
Jahrbuch Medienpädagogik 5. Evaluation und Analyse

Zweitveröffentlichung aus: Jahrbuch Medienpädagogik 5. (2005) Wiesbaden: VS Verlag für Sozialwissenschaften. Hrsg. v. Ben Bachmair, Peter Diepold und Claudia de Witt.

Evaluation von E-Learning: Checklisten, Kriterienkataloge oder Evaluationskonzepte?

Zum Stand der Bewertungsverfahren für E-Learning-Arrangements

Ulf-Daniel Ehlers

1. Qualität als Ziel

Qualität in Bildungsmaßnahmen – als Legitimation oder aus wissenschaftlichem Interesse – ist ein Thema von hoher Bedeutung, auch – und gerade – im E-Learning.¹ E-Learning, als eine noch relativ junge, nicht völlig etablierte, aber – so stellt sich immer häufiger heraus – kostenintensive Bildungsform, muss sich immer wieder der kritischen Frage nach der Qualität stellen, zumeist assoziiert mit Lernerfolg.² Vielfältige Instrumente, um die Qualität zu überprüfen oder zu entwickeln, begleitend oder im Nachhinein, sind bislang konzipiert worden (vgl. Ehlers/Pawlowski/Goertz 2003).

Der Artikel gibt einen Überblick zum derzeitigen Stand der Dinge in Bezug auf zwei dieser Verfahrensweisen: zum einen geht es dabei um die Beurteilungen anhand von Kriterienkatalogen bzw. Checklisten und zum anderen um Evaluationsverfahren. Obwohl vielfach bereits umfassendere Managementsysteme zum Einsatz kommen, die die Qualität bei Anbietern kontinuierlich verbessern sollen, haben beide Verfahren eine steigende Bedeutung – da auch Managementverfahren immer wieder auf Kriterien oder Evaluationsverfahren zurückgreifen.

Betrachtet man die Forschungsaktivitäten im Bereich der Qualitätsforschung beim E-Learning, so wird deutlich, dass unterschiedliche „konjunkturelle Moden“ auszumachen sind.³ Zum einen wurden seit Mitte der 1960er Jahre mit

1 E-Learning ist kein wissenschaftlicher Begriff. Der Begriff kam in den letzten Jahren (etwa seit 1999) als ein Neologismus der Werbeindustrie auf. Er umfasst alle Formen des Lernens mit Hilfe elektronischer Medien, sowohl online als auch offline. Es ist eine Lernform, bei der die neuen Informations- und Kommunikationsmedien (Computer und Internet) in Lernarrangements eingebunden werden, entweder zur Unterstützung des Lernprozesses (als sogenannte „hybride“ Lernarrangements) oder als ausschließliche Form der Vermittlung (vgl. Ehlers 2003).

2 Zur Bedeutung von Qualität im E-Learning vgl. auch Ehlers/Holmer/Gerteis/Jung (2003).

3 Dabei bedeutet das Aufkommen neuer Ansätze nicht zwangsläufig das Verwerfen von bereits bestehenden Methoden. So standen zwar zunächst kriteriengestützte Expertenbeurteilungsverfahren im Vordergrund, dann spezifische prozessorientierte Qualitätssicherungsverfahren, wie zum Beispiel spezielle Evaluationsverfahren oder Benchmarks für E-Learning. Die schließlich an Bedeutung gewinnenden generischen Qualitätsmanagementansätze und Standardisierungsversuche greifen aber teilweise Elemente der anderen Qualitätsansätze wieder auf beziehungsweise integrieren sie.

dem Aufkommen des computerunterstützten Unterrichts (CUU) Qualitätskriterien für diesen Unterrichtsansatz entwickelt. Diese wurden im Laufe der Zeit weiterentwickelt und stellen heute sogenannte produktbezogene, normative Bewertungs-, Entwicklungs- und Auswahlinstrumente für Lernplattformen, Lernsoftware oder E-Learning-Angebote dar. Zum anderen spielt das Entstehen von speziellen empirischen Evaluationsverfahren für die Beurteilung von Lernsituationen, in denen E-Learning eingesetzt wird, eine wichtige Rolle. Nicht das Produkt alleine steht hier zur Disposition, sondern der Lernprozess insgesamt. Damit rückt auch der Lerner in den Fokus der Analyse. Streng genommen gilt die Bewertung von E-Learning-Angeboten mit Hilfe von Qualitätskriterienkatalogen auch als eine Form der Qualitätsevaluation, beispielsweise im Rahmen einer sogenannten Expertenbeurteilung. Da sie sich aber im konzeptuellen Ansatz und der Durchführung deutlich von eher prozessbezogenen Evaluationsansätzen unterscheidet, werden beide Arten gesondert voneinander aufgeführt.

Eine weitere Entwicklung ist in prozessorientierten Ansätzen zu sehen, die die beiden vorherigen Konzepte jedoch in Teilen integriert und im Sinne übergreifender Managementansätze versucht, sowohl Standards als auch Evaluationsverfahren einzusetzen, um die bestehenden Lehr- und Lernprozesse (zusätzlich auch die Administrationsprozesse u.a.) begleitend zu überwachen und ggf. zu optimieren. Die beschriebenen Entwicklungsstränge stehen jedoch nicht unverbunden nebeneinander oder stellen eine aufeinander aufbauende chronologische Entwicklung dar, sondern sind vielmehr miteinander verzahnt. Der folgende Artikel hat das Ziel, einerseits einen Überblick über die derzeit bestehenden Instrumente im Bereich der Kriterienkataloge und Evaluationsverfahren zu geben und andererseits die Möglichkeiten, aber auch die Grenzen der jeweiligen Verfahren aufzeigen.

2. Beurteilung von E-Learning anhand von Qualitätskriterien

Bewertungsinstrumente auf Basis von Qualitätskriterienkatalogen sind deswegen so populär, weil sie Einschätzungen über die Lernwirksamkeit, die eigentlich nur über aufwändige empirische Verfahren erhoben werden können, relativ einfach zu ermöglichen scheinen. Dadurch bieten sie auch Laien die Möglichkeit, ohne vorhergehende empirische Studien Einschätzungen über die Qualität eines Lernarrangements oder einer Lernsoftware vorzunehmen. Als Qualitätskriterium wird dabei ein Merkmal einer Lernsoftware bezeichnet, dessen Lernwirksamkeit in einer Validitätsstudie nachgewiesen wurde. Kohrt (1995) fasst diesen Sachverhalt folgendermaßen:

„Ein potentielles Qualitätskriterium (das also in einer Checkliste auftauchen könnte) kann jedes isolierte Merkmal eines Lehrprogramms oder seiner Begleitmaterialien sein, das sich messbar positiv bzw. negativ auf den Lernerfolg auswirkt“ (ebenda, S. 6).

Viele Qualitätskriterienkataloge enthalten jedoch Kriterien, zu denen keine expliziten Validitätsstudien vorliegen, sondern über die lediglich vermutet wird, sie seien lernwirksam. Die Entstehung von Qualitätskriterienkatalogen kann mit Gräber (1996) in insgesamt vier Generationen von Bewertungsinstrumenten eingeteilt werden (siehe Abbildung 1). Sie unterscheiden sich sowohl hinsichtlich ihrer zeitlichen Entstehung als auch bezüglich ihres Funktionsumfangs. Im Hinblick auf den Funktionsumfang ist zu differenzieren zwischen einfachen Zusammenstellungen von Qualitätskriterien, die als Anhaltspunkt für die Auswahl von Lernsoftware für Laien gedacht sind, nach computergestützten Bewertungssystemen, die Einschätzungen zu unterschiedlichen Bereichen (bei-

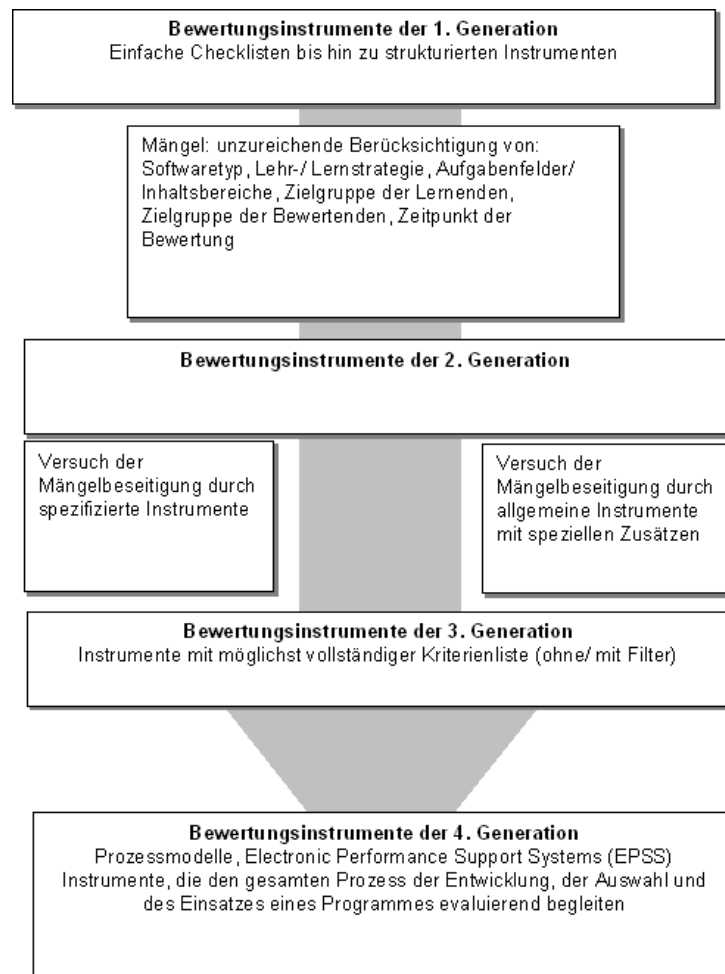


Abb. 1: Vier Generationen von Kriterienkatalogen (nach Gräber 1996).

spielsweise didaktische Gestaltung, Bildschirmpräsentation u.ä.) abfragen und am Schluss einen Punktwert liefern, mit dem die bewertete Software gegenüber anderen verglichen werden kann⁴, bis hin zu vollständig datenbankgestützten Kriteriensammlungen. Letztere stellen unter Zuhilfenahme unterschiedlicher Filter Kriterien entweder für die Auswahl, für die Entwicklung oder die Beurteilung von Lernsoftware bereit.⁵

Gräber (1996, S. 17ff.) ordnet einfache Checklisten, aber auch strukturierte, komplexere Instrumente der ersten Generation von Bewertungsinstrumenten zu. Er führt aus, dass diese charakterisiert sind durch eine unzureichende Berücksichtigung des Softwaretyps, der Lehr-Lernstrategie, der Aufgabenfelder beziehungsweise Inhaltsbereiche, der Zielgruppen der Lerner sowie der Beurteiler und des Zeitpunktes der Bewertung. Winship (1988, S. 371) weist darauf hin, dass es für unterschiedliche Verwendungszwecke von Lernsoftware auch unterschiedliche Beurteilungskriterien geben sollte.

Auf diese Forderung wurde in der zweiten Generation von Bewertungsinstrumenten reagiert, indem Instrumente von vornherein für bestimmte Softwaretypen entwickelt wurden. Insgesamt wird seit Anfang der 1990er Jahre verstärkt darauf hingewiesen, dass „verschiedene Lernformen [...] unterschiedliche Computerprogramme [benötigen, d. Autor], deren Verwendungszwecke demzufolge differenziert betrachtet werden müssen“ (Mandl et al. 1992, S. 15ff.). Mandl et al. (ebenda) unterscheiden nach Kanselaar (1992) vier Lernformen, denen sie vier Computerprogrammtypen beordnen (siehe Abbildung 2). Diesen können wiederum spezielle Bewertungsinstrumente zugeordnet werden, die in die sogenannte zweite Generation von Bewertungsinstrumenten fallen.

Neben Bewertungsinstrumenten für verschiedene Softwaretypen und Lernstrategien wurden auch spezifischere Bewertungsinstrumente entwickelt, unter anderem für spezielle Aufgabenfelder (zum Beispiel Instrument zur Bewertung von Software zum Lesen lernen von Squires/McDougall 1994, S. 127), für verschiedene Zielgruppen von Lernern (etwa im Grundschulbereich: Schwarz/Lewis 1989 oder in der Erwachsenenbildung: Dejoy/Mills 1989), für verschiedene Zielgruppen von Bewertenden (beispielsweise computerunerfahrene Lehrer: EPASoft von Gräber et al. 1992; für computererfahrene Lehrer: Reeves/Harmon 1994 oder für Lerner: Akademie des Deutschen Beamtenbundes 1986).

Neben solchen Bewertungsinstrumenten für spezifische Zwecke entstanden weitere, die spezifische *und* allgemeine Teile verbinden sollten. Sie sind insgesamt allgemein gehalten, setzen aber für bestimmte Bereiche spezifische Schwerpunkte. Hierunter fallen zum Beispiel Kriterienkataloge, die zum einen aus einem allgemeinen Teil bestehen und zum anderen weitere Teilbereiche enthalten, die sich spezifischen Bereichen des Lernens mit Lernsoftware wid-

4 Ein Beispiel für ein solches Bewertungsinstrument ist der vom Arbeitskreis der deutschen Automobilindustrie entwickelte Kriterienkatalog AKAB.

5 Ein Beispiel für ein solches umfassendes Bewertungsinstrument stellt der datenbankgestützte Kriterienkatalog MEDA dar.

Lernform	Computerprogramm	Bewertungsinstrumente (Beispiel)
Lernen als Wiederholen und Memorieren	Übungsprogramm (etwa Drill- & Practice-Programme zum Vokabellernen)	„Drill and Practice Software Evaluation“, G. Bitter, A. Camuse (1994), Arizona State University
Lernen als interaktiver und konstruktiver Prozess	Tutorielle Programme	„Tutorial Software Evaluation“, G. Bitter/A. Camuse (1994), Arizona State University
Lernen als explorativer und entdeckender Prozess	Simulationsprogramme	„Simulation Software Evaluation“, G. Bitter/A. Camuse (1994), Arizona State University
Lernen als Rekonstruktionsprozess	Cognitive Tools (wie etwa Textverarbeitungswerkzeuge oder Autorenwerkzeuge, sogenannte Teachware [Mandl et al. 1992, S. 498])	„Word processing Software Evaluation“ G. Bitter/A. Camuse (1994), Arizona State University oder „Leistungsmerkmalskatalog für Autorenwerkzeuge“, Siemens AG

Abb. 2: Lernformen, Computerprogrammtypen und dazu kongruente Bewertungsinstrumente der 2. Generation (Lernformen und Lernprogramme vgl. Mandl et al. 1992, S. 16; Zuordnung der Bewertungsinstrumente vgl. Gräber 1996, S. 19)

men. Dies können zum Beispiel der Bereich affektiver Lernprozesse (vgl. etwa Martin 1998) oder spezielle Fragen zu einzelnen Lernsoftwaretypen (vgl. etwa Blease 1986) sein. Reay (1985, S. 79) entwickelt einen Kriterienkatalog für einen zweistufigen Bewertungsprozess. Zunächst bewertet ein Experte (in diesem Fall ein Lehrer) das Lernprogramm alleine und dann wird ein Schüler beim Lernen mit der Software beobachtet.

Zusammenfassend wird deutlich, dass die zweite Generation von Bewertungsinstrumenten aus sehr komplexen Instrumenten bestand, die für spezifische Fragestellungen ausgelegt waren. Eine differenziertere Bewertungsmöglichkeit war so zwar möglich, andererseits führte diese Entwicklung auch zu einer unüberschaubaren Menge an Bewertungsinstrumenten, da teilweise sogar für einzelne Programme eigene Kriterienkataloge zur Bewertung entwickelt wurden (vgl. Gräber 1996, S. 20). Eine intersubjektive Vergleichbarkeit war daher teilweise nicht mehr gegeben.

Die dritte Generation von Bewertungsinstrumenten versuchte dieser Instrumentenvielfalt entgegenzuwirken, aber trotzdem die Differenzierungskraft beizubehalten, die zuvor erzielt worden war. So zählt Gräber (ebenda, S. 22) zur dritten Generation von Bewertungsinstrumenten solche, die versuchen, in *einem* Instrument alle Ansprüche und Gegebenheiten zu berücksichtigen. Darunter fallen Instrumente wie die „Große Prüfliste von Lernsoftware“ (GPL), die Thomè (1988) im Rahmen einer Dissertation aus einer Synopse von zwölf Kriterienkatalogen erstellte. Dieser Katalog enthält 221 Bewertungssaspekte, die in

unterschiedliche Bewertungskategorien geordnet werden (vgl. Gräber 1996). Ein anderes Instrument, MEDA (Gräber 1991), stellt in Form einer Datenbank einen Katalog von über 300 Fragen zur Verfügung, die bei der Beschreibung, Analyse und Bewertung didaktischer Software eingesetzt werden können. Der Beurteiler muss in einer Dreistufenhierarchie seine eigene Bewertungsintention und Zielvorstellungen wählen. Am Schluss bekommt er eine Liste, die für seine jeweiligen Zwecke relevante Fragen präsentiert.

Die vierte Generation von Bewertungsinstrumenten benennt Gräber als prozessorientierte Instrumente (ebenda). Sie fokussieren nicht allein auf eine punktuelle Überprüfung von Softwaremerkmalen, sondern sollen eine Orientierungshilfe für Bildungsmöglichkeiten in unterschiedlichen Kontexten sein. Die Bewertungskriterien beziehen sich hier darauf, den gesamten Prozess der Entwicklung, der Auswahl und des Einsatzes eines Programmes begleitend zu evaluieren. Als umfangreichste Arbeit in diesem Zusammenhang kann die sogenannte „Evaluator’s Toolbox“ (Reeves 1992) angesehen werden. Sie stellt ein sogenanntes Electronic Performance Support System (EPSS⁶) dar. Die Toolbox enthält elektronische Informationen, Beispiele, Checklisten und Fragebögen zur Evaluation von allen Schritten der Entwicklung interaktiver multimedialer Programme. Alle Schritte werden evaluiert und erst auf Basis der Ergebnisse wird der jeweils nächste Entwicklungsschritt geplant. Auch das Projekt Evaluationsnetz verfolgt das Konzept, für unterschiedliche Phasen der Konzeption, des Designs und der Durchführung angemessene Kriterien zu spezifizieren. [www.evanetz.de].

Insgesamt kann der Schluss gezogen werden, dass sich die Bewertungsinstrumente in zweierlei Hinsicht weiterentwickelt haben: Während die erste Generation an Bewertungsinstrumenten noch hauptsächlich technische Aspekte fokussierte, wurden nach und nach immer stärker pädagogisch-didaktische Elemente berücksichtigt. Der zweite Entwicklungsstrang ist in Richtung einer zunehmend stärkeren Differenzierung von Kontexten zu beobachten. Eine fundierte, theoriegeleitete und empirisch validierte Bewertungsgrundlage bieten die wenigsten derzeit verfügbaren Instrumente. Unter anderem daran schließt sich auch die Kritik an, die zur Bewertung mittels Katalogen von Qualitätskriterien oder Checklisten geäußert wird.

Zunächst weist Meier (1995) darauf hin, dass in vielen Qualitätskriterienkatalogen überwiegend Kriterien aus dem Bereich „Gestaltung der Bildschirmoberflächen“ oder „Technik des Programmablaufs“ enthalten sind, didaktische Kriterien also oftmals unterrepräsentiert sind. Aus Metastudien zur Lernwirksamkeit von multimedialen Lernarrangements geht jedoch hervor, dass gerade

6 Gerry (1991) definiert die Funktionen eines EPSS folgendermaßen: „A system that provides electronic task guidance and support to the user at the moment of need. EPSS can provide application help, reference information, guided instructions and/or tutorials, subject matter expert advice and hints on how to perform a task more efficiently. An EPSS can combine various technologies to present the desired information. The information can be in the form of text, graphical displays, sound, and video presentations“ (ebenda).

die didaktischen Konzepte, die im E-Learning umgesetzt sind und die zu Grunde liegenden Lernarrangements und -situationen günstigere Einflüsse auf den Lernprozess haben als die verwendete „delivery technology“ (vgl. Weidenmann 1997; Kulik/Kulik 1991, 1994).

In der Literatur werden darüber hinaus vier Argumente gegen die Verwendung von Kriterienkatalogen zur Beurteilung der Qualität von Lernsoftware genannt, die wie folgt zusammengefasst werden können:

- 1) **Mangelnde Beurteilerübereinstimmung bei der Quantifizierung von Qualitätskriterien:** Fricke (2000, S. 76) weist darauf hin, dass ein Merkmal eines Lernprogrammes dann als valide – und damit als lernwirksam – angesehen wird, wenn eine (zumindest geringe) Korrelation zwischen Ausprägungsgrad des Merkmals und dem Lernergebnis zu erkennen ist. Reduziert sich die Messgenauigkeit des Merkmals, so reduziert sich auch seine Validität. Bei der Quantifizierung von Merkmalen in Bezug auf eine Lernsoftware kommt es oftmals zu großen Diskrepanzen zwischen den Beurteilern. In diesem Fall wird von einer mangelnden Beurteilerübereinstimmung gesprochen. Fricke (ebenda) berichtet von einem Versuch, bei dem sechs Experten ein und dasselbe Lernprogramm anhand des oben beschriebenen Instruments MEDA unabhängig voneinander bewerten sollten. Weder war eine hinreichende Übereinstimmung in Bezug auf die Bewertung der Relevanz der Kriterien zu beobachten noch in Bezug auf die Einschätzung der Ausprägung der Kriterien bei der vorliegenden Lernsoftware.
- 2) **Geringe praktische Signifikanz der Qualitätskriterien:** Bei der Beurteilung von Lernsoftware ist festzustellen, dass Kriterien, die objektiv gut operationalisierbar und damit gut messbar sind, wie zum Beispiel Fragen nach dem Vorhandensein einer Hilfeschnittfläche in einem Lernprogramm, oftmals nur geringe Auswirkungen auf den tatsächlichen Lernerfolg haben. Merkmale einer Lernsituation hingegen, die in der Regel nur schwer operationalisiert werden und unter Gesichtspunkten der Beurteilerübereinstimmung auch nicht valide gemessen werden können, wie zum Beispiel der Enthusiasmus eines Lerner oder Lehrers, scheinen einen sehr hohen Einfluss auf den Lernerfolg zu haben (vgl. Fricke 2000). Das bedeutet, dass viele Merkmale, die eine hohe Messgenauigkeit aufweisen, nur geringe Validitäten in Bezug auf das Lernergebnis haben. Umgekehrt konnten mit Merkmalen, die nur eine geringe Messgenauigkeit bieten, bessere Voraussagen in Bezug auf das Lernergebnis getroffen werden (vgl. Rosenshine/Furst 1971). Fricke (ebenda) betont die Notwendigkeit, empirisch überprüfte Lehr-Lerntheorien zur Erklärung dieser Paradoxien heranzuziehen, um eine unbefriedigende Zusammenstellung von Kriterien mit augenscheinlicher, aber unbewiesener Validität („face validity“) zu vermeiden.
- 3) **Differentielle Methodeneffekte bei Qualitätskriterien:** Das Prinzip von differentiellen Methodeneffekten nach Cronbach und Snow (1971) lässt sich auch auf die Erforschung der Effektivität eines Lernarrangements anwenden. Legt man das von Fricke (1995) vorgeschlagene Paradigma zur Konstruktion und

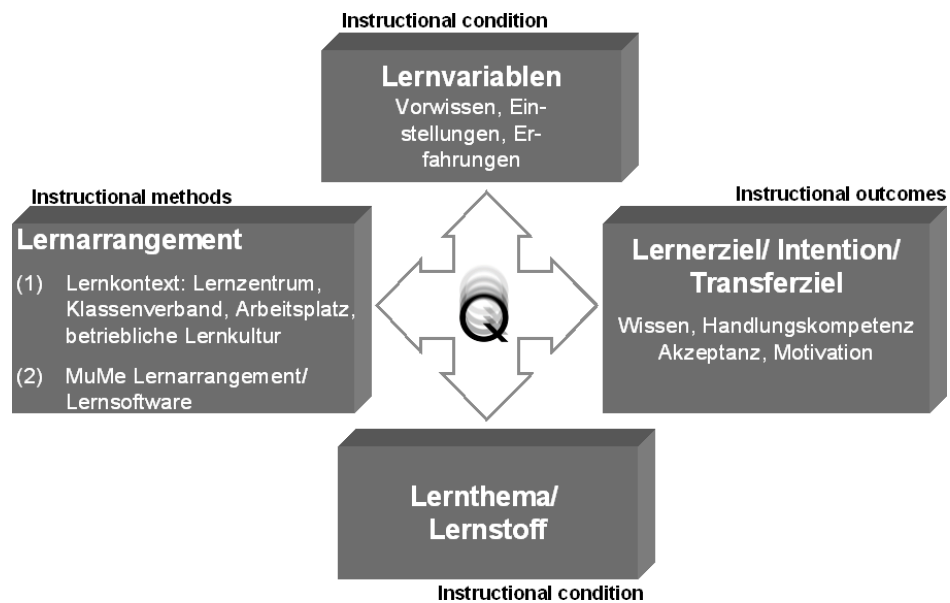


Abb. 3: Konstituierende Faktoren von Lehr-Lernumgebungen (nach Fricke 1995).

Evaluation multimedialer Lehr-Lernumgebungen zu Grunde (vgl. Fricke 1995, S. 405), dann besteht ein Lernarrangement aus vier Hauptfaktoren: (a) die multimediale Lernumgebung, (b) die Lernervariablen, (c) das Lerntbema und (d) das Lernergebnis (siehe Abbildung 3).

Die Bewertung von Lernarrangements mit Kriterienkatalogen folgt in der Regel dem Paradigma der Wirkungsforschung, das heißt, dass das Lernergebnis als abhängige Variable und die anderen Variablen als unabhängige Variablen angesehen werden. Die Ausprägungen der zu Grunde liegenden Kriterien bestimmen dann die Wirkung der unabhängigen Variablen auf die abhängige Variable, also das Lernergebnis.

Differentielle Methodeneffekte wirken dabei nun auf die Weise, dass die Zusammenhänge von unabhängigen und abhängigen Variablen geschwächt werden. Die Ursache hierfür liegt darin, dass die unabhängigen Variablen eines Lernarrangements nicht nur auf die abhängigen Variablen Einfluss nehmen, sondern sich auch untereinander beeinflussen. So sind in Untersuchungen Wechselwirkungen zwischen Lernumgebung und Lernstoff (vgl. Jung 1994), Lerner und Lehrstoff (vgl. Fricke 1989) und den Merkmalen von Lernumgebungen untereinander (vgl. Fricke 1989; Meier/Baratelli 1991) nachgewiesen worden.

Fricke (2000) kommt zu dem Schluss, dass die Bewertung einer Bildungssoftware nur aus ganzheitlicher Sicht erfolgen kann. Die aufgezeigten Wechselwirkungen führen dazu, dass es so etwas wie *die* Effektivität eines Lernprogrammes nicht geben kann – und auch nicht *die* Validität von Kriterien. Berücksichtigt man ein Lernarrangement ganzheitlich, so kann ein Kriterium einer Lernsoftware in einem Fall ein „positives“ Qualitätskriterium sein und in einem anderen Fall – etwa bei einer anderen Zielgruppe – eher ein Nachteil.

4) Nichtberücksichtigung des Verwertungszusammenhangs einer Bildungssoftware:

Oftmals werden multimediale Lernprogramme als Substitut für konventionelle Fortbildungen eingesetzt. Dabei besteht die Hoffnung, Kosten zu sparen, da die Mitarbeiter am Arbeitsplatz lernen und keine – oder zumindest nur geringere – Ausfallzeiten entstehen und keine Reisekosten anfallen (zu Kosten und Nutzen beim E-Learning vgl. Gröbriel 2003). Es ist unstrittig, dass E-Learning in der beruflichen Weiterbildung nur dann erfolgreich eingesetzt werden wird, wenn es Kostenvorteile mitbringt – oder zumindest bei gleichen Kosten eine viel versprechendere Alternative zu konventionellen Qualifizierungsmaßnahmen darstellt. Genau diese Aspekte werden in den meisten Kriterienkatalogen nicht mitberücksichtigt. Dazu kommt, dass in Studien gezeigt werden konnte, dass der Einsatz von Lernprogrammen in Betrieben oftmals zwar zu den gleichen Lerneffekten führt, aber insgesamt die Arbeitsorganisation und den Betriebsablauf so effektiviert, dass damit Kosteneinsparungen erzielt werden konnten (vgl. Brinker 1991). Anders ausgedrückt scheint es von Bedeutung, dass die Variable Lernergebnis in Fricke's Modell (siehe Abbildung 3) noch um Merkmale der betriebswirtschaftlichen Kosten-Nutzen-Rechnung erweitert wird.

Insgesamt kann der Schluss gezogen werden, dass es gravierende Argumente gegen die *ausschließliche* Verwendung von Kriterienkatalogen zur Bewertung der Qualität von Lernsoftware gibt. Die Beurteilerübereinstimmung kann dabei zwar durch gezieltes Training verbessert werden (vgl. dazu auch Flanders 1970) und auch Aspekte der betrieblichen Kosten-Nutzen-Rechnung können in Qualitätskriterien zumindest ansatzweise einfließen. Schwerer wiegen jedoch die Kritikpunkte der „geringen praktischen Signifikanz“ von Kriterien und der „differentiellen Methodeneffekte“.

In Bezug auf einen diskursiven Qualitätsbegriff kann eine deutliche Abgrenzung vorgenommen werden. Während ein am Lernenden orientierter Qualitätsbegriff sich auf den ausgehandelten Lernprozess aus Sicht des lernenden Subjektes bezieht und Qualität damit an Bedingungen und Prozessen dieser Aushandlung beziehungsweise dieses Erbringungsverhältnisses festmacht, liegt bei einer Bewertung durch Qualitätskriterien ein von vornherein normativ fixierter Qualitätsbegriff vor.

3. Evaluationsverfahren für E-Learning-Arrangements

Eine weitere Möglichkeit, Qualität beim E-Learning zu sichern beziehungsweise zu bewerten, besteht in der Anwendung von Evaluationsverfahren für E-Learning-Arrangements. Der Begriff Evaluation ist sehr vielschichtig und kann sehr unterschiedlich definiert werden. Die Definitionsvielfalt zeigt sich in der Literatur (Fricke [2002] führt eine umfangreiche Sammlung von Literaturstellen mit unterschiedlichen Evaluationsbegriffen auf: Stufflebeam 1969, 1972,

1994; Wulf 1972; Stiefel 1974; Lange 1983; Prell 1981 und 1986; Gerl/Pehl 1983; Wittmann 1987; Fricke 1986; Wottawa 1986; Lösel/Nowack 1987; Will et al. 1987; Rossi et al. 1988; Wottawa/Thierau 1990; Berendt/Stary 1993; Götz 1993; Seidel/Park 1994).

E-Learning kann aus den verschiedensten Perspektiven evaluiert werden. Die Aufarbeitung des Forschungsstandes zu diesem Thema kann eingeteilt werden in verallgemeinerbare Ergebnisse, die vielfältige Evaluationsstudien bislang hervorgebracht haben und Evaluationskonzepte, die speziell für das E-Learning entwickelt sind. Im folgenden Abschnitt werden vor allem Evaluationskonzepte berücksichtigt, da die bisherige Evaluationsforschung nur wenige verallgemeinerbare Ergebnisse vorgebracht hat. Diesen Standpunkt vertritt auch Keil-Slavik (1999, S. 12), der beispielsweise die Situation im Bereich multimedialer Lehre in der Hochschule folgendermaßen einschätzt: „... zu jedem Befund lässt sich mindestens eine Studie mit einem gegensätzlichen Befund finden.“ Seiner Meinung nach könne das Ziel von Evaluation daher nicht sein, eine abschließende Bewertung vorzunehmen, sondern sie vielmehr evolutionär im Sinne ständiger Weiterentwicklung zu betreiben.

Der Begriff der Evaluation kann nicht allgemeingültig definiert werden, da er in unterschiedlichsten Kontexten und theoretischen Zusammenhängen verwendet wird. Evaluation geht aber über die reine Messung von einzelnen Variablen, zum Beispiel im Sinne empirischer Lehr-Lernforschung (siehe dazu Kapitel 4.3), hinaus und stellt in der Regel eine umfassende Bewertung von erhobenen Daten dar. Vier Elemente werden von Will et al. (1987) besonders hervorgehoben:

- „1. Evaluation ist ziel- und zweckorientiert. Sie hat primär das Ziel, praktische Maßnahmen zu verbessern, zu legitimieren oder über sie zu entscheiden.
2. Grundlage der Evaluation ist eine systematisch gewonnene Datenbasis über Voraussetzungen, Kontext, Prozesse und Wirkungen einer praxisnahen Maßnahme.
3. Evaluation beinhaltet eine wertende Stellungnahme, d.h. die methodologisch gewonnenen Daten werden auf dem Hintergrund von Wertmaßstäben unter Anwendung bestimmter Regeln bewertet.
4. Evaluation bezieht sich im Gegensatz zur personenbezogenen Leistungsfeststellung oder Testung auf einzelne Bereiche geplanter, durchgeführter oder abgeschlossener Bildungsmaßnahmen. Sie zielt also in der Regel nicht primär auf die Bewertung des Verhaltens (z.B. von Leistungen) einzelner Personen, sondern ist Bestandteil der Entwicklung, Realisierung und Kontrolle planvoller Bildungsarbeit“ (ebenda, S. 14).

Im Gegensatz zur Bewertung von E-Learning-Angeboten mit Qualitätskriterienkatalogen oder Checklisten steht bei der Evaluation nicht so sehr die Messung angebots-/produktbezogener Merkmale im Vordergrund, sondern die Be-

Wertung von Lernprozessen und das Erheben von Urteilen bzgl. Qualität, Wirkung (Akzeptanz, Lernerfolg) und wahrgenommenem Nutzen.

3.1 Grundlegende Evaluationsformen und -methoden

In der Evaluationsforschung wird allgemein unterschieden zwischen formativer und summativer Evaluation. Formative Evaluation dient in der Regel eher der Qualitätssicherung. Ihr Ziel ist die Ermittlung von Schwachstellen. Die Forschungen verlaufen entwicklungsbegleitend und dienen der Optimierung bereits während des laufenden Prozesses. Summative Evaluation dient dahingegen eher der Kontrolle von Qualität, Wirkung und Nutzen eines Bildungsangebotes. Der Evaluationszeitpunkt liegt daher zu meist am Ende eines Prozesses. Götz (1993, S. 105ff) stellt weitere Evaluationsmodelle vor, beispielsweise Input-Outputevaluation, Fremd- vs. Selbstevaluation, Teil- und Gesamtevaluation, subjektive und objektive Evaluation, direkte und indirekte Evaluation, intrinsische und extrinsische Evaluation, additive und integrierende Evaluation sowie interne und externe Evaluation. Eine weitere Unterscheidung trifft Tergan (in Schenkel 2000) im Bereich von Evaluation, indem er zwischen Prozess- und Produktevaluation differenziert. Es wird dabei zwischen solchen Evaluationen, bei denen der Evaluationsgegenstand eher die Prozesse einer Entwicklung sind, und solchen, die sich auf das fertige Produkt am Ende einer Entwicklung konzentrieren, unterschieden.

Spezielle Evaluationskonzepte für E-Learning benötigen dahingehend eine theoretische Grundlegung, *welche* Prozesse beim Lernen mit Medienunterstützung sich in *welcher* Weise beeinflussen. Eine theoretische Basis bietet hier beispielsweise die Theorie des Instruktionsdesigns von Reigeluth (1983). Nach Reigeluth gibt es drei Hauptkomponenten im Instruktionsprozess: Lernergebnisse („instructional outcomes“), Randbedingungen der Instruktion („instructional conditions“) und die Instruktionmethode („instructional methods“). Fricke (vgl. 1995, 2000, 2002) entwickelt auf Basis dieses Modells das bereits vorgestellte „Paradigma zur Konstruktion und Evaluation von Lehr-/Lernumgebungen“ (siehe auch Kapitel 2, Abbildung 3).

Das Modell kann deskriptiv oder präskriptiv gelesen werden (vgl. Fricke 1995, 2000, 2002). Im Sinne einer deskriptiven Theorie gibt das Modell Aufschluss darüber, welche Effekte (Outcomes) verschiedene Lehrmethoden bei gegebenen Randbedingungen haben. Bei präskriptiver Lesart fungieren nicht die Lehrmethoden, sondern die Lernergebnisse als unabhängige Variable. Eine deskriptive Theorie soll also dazu dienen, das Zustandekommen von Lernergebnissen bei bestimmten Lehrmethoden zu erklären, während eine präskriptive Theorie geeignete Lehrmethoden zum Erreichen bestimmter Lernergebnisse Vorschlägen soll.

Das Modell ermöglicht es auf diese Weise, zwischen unterschiedlichen Evaluationszielen und unterschiedlichen Komponenten eines Lernarrangements zu

differenzieren. In Fricke's Paradigma werden auch die Aspekte von Konstruktion einerseits und Evaluation andererseits zusammengeführt. Hier wird die Bedeutung von prozessbegleitender Evaluation für die Qualität eines Lernarrangements deutlich. Ausdrücklich weisen auch Reigeluth (1983) und Reinmann-Rothmeier et al. (1994) auf die Bedeutung einer solchen formativen Evaluation auf allen Stufen der Konstruktion und Durchführung von E-Learning-Arrangements hin: „For the educational technology field, evaluation was now being viewed as an integral and ongoing part of the instructional development process“ (ebenda in Fricke 2002, S. 450).

3.2 Überblick über instruktionstheoretische Evaluationskonzepte für E-Learning

Kirkpatrick (1994) hat in einer Artikelserie „Techniken zur Evaluation von Trainingsprogrammen“ in den USA einen Evaluationsansatz formuliert, der zwischen vier Evaluationsebenen unterscheidet: Reaktionsebene, Lernebene, Verhaltensebene und Ergebnisebene. Kirkpatrick's Modell erleichtert den Zugang zu Evaluationen und präzisen Ergebnissen. Zu häufig werden Aussagen über die Qualität von Weiterbildungsmaßnahmen getroffen, ohne dass deutlich wird, ob sie sich auf die Reaktion der Lernenden, auf die Handlungsfähigkeit der Absolventen, auf die Kosten oder auf den Erfolg der Weiterbildung beziehen.

Daneben existiert eine Reihe von Evaluationskonzepten speziell für multimediale Lehr-Lernumgebungen. Reigeluth's (1983, S. 77ff.) Instruktionsdesign kann als Grundlage für die Evaluation von multimedialen Lehr-Lernumgebungen genutzt werden. Zur Konstruktion solcher Lernarrangements nach dem Paradigma des Instruktionsdesigns unterscheidet er fünf Stufen:

- 1) **Instructional Design:** Auswahl geeigneter Ziele und entsprechender Instruktions- und Lehrmethoden für das Erreichen eines bestimmten Zieles bei einer vorgegebenen Zielgruppe
- 2) **Instructional Development:** Entwicklung und Konstruktion konkreter Lehrmethoden (Reigeluth benutzt hier die Metapher des beziehfertigen Rohbaus eines Hauses)
- 3) **Instructional Implementation:** Genaue Anpassung der Lehrmethoden an die konkreten Rahmenbedingungen (Reigeluth benutzt hier die Metapher des „endgültigen Ausbaus eines Hauses“ nach den Wünschen eines Mieters)
- 4) **Instructional Management:** Konkrete Einsatzplanung eines Qualifizierungsprogrammes, zum Beispiel in einem Unternehmen
- 5) **Instructional Evaluation:** Hier geht es um „[...] understanding, improving, and applying methods for assessing the effectiveness and efficiency of all [...] activities“ (Reigeluth 1983, S. 9)

Im Evaluationsmodell von Ross und Morrison (vgl. Fricke 2002, S. 451) werden einige dieser Elemente aufgenommen. Formative und summative Ansätze, quantitative und qualitative Methoden sowie Ansätze, die sowohl aus dem Instruktionsdesign als auch aus der konstruktivistischen Lerntheorie stammen, sind hier kombiniert worden. Ihr Evaluationsmodell enthält vier Stufen:

- 1) **Needs Analysis:** Festlegung der Ziele und Fragen der Evaluationsstudie
- 2) **Methodology:** Auswahl der Evaluationsmethoden und Implementierung der Methoden auf fünf Stufen
 - **Programm Analysis:** Analyse der Lehrziele und Präsentationsformen des Programms
 - **Participant Analysis:** Festlegung der Personenstichprobe
 - **Evaluation design:** Aufstellen des Versuchsplans und Festlegung der Messmethoden (sowohl formative als auch summative; qualitative als auch quantitative Methoden)
 - **Instrumentation:** Zusammenstellung der Evaluationsinstrumente zu einem kohärenten Gesamtevaluationsplan unter Berücksichtigung der gegebenen Beschränkungen
 - **Implementation:** Festlegen des zeitlichen Verlaufs der Evaluation
- 3) **Data Analysis and Interpretation:** Auswertung und Aufbereitung der Daten
- 4) **Disseminating Results:** Erstellung und Übergabe eines Abschlussberichts

Ross und Morrison betonen, dass die Evaluation als ein iterativer Prozess gesehen werden kann, der zu einer fortwährenden Verbesserung des Programms führen kann. Ihr Vorschlag kann als Rahmenmodell für die Evaluation multimedialer Lehr-Lernarrangements angesehen werden und ist dementsprechend auch für E-Learning-Arrangements anwendbar. Fricke (2002) bündelt beide zuvor vorgestellten Evaluationsbeziehungsweise Konstruktionsansätze in einem Modell (Fricke 2002, S. 455ff.). Dazu erweitert er das von ihm nach Vorschlägen von Reigeluth entwickelte „Paradigma zur Konstruktion und Evaluation multimedialer Lehr-Lernumgebungen“ (siehe Abbildung 3) um zwei weitere Rahmenbedingungen des Lehrens und Lernens: zum einen handelt es sich dabei um die Rahmenbedingung Lehr-Lerntheorien, da die Bewertung von Lernarrangements nicht ohne die Berücksichtigung der implizit oder explizit zu Grunde liegenden Lerntheorie vorgenommen werden kann. Zum anderen ergänzt er das Modell um die Komponente allgemeiner Rahmenbedingungen, wie zum Beispiel politische Vorgaben, finanzielle Ressourcen, Produktionskosten, Vorgaben für Qualifikationen und Lehrziele, die seinem Verständnis nach ebenfalls eine wichtige Rahmenbedingung für die Bewertung von Lernarrangements darstellen (siehe Abbildung 4).

Je nachdem, ob mit einem präskriptiven oder einem deskriptiven Ansatz evaluiert wird, kommt er zu dem Schluss, dass sich insgesamt neun Evaluationsformen differenzieren lassen. Diese neun Evaluationsformen stellen sich als aufeinander aufbauende Evaluationsschritte für jeweils einen Teilschritt bei der Kon-

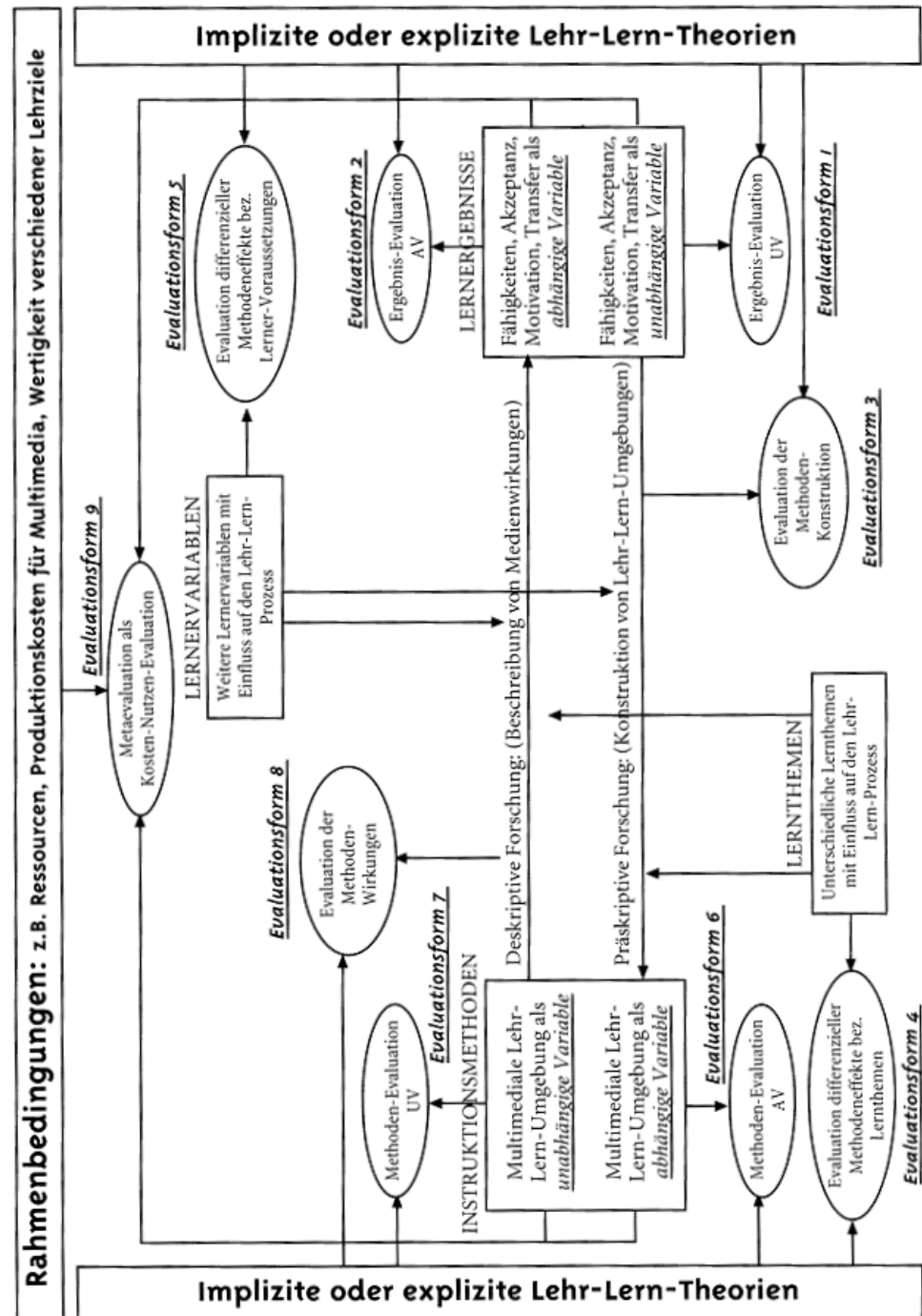


Abb. 4: Evaluationsformen auf Basis des Paradigmas zur Konstruktion und Evaluation multimedialer (Fricke 2002, S. 456).

struktion einer multimedialen Lehr-Lernumgebung dar (zusammengefasst nach Fricke 2002, S. 457ff):

- 1) Die **Ergebnisevaluation** (Ergebnis als unabhängige Variable): Überprüfung der Lehrziele mit den Lehr-Lerntheorien, die Lernergebnisse sind dabei unabhängige Variablen.
- 2) Die **Ergebnisevaluation** (Ergebnis als abhängige Variable): Überprüfung der tatsächlich erreichten Lernergebnisse anhand von Lehr-Lerntheorien, die Lernergebnisse sind hier abhängige Variablen.

Fricke (ebenda) unterscheidet im Weiteren zwischen Evaluation von Methodenkonstruktion, Evaluation von (differentiellen) Methodeneffekten (s. auch Kapitel 2), Methodenevaluation und Evaluation von Methodenwirkungen:

- 3) Evaluation der **Methodenkonstruktion**: Überprüfung der geplanten Lehrmethoden mit den zu Grunde liegenden Lehr-Lerntheorien, zumeist geschieht dies als entwicklungsbegleitende, optimierende Evaluation.
- 4) Überprüfung differenzieller **Methodeneffekte** bezüglich unterschiedlicher Lernthemen: Während der Methodenkonstruktion müssen Theorien und Erkenntnisse über differenzielle **Methodeneffekte** bezüglich unterschiedlicher Lernthemen berücksichtigt werden.
- 5) Überprüfung differenzieller **Methodeneffekte** bezüglich unterschiedlicher Lernvariablen: Während der Methodenkonstruktion müssen auch hierbei Theorien und Erkenntnisse über differenzielle Methodeneffekte bezüglich der Lernvariablen herangezogen werden.
- 6) **Methodenevaluation 1** (Methoden als abhängige Variable): Bei dieser Evaluationsform wird das fertig konstruierte Lernarrangement auf Übereinstimmung mit den zu Grunde liegenden Lehr-Lerntheorien überprüft, die Instruktionsmethoden fungieren dabei als abhängige Variable.
- 7) **Methodenevaluation 2** (Methoden als unabhängige Variable): Neue Instruktionsmethoden, die „ad-hoc“ in ein Lernarrangement eingebracht werden (beispielsweise vom Anwender) und nicht das Ergebnis eines Konstruktionsprozesses darstellen, können mit dieser Evaluationsform überprüft werden, indem sie als unabhängige Variable angesehen werden.
- 8) Evaluation der **Methodenwirkungen**: Hier werden die Medien- und Methodenwirkungen im Sinne der klassischen Medienwirkungsforschung evaluiert.
- 9) **Metaevaluation** als Kosten-Nutzen-Evaluation: Hier geht es nicht mehr um eine theoriegeleitete Konstruktion oder Überprüfung multimedialer Lehr-Lernumgebungen, sondern um eine Analyse der eingesetzten Ressourcen im Vergleich zu den Wertigkeiten der erreichten Leistungen (zum Beispiel Wissensbestände, Kompetenzzuwächse).

Die einzelnen Evaluationsformen zeigen, dass Evaluation während des gesamten Konstruktions- und Durchführungsprozesses auf ein multimediales Lehr-Lernarrangements angewendet werden kann. Fricke (2002) betont die Bedeutung der Orientierung und Überprüfung des Evaluationsprozesses an vorhande-

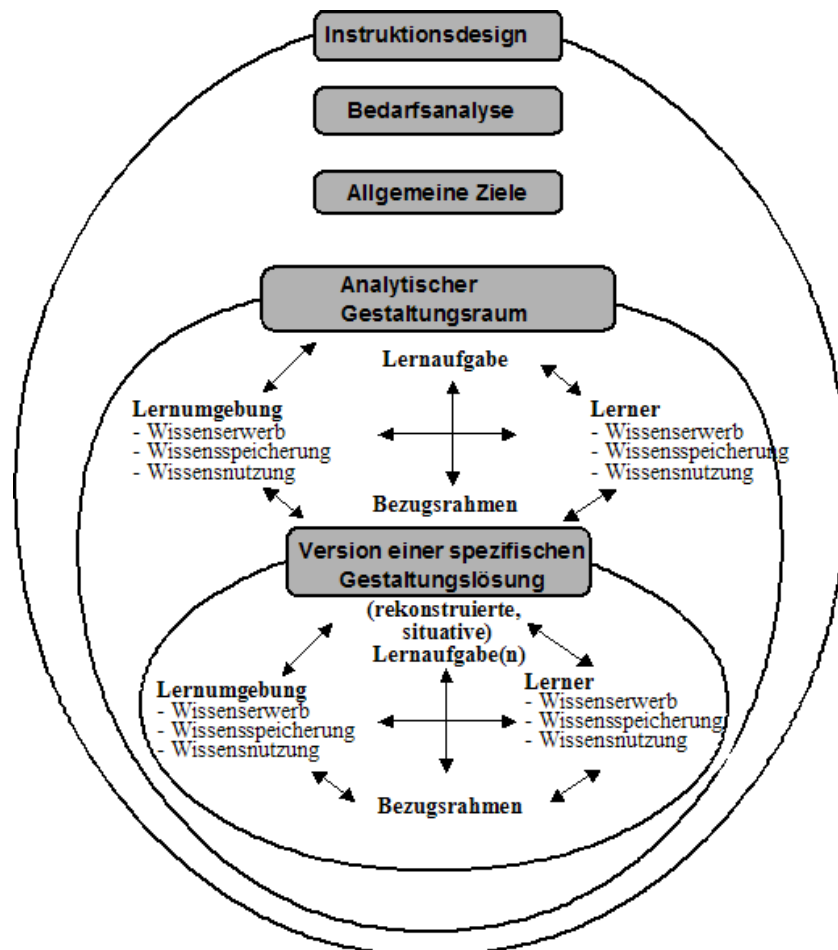


Abb. 5: Evaluationsmodell Dresdner Ei (Schott 2000).

nen Lehr-Lerntheorien. Wie die oben aufgeführten Evaluationsformen zeigen, ist an mindestens neun Stellen im Konstruktionsprozess eine solche Überprüfung sinnvoll. Die neun Evaluationsformen können so der Qualitätssicherung im Sinne einer durchgehenden Orientierung an gleich bleibenden Bestimmungsgrößen dienen.

Ein weiterer Evaluationsansatz, der ebenfalls auf einer Instruktionstheorie, der sogenannten „Universal Constructive Instructional Theory“ (UCIT, vgl. Dijkstra et al. 1997; Tennyson et al. 1997), aufbaut, liegt von Schott (2000) vor. Die sogenannte „Effiziente Lern- und Informations-System-Evaluation“ (ELISE) besteht aus zehn Ablaufschritten und stellt ein System an Bewertungsschritten dar, das eine strukturierte Evaluation ermöglicht. Es kombiniert Aspekte des Instruktionsdesigns und der Instruktionstheorie UCIT in einem Modell, dem sogenannten „Dresdner Ei“ (siehe Abbildung 5).

Dieses Modell stellt ebenfalls eine Verbindung von Konstruktionsprozessen und Evaluationsschritten dar. Der Konstruktionsprozess gemäß „Dresdner Ei“ geht von einer Bedarfsanalyse aus, formuliert ein Gesamtziel, prüft analytisch den möglichen Gestaltungsraum anhand der Komponenten Lernaufgabe, mögliche Lernumgebung, Lerner und Bezugsrahmen und realisiert dann ein Instruktionsvorhaben in Form einer spezifischen situativen Lernlösung (siehe Abbildung 5).

Der Evaluationsprozess geht nun den umgekehrten Weg. Vom Endprodukt her beginnend, ist die Evaluation zunächst mit einem fertigen multimedialen Lern- beziehungsweise Informationssystem konfrontiert. (In der Terminologie der UCIT wird dies als *situativ rekonstruierte* Lernaufgabe, eingebettet in eine *situative rekonstruierte* Lernumgebung [siehe Abbildung 5] bezeichnet.) Die Frage nach den Alternativen zu diesen Aufgaben spannt nun zunächst ein analytisch-situatives Möglichkeiten-Grenzen-System auf.

Danach wird gefragt, ob mit der eingesetzten Software das angestrebte Gesamtziel erreicht wird und die Probleme, derentwegen die Software entwickelt wurde, auch damit gelöst werden. Die anfangs eingeforderte theoretische Grundlage des Evaluationsansatzes ist durch die enge Verknüpfung mit der UCIT gewährleistet, jedoch liegen bislang nur wenig praktische Erfahrungen mit dem Modell vor (vgl. Schott et al. 2000).

Ein solches ganzheitliches Evaluationsverständnis, wie in den oben beschriebenen Ansätzen, das sich nicht nur auf die Messung von erreichten Effekten bezieht, sondern während des ganzen Konstruktionsprozesses entwicklungsbegleitende Optimierung leisten kann, findet sich auch bei Tergan (in Schenkel 2000), Schenkel (2000) und Mandl/Reinmann-Rothmeier (1995), von denen ebenfalls dokumentierte Evaluationsansätze vorliegen. In Bezug auf Qualität wird deutlich, dass Evaluation als Methode der Qualitätssicherung fungieren kann. Sie liefert jedoch keine konkreten Dimensionen subjektiver Qualität, sondern stellt in den vorgestellten Ansätzen vielmehr jeweils einen methodischtheoretischen Rahmen dar, in dem Qualität bewertet werden kann.

3.3 *Ein subjektorientierter Evaluationsansatz*

Ein weiterer Evaluationsansatz von Zimmer und Psaralidis (2000) basiert explizit auf einer Subjektperspektive. Er wendet sich gegen eine Evaluation im Sinne einer reinen Medienwirkungsforschung und stellt vor allem das Ziel beruflichen Lernens in den Vordergrund, welches hier als Zuwachs beruflicher Handlungskompetenz definiert wird.

Zimmer und Psaralidis (2000) argumentieren in ihrem Beitrag, dass der Lernerfolg die Qualität einer Lernsoftware bestimme und dieser durch den Erwerb von Handlungskompetenz definiert sei. Bislang werde jedoch in vielen Evaluationsmodellen beziehungsweise Evaluierungen davon ausgegangen, dass die Qualität einer Lernsoftware den Lernerfolg bestimmt – und nicht umge-

kehrt. Die Autoren kritisieren solche Evaluationsmodelle als zu weit angelehnt an „Maschinenmodelle der Wirkungsforschung“ (ebenda). Wird Lernen aber als ein Prozess begriffen, bei dem der Lernende den Lernerfolg (Handlungskompetenz) in Ko-Produktion mit dem Lernangebot (Lehrender, Lernarrangement etc.) erst erstellt, so ist es schlicht nicht möglich, die Qualität von Lernsoftware im Sinne eines Handlungskompetenzzuwachses mit Qualitätskriterien im Vorhinein zu prognostizieren.⁷ Dieser Evaluationsansatz definiert die Lernsituation als eine ganzheitliche Situation, die nur in einer ganzheitlichen Betrachtung Aufschluss über die Wirkung der Lernsoftware liefern kann. Ziel ist es dabei, subjektwissenschaftliche Erkenntnisse zu erlangen, also Erkenntnisse über das Lernen vom Standpunkt des Subjekts aus (vgl. Holzkamp 1983, 1993; Zimmer 1987).

Beim instruktionstheoretischen Lernprozess wird der Lernerfolg als durch die Lernsoftware bewirkt angesehen. Die Lernsoftware passt sich dabei durch die Aktionen der Lerner an den Prozess der Erzeugung des Lernerfolges an und soll ihn individuell optimieren. Die Lernumgebung sowie die individuellen und organisatorischen Kontexte werden dabei als intervenierende Variablen berücksichtigt, denen gegenüber die Lerner forschungsmethodologisch als „ausgeliefert“ betrachtet werden.

Beim handlungsbestimmten Lernprozess wird dagegen forschungsmethodologisch von dem am eigenen Lernerfolg interessierten Subjekt ausgegangen, das mit dem Ziel des Kompetenzgewinns die Lernsoftware nutzt. Die Lernumgebung sowie die individuellen und organisatorischen Situationen werden dabei als Bereiche betrachtet, die den Lernerfolg fördern oder hemmen, aber durchaus auch vom Subjekt mitgestaltet oder kontrolliert werden können.

Um die primäre Evaluationsfrage beantworten zu können – welcher Lernerfolg auf das Lernen mit einer Lernsoftware zurückzuführen ist –, muss zunächst der Evaluationsgegenstand genau definiert werden. Der Gegenstand ist dabei nicht die Lernsoftware, sondern zum einen der unmittelbare Lernerfolg mit der Lernsoftware und der daraus folgende Zugewinn an subjektiver Handlungskompetenz, etwa in einer Arbeitssituation, und zum anderen der Lernprozess, der zu Lernerfolg und Kompetenzgewinn führt.

Methodologisch schlagen Zimmer und Psaralidis (2000) dafür die Methode der logischen Rekonstruktion vor. Es geht dabei darum, den begründeten Zusammenhang der Handlungskompetenzen und der zu ihnen führenden Handlungen in ihren Verbindungen, Brüchen, Gegensätzlichkeiten und Widersprüchlichkeiten zu rekonstruieren, und zwar zu einem bestimmten Zeitpunkt und in einer bestimmten Situation hinsichtlich konkreter Handlungsziele. Die logische Rekonstruktion unterstellt, dass alle Handlungen von Subjekten einen

⁷ Zimmer und Psaralidis (2000) lehnen sich dabei an die Definition des Lernprozesses von Holzkamp an: „Ein Lernprozess vollzieht sich immer im Subjekt durch manuelle und geistige Handlungen an dinglichen oder ideellen Gegenständen in kulturellen Zusammenhängen und Prozessen mit dem Ziel eines Gewinns an subjektiver Handlungskompetenz“ (vgl. Holzkamp 1983, S. 236).

erkennbaren Grund haben (vgl. Holzkamp 1983, S. 342ff.). Unter dieser Voraussetzung können die Dimensionen des Lernhandelns bestimmt und beschrieben werden.

Qualität ist demnach nicht eine per se vorhandene Eigenschaft einer Lernsoftware (Wirkungsforschung), sondern wird erst im Lernprozess konstruiert. Darüber hinaus wird der Lernprozess auch als situativ beeinflusst angesehen und als ganzheitliche Handlungssituation wahrgenommen. Subjektive Qualität beim E-Learning zu bestimmen heißt also auch, die ganzheitliche Handlungssituation zu berücksichtigen.

3.4 *Besonderheiten der Evaluation vernetzten Lernens*

Abschließend soll noch auf einen Beitrag von Baumgartner (1997b) eingegangen werden, der einige Besonderheiten der Evaluation des netzgestützten Lernens aufzeigt. Er geht von folgender These aus:

„Durch die besonderen Merkmale interaktiver Medien und der internationalen Netzwerke (Internet) ergeben sich für den Lernprozess neue Möglichkeiten, aber auch neue Herausforderungen. Das muss sich notwendigerweise auch in den Inhalten und Formen von Evaluationen widerspiegeln. Diese spezifischen Anforderungen wurden jedoch bisher beim Design von Evaluationen noch kaum berücksichtigt“ (Baumgartner 1997b, S. 10).

Baumgartner geht insbesondere auf fünf Besonderheiten interaktiver Lernarrangements ein, die Einfluss auf die Evaluation nehmen und im Folgenden zusammengefasst werden (ebenda).

- 1) **Zugang:** Durch die Möglichkeit des individuellen Zugangs zur Software gibt es keine einheitlichen Zeiten, keine gemeinsamen und öffentlich zugänglichen Örtlichkeiten für Lernprozesse mehr. Das Lernen findet potenziell in der Form von privatem Studium – oft auch in den eigenen vier Wänden –, teilweise unabhängig von Lehrkräften oder anderen Lernern, statt. E-Learning ermöglicht einen individuellen und freien Zugang zu den Lernmaterialien (anytime und anywhere). „Viele Methoden der traditionellen Evaluation [...], die für gruppenbasierten Unterricht adäquat sein können, sind für die Evaluation interaktiver Lernsoftware entweder nicht stimmig oder aus Gründen des Kosten- und/oder Zeitaufwands nicht brauchbar“ (ebenda, S. 6).
- 2) **Ausgangs- und Bedürfnisstruktur:** Beim E-Learning gibt es in der Gruppe der Lerner eine potenziell heterogene Ausgangssituation. Dies gilt hinsichtlich demographischer Komponenten (berufliche Stellung, Bildungsstand etc.), inhaltlicher Komponenten (Vorwissen, Kenntnisse etc.) und der Lernerfahrungen, die Lernende haben. Auf einheitliche Voraussetzungen für eine Lerngruppe kann nicht ohne weiteres zurückgegriffen werden. Während der traditionelle Gruppenunterricht auf relativ homogene Voraussetzungen angewiesen ist, ist

E-Learning, als Form des selbstgesteuerten Lernens, nicht mehr dieser Restriktion unterworfen. Zudem gilt: Was wann, wie lange, wie oft gelernt wird, liegt in der Kontrolle des Lernalers. Eine ähnlich heterogene Ausgangslage liegt zumeist auch hinsichtlich der *Bedürfnisstruktur* vor. Potenziell ist medial gestütztes Lernen auch in den Zielvorstellungen und den damit zusammenhängenden Motivationsstrukturen völlig offen. Diese unterschiedlichen Bedürfnisse beeinflussen aber nicht nur den Lernerfolg, sondern sind auch für seine Definition entscheidend. Die unterschiedlichen Zielstrukturen sind dabei vielfältig: eine private Fortbildung aus Interesse am Thema ist ebenso denkbar wie ein weiterer Berufsabschluss oder eine im Berufsalltag notwendig gewordene Weiterqualifizierung. „Für Evaluationen heißt dies, dass sie komplexer werden und sowohl die unterschiedlichen Ausgangssituationen als auch die individuellen Lernstrategien und -ziele reflektieren müssen“ (ebenda, S. 7).

- 3) **Wahlfreiheit des Angebotes:** Beim E-Learning findet durch die eingesetzten Medien eine scheinbare Trennung von Inhalten und Lernprozess statt. Diese Trennung, so argumentiert Baumgartner, sei jedoch nur scheinbar. Tatsächlich gehe es gerade nicht darum, eine inhaltliche, statische Qualität eines Materials zu evaluieren, sondern vielmehr darum, zu analysieren, wie weit es in der Lage ist, Lernprozesse durch Interaktionen anzustoßen und zu unterstützen. Dementsprechend seien auch die Check- und Prüflisten für Softwareevaluationen nur mit äußerster Vorsicht und eingeschränktem Wirkungsgrad anzuwenden (ebenda, S. 8).
- 4) **Lernsituation:** Baumgartner regt an, die besondere Situation des netzgestützten Lernens im Evaluationsdesign zu berücksichtigen. Insbesondere sollte eine Differenzierung der Möglichkeiten und Effekte der Infrastruktur (delivery technology) und der didaktischen Potenziale (instruction technology) berücksichtigt werden.
- 5) **Kosten:** Baumgartner verweist darauf, dass bei den Kosten- und Nutzenanalysen zwei Aspekte ins Kalkül zu ziehen sind, die seiner Meinung nach immer noch zu wenig Beachtung finden. Zum einen argumentiert er mit Levin dahingehend, nicht nur ökonomisch-finanzielle Kosten zu berücksichtigen, sondern auch soziale Kosten:

„In economic terms, a cost is a sacrifice of an option. If a resource is applied to one use, it is not available for another use. In a purely economic sense, the cost of any decision is the value of what is sacrificed with respect to the best alternative use of that resource“ (Levin 1981, S. 30).

Zum anderen sei zwischen Kosten-Nutzen-Analysen und Kosten-Wirksamkeits-Analysen zu unterscheiden. Während im ersten Fall inhaltlich unterschiedliche Teilgebiete (beispielsweise Bildungs- vs. Transportwesen) miteinander verglichen werden könnten, sei dies bei der Kosten-Wirksamkeits-Analyse nicht möglich: „So ist z.B. schon ein Vergleich zwischen Lese- und Rechenfertigkeiten nicht mehr zulässig, weil sie verschiedene Testkriterien erforderlich

machen, deren Punktwerte jedoch nicht mehr dasselbe aussagen“ (ebenda, S. 10). Baumgartner kommt zu dem Schluss, dass die großen individuellen Steuerungsmöglichkeiten des E-Learning es erfordern, vor allem qualitative Analysen bei der Evaluation als grundlegende Methode zu verwenden.

Ein weiterer, bislang wenig beachteter Aspekt schließt Baumgartners Ausführen ab. Hier rückt ins Blickfeld, dass nicht nur die bereits genannte Individualisierung des netzgestützten Lernens besondere Anforderungen bei der Evaluation bewirkt, sondern auch und gerade die Möglichkeit zur sozialen Interaktion eine besondere und neue Art und Weise erfordere, über Evaluation von gruppengestützten Lernprozessen nachzudenken. Lernen ist demnach ein interaktiver sozialer Prozess, der zwischen den beteiligten Subjekten ausgehandelt wird und ist in diesem Sinne situiert (zu situiertem Lernen [*situated learning*] vgl. Brown et al. 1988, Chaiklin/Lave 1993; Collins et al. 1989; Lave 1988; Lave/ Wenger 1990; Suchman 1988). Ein Lernerfolg ist demnach nicht in der kognitiven Leistung einer Einzelperson zu finden, sondern im gesamten Lernsystem.

„In die Sprache der Evaluationstheorie übersetzt heißt dies: Der Evaluand (d.h. die zu evaluierende Sache) ist die Gruppenperformance, die selbst wiederum situationsabhängig kontextgebunden, sozial verteilt ist und konstruiert (d.h. jeweils interpretativ ausgehandelt) wird“ (ebenda, S. 12).

Baumgartner lenkt mit seinem Beitrag das Augenmerk in der Debatte um die Evaluation netzgestützten Lernens sehr deutlich auf neue Herausforderungen, für die es erst noch Lösungen und Erfahrungen zu finden gilt.

4. Fazit: Qualitätsevaluation auf allen Ebenen

Bislang beschränkt sich die Erfolgskontrolle bei Weiterbildungsmaßnahmen oft auf die Befragung von Teilnehmern nach ihrer Zufriedenheit, bestenfalls vielleicht auf die Überprüfung der unmittelbaren Lernergebnisse (vgl. Bliesener 1997). Wie bereits dargelegt, gibt es jedoch auch für den Bereich des E-Learnings umfassende Evaluationsmodelle. Insgesamt gilt hier: Qualitätssicherung ist mehr als die Kontrolle unmittelbarer Lerneffekte. Es geht darum, *alle* Einflussfaktoren zu berücksichtigen: den Lerner, das Lernthema, die gewünschten Ergebnisse und die Lernumgebung, sowohl die technologische als auch die soziale Lernumgebung (Arbeitsplatz, betriebliche Lernkultur, private Lernsituation etc.). Eine effektive Qualitätssicherung bedeutet Evaluation in *jeder* Phase des Qualifizierungsprozesses: von der Planung, über die Entwicklung, Durchführung bis hin zur Transfersicherung.

Die Lernerposition und die den Lerner beeinflussenden Kontextfaktoren müssen in Maßnahmen zur Qualitätssicherung verstärkt berücksichtigt werden. Ziel des Lernens ist es, beim Lernenden einen Zuwachs an Handlungskompetenz zu erreichen. Das bedeutet, dass der primäre Qualitätsmaßstab der Zuwachs an Handlungskompetenz beim Lerner ist und nicht – überspitzt formu-

liert – die Güte der Lernsoftware, etwa prognostiv ermittelt anhand von Kriterienkatalogen.

Qualitätssicherung erschöpft sich daher nicht in guter Planung oder Vorbereitung, sondern muss sich auf alle Phasen des Qualifikationsprozesses erstrecken und den Lerner mit einbeziehen. Qualitätssicherung beginnt beim Qualitätsbewusstsein und endet bei der Sicherung der Transferleistungen, um die es ja letztlich geht. Entscheidend ist demnach auch ein maßgeschneidertes Evaluationskonzept, um die Qualität zu sichern. Das Evaluationskonzept muss dabei situationsgerecht angepasst werden.

Ein weit verbreiteter Irrtum ist es, zu glauben, dass E-Learning lediglich gut geplant und vorbereitet sein müsse, um den gewünschten Erfolg zu erzielen. Ein detailliertes Qualitätssicherungskonzept umfasst alle Phasen eines Qualifizierungsprozesses und geht über Standardevaluationsverfahren für Weiterbildungsmaßnahmen hinaus, indem es zusätzlich zu einer detaillierten Evaluation der Einsatzphase auch noch die Planungs- und Entwicklungsphase und auch die Auswirkungen der Maßnahme (z.B. ROI in der betrieblichen Bildung) sowie Administrations- und Managementprozesse mit einbezieht.

Literatur

- Akademie des Deutschen Beamtenbundes* (Hrsg.): Beurteilungskriterien für Lernprogramme. In: Akademie des Deutschen Beamtenbundes für die berufliche Fortbildung e.V. Info-Paket, Report Nr. 4, 1986.
- Baumgartner, P.*: Evaluation vernetzten Lernens: 4 Thesen. In: *Simon, H.* (Hrsg.): Virtueller Campus. Forschung und Entwicklung für neues Lehren und Lernen. Münster 1997b, S. 131–146.
- Berendt, B./Stary, J.* (Hrsg.): Evaluation zur Verbesserung der Qualität der Lehre und weitere Maßnahmen. Weinheim 1993.
- Blease, D.*: Evaluating Educational Software. London 1986.
- Bliesener, T.*: Evaluation betrieblicher Weiterbildung. In: *Schwuchow, K./Gutmann, J.* (Hrsg.): Weiterbildung Jahrbuch 1997. Düsseldorf 1997, S. 163–167.
- Brinker, T.*: Dialogvideo im Führungskräfte-Training. Eine Studie zur Effektivität und Akzeptanz. Frankfurt am Main 1991.
- Brown, J. S./Collins, A./Duguid, P.*: Cognitive Apprenticeship, Situated Cognition and Social Interaction, Technical report no. 6886. Cambridge 1988.
- Chaiklin, S./Lave, J.*: Understanding Practice: Perspectives on Activity and Context. New York 1993.
- Collins, A./Brown, J. S./Newman, S. E.*: Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In: *Resnick, L. B.* (Hrsg.): Knowing, learning, and instruction: Essays in honor of Robert Glaser. Hillsdale 1989, S. 453–494.
- Cronbach, L. J./Snow, R. E.*: Aptitudes and Instructional Methodes. New York 1977.
- Dejoy, J. K./Mills, H. H.*: Criteria for Evaluation Interactive Instruction materials for Adult Self-Directed Learners. In: Educational technology, 1989, S. 2, 29, 39–41.
- Dijkstra, S./Seel, N. M./Schott, F./Tennyson, R. D.* (Hrsg.): Instructional Design: International Perspectives. Volume II: Solving of Instructional Design Problems. Mahwah 1997.
- Ehlers, U.-D.*: Qualität beim E-Learning. Empirische Grundlegung und Modellkonzeption einer subjektorientierten Qualitätsentwicklung. Dissertation. Universität Bielefeld 2003.
- Ehlers, U.-D./Gerteis, W./Holmer, T./Jung, H.* (Hrsg.): E-Learning-Services im Spannungsfeld von Pädagogik, Ökonomie und Technologie. L3-Lebenslanges Lernen im Bildungsnetzwerk der Zukunft. Bielefeld 2003.
- Ehlers, U.-D./Pawlowski, J. M./Görtz, L.*: Qualität von E-Learning kontrollieren, Die Bedeutung von Qualität im E-Learning. In: *Hohenstein, A./Wilbers, K.*: Handbuch E-Learning, Expertenwissen aus Wissenschaft und Praxis: Strategien, Instrumente, Fallstudien. 2003.

- Flanders, N. A.: Analyzing Teaching Behaviour. Reading, 1970.
- Fricke, R.: Evaluation von Multimedia. In: Issing L.J./Klimsa, P. (Hrsg.): Information und lernen mit Multimedia. Weinheim 1995.
- Fricke, R.: Qualitätsbeurteilung durch Kriterienkataloge. Auf der Suche nach validen Vorhersagemodeln. In: Schenkel, P. (Hrsg.): Qualitätsbeurteilung multimedialer Lern- und Informationssysteme. Evaluationsmethoden auf dem Prüfstand. Nürnberg 2000.
- Fricke, R.: Evaluation von Multimedia. In: Issing, L.J./Klimsa, P. (Hrsg.): Information und lernen mit Multimedia. Weinheim 2002.
- Gerl, H./Pehl, K.: Evaluation in der Erwachsenenbildung. Bad Heilbrunn 1983.
- Götz, K.: Zur Evaluierung beruflicher Weiterbildung. Band 1: Theoretische Grundlagen. Weinheim 1993.
- Gräber, W.: Das Instrument MEDA – Ein Verfahren zur Beschreibung, Analyse und Bewertung didaktischer Software im berufsbildenden Bereich. In: Dick, A.: AUDIO VISUELL – neue Technologien: gelungene oder misslungene Medialisierung von lernen, Staatliche Landesbildstelle Hessen. Frankfurt am Main 1991.
- Gräber, W.: Kriterien und Verfahren zur Sicherung der Qualität von Lernsoftware in der beruflichen Weiterbildung. Kiel 1996.
- Gräber, W./Lauterbach, R.: EPASoft – Instrument zur Bewertung pädagogischer Software. Kiel 1992.
- Gröbhel, U.: Kosten und Nutzen des E-Learning an der Fachhochschule. [URL: <http://www.dwi.fhbb.ch/e-learning/>], (extracted 02/2003).
- Holzkamp, K.: Lernen. Subjektwissenschaftliche Grundlegung. Frankfurt, New York 1993.
- Jung H. M.: Multimedia in der Softwareschulung, Frankfurt am Main, 1994.
- Kaselaar, G.: Computer und Lernen: der Lehrer als Subjekt oder Objekt. In: Bogner, C./Burger, H./Weiß, K. (Hrsg.): Computer und Kulturtechniken. München 1992, S. 132-165.
- Kirkpatrick, D. L.: Evaluation Training Programs. The Four Levels. San Francisco 1994.
- Kohrt, L.: Probleme und Perspektiven der Evaluation computerunterstützter Instruktion. Arbeiten aus dem Seminar für Pädagogik der TU Braunschweig, Bericht Nr. 1/95. Braunschweig 1995.
- Kulik, C.-L./Kulik, J.: Effectiveness of Computerbased Instruction: An Update Analysis. In: Computers in Human Behaviour, 7, 1991, S. 75–94.
- Kulik, J.: Meta-analytic studies of findings on computer-based instruction. In: Baker, E. L./O'Neil, H. F. (Eds.): Technology Assessment in Education and Training. Hillsdale 1994.
- Lange, E.: Zur Entwicklung und Methodik der Evaluationsforschung in der Bundesrepublik Deutschland. Zeitschrift für Soziologie, 12, 1983, S. 253-270.
- Lave, J.: Cognition in Practice: Mind, mathematics, and culture in everyday life. Cambridge 1988.
- Lave, J./Wenger, E.: Situated Learning: Legitimate Peripheral Participation. Cambridge 1990.
- Mandl, H./Heinz, G./Renkl, A.: Lernen mit dem Computer. Empirisch Pädagogische Forschung in der BRD zwischen 1970 und 1990. Forschungsbericht Nr. 7. München 1992.
- Mandl, H./Reinmann-Rothmeier, G.: Unterrichten und Lernumgebungen gestalten. Forschungsbericht Nr. 60. München 1995.
- Martin, B. L.: A Checklist for Designing Instruction in the Affective Domain. Educational Technology, (29) 8, 1989, S. 7-15.
- Meier, A.: Qualitätsbeurteilung von Lernsoftware durch Kriterienkataloge. In: Schenkel, P./Holz, H. (Hrsg.) (1995): Evaluation multimedialer Lernprogramme und Lernkonzepte, BIBB-Reihe Multimediales Lernen in der Berufsbildung. Nürnberg 1995, S. 149-190.
- Meier, F./Baratelli, S.: Wissenspsychologische Evaluation selbstgesteuerten Lernens mit modernen Medien und rechnergestützten Instruktionen, Medienpsychologie, 3, 1991, S. 109-123.
- Prell, S.: Evaluation. In: Schiefele, H./Krapp, A. (Hrsg.): Handlexikon zur Pädagogischen Psychologie. München 1981, S. 116–120.
- Prell, S.: Evaluation. In: Sarges, W./Fricke, R. (Hrsg.): Psychologie für die Erwachsenenbildung- Weiterbildung. Göttingen 1986, S. 208–213.
- Reay, D. G.: Evaluating Educational Software in the Classroom. In: Reid, I./Rushton, J. (Hrsg.): Teachers, Computers and the Classroom. Manchester 1985, S. 184–195.
- Reeves, T. C.: Evaluating Interactive Multimedia. Athens 1992.
- Reeves, T. C./Harmon, S. W.: Systematic Evaluation Procedures for Interactive Multimedia for Education and Training. Paper presented at the Annual Meeting of the American Educational Research Association. Atlanta 1994.

- Reigeluth, Ch. M.*: Instructional design: What is it and why is it? In: *Reigeluth, Ch. M.* (Hrsg.): Instructional Theories and Models: An Overview of Their Current Status. Hillsdale 1983, S. 3–36.
- Reinmann-Rothmeier, G./Mandl, H.*: Lernen auf der Basis des Konstruktivismus: Wie Lernen aktiv und anwendungsorientierter wird. In: *Computer und Unterricht* 23/1996, S. 41–44.
- Rosenshine, B./Fürst, N.*: Research in Teacher Performance Criteria. In: *Smith, B. O.* (Hrsg.): Research in Teacher Education. A Symposium. Englewood Cliffs 1971.
- Rossi, P. H./Freemann, H. E./Eloffmann, G.*: Programm-Evaluation. Einführung in die Methoden angewandter Sozialforschung. Stuttgart 1988.
- Schenkel, P.*: Zum Thema „Qualität von Lernsoftware“. In: *Arbeitshilfen für die Erwachsenenbildung*, Nr. 4, 1998. [URL: <http://www.shuttle.de/pae/ah4-98/schenkel.htm>] (extracted 02/ 2003).
- Schenkel, P.* (Hrsg.): Qualitätsbeurteilung multimedialer Lern- und Informationssysteme. Evaluationsmethoden auf dem Prüfstand. Nürnberg 2000.
- Schenkel, P./Holz, El.* (Hrsg.): Evaluation multimedialer Lernprogramme und Lernkonzepte. Berichte aus der Berufsbildungspraxis. Nürnberg 1995.
- Schott, E./Krien, E./Sachse, S./Schubert, T.*: Evaluation von multimedialer Lernsoftware auf der Basis von ELISE (1.0). Ein Ansatz zu einer theorie-, adressaten- und anwendungsorientierten Methode zur Evaluation von multimedialen Lern- und Informationssystemen. In: *Schenkel, P.* (Hrsg.): Qualitätsbeurteilung multimedialer Lern- und Informationssysteme. Evaluationsmethoden auf dem Prüfstand. Nürnberg 2000.
- Schwarz, I./Lewis, M.*: Basic Concept Microcomputer Courseware: A Critical Evaluation System for Educators, Educational technology. 1989, S. 5, 29, 16–21.
- Seidel, R. J./Park, O.*: An historical perspective and a model for evaluation of intelligent tutoring systems. In: *Journal of Educational Computing Research*, 10, 1994, S. 103–128.
- Squires, D./McDougall, A.*: Choosing and Using Educational Software: A teachers' guide. London 1994.
- Stiefel, R. Th.*: Grundfragen der Evaluierung in der Management-Schulung. Lernen und Leistung. Frankfurt am Main 1974.
- Stufflebeam, D. L.*: Evaluation as enlightenment for decision-making. In: *Worthen, B. R./Sanders, J. R.* (Hrsg.): Educational evaluation: Theory and practice. Worthington 1973.
- Stufflebeam, D. L.*: Professional Standards for Educational Evaluation: United States. In: *Elusen, T./Posthethwaite, T. N.* (Hrsg.): The International Encyclopedia of Education. Oxford 1994.
- Stufflebeam, D. L.*: The relevance of the CIPP evaluation model for educational accountability. *SRIS Quarterly*, 5(1), 1972.
- Stufflebeam, D. A.*: The CIPP model for evaluation. In: *Stufflebeam, D. L./Madaus, G. F./Kellaghan, T.* (Hrsg.): Evaluation Models. Boston 2000.
- Suchman, Lr.*: Plans and Situated Actions: The Problem of Human/Machine Communication. Cambridge 1988.
- Tennyson, R. D./Schott, F./Seel, N. M./Dijkstra, S.* (Hrsg.): Instructional Design: International Perspectives. Volume I: Theories and Models of Instructional Design. Mahwah 1997.
- Thomè, D.*: Kriterien zur Bewertung von Lernsoftware. Berlin 1988.
- Weidenmann, B.*: „Multimedia“: Mehrere Medien, mehrere Codes, mehrere Sinneskanäle? In: *Unterrichtswissenschaft* 2/97, 1997, S. 197–206.
- Will, H./Winteler, A./Krapp, A.*: Von der Erfolgskontrolle zur Evaluation. In: *Will, H./Winteler, A./Krapp, A.* (Hrsg.): Evaluation in der beruflichen Aus- und Weiterbildung. Konzepte und Strategien. Heidelberg 1987, S. 11–42.
- Winship*, Software Review or Evaluation: Are They Both Roses Or Is One a Lemon? Paper presented at the Proceedings of the Australian Computer Education Conference. Perth 1988.
- Wittmann, W. W.*: Evaluationsforschung – Aufgaben, Probleme, Anwendungen. Berlin 1987. *Wottawa, El./Thierau, El.*: Lehrbuch Evaluation. Stuttgart 1990.
- Wulf, Ch.* (Hrsg.): Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen. München 1972.
- Zimmer, G./Psaralidis, E.*: Der Lernerfolg bestimmt die Qualität einer Lernsoftware! Evaluation von Lernerfolg als logische Rekonstruktion. In: *Schenkel, P./Tergan, S.-O./Lottmann, A.* (Hrsg.): Qualitätsbeurteilung multimedialer Lern- und Informationssysteme. Evaluationsmethoden auf dem Prüfstand. Nürnberg 2000.