten zu ermitteln oder *tieferliegende Formen der Wirkung* auf das Denken und/oder Handeln der Adressatinnen und Adressaten oder gar auf die Wirkung ihres Handelns zu richten. Abhängig von der zu untersuchenden Wirkungsstufe ist nicht nur, welche Methoden zum Einsatz kommen können, sondern auch, wann eine Evaluation stattfinden kann (vgl. Abschnitt 6).

In Tabelle 4 werden noch einmal die beschriebenen Gestaltungsdimensionen zusammengefasst, anhand derer über die Ziele, den Fokus und die Fragestellungen eines Evaluationsvorhabens nachgedacht werden kann.

Tabelle 4: Mögliche Gestaltungsdimensionen zur Beschreibung von Zielen, Fokus und Fragestellungen eines Evaluationsvorhabens

Gegenstandsbereich	Bedarf	Konzept	Prozess	Kontext	Ergebnisse	
Funktionslogiken	Legitimation	Kontrolle	Entwicklung	Forschung		
Analysefokus	Qualitätsanalyse	Wirkungsanalyse	Effizienzanalyse			
Analysetiefe	Wirkungseinschätzung	Wirkungsnachweis	Wirkungsrekonstruktion			
Stufen der Wirkung bzw. Wirksamkeit	Ereignis findet statt	Partizipation	Zufriedenheit	Denken	Handeln	Impact auf Handlungspraxis

8

Aufgaben für das Selbststudium

- Welche Merkmale charakterisieren ein Evaluationsvorhaben? Worin unterscheidet sich eine wissenschaftliche Evaluation von Alltagsvorstellungen von Evaluation?
- Welche unterschiedlichen Zwecke von Evaluation lassen sich differenzieren? Weshalb ist es wichtig, diese Unterscheidung zu treffen?
- Was unterscheidet prospektive, formative und summative Evaluation? Wovon hängt ab, welche Form der Evaluation zu wählen ist?
- Worin unterscheiden sich eine Evaluation im Hinblick auf die Qualität, Wirkung oder Effizienz eines Programms/Projekts?

9

Weiterführende Materialien und Literatur

 Lernmaterialien zu verwandten Themen

- **Prämissen einer wirkungsorientierten Projektgestaltung und -steuerung**, KOMWEID-Impulse, in Planung.
- **Wirkungsorientierung: Begriffe und Visualisierungen einer Wirkungsanalyse**, KOMWEID-Impuls, in Planung.
- **Verfahrensschritte der Wirkungsreflexion bzw. Evaluation**, KOMWEID-Impuls, in Planung.
- **Formen und Formate der Wirkungsreflexion**, KOMWEID-Impulse, in Planung.

📖 Literatur
zur Vertiefung

- Balzger, L.; Beywl, W. (2018):* evaluiert: erweitertes Planungsbuch für Evaluationen im Bildungsbereich. Bern: hep verlag.
- Kurz, B.; Kubek, D. (2021):* Kursbuch Wirkung. Das Praxishandbuch für alle, die Gutes noch besser tun wollen. Berlin: PHINEO.
- Kromrey, H. (2001):* Evaluation - ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. In: Sozialwissenschaften und Berufspraxis 24, 2001, 2, S. 105–132.
- Stockmann, R.; Meyer, W. (2014):* Evaluation. Eine Einführung. Opladen: Barbara Budrich. – Kapitel 2: Wissenschaftsbasierte Evaluation

📖 Quellennachweise

- Altfeld, S.; Schmidt, U.; Schulze, K. (2015):* Wirkungsannäherung im Kontext der Evaluation von komplexen Förderprogrammen im Hochschulbereich. In: Qualität in der Wissenschaft 9, 2015, 2, S. 56–63.
- Balzger, L.; Beywl, W. (2018):* evaluiert: erweitertes Planungsbuch für Evaluationen im Bildungsbereich. Bern: hep verlag.
- Bellmann, J.; Müller, T. (Hrsg.) (2011):* Wissen, was wirkt. Kritik evidenzbasierter Pädagogik. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Berger, P. L.; Luckmann, T. (1969):* Die gesellschaftliche Konstruktion der Wirklichkeit: Eine Theorie der Wissenssoziologie. Frankfurt am Main: Fischer.
- Bohnsack, R. (2020):* Qualitative Evaluationsforschung und dokumentarische Methode. In: Bohnsack, R. et. al. (Hrsg.): Dokumentarische Evaluationsforschung: Theoretische Grundlagen und Beispiele aus der Praxis: Verlag Barbara Budrich, S. 27–66.
- Bohnsack, R.; Nentwig-Gesemann, I. (Hrsg.) (2020):* Dokumentarische Evaluationsforschung: Theoretische Grundlagen und Beispiele aus der Praxis: Verlag Barbara Budrich.
- Chelimsky, E. (1997):* Thoughts for a New Evaluation Society. In: Evaluation 3, 1997, 1, S. 97–109.
- Ditzel, B. (2020):* Umgang mit Daten zur Qualität von Studium und Lehre zwischen Leistungsbewertung und Erkenntnisinteresse. Eckpunkte einer wissenschaftsgeleiteten Qualitätsentwicklung. In: Pohlenz, P. et. al. (Hrsg.): Qualitätssicherung im Student Life Cycle. Münster: Waxmann, 227-242.
- Ditzel, B. (2022):* Paradigmen und Paradoxien des Qualitätsmanagements an Hochschulen. Ein theoretisch informierter Blick hinter die Kulissen formaler Implementierung. In: Reinbacher, P. (Hrsg.): Qualität und Qualitätsmanagement im Universitäts- und Hochschulbetrieb: Paradoxien, Probleme, Perspektiven. Weinheim: Beltz Juventa, 57–108.
- Donabedian, A. (1980):* The Definition of Quality and Approaches to Its Assessment and Monitoring; Ann Arbor.
- Flick, U. (2009):* Qualitative Methoden in der Evaluationsforschung. In: Zeitschrift für Qualitative Forschung 10, 2009, 1, S. 9–18.
- Gollwitzer, M.; Jäger, R. S. (2009):* Evaluation kompakt. Weinheim: Beltz.
- Kardorff, E. v. (2012):* Qualitative Evaluationsforschung. In: Flick, U. et. al. (Hrsg.): Qualitative Forschung. Ein Handbuch. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag, S. 238–250.
- Kardorff, E. v.; Schönberger, C. (2010):* Evaluationsforschung. In: Mey, G. et. al. (Hrsg.): Handbuch Qualitative Forschung in der Psychologie. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 367–381.
- Kromrey, H. (2000):* Die Bewertung von Humandienstleistungen. Fallstricke bei der Implementations- und Wirkungsforschung sowie methodische Alternativen. In: Müller-Kohlenberg, H. et. al. (Hrsg.): Qualität von Humandienstleistungen: Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 19–57.

- Kromrey, H. (2001):* Evaluation - ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. In: Sozialwissenschaften und Berufspraxis 24, 2001, 2, S. 105–132.
- Kromrey, H. (2006):* Qualität und Evaluation im System Hochschule. In: Stockmann, R. (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Münster: Waxmann, S. 234–259.
- Kurz, B.; Kubeck, D. (2021):* Kursbuch Wirkung. Das Praxishandbuch für alle, die Gutes noch besser tun wollen. Berlin: PHINEO.
- Meyer, W. (2017):* Evaluation von und in Organisationen. In: Liebig, S. et. al. (Hrsg.): Handbuch Empirische Organisationsforschung. Wiesbaden: Springer Fachmedien Wiesbaden, S. 513–535.
- Schmidt, U. (2010):* Wie wird Qualität definiert? In: Winde, M. (Hrsg.): Von der Qualitätsmessung zum Qualitätsmanagement. Praxisbeispiele an Hochschulen. Essen: Edition Stifterverband, S. 10–17.
- Stockmann, R.:* Qualitätsmanagement und Evaluation - Konkurrierende oder sich ergänzende Konzepte, 2002. Saarbrücken.
- Stockmann, R. (2006):* Evaluation in Deutschland. In: Stockmann, R. (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Münster: Waxmann, S. 15–46.
- Stockmann, R.; Meyer, W. (2014):* Evaluation. Eine Einführung. Opladen: Barbara Budrich.
- Tietgens, H. (1986):* Allgemeiner Teil. In: Tietgens, H. (Hrsg.): Aufgaben und Probleme der Evaluation in der Erwachsenenbildung. Bonn: Deutscher Volkshochschul-Verband.
- Wachsmuth, E.; Hense, J. (2016):* Wirkmodelle zur Unterstützung der Evaluation komplexer Hochschulprojekte. In: Qualität in der Wissenschaft 10, 2016, 3-4, S. 80–87.
- Widmer, T. (2006):* Qualität der Evaluation - Wenn Wissenschaft zur praktischen Kunst wird. In: Stockmann, R. (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Münster: Waxmann, S. 85–112.
- Wolbring, T. (2014):* Kausalanalyse und Wirkungsevaluation. Potential Outcomes, Graphenmethodologie und ihre Anwendung am Beispiel der Bologna-Reform. In: Zeitschrift für Evaluation 13, 2014, 2, S. 243–270.
- Wolbring, T. (2016):* Evaluation, Kausalität und Validität. Wissenschaftliches Ideal und universitäre Praxis. In: Großmann, D. et. al. (Hrsg.): Evaluation von Studium und Lehre: Grundlagen, methodische Herausforderungen und Lösungsansätze. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 57–89.

