



Studientext
EVALUATION

Gabi Reinmann

Reinmann, G. (2012). Studententext Evaluation. München. URL: <http://lernen-unibw.de/studententexte>

2011: 1. Version
2012: 2. Korrigierte Version



Dieser Werk bzw. Inhalt ist unter einer Creative Commons-Lizenz lizenziert.
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Universität der Bundeswehr München
Fakultät für Pädagogik
Professur für Lehren und Lernen mit Medien
Prof. Dr. Gabi Reinmann
Werner-Heisenberg-Weg 39
85579 Neubiberg
E-Mail: reinmann.gabi@googlemail.com
Internet: <http://lernen-unibw.de>

INHALTSVERZEICHNIS

0	EINFÜHRUNG IN DEN STUDIENTEXT	02
0.1	Ziele des Studientextes	02
0.2	Aufbau des Studientextes	03
1	BEGRIFF DER EVALUATION UND EVALUATIONSFORSCHUNG	05
1.1	Was ist Evaluation?	05
1.2	Was hat Evaluation mit Qualitätsmanagement zu tun?	07
2.	ZWECKE UND INHALTE EINER EVALUATION	08
2.1	Wozu kann man evaluieren?	08
2.2	Was kann man evaluieren?	09
2.2.1	Evaluationsgegenstände	09
2.2.2	Bewertungskriterien	12
2.2.3	Bewertungsmaßstäbe	14
3.	DESIGN UND VERANTWORTLICHE EINER EVALUATION	15
3.1	Wie kann man evaluieren?	15
3.1.1	Evaluationsformen	15
3.1.2	Ordnung von Evaluationsformen	16
3.2	Wer kann evaluieren?	17
3.2.1	Evaluatoren	17
3.2.2	Ordnung von Evaluatoren	18
4.	METHODEN UND VERFAHRENSSCHRITTE EINER EVALUATION	20
4.1	Welche Methoden sind für Evaluationen geeignet?	20
4.1.1	Erhebungsmethoden	20
4.1.2	Auswertungsmethoden	21
4.2	Welche Schritte geht man in einem Evaluationsprojekt?	22
4.2.1	Planungsphase	22
4.2.2	Durchführungsphase	23
4.2.3	Verwertungsphase	24
5.	EVALUATION MEDIENGESTÜTZTER LERNANGEBOTE	25
5.1	Mediengestützte Lernangebote als Evaluationsgegenstand	25
5.1.1	Spezifizierung des Gegenstands	25
5.1.2	Gegenstandsspezifische Evaluationsarten	26
5.2	Usability-Tests als Evaluationsverfahren	28
6.	KOMMENTIERTE LITERATURÜBERSICHT	31
6.1	Monografien und Sammelbände	31
6.2	Kapitel aus Hand- und Lehrbüchern	33
6.3	Online verfügbare Dokumente, Portale und Web-Seiten	35
	LITERATUR	37

0 EINFÜHRUNG IN DEN STUDIENTEXT

0.1 ZIELE DES STUDIENTEXTES

Wie ist dieser Studientext entstanden?

Das Thema Evaluation spielt seit jeher entweder in Form eigener Veranstaltungen oder als Veranstaltungskomponente eine wichtige Rolle in meiner Lehre. Immer wieder war und bin ich auf der Suche nach einem Standardwerk, das ich den Studierenden als besonders geeignet für unsere (bildungswissenschaftlichen bzw. mediendidaktischen) Zwecke empfehlen kann. Immer wieder gebe ich diese Suche auf. Damit will ich nicht sagen, dass es keine guten Bücher gibt. Im Gegenteil: Aufgrund meiner sich ständig wiederholenden Suche stoße ich immer wieder auf interessante neue Lektüre. Für die mir wichtigen Ziele aber sind diese Funde in der Regel entweder zu umfangreich oder aber zu einseitig. In der Folge stelle ich stattdessen – wie das wohl viele machen – Reader mit einzelnen Artikeln oder Teilen aus weitgehend passenden Büchern zusammen und versuche, die Klammer zwischen diesen oft recht verschiedenen Fragmenten in Präsenzveranstaltungen durch Inputphasen (inklusive Folien) zu vermitteln. Im Herbst 2010 hatte ich im Rahmen einer Veranstaltung allerdings besonders deutlich das Gefühl, dass auf diesem Weg trotz aller Bemühungen das Verständnis seitens der Studierenden auf der Strecke geblieben ist. In der Folge habe ich den Versuch gemacht, meine Folien zu verbalisieren und möglichst einfach als Text verfügbar zu machen. Dabei bin ich zu einer neuen Struktur gekommen, mit der ich die Grundidee der Evaluation sowie mögliche Vorgehensweisen im Überblick deutlich machen will. Herausgekommen ist dieser knappe Studientext.

Was kann der Studientext leisten?

Der Studientext ersetzt kein Grundlagenwerk und keine Lehrveranstaltung – dazu sind die Auswahl und Ausführungen einzelner Aspekte der Evaluation eindeutig viel zu kurz. Der Text kann als *eine* Komponente in bildungswissenschaftlichen Einführungsveranstaltungen zum Thema Evaluation verwendet werden. Es handelt sich um eine relativ reduzierte Zusammenstellung der aus meiner Sicht wichtigsten Begriffe und Konzepte, um ein erstes Verständnis zu erlangen. Ich selbst verwende den Studientext im Rahmen einer Präsenzveranstaltung, die den Studierenden im ersten (kürzeren) Teil einige theoretische Grundlagen liefert und sie im zweiten (längeren) Teil bei einem eigenen kleinen Evaluationsprojekt anleitet. Die ersten Erfahrungen mit dem Studientext zeigen, dass dieser dabei hilft, so etwas wie ein mentales Modell vom Thema Evaluation zu entwickeln, und ein Grundgerüst für erste eigene „Gehversuche“ in Sachen Evaluation liefert.

Welche inhaltliche Breite wird abgedeckt?

Das Themenfeld Evaluation ist sehr weit und zunächst ist es naheliegend, eine Eingrenzung vorzunehmen. Diese würde bei unserem Lehrgebiet in der *Evaluation mediengestützter Lernangebote* liegen. Ich habe mich allerdings *dagegen* entschieden, diese Spezifizierung zu Beginn zu machen. Vielmehr habe ich mich dazu entschlossen, das Thema Evaluation trotz der damit verbundenen Schwierigkeiten aus folgenden Gründen zunächst *genereller* zu behandeln: (a) Die Evaluation mediengestützter Lernangebote lässt sich auf diese Weise besser in den Stand der Evaluationsforschung einordnen. (b) Viele inhaltsunabhängige Begriffe, Modelle und Methoden aus dem (interdisziplinären) Umfeld der Evaluation können so auch für die Evaluation mediengestützter Lernangebote genutzt werden. (c) Die speziell auf Medien ausgerichtete Literatur ist letztlich gering ausgeprägt; vorhandene Quellen beschränken sich häufig auf Usability-Fragen und beziehen die hier im Interesse stehenden pädagogisch-didaktischen Fragen weniger ein. Der erweiterte Fokus ermöglicht daher letztlich eine bessere Grundlage.

0.2 AUFBAU DES STUDIENTEXTES

Welche Aspekte der Evaluation werden behandelt?

In der Literatur werden verschiedenste Evaluationsmodelle vorgeschlagen. Diese „Modelle“ liefern in der Regel unterschiedliche Vorschläge für die Eingrenzung und Beschreibung von Evaluationsgegenständen und listen verschiedene Bewertungskriterien und -maßstäbe auf. Zudem werden unterschiedliche Designs und Methoden in diese Modelle integriert. Gerade für den Novizen in Sachen Evaluation sind die vorhandenen Modelle, die noch dazu alle auf unterschiedlichen logischen Ebenen liegen und unterschiedlich abstrakt bzw. konkret sind, eher verwirrend als hilfreich. Auch für praktische Fragen der Evaluation (und damit für erste eigene Evaluationsversuche) ist dies eher ungeeignet, weshalb in diesem Studientext eine Darstellung entlang der wichtigsten Fragen erfolgt, die es zu klären gilt, wenn man erstmals ein Evaluationsprojekt in Angriff nehmen will: (1) *Wozu* kann man evaluieren? (2) *Was* kann man evaluieren? (3) *Wie* kann man evaluieren? (4) *Wer* kann evaluieren? (5) *Welche* Methoden sind für Evaluationen geeignet? (6) *Welche* Schritte geht man in einem Evaluationsprojekt? Vorangestellt werden Klärungen zum Begriff der Evaluation inklusive einer kurzen Erörterung, in welchem Verhältnis die Evaluation zum Begriff des Qualitätsmanagements steht. Erst nach der Beantwortung der skizzierten W-Fragen werden die Besonderheiten der Evaluation mediengestützter Lernangebote sowie Usability-Tests als spezielle Form der Evaluation behandelt.

Wie ist der Studientext strukturiert?

Der Studientext gliedert sich in insgesamt sechs kurze Kapitel: *Kapitel 1* beschäftigt sich mit dem Begriff der Evaluation und Evaluationsforschung. *Kapitel 2* bündelt die beiden Fragen nach dem Wozu und Was und beschreibt die wichtigsten Zwecke und Inhalte von Evaluationen inklusive Ausführungen zu Bewertungskriterien und -maßstäben. *Kapitel 3* beschäftigt sich mit dem Design und den Verantwortlichen von Evaluationen und gibt damit Antworten auf die Fragen nach dem Wie und Wer. Dabei wird unter anderem dargelegt, welche Formen von Evaluation man unterscheiden kann und welche Personen als Evaluatoren in Frage kommen. Noch einmal genauer aufgeschlüsselt wird die Frage nach dem Wie in *Kapitel 4*, das Methoden und Verfahrensschritte von Evaluationen aufzeigt. Beschrieben wird das Vorgehen beim Umgang mit Daten für die Evaluation sowie beim Management verschiedener Evaluationsphasen. *Kapitel 5* widmet sich den Besonderheiten der Evaluation mediengestützter Lernangebote. Hier wird dann auch kurz auf die Usability-Forschung eingegangen. Das letzte Kapitel (*Kapitel 6*) schließlich bietet eine kommentierte Literaturübersicht an: Es wird eine Auswahl von Monografien, Sammelbänden, Kapiteln aus Hand- und Lehrbüchern sowie von online verfügbaren Dokumenten, Portalen und Web-Seiten vorgestellt. Die Kommentare sind subjektiver Natur – es handelt sich um „Bewertungen“ vor dem Hintergrund meiner Erfahrung speziell aus der Lehre zum Thema Evaluation. Wichtig ist dieses Kapitel, weil es aufzeigt, wie man sein Wissen erweitern und vertiefen kann. Die Quellen für die Ausführungen der Kapitel 2 bis 5 stammen zum großen Teil, aber nicht ausschließlich, aus der kommentierten Literaturübersicht.

1 BEGRIFF DER EVALUATION

1.1 WAS IST EVALUATION UND EVALUATIONSFORSCHUNG?

Wann ist eine Evaluation Evaluationsforschung?

Evaluation ist ein Begriff, der auch in der Alltagssprache gebraucht und dort für alle *bewertenden* Prozesse verwendet wird. Dieser Aspekt der Bewertung trifft auch auf die *Evaluationsforschung* bzw. auf Evaluation in einem wissenschaftlichen Sinne zu, was sich in verbreiteten Definitionen von Evaluation z.B. als "the process of determining the merit, worth, or significance of things" (Scriven, 1996, pp. 158) zeigt. In der Regel wird diese Umschreibung für die Evaluationsforschung allerdings als nicht ausreichend angesehen. Zum Aspekt der Bewertung hinzukommen müssen nämlich:

- ein systematisches *Vorgehen*, das transparent und nachvollziehbar ist und
- sich auf wissenschaftliche *Methoden* der Erhebung und Auswertung stützt,
- ein eingegrenzter, empirisch beschreibbarer *Gegenstand* einschließlich
- einer präzisen *Fragestellung*, die sich operationalisieren lässt, sowie
- *Personen* (Evaluatoren) mit angemessener Expertise und Legitimation.

Diese Merkmale gelten als Möglichkeit, die Evaluationsforschung von Evaluation als Vorgang ohne wissenschaftlichen Anspruch abzugrenzen. In diesem Studientext geht es nur um die Evaluationsforschung. Wird daher der Kürze wegen von Evaluation gesprochen, ist *Evaluationsforschung* gemeint.

Was zeichnet die Evaluationsforschung methodisch aus?

Evaluationsforschung lässt sich von anderen Forschungsdesigns *nicht* durch ihre Verwendung von Methoden unterscheiden. Grundsätzlich können für wissenschaftliche Evaluationen vielmehr *alle* Erhebungs- und Auswertungsmethoden herangezogen werden, die in den Sozialwissenschaften oder (bezogen auf das Thema Lehren und Lernen) in den Bildungswissenschaften genutzt werden, nämlich (siehe Abschnitt 4.1):

- Befragungen (mündlich, schriftlich, offen, geschlossen etc.),
- Beobachtungen (offen, verdeckt, strukturiert, unstrukturiert etc.),
- Tests (Wissens-, Intelligenz-, Einstellungstests etc.),
- Dokumentenanalysen (Sprache, Zahlen, analog, digital etc.),
- experimentelle Verfahren (echte Experimente, Quasi-Experimente) und
- Kriterienlisten/-kataloge (für Experten- oder Nutzerurteile).

Die letztgenannten Kriterienlisten bzw. -kataloge sind speziell bei der Evaluation von Medien ein gängiges Verfahren und werden nicht nur, aber besonders für Usability-Tests (siehe Abschnitt 5.2) eingesetzt. Zur Auswertung bieten sich – je nach Fragestellung und Art der Erhebung – sowohl quantitativ-statistische als auch qualitativ-interpretative Methoden an. Evaluationsforschung, so kann man folglich festhalten, unterscheidet sich von anderen Formen der Forschung *nicht* durch besondere Methoden, sondern allein durch den *Zweck*, eine Maßnahme oder Intervention zu *bewerten*.

Zu welchem Forschungstypus gehört die Evaluationsforschung?

Folgt man der Einteilung von wissenschaftlicher Forschung in Grundlagenforschung und angewandter Forschung (oder Anwendungsforschung), dann gilt die Evaluationsforschung als ein Teilbereich der *angewandten* Forschung: Auch wenn aus Evaluationsstudien wissenschaftliche Erkenntnisse resultieren können, sind Ziel und Anlage von Evaluationen in der Regel darauf ausgerichtet, einen *praktischen* Nutzen zu stiften und – im Bereich Bildung – Lernangebote und/oder deren Implementierung zu legitimieren oder zu verbessern. Damit sind in der Regel einige besondere Herausforderungen verbunden, die man in der grundlagenorientierten Forschung nicht antrifft:

- Die Komplexität des Untersuchungsfeldes muss möglichst erhalten bleiben, um praktisch relevante Folgerungen ziehen zu können.
- Außerwissenschaftliche Interessen beeinflussen die Wahl des Gegenstands und können auch das Verfahren tangieren und Aushandlungsprozesse erforderlich machen.
- Das Vorhaben ist in ein zeitliches Raster eingebunden und unterliegt immer auch praktischen Erfordernissen.

Gibt es eine konsensfähige Definition von Evaluation?

Eine gängige Definition, welche alle in diesem ersten Abschnitt genannten Aspekte prägnant umfasst, stammt von Wottawa (2006): Für ihn ist Evaluation „die systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung des Konzepts, des Designs, der Umsetzung und des Nutzens sozialer Interventionsprogramme. Evaluatoren nutzen sozialwissenschaftliche Forschungsmethoden, um die Art und Weise, in der Gesundheits-, Bildungs- und andere soziale Interventionsmaßnahmen durchgeführt werden, zu beurteilen und zu verbessern, angefangen bei der Planungsphase bis hin zur Entwicklung und Umsetzung eines Programms“ (Wottawa, 2006, S. 662). Diese Definition verweist zugleich auf den Umstand, dass Evaluationen an verschiedenen Punkten einer Intervention, eines Projekts etc. ansetzen *können*, nämlich an Zielen und Ausgangspunkten ebenso wie an Prozessen und Ergebnissen bzw. Wirkungen. Mit diesen verschiedenen Gegenständen der Evaluation (Frage nach dem Was) beschäftigt sich Abschnitt 2.2.1. Zuvor aber gilt es, die begrifflichen Klärungen abzuschließen und kurz darzulegen, in welchem Verhältnis die Evaluation zum Qualitätsmanagement steht, das in der Praxis (auch in der Bildungspraxis) einen hohen Stellenwert hat.

1.2 WAS HAT EVALUATION MIT QUALITÄTSMANAGEMENT ZU TUN?

Was versteht man unter Qualitätsmanagement?

Qualitätsmanagement ist ein Managementansatz, der sich im Kontext der Wirtschaft entwickelt hat (z.B. Reinmann, 2004): In den 1950er Jahren interessierte man sich lediglich dafür, ob ein Produkt verkaufsfähig war oder ob es nachgebessert werden musste. Mit Hilfe der Wahrscheinlichkeitsrechnung wurde Qualität berechenbar, indem man stichprobenartig die Produktqualität überprüfte. In den 1970er Jahren rückte die Qualität der Herstellungsprozesse in den Blickpunkt; es entwickelte sich die Idee der Qualitätssicherung mittels Normen. Ende der 1980er Jahre schließlich etablierte sich der Gedanke eines umfassenden Qualitätsmanagements; nun wurde neben Produkten und Dienstleistungen sowie Herstellungsprozessen auch das gesamte Umfeld der Organisation einbezogen. Die allmähliche Ablösung des Terminus Qualitätssicherung durch Qualitätsmanagement symbolisiert, dass es nicht mehr nur um normgerechte Produkte und die Einhaltung von Standards geht, sondern immer mehr um die Entwicklung neuer Qualitätsmerkmale und kontinuierliche Verbesserung (vgl. Wottawa, 2006). Qualitätsmanagement kann man heute als übergeordneten Begriff verstehen, der Qualitätssicherung und Qualitätsentwicklung als zwei Komponenten umfasst. Die damit verbundenen Verfahren im Umgang mit Qualität finden nicht nur in Unternehmen, sondern inzwischen auch in Non Profit-Organisationen und auf dem öffentlichen Sektor Verwendung.

Welche Rolle spielt die Evaluation im Qualitätsmanagement?

Das Gemeinsame zwischen der Evaluation und dem Qualitätsmanagement ist der Aspekt der Bewertung. Eine Evaluation kann so angelegt sein, dass sie der Qualitätssicherung oder -entwicklung dient. Im Unterschied zur Evaluationsforschung ist der Ansatz des Qualitätsmanagements allerdings breiter (Stockmann, 2007a): Speziell in Unternehmen sind beispielsweise auch das *Controlling* (komplexere Formen des Rechnungswesens), das *Benchmarking* (der Vergleich eigener Produkte oder Prozesse mit den besten Wettbewerben) oder *Audits* („Anhörungen“ zur Überprüfung, ob Anforderungen erfüllt werden) Elemente des Qualitätsmanagements. Im Gegensatz zu solchen und anderen Qualitätsmanagement-Verfahren sind Evaluationen in einem wissenschaftlichen Sinne nicht nur praktischen Erfordernissen, sondern zusätzlich wissenschaftlichen Standards (insbesondere der angewandten Forschung) unterworfen und können vor diesem Hintergrund immer *auch* der Erkenntnis dienen. Evaluation hat also auf jeden Fall etwas mit Qualitätsmanagement zu tun: Eine Evaluation kann im Rahmen des Qualitätsmanagements eingesetzt werden und dazu beitragen, Qualität zu sichern und zu entwickeln. Evaluation darf aber nicht mit Qualitätsmanagement gleichgesetzt werden.

2 ZWECK UND INHALTE EINER EVALUATION

2.1 WOZU KANN MAN EVALUIEREN?

Welche Evaluationszwecke werden unterschieden?

Verschiedene Autoren (z.B. Kromrey, 2005; Stockmann & Meyer, 2010) sind sich weitgehend darin einig, dass man mindestens drei Zwecke von Evaluationen unterscheiden kann, nämlich (a) Kontrolle und Legitimation, (b) Verbesserung und Weiterentwicklung sowie (c) Erkenntnis und wissenschaftlichen Fortschritt. Mitunter werden diese Zwecke auch als Leitkonzepte der Evaluation oder als Evaluationsparadigmen bezeichnet. Manchmal findet man auch vier Zwecke, wobei mit zunehmender Anzahl der Unterscheidungen die Trennschärfe kleiner wird. Kontrolle und Legitimation, Verbesserung und Weiterentwicklung sowie Erkenntnis und wissenschaftlicher Fortschritt schließen sich als Zwecke nicht aus. Kombinationen verschiedenster Art sind durchaus möglich. Dennoch ist es für die Evaluationspraxis wichtig, sich Rechenschaft über den *primären Zweck* abzulegen, da dieser Einfluss auf nachfolgende Entscheidungen über Evaluationsformen, Evaluatoren sowie Erhebungs- und Auswertungsmethoden haben kann.

Was bedeuten Kontrolle und Legitimation als Zweck?

Kontrolle und Legitimation bilden einen auch im Alltagsverständnis von Evaluation oft beabsichtigten Zweck. Die Implementation und der Erfolg einer Maßnahme werden erfasst und bewertet, um zu kontrollieren, ob vorab gesetzte Ziele erreicht wurden, oder um eine Investition zu legitimieren bzw. weitere Entscheidungen (z.B. über Erhalt, Ausbau oder Einstellung einer Intervention) zu treffen. Mit der Evaluation unter dem Leitkonzept „Kontrolle und Legitimation“ will man vorrangig einen Beitrag zur Planungsrationalität leisten. Es kommt in der Regel eine kritisch-rationale Forschungslogik mit tendenziell quantitativen Methoden (siehe Abschnitt 4.1) zum Einsatz. Viele zentral organisierte Lehrevaluationen dienen z.B. primär der Kontrolle und Legitimation.

Was bedeuten Verbesserung und Weiterentwicklung als Zweck?

Verbesserung und Weiterentwicklung sind für die Bildungspraxis und deren Qualitätsbestrebungen ein wesentlicher und häufig angestrebter Evaluationszweck. Das zugrundeliegende Konzept, ausgelöste Prozesse und erreichte Ergebnisse werden erfasst und bewertet, um Ansatzpunkte für notwendige Modifikationen zu finden. Oft beginnt die Evaluation hier bereits in der Planungsphase eines Projekts und integriert einen Dialog mit den Beteiligten. Mit der Evaluation unter dem Leitkonzept „Verbesserung und Weiterentwicklung“ will man vorrangig Problemlösungen optimieren. Es kommt in der Regel eine sozial-konstruktivistische Forschungslogik mit tendenziell qualitativen Methoden (siehe Abschnitt 4.1) zum Einsatz. Evaluationen im Kontext der Entwicklung neuer Bildungsmedien verfolgen z.B. primär den Zweck der Verbesserung und Weiterentwicklung.

Was bedeuten Erkenntnis und wissenschaftlicher Fortschritt als Zweck?

Erkenntnis und wissenschaftlicher Fortschritt stellen einen dritten Evaluationszweck dar, der zu eher praktisch orientierten Zwecken hinzutreten kann, in der Regel aber nicht *allein* verfolgt wird. Dies wäre auch infolge der Zuordnung der Evaluationsforschung zur angewandten Forschung wenig konsistent. Allerdings trifft man hier in der Literatur durchaus auf unterschiedliche Auffassungen. Unter Nutzung von Vorgehensweisen der Feldforschung oder Entwicklungsforschung ist eine Kombination von Erkenntnis und Kontrolle oder von Erkenntnis und Verbesserung durchaus möglich. Mit der Evaluation unter dem Leitkonzept „Erkenntnis und wissenschaftlicher Fortschritt“ will man vorrangig die Wissensbasis einer wissenschaftlichen Disziplin erweitern. Prinzipiell können eine kritisch-rationale ebenso wie eine sozial-konstruktivistische Forschungslogik und damit quantitative und qualitative Methoden zum Einsatz kommen (vgl. Flick, 2006).

2.2 WAS KANN MAN EVALUIEREN?

2.2.1 EVALUATIONSGEGENSTÄNDE

Welche Gegenstände eignen sich zur Evaluation?

Man kann das „Was“ einer Evaluation als Evaluationsgegenstand (analog zum Forschungsgegenstand) bezeichnen. Prinzipiell gibt es bei der Wahl von Evaluationsgegenständen kaum Einschränkungen: Gesetze, Produkte, Dienstleistungen, Organisationen, Personen, Evaluationen selbst etc. können evaluiert und damit zum Evaluationsgegenstand werden. Im Kontext des mediengestützten Lehrens und Lernens können Medien an sich (z.B. Learning und Content Management-Systeme), digitale Lernangebote zum Selbstlernen (z.B. Computer- und Web-Based Trainings), Blended Learning-Angebote (also Seminare mit Präsenz- und E-Learning-Elementen), didaktische Konzepte hinter Selbstlern- oder Blended Learning-Angeboten, E-Learning-Strategien, Abteilungen zur Gestaltung von Medien und medialen Lernangeboten und vieles mehr zu Evaluationsgegenständen gemacht werden. Um Evaluationsgegenstände (bisher neutral als Maßnahme oder Intervention bezeichnet) klarer beschreiben zu können, gibt es in der Literatur verschiedene begriffliche Vorschläge und Einteilungen, die auch in der Praxis hilfreich sein können.

Eignet sich die Größe einer Maßnahme als Einteilungskriterium?

Wenn Evaluationsgegenstände prinzipiell nicht eingeschränkt sind, liegt auf der Hand, dass sich diese in ihrer Größe, ihrem Umfang und ihrer Komplexität erheblich unterscheiden können. Größe, Umfang und Komplexität sind denn auch in der Tat eine erste Möglichkeit, Evaluationsgegenstände zu ordnen (Stockmann & Meyer, 2010): So kann man z.B. Einzelmaßnahmen, Projekte, Programme und Strategien (Policy) unterscheiden, die man jeweils evaluieren kann, dabei aber vor recht unterschiedlichen Herausforderungen steht:

- Mit einer *Einzelmaßnahme* ist eine konkrete, gut abgrenzbare Aktivität oder ein Produkt gemeint: z.B. ein Learning Management Systems (LMS) als Produkt oder der Einsatz eines LMS in einer universitären Veranstaltung als Aktivität. Es handelt sich hier um die kleinste Interventionseinheit.
- Mit *Projekt* wird ein Bündel von Einzelmaßnahmen bezeichnet: z.B. die Nutzung eines LMS in Kombination mit einer Veränderung der Veranstaltungsdidaktik und der Prüfungsform, bei der mehrere Veranstaltungen einbezogen werden.
- Unter einem *Programm* versteht man eine Reihe aufeinander bezogener Projekte: z.B. die Einführung digitaler Medien in Lehrveranstaltungen zusammen mit der Nutzung digitaler Medien in der Verwaltung und in der Forschung in Form einzelner, aber zusammenhängender Teilprojekte.
- Mit *Policy* schließlich ist eine spezifische Handlungsstrategie oder „Vision“ mit Angabe von Zielen einer Organisation gemeint: z.B. eine E-Learning-Strategie oder ein neues Leitbild an einer Hochschule.

Gibt es noch andere Einteilungsmöglichkeiten?

Neben diesem inhaltlich recht unspezifischen Vorschlag, Evaluationsgegenstände nach ihrer Größe und Komplexität einzuteilen, werden speziell im *Bildungskontext* noch weitere Spezifizierungen zur genaueren Eingrenzung, Beschreibung und damit auch Ordnung von Evaluationsgegenständen empfohlen. Eine dieser Empfehlungen besteht darin, einzelne Komponenten eines Evaluationsvorhabens als potenzielle Evaluationsgegenstände zu definieren. Dabei handelt es sich um (a) die *Bedingungen*, unter denen eine Maßnahme implementiert wird, (b) die *Maßnahme an sich* in Form eines Konzepts oder Produkts, (c) die *Implementation* (Umsetzung) einer Maßnahme und (d) die mit einer Maßnahme erzielten *Wirkungen*. Diese Einteilung ist weit verbreitet und findet sich auch in vielen Evaluationsmodellen. Oft werden daraus verschiedene Evaluationsformen abgeleitet, die dann aber genau genommen verschiedene Evaluationsgegenstände definieren (siehe auch Abschnitt 5.1.2).

Was bedeutet es, wenn Bedingungen evaluiert werden?

Es ist allgemein bekannt, dass speziell in sozialen und Bildungsprojekten und -programmen nicht nur die Maßnahmen an sich einflussreich sind, sondern dass auch die Bedingungen, unter denen eine Maßnahme zum Einsatz kommt, eine wichtige Rolle z.B. für den Erfolg oder Misserfolg einer Maßnahme spielen. Zu den Bedingungen gehören zunächst einmal allgemeine soziale, politische und kulturelle, aber auch konkrete situationale Bedingungen. Diese bilden den Kontext einer Maßnahme. Evaluiert man vor allem diese Kontextbedingungen, kann man das als *Kontextevaluation* bezeichnen. Davon separat betrachten kann man die gezielt in die Maßnahme mit einfließenden Ressourcen und Voraussetzungen. Das bezeichnet man meist als den Input und meint damit die zusätzlich zum gegebenen Kontext einfließenden Faktoren. Evaluiert man diese, spricht man entsprechend von einer *Inputevaluation*.

Was bedeutet es, wenn Maßnahmen an sich evaluiert werden?

Mit der „Maßnahme an sich“ ist das gemeint, was implizit bei der Einführung in das Thema Evaluation wie auch in den meisten Evaluationsdefinitionen gemeint ist, wenn von Interventionen oder eben Maßnahmen die Rede ist und man den Evaluationsgegenstand noch nicht genauer eingegrenzt hat. Speziell in Bildungskontexten können hier das einer Maßnahme zugrundeliegende Konzept und/oder das zu implementierende didaktische oder technische Produkt gemeint sein. Evaluiert man das Konzept einer Maßnahme, kann man dies als *Konzeptevaluation* bezeichnen. Ist dagegen das sichtbare Produkt (z.B. ein LMS, eine konkrete Lernumgebung, ein Lehrtext oder ein Aufgabenpool) Gegenstand der Evaluation, handelt es sich um eine *Produktevaluation*. Wichtig ist, dass es hier um die Konzepte und Produkte an sich geht und (noch) nicht um deren Einsatz und die damit erzielten Wirkungen.

Was bedeutet es, wenn die Implementation evaluiert wird?

Unter Implementation versteht man die Umsetzung eines Konzepts oder den Einsatz eines Produkts. Bei der Implementation hat man folglich Prozesse im Blick. Man bezeichnet Evaluationen, die speziell die Implementation einer Maßnahme genauer betrachten, daher auch als *Prozessevaluation*. Eine Prozessevaluation lässt sich noch weiter differenzieren: Genau genommen nämlich kann man (a) die Prozesse bei der Planung (noch vor der eigentlichen Umsetzung), (b) die Prozesse direkt bei der Umsetzung (also die Implementation im engeren Sinne) und/oder (c) die Prozesse bei der Überprüfung der Implementation (Evaluation) evaluieren. Letzteres mutet auf den ersten Blick seltsam an (eine Evaluation der Evaluation). Da aber Evaluationen störanfällig sind, ist es durchaus sinnvoll, diese Prozesse im Bedarfsfall ebenfalls zu überprüfen.

Was bedeutet es, wenn Wirkungen evaluiert werden?

Zu den Wirkungen einer Maßnahme kann man alle möglichen Resultate und Effekte zählen, die man wiederum nach mehreren Aspekten einteilen kann. Eine solche Einteilung ist aus verschiedenen Perspektiven möglich: Man kann z.B. (a) Wirkungen auf Personen von Wirkungen auf Organisationen unterscheiden, (b) quantitative (zählbare) Ergebnisse von qualitativen Ergebnissen, (c) kurzfristige Wirkungen von langfristigen Wirkungen und (d) beabsichtigte Wirkungen von nicht-intendierten Wirkungen etc. Sind Resultate und Effekte dieser Art Gegenstand der Evaluation, bezeichnet man das meist als *Wirkungsevaluation*, manchmal auch als *Ergebnisevaluation*. Oft werden Wirkungen zusammen mit anderen Gegenständen evaluiert, beispielsweise zusammen mit Konzepten und Produkten, die man dann zusätzlich von Experten bewerten lässt.

2.2.2 BEWERTUNGSKRITERIEN

Was versteht man unter Bewertungskriterien?

Für eine Evaluation ist es essenziell, dass der Evaluationsgegenstand genau beschrieben und eingegrenzt wird, da alle weiteren Entscheidungen von dieser Beschreibung und Eingrenzung abhängen. Da sich die Evaluationsforschung gegenüber anderen Forschungstypen durch den Akt des Bewertens auszeichnet, kann man die Frage nach dem Was nicht nur mit der Angabe des Evaluationsgegenstands an sich beantworten. Vielmehr muss man sich auch Gedanken dazu machen, *in Bezug auf was genau* der nun beschriebene und eingegrenzte Gegenstand bewertet werden soll: Man kann beispielsweise eine Einzelmaßnahme wie ein LMS als Produkt daraufhin analysieren, ob es von den Nutzern akzeptiert wird oder ob es einen bestimmten Bedarf deckt. Ein Programm wie die Einführung digitaler Medien in mehrere Fakultäten einer Universität kann man daraufhin evaluieren, wie wirksam das für den Lernerfolg ist oder ob erzielte Effekte auch nachhaltig sind. Akzeptanz und Bedarf bzw. Wirksamkeit und Nachhaltigkeit wären dann die Evaluationskriterien. Aufgrund der Vielfalt möglicher Evaluationsgegenstände gibt es *keinen* feststehenden Kanon an Bewertungskriterien. Im Kontext der Bildungswissenschaften haben sich aber eine Reihe zumindest typischer Kriterien etabliert. Dabei ist die Unterscheidung zwischen Evaluationsgegenständen und Bewertungskriterien nicht immer trennscharf: Sind z.B. Lerneffekte, also Wirkungen, der Evaluationsgegenstand, dann sind Bewertungskriterien wie Akzeptanz und Lernerfolg logisch schwer vom Gegenstand zu trennen. Dies ist nur möglich, indem man z.B. in der Akzeptanz und im Lernerfolg eine *Wertung* (nämlich eine positive Wirkung) sieht.

Welche Bewertungskriterien kann man heranziehen?

Prinzipiell ist man bei der Formulierung von Evaluationskriterien frei: Sowohl Auftraggeber (im Falle einer Fremdevaluation) als auch die Evaluatoren selbst können die Bewertungskriterien letztlich beliebig festlegen, wobei in einer wissenschaftlichen Evaluation die Kriterien sinnvoll und nachvollziehbar zu begründen sind. Die hier aufgenommenen Evaluationskriterien sind nur eine Auswahl, die sich speziell für die Evaluation mediengestützter Lernangebote eignet bzw. häufig anzutreffen ist. Selbstredend müssen die Kriterien zum Gegenstand passen, dem angestrebten Zweck der Evaluation entsprechen und prinzipiell für eine empirische Überprüfung geeignet sein. Ich habe versucht, die Bewertungskriterien einzuteilen, um die Übersicht zu erleichtern: (a) Kriterien, die typische Bedingungsfaktoren eines (erfolgreichen) Bildungs- oder Lernprozesses im Blick haben (Akzeptanz, Lernerfolg, Lerntransfer und das wahrgenommene Verhältnis von Aufwand und Nutzen), (b) Kriterien, die eher Aspekte der betroffenen Bildungsorganisation fokussieren (Bedarfsdeckung, Nachhaltigkeit und Kosten-Ertrag-Bilanz), sowie (c) Kriterien, die sich auf Qualitätsaspekte bei der Entwicklung und Umsetzung einer Maßnahme beziehen (Konzept-, Produkt- und Implementationsqualität).

Wie sehen lernbezogene Bewertungskriterien aus?

Ein Lernangebot, das nicht akzeptiert wird, ist nahezu wertlos, weshalb *Akzeptanz* in vielen Evaluationen im Bereich der Bildung ein wichtiges Bewertungskriterium darstellt. Akzeptanz als Kriterium ist mit der Frage verbunden, wie eine Intervention, ein Projekt, ein Programm oder eine Policy (im Folgenden wieder zusammenfassend als Maßnahme bezeichnet) angenommen wird bzw. wie hoch (oder niedrig) die Zufriedenheit damit ist. Des Weiteren ist man in der Regel daran interessiert, dass eine Maßnahme effektiv ist, also im Falle einer Bildungsmaßnahme ein *Lernerfolg* resultiert. Lernerfolg (Effektivität) als Kriterium ist mit der Frage verbunden, welche Effekte eine Maßnahme hat und inwieweit sie dem Ziel entsprechen; dabei sind auch Neben- oder Folgewirkungen zu berücksichtigen. Eine Bildungsmaßnahme soll zudem nicht nur unmittelbar prüfbare Lernerfolge bewirken, sondern auch dazu beitragen, dass das Gelernte außerhalb der Bildungssituation angewandt und übertragen wird. *Lerntransfer* als Kriterium ist mit der Frage verbunden, ob erzielte Erfolge auf andere Kontexte übertragen werden können. Schließlich kann man auch die *individuelle Aufwand-Nutzen-Bilanz* als lernbezogenes Bewertungskriterium mit aufnehmen. Diese ist mit der Frage verbunden, wie hoch oder niedrig bei einer Maßnahme der persönliche Nutzen im Vergleich zum Aufwand ist.

Wie sehen organisationale Bewertungskriterien aus?

Nicht alle Bewertungskriterien, auch nicht bei der Evaluation von Bildungsmaßnahmen, beziehen sich auf das Lernen und den Lernerfolg. Wenn z.B. ein Programm evaluiert wird, kann es wichtig sein, darauf zu achten, dass damit auch ein Bedarf gedeckt wird. *Bedarfsdeckung* als Kriterium ist mit der Frage verbunden, ob bzw. inwieweit die Maßnahme ein Problem löst, eine Lücke füllt oder anderweitig zu rechtfertigen ist. Das Pendant zum Lerntransfer auf der individuellen Ebene ist eine generelle *Nachhaltigkeit*. Nachhaltigkeit als Kriterium ist mit der Frage verbunden, ob und inwieweit eine Maßnahme auf langfristigen Nutzen ausgelegt ist, Wirkungen also über aktuelle Effekte hinausgehen. Ein häufig verwendetes organisationales Bewertungskriterium ist die *organisationale Kosten-Ertrag-Bilanz* (Effizienz). Dieses Kriterium ist mit der Frage verbunden, wie hoch oder niedrig bei einer Maßnahme der Ertrag im Vergleich zu den investierten Kosten ist.

Inwiefern können Qualitätsaspekte zu Bewertungskriterien werden?

Wenn Maßnahmen an sich (Konzepte oder Produkte) oder deren Implementation evaluiert werden, können neben den schon genannten Bewertungskriterien auch deren Eigenschaften (Qualität) zum Bewertungskriterium gemacht werden. *Konzeptqualität* als Kriterium ist mit der Frage verbunden, wie gut oder schlecht (begründet) das (inhaltliche, didaktische und technische) Konzept einer Maßnahme ist. *Produktqualität* als Kriterium ist in der Regel mit der Frage verbunden, wie angemessen ein technisches System/Werkzeug oder ein für sich stehendes Lernangebot (didaktisches Produkt) ist. *Implementationsqualität* als Kriterium ist mit der Frage verbunden, wie gut bzw. passend (oder schlecht bzw. unpassend) eine Maßnahme umgesetzt wurde oder wird.

2.2.3 BEWERTUNGSMABSTÄBE

Was sind Bewertungsmaßstäbe?

Zur Bewertung von Maßnahmen benötigt man neben Bewertungskriterien immer auch Bewertungsmaßstäbe, denn: Die Frage, wie z.B. Akzeptanz, ein Lernerfolg oder Lerntransfer, eine Bedarfsdeckung oder Produktqualität zu interpretieren, also positiv, neutral oder negativ einzuschätzen sind, lässt sich niemals absolut bestimmen. Vielmehr benötigt man *Bezugspunkte*, mit denen man erhobene Evaluationsdaten vergleichen kann. Solche Bezugspunkte zum Vergleich, um zu einer Bewertung kommen zu können, nennt man Bewertungsmaßstäbe. Zur eigentlichen Bewertung, die eine Evaluation erfordert, kommt man also immer nur durch Vergleiche. Dabei kann man zwei Gruppen von Vergleichen unterscheiden: Vergleiche, bei denen man Ähnliches zueinander in Beziehung setzt (zeitliche, örtliche und Systemvergleiche) und Vergleiche, bei denen die Vergleichsmaßstäbe extra konstruiert wurden (Vergleiche mit selbst gesetzten Zielen und bereits bestehenden Standards).

Was sind zeitliche, örtliche und Systemvergleiche?

Der *zeitliche Vergleich* als Maßstab setzt die erzielten Evaluationsergebnisse mit Ergebnissen bzw. Leistungen in der Vergangenheit in Beziehung. Die Frage ist also: Wie war es früher oder vorher und wie ist es jetzt? Der *örtliche Vergleich* als Maßstab setzt die erzielten Evaluationsergebnisse mit Ergebnissen bzw. Leistungen in anderen Abteilungen, regionalen, nationalen oder internationalen Einrichtungen in Beziehung. Die Frage ist also: Wie ist es woanders und wie ist es hier? Der *Systemvergleich* als Maßstab kommt vor allem bei der Evaluation von technischen Systemen oder Werkzeugen zum Einsatz und setzt die erzielten Evaluationsergebnisse mit Ergebnissen bzw. Leistungen anderer, vergleichbarer Systeme oder Werkzeuge in Beziehung. Die Frage ist also: Was leistet System A im Vergleich zu System B, C etc.?

Was sind Vergleiche mit gesetzten Zielen und Standards?

Der *Vergleich mit gesetzten Zielen* als Maßstab setzt die erreichten Evaluationsergebnisse mit den Ergebnissen bzw. Leistungen in Beziehung, die man selbst als Ziel oder Soll gesetzt bzw. erwartet hat. Die Frage ist also: Wie waren die Ziele und welche davon sind erreicht worden bzw. wie weit hat man sich diesen angenähert? Der *Vergleich mit Standards* als Maßstab setzt die erreichten Evaluationsergebnisse mit den Ergebnissen bzw. Leistungen in Beziehung, die als akzeptierte Standards gelten, sofern diese vorhanden sind. Wann allerdings ein gesetztes Ziel bereits ein Standard ist, lässt sich immer nur in der jeweiligen Fach-Community klären. Die Frage ist also: Wie lauten die Standards und wie gut werden diese erfüllt? In Bezug auf Standards lassen sich Minimal- und Optimal-Standards unterscheiden: Bei Minimal-Standards fragt man danach, ob die erreichte Leistung so ist, dass sie Mindestansprüchen genügt. Bei Optimal-Standards geht es darum, ob die erreichte Leistung im Vergleich zu einem Optimalwert so hoch wie möglich ist.

3 DESIGN UND VERANTWORTLICHE EINER EVALUATION

3.1 WIE KANN MAN EVALUIEREN?

3.1.1 EVALUATIONSFORMEN

Was ist hier mit Evaluationsformen gemeint?

Die Frage nach dem Wie in einer Evaluation ist zunächst einmal unspezifisch und kann sich auf das Design (im Sinne der Versuchsanordnung) ebenso wie auf die Methoden und Projektschritte bei einer Evaluation beziehen. Im Folgenden wird die Frage nach dem Wie unter Rückgriff auf *typische Evaluationsformen* beantwortet, die man in der Literatur findet. Der Schwerpunkt wird auf die Unterscheidung zwischen formativen und summativen Evaluationen einerseits und internen und externen Evaluationen andererseits gelegt (Scriven, 1991). In der Literatur werden häufig auch Spezifizierungen des Gegenstandes als verschiedene Evaluationsformen bezeichnet: z.B. Inputevaluation, Prozessevaluation, Produktevaluation oder Wirkungsevaluation. Dies ist auf der einen Seite durchaus nachvollziehbar, weil die unterschiedlichen Gegenstände Einfluss auf das Vorgehen haben. Auf der anderen Seite erweist sich genau das oft als verwirrend speziell für Novizen, weshalb in dieser Einführung die skizzierte Einschränkung auf summative, formative, interne und externe Evaluationen als verschiedene Evaluationsformen erfolgt und die genannten weiteren „Evaluationsformen“ beim Abschnitt zum Evaluationsgegenstand bereits besprochen wurden (vgl. Abschnitt 2.2).

Wie entscheidet man sich für eine Evaluationsform?

In den meisten Fällen ist es so, dass der Evaluationszweck sowie der Evaluationsgegenstand bereits einen „Pfad“ gelegt haben, auf dem man weitergeht; man hat dann nur mehr wenig Spielraum, diesen zu verlassen: Besteht z.B. der primäre Zweck darin, eine Bildungsmaßnahme zu kontrollieren, wird man sich für eine Fremdevaluation entscheiden, hat aber noch Spielraum in der Entscheidung für eine interne oder externe Evaluation. Umgekehrt gilt, dass jemand, der den Zweck verfolgt, eine Bildungsmaßnahme zu verbessern, eine interne Evaluation bevorzugen wird und im Falle eines komplexeren Entwicklungsszenarios zunächst eine Selbstevaluation durchführen wird. Auch der Gegenstand spielt bei der Wahl der Evaluationsform eine große Rolle: Ein großes hochschulpolitisches Programm, in das viel Geld investiert wurde, verlangt nach einer externen Evaluation und wohl auch nach einer Fremdevaluation. Ein kleines selbst finanziertes Projekt dagegen wird man lieber intern evaluieren wollen und sich dann für eine Selbst- oder Fremdevaluation entscheiden. Es gibt also einige „Pfadentscheidungen“ in einem wissenschaftlichen Evaluationsvorhaben, ohne dass diese allerdings determiniert sind.

3.1.1 ORDNUNG VON EVALUATIONSFORMEN

Was genau ist eine summative Evaluation?

Eine summative Evaluation lässt sich in der Regel durch mehrere Merkmale kennzeichnen: Sie findet nach Beendigung einer Maßnahme statt (*zeitliches Merkmal*) und bezieht sich in der Regel auf einen Gegenstand, der bereits fertig entwickelt ist (*Merkmal Entwicklungsstand*). Die Adressaten einer summativen Evaluation sind vorrangig (aber nicht ausschließlich) diejenigen Personen, die über eine Maßnahme (z.B. deren Fortführung oder Einstellung) zu entscheiden haben, ohne direkt Akteure der Maßnahme zu sein (*Merkmal Adressat*). In der Folge ist der Zweck der Evaluation häufig die Kontrolle und/oder Legitimation (*Merkmal Zweck*), wobei ein zusätzlicher Erkenntnisgewinn wissenschaftlicher Art nicht ausgeschlossen ist. Summative Evaluationen gelten als zusammenfassend, bilanzierend und ergebnisorientiert (Stockmann & Mayer, 2010). Das Pendant zur summativen Evaluation ist die formative Evaluation. Summative Evaluationen sind im Kontext des mediengestützten Lehrens und Lernens vor allem dann wichtig, wenn hohe Investitionen (etwa in teure Technik) getätigt wurden. Summativen Charakter haben außerdem vor allem solche Evaluationen, bei denen z.B. verschiedene Systeme oder Lernangebote miteinander verglichen werden, um für einen bestimmten Kontext oder Zweck *eines* auszuwählen (vgl. Glowalla, Herder, Süße & Koch, 2009).

Was genau ist eine formative Evaluation?

Eine formative Evaluation lässt sich ebenfalls durch mehrere Merkmale kennzeichnen: Sie findet während einer Maßnahme statt (*zeitliches Merkmal*) und bezieht sich in der Regel auf einen Gegenstand, der sich noch in der Entwicklung befindet (*Merkmal Entwicklungsstand*). Die Adressaten einer formativen Evaluation sind die Akteure selbst, also diejenigen Personen, die eine Maßnahme planen, gestalten und/oder durchführen und auf diesem Wege die Maßnahme modifizieren können (*Merkmal Adressat*). In der Folge ist der Zweck der Evaluation in der Regel die Entwicklung und/oder Verbesserung (*Merkmal Zweck*), wobei auch hier ein zusätzlicher Erkenntnisgewinn wissenschaftlicher Art nicht ausgeschlossen ist. Formative Evaluationen gelten als aktivgestaltend, kommunikationsfördernd und prozessorientiert (Stockmann & Mayer, 2010). Das Pendant zur formativen Evaluation ist die summative Evaluation. Formative Evaluationen sind im Kontext des mediengestützten Lehrens und Lernens häufig: Vor allem dann, wenn innovative und damit in der Regel aufwändige Lernangebote entwickelt werden, dienen formative Evaluationen dazu, diese in Pilotphasen schrittweise zu optimieren. In diesen (dem „Echtbetrieb“) vorgeschalteten Phasen werden die Nutzerzahlen in der Regel klein gehalten (vgl. Glowalla et al., 2009).

Was genau ist eine interne Evaluation?

Eine interne Evaluation ist eine Evaluation, bei welcher der Evaluator bzw. das Evaluationsteam aus der Organisation stammt, in der eine Maßnahme durchgeführt, geplant und umgesetzt wird. Das muss nicht zwingend mit einer Selbstevaluation zusammenfallen (siehe Abschnitt 3.2). Möglich ist auch, dass in einer internen Evaluation eine Fremdevaluation stattfindet, nämlich z.B. dann, wenn die Organisationsleitung die Maßnahme einer eigenen Abteilung evaluiert. Das Pendant zur internen Evaluation ist die externe Evaluation. Interne Evaluationen haben sowohl Vor- als auch Nachteile (Stockmann, 2007b, S. 61): Die *Vorteile* einer internen Evaluation bestehen darin, dass der zeitliche und finanzielle Aufwand in der Regel vergleichsweise überschaubar bleibt, eine hohe Sachkenntnis zum Evaluationsgegenstand vorhanden ist und eine große Chance besteht, dass die Evaluationsergebnisse auch umgesetzt werden. Die *Nachteile* einer internen Evaluation sind, dass eine gewisse „Betriebsblindheit“ nicht ausgeschlossen werden kann, die Evaluatoren nicht zwingend methodisch kompetent sind und möglicherweise nicht distanziert und unabhängig arbeiten und urteilen können.

Was genau ist eine externe Evaluation?

Eine externe Evaluation ist eine Evaluation, bei welcher der Evaluator oder das Evaluationsteam nicht zur Organisation gehört, in der eine Maßnahme durchgeführt, geplant und umgesetzt wird. Eine externe Evaluation ist immer eine Fremdevaluation. Das Pendant zur externen Evaluation ist die interne Evaluation. Auch externe Evaluationen haben Vor- und Nachteile (Stockmann, 2007b, S. 61): Die *Vorteile* einer externen Evaluation bestehen darin, dass man die Evaluationskompetenz der Evaluatoren ebenso wie deren Unabhängigkeit leichter sicherstellen kann, dass sie meist als glaubwürdig erlebt werden und auf diesem Wege Reformer in einer Organisation gut unterstützen können. Die *Nachteile* einer externen Evaluation sind, dass zum Evaluationsgegenstand möglicherweise wenig Sachkenntnis vorhanden ist, dass die Arbeit der externen Evaluatoren Abwehrreaktionen hervorruft und die Evaluationsergebnisse nicht anerkannt und/oder umgesetzt werden.

3.2 WER KANN EVALUIEREN?

3.2.1 EVALUATOREN

Wer kann ein Evaluator sein?

Evaluatoren sind Personen, die eine Evaluation durchführen, also diese planen, indem sie (eventuell zusammen mit dem Auftraggeber) die Ziele festlegen und den Gegenstand spezifizieren, eine oder mehrere Evaluationsformen auswählen, sich für bestimmte Erhebungs- und Auswertungsmethoden entscheiden, die Evaluation schließlich durchführen und darüber Bericht erstatten.

Man kann Evaluatoren danach unterteilen, ob sie (a) die zu evaluierende Maßnahme auch selbst zu verantworten haben oder durchführen (Selbstevaluation) oder (b) dies nicht der Fall ist (Fremdevaluation). Immer dann, wenn man auch Personen außerhalb des eigentlichen Evaluationsteams heranzieht, um ein Maßnahme zu bewerten, kommen dazu prinzipiell entweder (c) Nutzer (etwa von mediengestützten Lernangeboten) oder (d) Experten (etwa für Medientechnik und/oder -didaktik) in Frage (vgl. auch Brandt, 2007). Im weitesten Sinne sind diese Personen auch „Evaluatoren“, aber in einer anderen Rolle.

Wozu braucht man Evaluationskompetenz?

In allen Fällen (Selbst- oder Fremdevaluation, Evaluation durch Nutzer oder Experten) ist zu beachten, welche Kompetenzen diese Personen wofür genau mitbringen (Ist-Zustand) und welche sie aus welchen Gründen mitbringen sollten (Soll-Zustand). Speziell bei Selbstevaluationen gilt es, kritisch zu überprüfen, ob das eigene Wissen und Können wirklich ausreicht, und wie man sich im Bedarfsfall bei seinem Evaluationsvorhaben unterstützen lassen könnte. Bei Fremdevaluationen wird man zwar in der Regel auf Experten zugreifen; der Kontext (z.B. Bildungs- und Medienkontext) kann aber durchaus besondere Kompetenzen verlangen, die man vorher sicherstellen sollte. Werden Nutzer als Evaluatoren tätig, ist es wichtig, im Blick zu haben, was diese als „Alltagsevaluatoren“ gut beurteilen können und was nicht bzw. von welchen Faktoren sie sich (wahrscheinlich) bei ihrer Bewertung leiten lassen. Dies gilt im Prinzip auch bei Experten, bei denen darauf zu achten ist, dass eine deutliche Passung zwischen ihrer Expertise einerseits und dem Gegenstand und den Zielen der Evaluation andererseits besteht.

3.2.2 ORDNUNG VON EVALUATOREN

Was genau ist eine Selbstevaluation?

Bei der Selbstevaluation führen diejenigen Personen, welche die Maßnahmen planen und/oder umsetzen (Akteure), auch die Evaluation durch, erheben also selbst Daten, werten diese aus und ziehen daraus Folgerungen. Eine Selbstevaluation ist immer eine interne Evaluation. Trotzdem sind die beiden Begriffe (wie dies oft zu lesen ist) *nicht* gleichzusetzen, da sie sich auf verschiedene Dimensionen beziehen. Selbstevaluationen bieten sich im Rahmen von formativen Evaluationen an, wenn sich Entwicklungen noch in einem frühen Stadium befinden und eher wenige Ressourcen vorhanden sind. Selbstevaluationen können aber auch integraler Bestandteil von Qualitätsmanagementkonzepten sein (Hense & Mandl, 2006). Selbstevaluationen sind einerseits fehleranfällig, da man lieber eigene Erfolge als Misserfolge wahrnimmt. Sie können aber andererseits einen hohen Lern- und Veränderungseffekt auslösen.

Was genau ist eine Fremdevaluation?

Bei der Fremdevaluation führen Personen die Evaluation durch, die nicht an der Planung und/oder Umsetzung der zu evaluierenden Maßnahme beteiligt sind. Akteure und Evaluatoren sind hier klar getrennte Personengruppen. Eine Fremdevaluation ist im Rahmen sowohl einer externen als auch einer internen Evaluation möglich. Umgekehrt gilt aber, dass eine externe Evaluation immer eine Fremdevaluation ist. Fremdevaluationen gelten einerseits als „objektiver“, weil man die Fehleranfälligkeit von Selbstevaluationen vermeidet. Andererseits können sie aber auch Abwehrreaktionen auslösen. Es besteht die Möglichkeit Selbst- und Fremdevaluationen gezielt miteinander zu verbinden. Wenn allein von Evaluation die Rede ist, ist in den meisten Fällen eine Fremdevaluation gemeint, bei der zwischen Evaluatoren und Akteuren (oder Evaluierten) *keine* Personalunion besteht.

Was genau ist eine Evaluation durch Nutzer?

Speziell bei der Evaluation von mediengestützten Lernangeboten ist häufig der Nutzer als bewertende Person – und damit in gewissem Sinne als Evaluator – gefragt. Die Besonderheit bei dieser Betrachtungsweise der „Evaluation durch Nutzer“ besteht darin, dass dieser „Evaluator“ die Evaluation *nicht* plant, konzipiert und auswertet. Vielmehr wird er um sein Nutzerurteil, also eine subjektive Einschätzung etwa der Qualität einer Maßnahme, gebeten und fungiert damit als Datenquelle für die Evaluation. Für diese Nutzerurteile werden in der Regel Kriterienlisten bzw. -katalogen zur Verfügung gestellt. Relativ häufig kommen Nutzerurteile im Rahmen von Usability-Tests zum Einsatz (vgl. Abschnitt 5.2). Prinzipiell können aber auch andere Aspekte der Qualität einer Maßnahme auf diesem Wege analysiert und bewertet werden (weshalb man auch von Qualitätsanalysen spricht).

Was genau ist eine Evaluation durch Experten?

Neben dem (potenziellen) Nutzer werden häufig Experten für die Evaluation von mediengestützten Lernangeboten herangezogen und um Expertenurteile gebeten. Auch der Experte ist bei dieser Betrachtungsweise der „Evaluation durch Experten“ *nicht* derjenige, der die Evaluation plant, konzipiert und auswertet, sondern stellt eher eine Datenquelle dar. Für Experteneinschätzungen werden in der Regel ebenfalls Kriterien vorgegeben. Je nach Art der Expertise dieser „Evaluatoren“ werden mit solchen Expertenurteilen bzw. expertengestützten Qualitätsanalysen (auch als Expertenratings bezeichnet) neben der Usability vor allem inhaltliche und didaktische Aspekte der Qualität von Maßnahmen eingeschätzt.

4 METHODEN UND VERFAHRENSSCHRITTE EINER EVALUATION

4.1 WELCHE METHODEN SIND FÜR EVALUATIONEN GEEIGNET?

Hat man ein breites Wissenschaftsverständnis, das eine Pluralität von wissenschaftlichen Strategien und Methoden (zur Datenerhebung und -auswertung) umfasst, gilt auch für die Evaluationsforschung, dass prinzipiell *alle* Methoden zum Einsatz kommen können (vgl. auch Abschnitt 1). Entscheidend für die Methodenwahl sind das Evaluationsziel und die dahinter stehende Fragestellung, aber auch der Evaluationsgegenstand und die verfügbaren (finanziellen, zeitlichen, personellen) Ressourcen und Rahmenbedingungen. Als angewandte Forschung muss die Evaluationsforschung die Umfeld-Bedingungen in jedem Fall mitberücksichtigen. Bei der Methodenfrage ist es sinnvoll, Erhebung und Auswertung zunächst einmal getrennt voneinander zu betrachten. Mit allen unten stehenden Erhebungsformen lassen sich sowohl numerische als auch nicht-numerische Daten erheben, die dann allerdings entsprechend ihrer Datenqualität unterschiedlich auszuwerten sind.

4.1.1 ERHEBUNGSMETHODEN

Wie kann man Daten erheben?

Es gibt unzählige Instrumente, um Daten zu erheben, die sich aber in der Regel alle einer oder mehreren der folgenden *Grundformen* der Datenerhebung zuordnen lassen. Durch verschiedene Ausprägungen auf mehreren Dimensionen, mit denen man jede dieser Grundformen charakterisieren kann, entsteht eine Vielzahl verschiedener Erhebungsmethoden:

- *Befragungen* mit den Dimensionen: (a) mündlich (face-to-face, Telefon, Audiokanal am Computer) oder schriftlich (paper-pencil, digital), (b) offen, halboffen oder geschlossen, (c) einzeln (z.B. Interview) oder in der Gruppe (z.B. Gruppendiskussion).
- *Beobachtungen* mit den Dimensionen: (a) offen (mit Wissen der Beobachteten) oder verdeckt (ohne Wissen der Beobachteten), (b) teilnehmend oder nicht-teilnehmend, (c) unstrukturiert, teilstrukturiert oder strukturiert.
- *Tests* mit den Dimensionen: (a) standardisiert (z.B. Intelligenztests) oder nicht-standardisiert (eigene Wissenstests), (b) bezogen auf Wissen, Kompetenz, Motivation, Einstellung oder andere Bereiche.
- *Dokumentenanalysen* mit den Dimensionen (a) Sprachdokumente oder Zahlendokumente (Statistiken für eine Sekundäranalyse), (b) Text-, Audio- oder Video-Format der Dokumente, (c) analoge oder digitale Dokumente, (d) quantitative oder qualitative Analyseverfahren.
- *Experimentelle Verfahren* mit den Dimensionen: (a) echtes Experiment oder Quasi-Experiment, (b) Laborexperiment oder Feldexperiment.
- *Kriterienlisten* mit den Dimensionen: (a) für Expertenurteile oder für Nutzerbeurteilungen, (b) offen, teilstrukturiert oder geschlossen.

Wie unterscheiden sich diese Erhebungsmethoden?

Die aufgelisteten Erhebungsmethoden unterscheiden sich zum Teil erheblich und beeinflussen auf unterschiedliche Weise das, was man evaluieren will. Das wird besonders deutlich, wenn man die Methoden danach einteilt, ob es sich um reaktive oder non-reaktive Methoden handelt: *Reaktive Methoden* sind solche, bei denen man mit der Erhebung bzw. Erfassung von Daten bereits einen Einfluss als Evaluator ausübt. Dies geschieht bei allen Formen der Befragung, allen Tests und Kriterienlisten, Laborexperimenten sowie bei Beobachtungen, in denen der Beobachter als Externer fungiert und die Beobachteten das wissen. In all diesen Settings reagieren die für die Evaluation relevanten Personen auf die eingesetzten Erhebungsmethoden, was dazu führen kann, dass neue oder andere Effekte ausgelöst werden als dies ohne die Erhebung der Fall gewesen wäre. *Non-reaktive Methoden* dagegen sind solche, bei denen man mit der Erhebung bzw. Erfassung von Daten als Evaluator *keinen* direkten Einfluss ausübt: Dies ist der Fall bei verdeckten Beobachtungen sowie bei allen Dokumentenanalysen. Beim Einsatz dieser Methoden erfolgt seitens des Evaluators keine Intervention speziell zu Erhebungszwecken, auf welche die für die Evaluation relevanten Personen reagieren könnten.

4.1.2 AUSWERTUNGSMETHODEN

Kommt nach der Erhebung gleich die Auswertung?

Zwischen der Erhebung und Auswertung liegt die Aufbereitung der erhobenen Daten – ein Schritt, der gerne unterschätzt wird (Hug & Poscheschnik, 2010, S. 133 f). Die erhobenen Daten können in unterschiedlicher Form vorliegen: als Texte, Zahlen, Audio- oder Videoaufnahmen etc. Diese Daten müssen vor der eigentlichen Auswertung aufbereitet werden. Dazu gehört, dass man die Daten z.B. im Hinblick auf mögliche Fehlerquellen sichtet, eventuell auswählt (weil man nicht alles auswerten will), neu ordnet oder in eine andere Form bringt. Letzteres ist besonders häufig bei Interviews der Fall: Hier gilt es, die beispielsweise als Audioaufnahmen vorliegenden Interviewdaten vor der Auswertung zu transkribieren. Bei der Anwendung qualitativer Auswertungsmethoden gehen Aufbereitung und Auswertung fließend ineinander über.

Wie kann man Daten auswerten?

Wie man erhobene Daten auswertet, ist zum einen vom Zweck und Ziel der Untersuchung sowie ihrer Fragestellung abhängig. Zum anderen beeinflusst die Qualität der Daten die Auswertung (vgl. z.B. Schlömerkemper, 2010):

- Für die *Auswertung numerischer Daten* (bzw. quantitativer Informationen) werden statistische Auswertungsmethoden herangezogen. Hier kommen häufig Methoden zum Einsatz, mit denen man (a) Kennwerte für einzelne Merkmale ermittelt (z.B. Häufigkeiten, zentrale Tendenzen, Streuungen etc.), (b) Beziehungen zwischen Variablen und Personen berechnet (z.B. Kovarianzen, Korrelationen, Regressionen etc.) und (c) statistische Befunde beurteilt (etwa mit der Berechnung der statistischen Signifikanz oder der Effektstärke).

- Für die *Auswertung nicht-numerischer Daten* (bzw. qualitativer Informationen) werden hermeneutisch-interpretative Auswertungsmethoden herangezogen. In der Regel kommen dabei Methoden zum Einsatz, die sich dazu eignen (a) Textmaterial (z.B. Interview-Transkripte oder bereits bestehende Dokumente) explorativ oder anhand vorgegebener Fragestellungen zu analysieren oder (b) Phänomene (z.B. Beobachtungsergebnisse oder bestehende Dokumente) unter Einbezug eines Vorverständnisses oder möglichst unvoreingenommen zu deuten.

Was kommt nach der Auswertung?

Nachdem Daten erhoben, aufbereitet und ausgewertet wurden, geht es darum, diese zu interpretieren bzw. zu diskutieren. Dieser Punkt ist in der Evaluationsforschung besonders wichtig, denn es geht ja per definitionem um eine Bewertung. An der Stelle wird deutlich, warum neben der Eingrenzung des Evaluationsgegenstands und der genauen Festlegung von Bewertungskriterien die *Bewertungsmaßstäbe* (vgl. Abschnitt 2.2.3) so wichtig sind. Sie nämlich liefern die entscheidende Grundlage für eine Interpretation der Daten, die wiederum die Basis für die Bewertung bildet, die am Ende der gesamten Evaluation erforderlich ist.

4.2 WELCHE SCHRITTE GEHT MAN IN EINEM EVALUATIONS-PROJEKT?

Versucht man, übliche Evaluationsprojekte in Phasen und Schritte zu unterteilen, so lassen sich (in der Literatur mit variierenden Bezeichnungen) in der Regel drei Phasen unterscheiden: eine Planungsphase, eine Durchführungsphase und eine Verwertungsphase (vgl. Stockmann, 2007).

4.2.1 PLANUNGSPHASE

Welche Schritte kennzeichnen die Planungsphase?

In der Planungsphase geht es darum, das Evaluationsvorhaben einzugrenzen und ein Evaluationskonzept zu erarbeiten. Im Einzelnen sollten folgende Schritte gegangen werden: (1) Den Evaluationszweck und einzelne Ziele festlegen. (2) Den Evaluationsgegenstand spezifizieren. (3) Entsprechend der Ziele eine oder mehrere konkrete Fragestellungen formulieren. (4) Bewertungskriterien und -maßstäbe festlegen. (5) Sich für die Evaluationsform (formativ oder summativ, intern oder extern) entscheiden, falls dies nicht schon vorgegeben ist. (6) Die Evaluatoren bestimmen. (7) Die bestehende Ressourcenausstattung prüfen. (8) Erhebungs- und Auswertungsmethoden festlegen. (9) Einen Ablaufplan für die Evaluation erstellen. (10) Einen Zeit-, Personal- und Budgetplan erarbeiten. Diese Form von Planung bei Evaluationsvorhaben unterscheidet sich nicht prinzipiell von anderen Projektplanungen. Ausgesprochen wichtig für *alle* weiteren methodischen Entscheidungen sind Schritt (2) bis (4), also die genaue Spezifizierung des Evaluationsgegenstands, die Fragestellung sowie die Bewertungskriterien und -maßstäbe.

Welche Probleme können in der Planungsphase auftreten?

Es kann schwierig sein, den primären Zweck einer Evaluation zu bestimmen, wenn (wie es oft der Fall ist) mehrere verschiedene Ziele verfolgt werden. Lösen kann man dieses Problem, indem man genau diese Schwierigkeit festhält und transparent macht. Die größte Stolperfalle bei der Planung besteht darin, dass man den Evaluationsgegenstand zu wenig spezifiziert. Meist sind dann auch die Bewertungskriterien und -maßstäbe unklar. In der Folge ergeben sich Probleme sowohl bei der Erhebung als auch bei der Auswertung von Daten. Schwierigkeiten können auch bei der Wahl der Evaluatoren auftreten, wenn im Rahmen eines Evaluationsvorhabens verschiedene Interessen konkurrieren. Schließlich ist es ein häufiger Fehler in der Planung, dass die bestehende Ressourcenausstattung für den erstellten Zeit-, Personal- und Budgetplan an sich zu gering ist: In der Folge hat man bei der Durchführung mit Zeitproblemen, eventuell auch mit finanziellen Problemen zu kämpfen.

4.2.2 DURCHFÜHRUNGSPHASE

Welche Schritte kennzeichnen die Durchführungsphase?

In der Durchführungsphase geht es darum, seinen Plan zu realisieren. In dieser Phase müssen im Einzelnen die erforderlichen Instrumente für die Erhebung und Auswertung von Daten ausgewählt oder entwickelt werden, um anschließend die angestrebten Daten zu erheben und auszuwerten. Im Einzelnen sollten folgende Schritte gegangen werden: (1) Einen Evaluationsleitfaden für den Projektablauf erarbeiten. (2) Einen Plan für die Erhebung der Daten machen. (3) Die Erhebungsinstrumente auswählen oder entwickeln und testen (Pretest). (4) Die Zielgruppe, die man untersucht, bzw. die erforderlichen Datenquellen festlegen. (5) Die Bewertungskriterien und -maßstäbe ausformulieren. (6) Die Datenerhebung durchführen. (7) Die Daten aufbereiten und auswerten. (8) Die Ergebnisse bewerten, diskutieren und Empfehlungen ableiten.

Welche Probleme können in der Durchführungsphase auftreten?

Zunächst einmal kann man festhalten, dass es selten der Fall ist, einen Plan im Rahmen einer Evaluation vollständig und ohne Abstriche in die Tat umzusetzen. Es sind nicht zwingend Planungsfehler, die dazu führen, dass die Durchführung in manchen Punkten anders als nach Plan verläuft, denn: Vor allem Forschungen im Feld (und genau das ist bei Evaluationen ja meistens der Fall) müssen immer mit unerwarteten Ereignissen rechnen. Dennoch gibt es ein paar Stolperfallen, die man reduzieren oder vermeiden kann, wenn man diese kennt oder darauf vorbereitet ist: Ein häufiger Fehler ist z.B. die mangelnde Überprüfung von Erhebungsinstrumenten (Pretest). Wenn man erst in der eigentlichen Durchführung feststellt, dass das Aufnahmegerät nicht funktioniert oder die gestellten Fragen nicht verstanden werden (um nur Beispiele zu nennen), ist es für Korrekturen zu spät. Unterschätzt wird zudem, eine geeignete Zielgruppe zu finden und für die Evaluation auch zu gewinnen, sodass es an der Stelle schnell zu Zeitproblemen kommen kann.

4.2.3 VERWERTUNGSPHASE

Welche Schritte kennzeichnen die Verwertungsphase?

In der Verwertungsphase – einer Phase, die für Evaluationen typisch und wichtig ist – gilt es, die Evaluationsergebnisse und abgeleitete Empfehlungen zu präsentieren, Maßnahmen zur Nutzung der Ergebnisse und Empfehlungen in die Wege zu leiten und die Qualität der eigenen Evaluation sicherzustellen. Im Einzelnen sollten folgende Schritte gegangen werden: (1) Einen *Evaluationsbericht* mit klarer Struktur und Zielgruppenorientierung verfassen. (2) Ergebnisse und Empfehlungen in mündlichen *Präsentationen* vorstellen. (3) Durch Maßnahmen wie ansprechende Web-Seiten, Publikationen, Tagungsbeiträge etc. zur *Verbreitung* der Evaluationsergebnisse beitragen. (4) Eventuell die *Umsetzung* der Evaluationsergebnisse und Empfehlungen unterstützen. (5) Die Qualität der eigenen Evaluation z.B. durch Verwendung der Standards der Deutschen Gesellschaft für Evaluation (DeGEval, 2004) sicherstellen, die vier übergeordnete Leitstandards (mit insgesamt 25 Einzelstandards) postuliert: (a) Nützlichkeit, (b) Durchführbarkeit, (c) Fairness und (d) Genauigkeit.

Was genau ist ein Evaluationsbericht?

Wichtig ist, sich vor Augen zu halten, dass ein Evaluationsbericht die Evaluationsergebnisse an diejenigen kommunizieren soll, die anhand der Resultate weitere Entscheidungen treffen (in Abhängigkeit vom Zweck der Evaluation). Der Evaluationsbericht sollte daher systematisch und präzise entsprechend der Regeln wissenschaftlichen Schreibens (z.B. Kruse, 2010), aber auch allgemein verständlich formuliert sein. Es ist empfehlenswert, den Bericht folgendermaßen zu strukturieren: (1) Vorangestellte Zusammenfassung (Executive Summary). (2) Kontext der Evaluation: Zweck, Ziele und Fragestellung(en). (3) Evaluationsgegenstand. (4) Evaluationsart (Design) und Methoden. (5) Ergebnisse (chronologisch oder thematisch gegliedert). (6) Diskussion und Empfehlungen. Die Ergebnisse sollten nüchtern und ohne Wertung sowie anschaulich (bei statistischen Daten ergänzt durch geeignete Grafiken) dargestellt werden. Die Diskussion sowie die darauf aufbauenden Empfehlungen müssen nachvollziehbar sein und einen deutlichen Bezug zu den Ergebnissen aufweisen. Dasselbe gilt für die mündliche Berichterstattung in Form einer Präsentation.

Welche Probleme können in der Verwertungsphase auftreten?

Es versteht sich von selbst, dass man Probleme bekommt, wenn der schriftliche oder mündliche Evaluationsbericht ungenau, unvollständig oder unsachlich ist. Aber selbst bei korrekter Berichterstattung muss man damit rechnen, dass Evaluationsergebnisse und abgeleitete Empfehlungen auf Kritik stoßen und/oder abgelehnt werden. Typische Kritikmuster nach Stockmann (2010, S. 176 ff.) sind: (a) Die Ergebnisse werden als „schon bekannt“ beurteilt. (b) Das Design und die Methoden werden in Zweifel gezogen. (c) Die Ergebnisse werden schlicht geleugnet („Es kann nicht sein, was nicht sein darf“). (d) Es erfolgt eine akribische Fehlersuche. (e) Die Durchführung wird bemängelt.

5 EVALUATION MEDIENGESTÜTZTER LERNANGEBOTE

5.1 MEDIENGESTÜTZTE LERNANGEBOTE ALS EVALUATIONS- GEGENSTAND

5.1.1 SPEZIFIZIERUNG DES GEGENSTANDS

Was kennzeichnet mediengestützte Lernangebote als Gegenstand?

Die meisten der bislang dargestellten Inhalte gelten für Evaluationen im sozialwissenschaftlichen und bildungswissenschaftlichen Bereich generell. Einige der angeführten Konkretisierungen haben sich zudem bereits auf den Kontext „Lehren und Lernen mit (digitalen) Medien“ bezogen. Dennoch erscheint es sinnvoll, in einem letzten Schritt noch einmal auf die Besonderheiten der Evaluation mediengestützter Lernangebote einzugehen, die sich aus dem Gegenstand, also dem Lehren und Lernen einerseits und dem Einsatz (in der Regel digitaler) Medien andererseits ergibt. Evaluation im Kontext des Lehrens und Lernens mit (digitalen) Medien ist ein Oberbegriff für ganz verschiedene Evaluationen, nämlich: (a) die Evaluation von technischen Systemen und Werkzeugen oder von Selbstlernangeboten wie CBTs und WBTs, also *Einzelmaßnahmen*, (b) die Evaluation von mediengestützter Lehre (in dem Sinne, dass digitale Medien und neue didaktische Konzepte in den Aus- oder Fortbildungsalltag integriert werden) oder mediengestützter Angebote etwa zu Präventions- oder Förderzwecken außerhalb der Lehre, was in der Regel die Komplexität von *Projekten* hat. (c) Darüber hinaus kann man natürlich auch ganze *Programme* und *Policies*, die aus Maßnahmen und Projekten resultieren oder diesen übergeordnet sind, evaluieren, worauf aber in diesem Papier nicht weiter eingegangen wird (vgl. z.B. Seufert & Euler, 2005).

Gibt es noch andere Ordnungsmöglichkeiten?

Neben dieser Ordnung nach Umfang und Komplexität möglicher Gegenstände bei der Evaluation mediengestützten Lehrens und Lernens ist es für praktische Belange hilfreich, die unter Abschnitt 2.2.1 dargelegte Differenzierung heranzuziehen. Danach kann man (a) die *Bedingungen* (Kontext und Input) analysieren, unter denen ein Lernangebot implementiert wird, (b) das *Lernangebot an sich* in Form eines Konzepts oder Produkts, (c) die *Implementation* (Umsetzung) eines Lernangebots und (d) die damit erzielten *Wirkungen*. In der Folge kann man je nach Entscheidung bei der Festlegung des Evaluationsgegenstands von Kontextevaluation, Inputevaluation, Konzeptevaluation, Produktevaluation, Prozessevaluation und/oder Wirkungsevaluation sprechen. Auf diese Weise signalisiert man bereits mit der Evaluationsbezeichnung, welchen Gegenstands bzw. welchen Aspekt des Gegenstands man bei der Evaluation vorrangig im Blick hat.

5.1.2 GEGENSTANDSSPEZIFISCHE EVALUATIONSARTEN

Was ist eine Kontextevaluation?

Zum Kontext des Lehrens und Lernens mit Medien gehören Faktoren oder Aspekte, die meist nicht direkt oder zeitnah verändert und beeinflusst werden können. Vielmehr stellen diese den Rahmen für die im Interesse stehende Maßnahme dar und können daher Gegenstand einer Kontextevaluation sein. Vor allem dann, wenn es z.B. darum geht, Transferprojekte zu evaluieren, bei denen man bereits bestehende Produkte oder Konzepte in neue Kontexte implementieren will, können Kontextevaluationen relativ hilfreich sein. Dabei kann man Folgendes berücksichtigen: (a) die *Lehr-Lernkultur* in einer Organisation oder Teilbereichen einer Organisation im Sinne von Auffassungen zum Lehren und Lernen, Einstellungen gegenüber digitalen Medien und deren Nutzung für Bildungsprozesse, Gewohnheiten und Routinen in der Lehr-Lernpraxis, explizite Leitbilder und ähnliches, (b) die *räumlichen Gegebenheiten* (Gebäude, Räume für unterschiedliche Lehr-Lernprozesse, Mobiliar etc.) inklusive der damit verbundenen Atmosphäre, (c) die *institutionellen Gegebenheiten* mit ihren Restriktionen und Freiräumen z.B. was Lehrpläne, Zeitstrukturen, personelle und technische Grundausstattung etc. betrifft.

Was ist eine Inputevaluation?

Zum Input in das Lehren und Lernen mit Medien gehören Faktoren oder Aspekte, die nicht bereits integraler Bestandteil eines Konzepts oder Produkts sind, aber von außen zusätzlich in die Maßnahme einfließen. Anders als beim Kontext handelt es sich hier nicht um den Rahmen, sondern um Ressourcen und Voraussetzungen, die man gezielt in die Intervention mit hineinnimmt oder hineinnehmen muss (wobei die Übergänge zum Kontext fließend sind). In der Regel sind folgende Inputvariablen bei der Evaluation mediengestützter Lernangebote wichtig: (a) die *Lernvoraussetzungen* der Personen, für die das Lernangebot bestimmt ist, wozu z.B. inhaltliches Vorwissen, überfachliche (Lern-)Fähigkeiten, Vorerfahrungen in der Mediennutzung, aber auch Einstellungen und Motive gehören können, (b) die *technischen Voraussetzungen* bei den beteiligten Personen und Organisationen, in denen das Lernangebot umgesetzt werden soll, (c) die *personellen und zeitlichen Ressourcen*, die in die Umsetzung des mediengestützten Lernangebots einfließen.

Was ist eine Konzeptevaluation?

Mit dem Konzept für ein Lehren und Lernen mit Medien sind Entwürfe und Pläne – man könnte auch sagen Designs – gemeint, die in gewisser Weise ein Soll bzw. Ziele definieren und sowohl diese als auch den Weg dorthin möglichst begründen. Man kann ein solches Konzept analytisch in die nachfolgenden Teilkonzepte untergliedern, wobei man diese sinnvollerweise miteinander kombiniert und aufeinander bezieht:

(a) Das *technische Konzept* legt die für das Lernangebot verwendeten Systeme und Werkzeuge dar und erläutert deren Funktionen für Lern-, Betreuungs- und/oder Assessmentprozesse. (b) Das *didaktische Konzept* kann aus mehreren Elementen bestehen: aus einem Vermittlungskonzept (Inhalte), einem Aktivierungskonzept (Aufgaben und Interaktionen) und einem Betreuungskonzept: Beim *Vermittlungskonzept* sind entsprechend die Inhaltsauswahl und -sequenzierung, die Korrektheit der Inhalte sowie die (mediale) Aufbereitung von Interesse. Beim Aktivierungskonzept geht es um Aufgabentypen (offen-geschlossen; individuell-kooperativ; gegenstandsorientiert-problemorientiert etc.), Instruktionen und Interaktionsmöglichkeiten. Das *Betreuungskonzept* erläutert (z.B. im Zusammenhang mit dem Aufgabenkonzept) den Einsatz von Moderatoren, Tutoren, Coachs etc., den Umfang und die Art vorgesehener Rückmeldungen, den technischen Support etc. (c) Das *Assessment-Konzept* legt dar, wie welche Prüfungen oder Lernerfolgskontrollen erfolgen sollen und in welcher Beziehung diese zur Didaktik stehen, was speziell in institutionalisierten Kontexten eine wichtige Rolle spielt.

Was ist eine Produktevaluation?

Als Produkte für ein Lehren und Lernen mit Medien kommen technische Systeme und Werkzeuge (technische Produkte) und „Stand-Alone“-Lernangebote in Frage, die der Lernende ohne weitere Interaktionen mit anderen Personen oder Unterstützung bearbeiten kann (didaktische Produkte): (a) *Technische Produkte* sind z.B. LMS, CMS, E-Portfolio-Software, aber auch Anwenderprogramme aller Art etwa für Konstruktionsprozesse beim Lernen. (b) *Didaktische Produkte* sind klassischerweise CBTs und WBTs aller Art – unabhängig davon, ob sie eher der Informationsvermittlung dienen, Interaktionsmöglichkeiten anbieten oder umfängliche Simulationen zulassen.

Was ist eine Prozessevaluation?

Mit Prozessen im Rahmen der Evaluation von mediengestützten Lernangeboten sind meistens Umsetzungsprozesse, also der Vorgang der Implementierung eines Konzepts im engeren Sinne gemeint. Darüber hinaus lassen sich aber auch Prozesse der Planung sowie der Evaluation (was im weitesten Sinne auch zur Implementation gehört) evaluieren. Das bedeutet im Einzelnen Folgendes: (a) *Planungsprozesse* sind solche, die man bei der Vorbereitung und Entwicklung eines medialen Lernangebots braucht. Diese zu kennen und zu bewerten ist vor allem dann wichtig, wenn man noch öfter und/oder in anderen Kontexten die gleichen oder ähnliche Planungen und Entwicklungen umsetzen will. (b) *Implementationsprozesse* (i.e.S.) sind solche, die bei der Umsetzung eines Konzepts oder beim Einsatz eines Produkts ablaufen und mehr oder weniger große Abweichungen vom Planungssoll erforderlich machen. Diesen kommt eine besonders große Bedeutung zu, denn das beste mediale Lernangebot nützt wenig, wenn es sich nicht oder nur schlecht umsetzen lässt. (c) *Evaluationsprozesse* sind nicht nur die, die man mit einer wissenschaftlichen Evaluation anstößt. Man kann auch alle Kontroll- und Prüfverfahren darunter subsumieren (z.B. alle Qualitätsmanagementmaßnahmen), die im Zuge der Umsetzung eines mediengestützten Lernangebots zum Einsatz kommen.

Was ist eine Wirkungsevaluation?

Die zu evaluierenden Wirkungen des Lehrens und Lernens mit Medien stehen häufig im Vordergrund und sind ganz besonders vielfältig. Es kann praktisch hilfreich sein, sich verschiedene Wirkungsgruppen als Dimensionen dichotom vorzustellen, um die Spannweite der möglichen Effekte, die man evaluieren kann, deutlich zu machen. Dabei lassen sich diese Dimensionen selbstverständlich miteinander kombinieren, weil sie verschiedene Aspekte von Wirkungen herausstellen: (a) Wirkungen auf *Personen* versus Wirkungen auf *Organisationen*: Wirkungen auf Personen sind z.B. Veränderungen im Wissen, Können, Meinen oder auch Fühlen; man kann auch bestimmte (personale) Leistungen als Wirkungen medialer Lernangebote definieren. Wirkungen auf Organisationen sind z.B. Veränderungen in der Lehr-Lernkultur, in der Bereitstellung oder Aufbereitung von Inhalten, in der Ausbildung von Lehrpersonal etc. (b) *Quantitative* (zählbare) Ergebnisse bzw. Wirkungen versus *qualitative* Ergebnisse bzw. Wirkungen: Quantitative Wirkungen sind z.B. Teilnehmerzahlen, die Anzahl von Abschlüssen, Abbruch-Quoten, produzierte Lerninhalte etc., die an sich zählbar sind und ohne weitere Zwischenschritte als Output verwendet werden können. Qualitative Wirkungen sind z.B. alle personalen Wirkungen auf Wissen, Können, Meinen, Fühlen, die man nicht direkt messen, sondern nur über bestimmte Indikatoren, Tests oder andere Formen der Befragung erschließen kann. (c) *Kurzfristige* Wirkungen versus *langfristige* Wirkungen: Kurzfristige Wirkungen sind z.B. unmittelbar festzustellende Lernergebnisse. Langfristige Wirkungen sind z.B. später erkennbare Transferergebnisse. (d) *Beabsichtigte* versus *nicht-intendierte* Wirkungen: Beabsichtigte Wirkungen sind die (in der Regel positiven) Effekte, die mit dem Lernangebot erzielt werden sollten. Nicht-intendierte Wirkungen sind positive oder negative „Nebeneffekte“.

5.2 USABILITY-TESTS ALS EVALUATIONSVERFAHREN

Was bedeutet Usability?

Usability-Tests sind eine Evaluationsform, die für das mediengestützte Lehren und Lernen spezifisch ist. Usability-Tests beziehen sich auf *Produkte*, also auf technische oder didaktische Produkte. Diese Produkte fungieren für sich betrachtet als Einzelmaßnahmen, können aber natürlich Teil größerer Projekte oder Programme sein, die evaluiert werden. Der Begriff der Usability hat sich auch im Deutschen weitgehend durchgesetzt, obschon man dazu auch Nutzbarkeit, Nützlichkeit, Benutzerfreundlichkeit, Gebrauchstauglichkeit und ähnliches sagen könnte. Usability ist ein (relationales) Konstrukt, das beschreibt, wie adäquat ein Produkt in der Handhabung zu den Bedürfnissen, Fähigkeiten und Fertigkeiten sowie Wünschen der Nutzer passt (Niegemann, 2008). Es gibt zahlreiche Handreichungen für eine „gute Usability“ sowie dazugehörige Kriterienlisten bzw. -kataloge (also Listen bzw. Kataloge mit Bewertungskriterien), anhand derer man die Usability überprüfen kann. Den wenigstens dieser Vorschläge liegen allerdings theoretische oder empirische Erkenntnisse zugrunde (Kerkau, 2009).

Was sind zentrale Usability-Kriterien?

Trotz der Vielfalt an Usability-Kriterien gibt es immerhin eine gewisse Einigkeit zwischen verschiedenen Autoren und Richtlinien, was zentrale Kriterien sind. Diese tauchen auch in den meisten Listen oder Katalogen auf (wobei die Formulierungen variieren), die man für Evaluationen im Lehr-Lernkontext verwendet. Dabei handelt es sich um folgende Kriterien:

- die *Effektivität* gemeint als Ausmaß, in dem ein Produkt bzw. System oder Werkzeug den Nutzer unterstützt, sein Ziel zu erreichen,
- die *Effizienz* gemeint als Verhältnis von Ressourceneinsatz (vor allem Zeit, Geld, Anstrengung) zum erreichten Ergebnis,
- die *Zufriedenheit* gemeint als Ausmaß, in dem die Erwartungen des Nutzers erfüllt werden und dieser das Produkt akzeptiert,
- die *Erlernbarkeit* gemeint als zeitlicher Aufwand, den man braucht, um mit dem Produkt zurechtzukommen, sowie
- die *Fehlerrobustheit* gemeint einerseits als eine geringe Fehlerrate des Systems und andererseits als Ausmaß der Toleranz eines Systems gegenüber Fehlern und/oder Hilfen zur leichten Fehlerkorrektur.

Weitere Kriterien, die z.B. in Qualitätsmanagement-Normen aufgenommen wurden und häufig verwendet werden, sind die Selbstbeschreibungsfähigkeit eines Systems, die Erwartungskonformität (das System verhält sich so, wie vom Nutzer angenommen) und die Individualisierbarkeit.

Wie überprüft man die Usability?

In der Literatur werden häufig zwei verschiedene Arten der Überprüfung der Usability technischer Produkte bzw. Systeme oder Werkzeuge unterschieden. Der Unterschied liegt dabei vor allem in den Bewertungsmaßstäben: (a) Bei einer *vergleichenden* Überprüfung werden verschiedene Produkte miteinander verglichen, um z.B. die für einen bestimmten Anwendungskontext beste Lösung herauszufinden (was man z.B. für LMS, CMS oder E-Portfolio-Systeme bereits in größerem Umfang gemacht hat; z.B. Baumgartner, Häfele & Mayer-Häfele, 2002; Schulmeister, 2005). (b) Bei einer *bewertenden* Überprüfung konzentriert man sich auf ein einzelnes System oder ein einzelnes Werkzeug und analysiert, in welchem Ausmaß vorher definierte Usability-Kriterien erfüllt sind (was meist der Fall ist, wenn man eine bestimmte Software-Anwendung in einem didaktischen Kontext einsetzt). Natürlich ist auch ein Vergleich eine Bewertung, und Bewertungen in Evaluationen basieren immer auf Vergleichen mit etwas (vgl. Abschnitt 2.2.3). Die Begriffe „vergleichend“ und „bewertend“ sind hier also in einem engeren Sinne zur Differenzierung verschiedener Bewertungsformen zu verstehen. Kriterienlisten bzw. -kataloge unterscheiden sich bisweilen kaum von einem Fragebogen: Sie enthalten Fragen oder Aussagen zum Einschätzen. Es ist nicht die Form der Listen oder Kataloge, sondern deren Einsatzzweck, die sie spezifisch für die Usability-Forschung machen. Dieser Einsatzzweck besteht darin, dass ein Experte oder Nutzer ein Qualitätsurteil über ein Produkt abgibt (und nicht z.B. Einschätzungen zum eigenen Lernprozess, Lernerfolg oder ähnliches).

Wer überprüft die Usability?

Die Evaluatoren können bei Usability-Tests sowohl Experten (Expertenurteile) als auch Nutzer (Nutzerbeurteilung) sein, was jeweils Vor- und Nachteile hat und eher ergänzend zu sehen ist. Die Art der eingesetzten Methoden unterscheidet sich allerdings in der Regel deutlich zwischen Usability-Tests durch Experten und durch Nutzer. (a) Für *Expertenurteile* werden sogenannte „heuristische Evaluationen“ bevorzugt, womit Einschätzungen anhand von Leitlinien gemeint sind, für die man zwischen drei bis fünf Experten heranzieht. Man kann Experten aber auch bitten, das Nutzerverhalten bei der Bewältigung bestimmter Aufgaben mit dem System zu simulieren und dann Einschätzungen abzugeben ("cognitive walkthrough"). In beiden Fällen vertraut man stark auf die Aussagekraft subjektiver Urteile. (b) Für *Nutzerbeurteilungen* kommen ebenfalls Befragungen in Frage, die die wichtigsten Usability-Kriterien als Items einsetzen. Nutzer lassen sich aber auch beim Umgang mit einem System beobachten; zudem sind Teile des Verhaltens in der Regel durch Tracking-Verfahren im System selbst festzustellen. Eine weitere, oft verwendete Methode besteht darin, dass Nutzer beim Umgang mit dem System laut denken. Daneben kommen zunehmend Eye-Tracking-Verfahren zum Einsatz, bei denen die Blickbewegungen von Nutzern aufgezeichnet werden.

6 KOMMENTIERTE LITERATURÜBERSICHT

6.1 MONOGRAFIEN UND SAMMELBÄNDE

Reischmann, J. (2006). *Weiterbildungsevaluation. Lernerfolge messbar machen*. Augsburg: ZIEL Verlag.

Reischmanns Buch ist eine vergleichsweise einfache, verständliche und praxisorientierte Einführung, die auch für Anfänger ohne Vorkenntnisse gut geeignet ist. Das Buch behandelt die wichtigsten Fragen der Evaluation, schafft ein solides Grundverständnis und ist dabei auf die Weiterbildung bzw. Erwachsenenbildung spezialisiert. Medien werden kaum thematisiert, sodass Fragen zur Evaluation mediengestützten Lehrens und Lernens nicht beantwortet werden können.

Stockmann, R. & Meyer, W. (2010). *Evaluation. Eine Einführung*. Opladen: Budrich.

Stockmanns und Meyers Buch ist insofern einführend, als dass alle wichtigen Aspekte der Evaluation relativ ausführlich behandelt werden, beim Leser aber bereits gewisse Vorkenntnisse voraussetzen. Das Buch ist eher komplexer Natur und stärker theoretisch als praktisch orientiert. Es handelt sich um eine sozialwissenschaftlich geprägte Einführung, die *nicht* auf Fragen des (mediengestützten) Lehrens und Lernens spezialisiert ist, aber eine gute und reflektierte Grundlage für das Verständnis von Evaluation liefert.

Gollwitzer, M & Jäger, R. (2009). *Evaluation kompakt*. Weinheim: BeltzPVU.

Auch das Buch von Gollwitzer und Jäger versteht sich als „kompakte“ Einführung, wobei der Akzent auf der psychologischen Evaluationsforschung liegt. Vor diesem Hintergrund werden experimentelle und quantitative Verfahren bevorzugt behandelt; sie machen einen großen Teil des Buches aus. Dieser Teil ist ohne Grundkenntnisse nicht einfach zu verstehen, bietet aber einen guten Überblick. Alternative Ansätze und Methoden werden allerdings ausgeklammert. Bildungsbezogene Beispiele kommen vor, stehen aber nicht im Zentrum.

Flick, U. (Hrsg.) (2006). *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen*. Reinbek bei Hamburg: Rowohlt.

Bei diesem Buch von Flick handelt es sich um einen Herausgeberband, in dem eine Reihe von Artikeln speziell zu qualitativen Evaluationsansätzen und -methoden zusammengestellt werden. Das Buch ist so gesehen komplementär zur Einführung von Gollwitzer und Jäger. Die einzelnen Beiträge sind tendenziell eher theoretisch als praktisch orientiert und nicht speziell auf bildungswissenschaftliche, sondern genereller auf sozialwissenschaftliche Fragen ausgerichtet. Methodisch wird ein sehr guter Überblick über qualitative Strategien und Verfahren in der Evaluation gegeben.

Stockmann, R. (Hrsg.) (2007). *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*. Münster: Waxmann.

Der Sammelband von Stockmann liefert umfangreiche grundlegende, aber auch spezielle Inhalte zur Evaluation aus dem sozialwissenschaftlichen Bereich, die entlang praktischer Fragen, Schritte und Abläufe angeordnet sind. Auf diesem Wege gelingt eine relativ praxisnahe Darstellung, die die Umsetzung der Inhalte erleichtern kann. Hervorzuheben sind unter anderem Hinweise zur Berichterstattung (in Form von Berichten und Präsentationen) am Ende einer Evaluation.

Meister, D., Tergan, S.-O. & Zentel, P. (Hrsg.) (2004). *Evaluation von E-Learning. Zielrichtungen, methodologische Aspekte, Zukunftsperspektiven*. Münster: Waxmann.

Beim Buch von Meister et al. handelt es sich ebenfalls um einen Sammelband, der nicht ausschließlich, aber vorrangig praktisch orientierte Beiträge zusammenstellt. Viele der Buchbeiträge sind Projektdarstellungen, sodass die Inhalte einen gewissen Grad an Anschaulichkeit aufweisen. Lange Zeit war dieses Buch auf dem deutschsprachigen Sektor eines der wenigen, die das Thema Evaluation mit explizitem Bezug zu Lehren und Lernen mit digitalen Medien behandelt hat, *ohne* sich auf Usability-Forschung zu begrenzen. Bei der Lektüre ist darauf zu achten, dass der Band bereits 2004 publiziert wurde, die Medienbeispiele also etwas veraltet sein können.

Mayer, H.O. & Kriz, W. (Hrsg.) (2010). *Evaluation von eLernprozessen*. München: Oldenbourg.

Auch dieser Sammelband von Mayer und Kriz bezieht sich explizit auf die Evaluation im Bereich Lehren und Lernen mit digitalen Medien und stellt vor diesem Hintergrund eine wichtige, vor allem aktuelle Ergänzung oder Alternative zum Buch von Meister et al. dar. Die Beiträge sind teils theoretisch, teils praktisch ausgerichtet und weisen im Hinblick auf das inhaltliche Niveau relativ große Diskrepanzen auf. Es sind daher nicht alle Texte gleichermaßen empfehlenswert. Aus dem Buch lassen sich aktuelle Beispiele entnehmen, die relativ viele Evaluationsziele und -aspekte abdecken.

6.2 BEITRÄGE IN HAND- UND LEHRBÜCHERN

Wottawa, H. (2006). Evaluation. In A. Krapp und B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 659-687). Weinheim: BeltzPVU.

Wottawas Evaluationskapitel aus einem Lehrbuch skizziert in aller Kürze die wichtigsten Aspekte der Evaluation aus einer pädagogisch-psychologischen Perspektive und weist damit wichtige bildungswissenschaftliche Bezüge auf, ohne sich aber explizit mit Medien zu beschäftigen. Der Beitrag umfasst mehrere, kurze Beispiele aus verschiedenen Kontexten, was die Anschaulichkeit erhöht und auf diesem Wege einen ersten knappen Einstieg in das Evaluationsthema ermöglicht. Im Text wird auch kurz der Zusammenhang zum Qualitätsmanagement hergestellt.

Ditton, H. (2009). Evaluation und Qualitätssicherung. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 607-623). Wiesbaden: VS-Verlag.

Dittons Kapitel zur Evaluation aus einem umfangreichen, interdisziplinär ausgerichteten Handbuch liefert einen Überblick zu Fragen der Qualitätssicherung aus Sicht der Bildungsforschung. Der Beitrag arbeitet die Zusammenhänge zwischen Begriffen bzw. Konzepten des Qualitätsmanagements einerseits und der Evaluation andererseits heraus. Der Fokus liegt auf der Frage, was Bildungsqualität ist und wie (und unter welchen Schwierigkeiten) diese erfasst werden kann.

Hessel, S. (2008). Evaluation multimedialen Lernens. In H. Niegemann et al. (Hrsg.), *Kompodium multimedialen Lernens* (S. 395-418). Berlin: Springer.

Hessels Kapitel aus einem Kompodium zum E-Learning bzw. zum Lernen mit Multimedia liefert eine Überblick über einige zentrale Fragen der Evaluation speziell von mediengestützten Lernangeboten inklusive checklistenartiger Hinweise und Empfehlungen. Der Beitrag ist theoretisch wenig fundiert, aber für praktische Fragen infolge des klaren Medienbezugs für einen Einstieg in die Evaluation mediengestützten Lehrens und Lernens durchaus hilfreich.

Glowalla, U., Herder, M., Süße, C., & Koch, N. (2009). Methoden und Ergebnisse der Evaluation elektronischer Lernangebote. In L. J. Issing & P. Klimsa (Hrsg.), *Online-Lernen. Handbuch für Wissenschaft und Praxis* (S. 309-328). München: Oldenbourg.

Glowallas et al. Kapitel zur Evaluation elektronischer Lernangebote stammt aus einem aktuellen Handbuch zum Online-Lernen und verfolgt einen ähnlichen Zweck wie das Evaluationskapitel aus dem oben stehenden Kompodium. Es wird ein Überblick über relevante Fragen der Evaluation mediengestützten Lehrens und Lernens gegeben und mit einer Reihe von Beispielen angereichert. Neben dem Praxisbezug wird versucht, theoretisch begründbare Ordnungen herzustellen.

Niegemann, H. (2008). Usability. In H. Niegemann et al. (Hrsg.), *Kompendium multimedialen Lernens* (S. 419-453). Berlin: Springer.

Niegemanns Kapitel aus seinem Kompendium zum E-Learning bzw. zum Lernen mit Multimedia führt in die Usability-Forschung ein und gibt einen kompakten Überblick über wesentliche Modelle, Kriterien und Vorgehensweisen bei Usability-Tests. Vorgestellt werden verschiedene Messansätze und Methoden sowie der typische Ablauf beim Usability-Testing. Als Einstieg und Überblick ist dieses Kapitel zum Thema Usability vergleichsweise gut geeignet.

Kerkau, F. (2009). Usability-Testing zur Qualitätssicherung von Online-Lernangeboten. In L. J. Issing & P. Klimsa (Hrsg.), *Online-Lernen. Handbuch für Wissenschaft und Praxis* (S. 329-337). München: Oldenbourg.

Kerkaus Kapitel zur Usability aus dem oben bereits genannten Handbuch zum Online-Lernen beschreibt sehr knapp den theoretischen Rahmen von Usability-Tests und geht auf gängige Erhebungsmethoden ein. Auch die Voraussetzungen und die zu erwartenden Ergebnisse werden kurz diskutiert. Als Ergänzung zur umfassenderen Einführung in das Thema Usability ist der Beitrag durchaus nützlich.

Braun, D. (2008). Evaluation und unintendierte Effekte – eine theoretische Reflexion. In H. Matthies & D. Simon (Hrsg.), *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen* (S. 103-124). Wiesbaden: VS Verlag.

Brauns Kapitel zu Nebeneffekten der Evaluation stammt aus einem Band, der sich speziell der Wissenschaftsevaluation widmet. Der Beitrag nähert sich dem Thema der Evaluation eher aus einer politischen Sicht und bewegt sich ausschließlich auf dem Kontext von Wissenschaft und Forschung als Evaluationsgegenstand. Die skizzierten ungewollten Effekte können aber auch in anderen Kontexten eine Rolle spielen. Zudem sensibilisiert der Beitrag für die politische Dimension, die es bei allen Evaluationen geben kann.

6.3 ONLINE VERFÜGBARE DOKUMENTE UND WEB-SEITEN

Kromrey, H. (2005). *Evaluation – ein Überblick*. In H. Schöch (Hrsg.), *Was ist Qualität? Die Entzauberung eines Mythos* (S. 31-85). Berlin: Wissenschaftlicher Verlag.

URL: [http://www.profkromrey.de/Kromrey_Evaluation - ein Überblick.pdf](http://www.profkromrey.de/Kromrey_Evaluation_-_ein_Ueberblick.pdf)

In Kromreys online verfügbaren Beitrag aus einem Sammelband zum Thema Qualität erhält man einen prägnanten, aber umfassenden sozialwissenschaftlich geprägten Überblick über Evaluation und Evaluationsforschung. Ergänzt werden die Ausführungen mit Beispielen aus der Evaluation an Hochschulen. Es wird zwar kein Medienbezug hergestellt, dennoch ist der Text hilfreich, um typische Probleme zu verstehen, die auch bei der Evaluation mediengestützten Lehrens und Lernens auftreten können.

Hense, J. & Mandl, H. (2006). *Selbstevaluation als Ansatz der Qualitätsverbesserung von E-Learning-Angeboten*. München: LMU Forschungsberichte.

URL: http://epub.ub.uni-muenchen.de/947/1/Forschungsbericht_184.pdf

Henses und Mandls online verfügbarer Forschungsbericht konzentriert sich speziell auf das Feld des mediengestützten Lehrens und Lernens und geht der Frage nach, welche Rolle und welches Potenzial allem voran der Selbstevaluation zukommt. Der Text gibt mit dieser Eingrenzung auf die Möglichkeiten und Grenzen der Selbstevaluation interessante Impulse für die Praxis der Evaluation mediengestützter Lernangebote.

Portal der Gesellschaft für Evaluation e.V.

URL: <http://www.degeval.de/>

Das Portal der Gesellschaft für Evaluation e.V. ist ein deutschsprachiges Portal und gibt Auskunft bzw. liefert einen Überblick über weitere Gesellschaften und Verbände nationaler und internationaler Art. Darüber hinaus werden Leitfäden für verschiedene Evaluationskontexte bereitgestellt. Eine thematische Eingrenzung von Evaluationskontexten (z.B. auf Bildung oder Lehren und Lernen) besteht nicht; das Portal ist übergreifend ausgerichtet.

Portal der American Evaluation Association

URL: <http://www.eval.org>

Das Portal der American Evaluation Association ist ein englischsprachiges Portal und informiert umfassend über zahlreiche Communities, Zeitschriften, Fortbildungen etc. Unter anderem enthält das Portal eine eigene E-Library mit digitalen Ressourcen. Auch dieses Portal ist thematisch nicht auf Evaluationskontexte (z.B. auf Bildung oder Lehren und Lernen) eingegrenzt, sondern will alle möglichen Evaluationskontexte ansprechen.

Evaluationsmodul auf der Web-Seite E-Didaktik

URL: <http://www.zhw.uni-hamburg.de/edidakt/projekt.php>

Bei diesem Angebot der Universität Hamburg handelt es sich um ein Modul zum Thema Evaluation im Rahmen einer Web-Seite zur E-Didaktik. Die Seite ist das Ergebnis eines abgeschlossenen Projekts und liefert Informationen zum Lehren und Lernen mit digitalen Medien generell. Das Evaluationsmodul gibt einen knappen Überblick über wichtige Evaluationsaspekte und liefert vor allem auch eine Reihe weiterführender Links.

Web-Seite Evaluation aus einem Dissertationsprojekt

URL: <http://www.evaluation.de>

Bei diesem Angebot handelt es sich um eine Web-Seite zur Evaluation mit Fokus auf Bildungskontexte, die von einem einzelnen Wissenschaftler sehr sorgfältig erstellt und gepflegt wird (Jan Hense). Es werden sowohl zahlreiche Informationen zur Evaluation (auch mit Bezug auf Medien) angeboten (inklusive eines Glossars) als auch weiterführende Links. Die Seite ist das Ergebnis eines Evaluationsprojekts und entsprechend abgeschlossen.

LITERATUR

- Baumgartner, P., Häfele, H. & Maier-Häfele, K. (2002). *Auswahl von Lernplattformen. Marktübersicht - Funktionen – Fachbegriffe*. Innsbruck: StudienVerlag.
- Brandt, T. (2007). Sozialer Kontext der Evaluation. In R. Stockmann (Hrsg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (S. 164-194). Münster: Waxmann.
- Braun, D. (2008). Evaluation und unintendierte Effekte – eine theoretische Reflexion. In H. Matthies & D. Simon (Hrsg.), *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen* (S. 103-124). Wiesbaden: VS Verlag.
- Deutsche Gesellschaft für Evaluation (DeGEval). (2004). *Empfehlungen zur Anwendung der Evaluationsstandards der DeGEval im Handlungsfeld der Selbstevaluation*. Deutsche Gesellschaft für Evaluation.
- Ditton, H. (2009). Evaluation und Qualitätssicherung. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 607-623). Wiesbaden: VS-Verlag.
- Flick, U. (Hrsg.) (2006). *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen*. Reinbek bei Hamburg: Rowohlt.
- Glowalla, U., Herder, M., Süße, C. & Koch, N. (2009). Methoden und Ergebnisse der Evaluation elektronischer Lernangebote. In L. J. Issing & P. Klimsa (Hrsg.), *Online-Lernen. Handbuch für Wissenschaft und Praxis* (S. 309-328). München: Oldenbourg.
- Gollwitzer, M. & Jäger, R. (2009). *Evaluation kompakt*. Weinheim: Beltz-PVU.
- Hense, J. & Mandl, H. (2006). *Selbstevaluation als Ansatz der Qualitätsverbesserung von E-Learning-Angeboten*. München: LMU Forschungsberichte.
- Hessel, S. (2008). Evaluation multimedialen Lernens. In H. Niegemann et al. (Hrsg.), *Kompendium multimedialen Lernens* (S. 395-418). Berlin: Springer.
- Hug, T. & Poscheschnik, G. (2010). *Empirisch Forschen*. Konstanz: UVK UTB.
- Kerkau, F. (2009). Usability-Testing zur Qualitätssicherung von Online-Lernangeboten. In L. J. Issing & P. Klimsa (Hrsg.), *Online-Lernen. Handbuch für Wissenschaft und Praxis* (S. 329-337). München: Oldenbourg.
- Kruse, O. (2010). *Lesen und Schreiben*. Konstanz: UVK UTB.
- Kromrey, H. (2005). Evaluation – ein Überblick. In H. Schöch (Hrsg.), *Was ist Qualität? Die Entzauberung eines Mythos* (S. 31-85). Berlin: Wissenschaftlicher Verlag.
- Mayer, H.O. & Kriz, W. (Hrsg.) (2010). *Evaluation von eLernprozessen*. München: Oldenbourg.
- Meister, D., Tergan, S.-O. & Zentel, P. (Hrsg.) (2004). *Evaluation von E-Learning. Zielrichtungen, methodologische Aspekte, Zukunftsperspektiven*. Münster: Waxmann.

- Meyer, W. (2007). Datenerhebung: Befragungen – Beobachtungen – Nicht-reaktive Verfahren. In R. Stockmann (Hrsg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (S. 223-277). Münster: Waxmann.
- Niegemann, H. (2008). Usability. In H. Niegemann et al. (Hrsg.), *Kompendium multimedialen Lernens* (S. 419-453). Berlin: Springer.
- Reinmann, G. (2004). Qualitätsmanagement in Schulen: Eine Einführung. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung* (S. 145-165). Hohengehren: Schneider.
- Reischmann, J. (2006). *Weiterbildungsévaluation. Lernerfolge messbar machen*. Augsburg: ZIEL Verlag.
- Schlömerkemper, J. (2010). *Konzepte pädagogischer Forschung*. Bad Heilbrunn: Julius Klinkhardt.
- Schulmeister, R. (2005). *Lernplattformen für das virtuelle Lernen. Evaluation und Didaktik*. München: Oldenbourg.
- Scriven, M. (1991). *Evaluation Theasaurus*. Newbury Park, CA: Sage.
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, 17, 151-161.
- Seufert, S. & Euler, D. (2005). *Nachhaltigkeit von eLearning- Innovationen: Fallstudien zu Implementierungsstrategien von eLearning als Innovationen an Hochschulen*. SCIL-Arbeitsbericht 4. St. Gallen: SCIL.
- Stockmann, R. & Meyer, W. (2010). *Evaluation. Eine Einführung*. Opladen: Budrich.
- Stuckmann, R. (2010). Evaluationsprozess. In R. Stockmann & W. Meyer (Hrsg.), *Evaluation. Eine Einführung* (S. 159-189). Opladen: Budrich.
- Stockmann, R. (2007a). Konkurrierende und komplementäre Ansätze zur Evaluation. In R. Stockmann (Hrsg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (S. 71-107). Münster: Waxmann.
- Stockmann, R. (2007b). Einführung in die Evaluation. In R. Stockmann (Hrsg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (S. 24-70). Münster: Waxmann.
- Stockmann, R. (Hrsg.) (2007). *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*. Münster: Waxmann.
- Wottawa, H. (2006). Evaluation. In A. Krapp und B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 659-687). Weinheim: BeltzPVU.