

Peter Baumgartner

***Evaluation mediengestützten Lernens
Theorie – Logik – Modelle***

Baumgartner, P. (1999). Evaluation mediengestützten Lernens. Theorie - Logik - Modelle. In: Projektevaluation in der Lehre - Multimedia an Hochschulen zeigt Profil(e). M. Kindt. Münster, Waxmann. 7: 61-97.

1. Einleitung

Wenn ich vom allgemein gehaltenen Titel dieses Artikels ausgehe, so möchte ich die dahinter stehende Fragestellung in zweierlei Hinsicht konkretisieren:

- Was genau ist unter einer Evaluation des Lernens zu verstehen? und
- Worin bestehen die Spezifika einer Evaluation *mediengestützten* Lernens?

Wenn ich mich hier nun relativ ausführlich mit Definitions- und Abgrenzungsfragen beschäftige, so geht es mir dabei nicht in erster Linie darum, eine operationale, möglichst umfassende theoretische Definition zu geben. Obwohl sich dies auch – sozusagen als ein praktisches Nebenprodukt meiner Überlegungen – ergeben wird. Vielmehr möchte ich zeigen, daß ein bestimmtes Verständnis von Evaluation und von Lernen ganz entscheidend dafür ist, *wie* eine Evaluation des Lernens betrieben wird. Jede bestimmte theoretische Sichtweise hat zahlreiche und – wie ich zeigen werde – deutlich unterschiedliche Auswirkungen auf den Forschungsprozeß (relevante Fragestellung zur Evaluationsforschung) aber auch auf das Forschungshandeln (angewandte bzw. verfolgte Methodik) selbst. So hat eine bestimmte Charakterisierung von Evaluation Einfluß auf

- die relevanten Themen und theoretischen Fragestellungen (in meiner Sichtweise z.B. vermehrte Bedeutung philosophischer Fragestellungen wie Ethik und Werttheorie gegenüber der oft vorherrschenden sozialwissenschaftlichen Methodenlehre);
- die notwendigen Qualifikationen und damit auf die Gewichtung der Themen in der Ausbildung von EvaluatorInnen (in meiner Sichtweise z.B. verstärkte Beachtung von Interventionsstrategien gegenüber abstrakt-theoretischer wissenschaftlicher Analyseverfahren);
- die Art und Weise wie Daten produziert werden, wer bei der Datenproduktion einbezogen wird, wie und an wen über die Ergebnisse berichtet wird (in meiner Sichtweise nicht bloß ein Bericht an das Management, sondern Einbeziehung der „Betroffenen“ und der „Konsumenten“);
- jene Kriterien, die selbst wiederum auf die Beurteilung der durchgeführten Evaluation angewendet werden (sogenannte Meta-Evaluation).

Meiner Ansicht nach müssen vorerst grundsätzliche erkenntnistheoretische Fragestellungen geklärt werden, erst dann können verschiedene Evaluationsansätze in ihrer Sinnhaftigkeit beurteilt und die dabei auftretenden methodischen Probleme zu lösen versucht werden. Darauf aufbauend werde ich auf die spezifische Fragestellung dieses Artikels – Evaluation mediengestützten, Lernens – näher eingehen.

2. Zur Definition des Evaluationsbegriffes

Bei der Sichtung relevanter Literatur ist mir aufgefallen, daß praktisch jeder Autor/jede Autorin sich der Bedeutung dieser Definitionsversuche bewußt ist. Während bei vielen anderen Themenstellungen die grundsätzliche Position oft nur implizit angesprochen wird, finden wir in der Evaluationsliteratur praktisch immer einen Vorspann mit expliziten Definitionsversuchen. Die Ursache dürfte nicht nur in der relativen Neuheit dieses Forschungs- und Praxisfeldes liegen¹, sondern auch in der nahezu explosionsartigen Ausarbeitung unterschiedlicher Modelle, die zu einem eigenen Forschungsgebiet „Taxonomie von Evaluationsansätzen“ mit dazugehörigen Abgrenzungsversuchen geführt haben.

2.1. *Evaluation läßt sich nicht definieren: die relativistische Position*

Evaluation – more than any science – is what people say it is; and people currently are saying it is many different things.“ (Glass & Ellet 1980:211, zitiert nach Shadish, Cook, und Leviton 1991:30)

Dieses Zitat stellt quasi die Reinform der relativistischen Position dar: Evaluation ist alles und nichts. Der Hintergrund für diese Ansicht – die durch das Fehlen jeglichen Definitionsversuches charakterisiert ist – liegt darin, daß sowohl die Ansichten zur Evaluation als auch die zu beobachtende Praxis der Evaluation mannigfaltig ist. Bei Befragungen sowohl von evaluierten Personen und EvaluationsexpertInnen hat sich ein äußerst vielfältiges und uneinheitliches Bild ergeben. Unter Evaluation des Lernerfolges werden ganz unterschiedliche Prozesse bzw. Produkte verstanden (vgl. Will, Winteler, und Krapp 1987), wie z.B.:

- die quantitative Bestimmung von Lernergebnissen am Ende eines Kurses
- die sporadische oder systematische Rückmeldung zur Verbesserung des Unterrichts an die Dozenten/Trainer
- die wissenschaftliche Begleituntersuchung eines Reform- oder Pilotprojekts, Modellversuchs etc.
- die ökonomisch-quantitative Kostenanalyse eines Ausbildungsmodells.

Ich persönlich halte die relativistische Position, die sich auf ein konstruktivistisches Verständnis von Wissenschaft beruft, nicht nur für äußerst unbefriedigend, sondern auch für falsch: Erstens werden damit wissenschaftliche Forschungsfragen wie z.B. „Wohin soll sich die Evaluationspraxis entwickeln? Welche Evaluationsmodelle sind wissenschaftstheoretisch zu vertreten?“ ausgeblendet und unterbunden. Andererseits wird die wesentliche Aufgabe des konstruktivistischen Diskurses (um im Paradigma des konstruktivistischen Wissenschaftsverständnisses zu bleiben), der zur Ausarbeitung einer einheitlichen Sichtweise und Praxis führen soll, negiert.

In einer abgeschwächten Form findet sich die relativistische Position im Aufzeigen der unterschiedlichen theoretischen Ansätze und praktischen Verfahrensweisen und in der kritiklosen Auflistung von unterschiedlichen Merkmalen. Evaluation sei eben sehr vielschichtig, komplex und nicht so recht zu definieren bzw. von anderen Gebieten abzugrenzen.

¹Obwohl manche AutorInnen glauben, daß die Evaluationsproblematik bereits bei Plato oder gar in China 2000 vor unserer Zeitrechnung zu orten ist, wird der Beginn im allgemeinen mit den Arbeiten von Abraham Flexner zum amerikanischen öffentlichen Schulsystem Mitte der 20er-Jahre datiert. Theoretisch relevante Bemühungen beginnen überhaupt erst mit Ralph Tyler zu Beginn der 30er-Jahre (vgl. Worthen und Sanders 1987:11-20; Madaus, Scriven, und Stufflebeam 1983:3-18; Stufflebeam und Shinkfield 1985:1-44 und Götz 1993:27-36).

2.2. *Evaluation als quantitatives Analyseverfahren*

Manchmal wird in der Literatur Evaluation implizit mit der Konstruktion und Auswertung von Tests (oder anderer quantitativer Meßverfahren) gleichgesetzt. Ein typisches Beispiel dafür stellt das Lehrbuch „Measurement and Evaluation in Psychology and Education“ (Thorndike et al. 1991) dar: Während im Titel sowohl Meßverfahren als auch Evaluation gleichrangig genannt werden, handelt der Standardtext mit einer einzigen Ausnahme von statistischen Problemen quantitativer Meßverfahren wie Auswertung, Normierung, Reliabilität und Validität. Bei dem erwähnten Sonderfall handelt es sich bloß um ein kleines Kapitel mit 15 – von insgesamt 500 – Seiten über Leistungs- und Produktevaluation. Doch selbst hier werden überwiegend methodische Probleme wie die Konstruktion von Skalen, Durchführung systematischer Beobachtung etc. behandelt.

In meiner Kritik dieser Sichtweise möchte ich nicht nur das Fehlen qualitativer Verfahren anführen, das alleine würde zu kurz greifen. Bei der Evaluation geht es nicht bloß um eine systematische (genaue, umfassende, relevante usw.) Erfassung von Daten, sondern ganz wesentlich um ihre Bewertung (vgl. den englischen Wortstamm „value“ = Wert).

...evaluation is a much more comprehensive and inclusive term than measurement. Evaluation includes both qualitative and quantitative descriptions of pupil behavior plus value judgments concerning the desirability of that behavior. Measurement is limited to quantitative description of pupil behavior. It does not include qualitative descriptions nor does it imply judgments concerning the worth or value of the behavior measured. The following diagrams clearly show the relationship between measurement and evaluation:

Evaluation = Quantitative descriptions of pupils (measurement) + Value judgments
Evaluation = Qualitative descriptions of pupils (nonmeasurement) + Value judgments
(Gronlund 1976:6)

Unter Bewertung ist jedoch hier nicht bloß eine statistische Interpretation gemeint, sondern vor allem die Entscheidung darüber, ob und wie weit der *Evaluand* (die evaluierte Sache, der evaluierte Prozeß etc.) den Wünschen bzw. Vorstellungen entspricht oder aber modifiziert bzw. gar eingestellt werden soll. Die Begriffe „Entscheidung“ und „Wunsch“ zeigen mit aller Deutlichkeit Gebiete auf, die weder durch die Statistik noch – wie ich weiter unten ausführlich zeigen werde – durch eine umfassende Methodenlehre (Methodologie) abgedeckt sind.

2.3. *Evaluation als eine Anwendung von Methoden (Methodenlehre)*

Sieht man sich Evaluationsliteratur durch, so erkennt man, daß die überwiegende Mehrzahl der Bücher und Artikel sich mit der Beschreibung, Diskussion und Umsetzung von (sozialwissenschaftlichen) Methoden beschäftigt². Obwohl nach dieser Ansicht nun auch qualitative Verfahren einbezogen werden und die Zusammenhänge dieser Methoden mit Fragestellung, Umsetzungs- und Auswertungsproblemen diskutiert werden, greift auch dieser Definitionsversuch meiner Meinung nach zu kurz.

Evaluation bloß als ein Bündel von Methoden zu betrachten, würde es mit untergeordneten Spezialgebieten wie Feldforschung, experimentelles Design oder Psychometrie gleichsetzen. Demgegenüber möchte ich festhalten, daß Evaluation nicht nur ein selbständiges wissenschaftliches (multidisziplinäres) Feld darstellt, sondern auch einen eigenen speziellen Inhalt (nämlich: *bewerten*) und eine eigene innere Logik besitzt.

²Obiger Kritikpunkt an der impliziten Gleichsetzung von quantitativen Analyseverfahren und Evaluation kann dabei als ein einschränkender Sonderfall der Sichtweise „Evaluation = Methodenlehre“ aufgefaßt werden.

Die Gleichsetzung von Evaluation und Methodenlehre läuft letztlich auf eine bloß „objektive“ bzw. „intersubjektive“ Beschreibung eines Sachverhaltes (des Evaluanden) hinaus. Nach dem Motto „give them just the facts“ wird die eigentliche Aufgabe der Evaluation, die *Bewertung* der Fakten bzw. des Datenmaterials, nicht durchgeführt.

As a specialty, evaluation is most like methodological specialties – ethnography, psychometrics, experimental design, or survey research...Evaluation may be the broadest methodological specialty. Its theory includes a vast array of decisions about the shape, conduct, and effects of an evaluation. To inform evaluators about choosing methods, it needs to discuss philosophy of science, public theory, value theory, and theory of use.

Without its unique theories, program evaluation³ would be just a set of loosely conglomerated researchers with principal allegiances to diverse disciplines, seeking to apply social science methods to studying social programs. Program evaluation is more than this, more than applied methodology. Program evaluators are slowly developing a unique body of knowledge that differentiates evaluation from other specialties while corroborating its standing among them. (Shadish, Cook, und Leviton 1991:31)

Faßt man Evaluation daher – wie im Zitat angedeutet – umfassender, so ergeben sich unter Einschluß der allgemein akzeptierten Methodenfrage vor allem fünf große inhaltliche Aufgabengebiete der Evaluationsforschung:

- Welche inhaltlichen Probleme deckt der Evaluand ab? Kann das durch den Evaluanden abgedeckte Bedürfnis anders besser befriedigt werden?
- Wie werden gültige Fakten zur Analyse und Bewertung gewonnen? (Erkenntnistheorie, Wissenschaftstheorie)
- Wie ist der Evaluand zu bewerten? (Werttheorie)
- Wie können die Ergebnisse der Evaluation umgesetzt werden? (Theorie über gesellschaftlichen Wandel)
- Welches pragmatische Design soll für die Evaluation gewählt werden? (Angewandte Methodologie)

2.4. Evaluation als Verbesserung praktischer Maßnahmen (Treatments)

Sehr häufig wird der Sinn von Evaluationen in der Entwicklung von Verbesserungsvorschlägen gesehen. Dabei werden Maßnahmen zur Verbesserung des Evaluanden nicht bloß als ein (mögliches) Kriterium für gelungene Evaluationen definiert, sondern quasi als das wesentliche Ziel jeglicher Evaluationsbemühungen gesehen. Diese pragmatische Position argumentiert mit Erfordernissen aus der Praxis und setzt Evaluationen mit Verbesserungsvorschlägen gleich.

Gegen diese einschränkende Sichtweise ist zweierlei einzuwenden: Evaluationen können auch für bloße „go/stop“-Entscheidungen sinnvoll durchgeführt werden: Soll z.B. eine bestimmte Maßnahme fortgeführt oder abgebrochen werden? Andererseits ist festzuhalten, daß die Entwicklung von Verbesserungsvorschlägen eine ganz andere Logik impliziert als sie Evaluationen im allgemeinen innewohnen.

In erster Linie geht es bei Evaluationen nämlich um die Erstellung und Zuweisung eines Werturteils (Evaluand = gut/schlecht, wertvoll/wertlos). Bei der detaillierten Analyse von Mängeln handelt es sich jedoch oftmals um ein anderes fachliches Themengebiet, als dies bei der Evaluation selbst der Fall ist. Es werden daher ganz andere Inhalte und demgemäß auch andere Kenntnisse angesprochen. So kann z.B. eine vergleichende Produktevaluation von Computermonitoren zu klaren Ergebnissen kommen

³Gemeint sind hier Programme im Sinne von sozialen Maßnahmenpaketen oder Projekten (also Gesundheits- Bildungsprogramme...) und nicht etwa Software. Die im obigen Zitat implizierte Gleichsetzung von „Evaluation = Programm-Evaluation“ wird von mir nicht akzeptiert, weil es auch andere Evaluationsfelder (Produkt-, Performance-, Personalevaluation etc.) gibt. Ich gehe aber in diesem Artikel auf die Besonderheiten dieser unterschiedlichen Felder nicht näher ein.

und Mängel bestimmter Markenprodukte eindeutig feststellen (z.B. zu hohe Strahlungsintensität). Wie und ob diese Mängel jedoch bei diesem Produkt behoben werden können oder sollen, ist eine ganz andere Sache und verlangt vielleicht eine weitere Studie und/oder Laborexperimente mit ganz anderen inhaltlichen Kompetenzen, als sie von EvaluationsexpertInnen dieser Produktevaluation benötigt werden.

2.5. *Evaluation als angewandte Sozialforschung*

Vor allem im pädagogischen Bereich handelt es sich hier um eine der häufigsten Ansichten, um nicht zu sagen, um die Standardmeinung. Evaluation wird dabei bloß als ein (weiteres) Anwendungsgebiet der Sozialforschung gesehen. So heißt es im wohl bekanntesten und am weitest verbreiteten Standardlehrbuch zur Evaluation:

We begin this volume with a simple definition of evaluation, or evaluation research (we will use the terms interchangeably): *Evaluation research is the systematic application of social research procedures for assessing the conceptualization, design, implementation, and utility of social intervention programs.* In other words, evaluation researchers (evaluators) use social research methodologies to judge and improve the ways in which human services policies and programs are conducted, from the earliest stages of defining and designing programs through their development and implementation. (Rossi und Freeman 1989:18. Hervorhebungen im Original, vgl. auch die deutsche Übersetzung Rossi, Freeman, und Hofmann 1988:3)

Untersucht man diese Argumentation etwas genauer, so zeigen sich in der Definition von „Evaluation als angewandte Sozialforschung“ zwei Schritte der Reduktion:

- Zuerst wird Evaluation auf Programm-Evaluation (also auf Bildungs-, Gesundheits-, Sozialhilfe- oder andere soziale Interventionsmaßnahmen) eingeschränkt. Damit werden aber andere Evaluationsfelder mit ganz unterschiedlichen Zugängen, wie z.B. Produktevaluation, aus der Definition ausgeklammert.
- Mit der Betonung von *angewandter* Sozialforschung werden grundsätzliche philosophische oder ethische Fragestellungen (Erkenntnistheorie, Werttheorie) nicht besonders ausgewiesen, sondern nur mehr unter der Rubrik „Sozialwissenschaft“ behandelt. Nun wird aber durch das Weber'sche Postulat der Wertfreiheit (Weber 1988a; Weber 1988b) gerade eine Trennung von Beschreibung und Werturteil in den Sozialwissenschaften gefordert. Die Aufhebung dieser Trennung sehe ich jedoch gerade für eine Definition von Evaluation als wesentlich an!

Obwohl ich mich selbst in erster Linie als Sozialwissenschaftler sehe, halte ich die Vorstellung, Evaluationen bloß als ein (weiteres) Einsatzgebiet angewandter Sozialforschung zu sehen, sowohl für die Selbständigkeit als auch für die weitere Entwicklung dieser jungen Disziplin für verhängnisvoll. Statt Evaluation bloß als spezifischen Anwendungsfall von Sozialforschung zu begreifen, sollten sie vielmehr als eine eigene Hilfswissenschaft (tool discipline) ähnlich der Statistik, Logik, oder Designwissenschaft betrachtet werden (vgl. dazu ausführlich Scriven 1991c, insbesondere 13-26).

Nur unter dieser erweiterten Sichtweise kann sich die Evaluationsforschung von einem „bag of tricks and bricks“ (a.a.O., 24) zu einem eigenständigen intellektuellen Konstrukt in der komplexen Wissenschaftstaxonomie entwickeln. Das schließt mit ein, daß es einer inhaltlichen Besonderheit bedarf, die die Bezeichnung und Abgrenzung einer eigenständigen Disziplin rechtfertigt⁴. Erst damit läßt sich eine adäquate eigenständige Methodologie identifizieren, die in Zusammenhang mit einer wachsenden Datenbasis von Erkenntnissen, allgemeinen Prinzipien und Theorien auch die Entwicklung eine Metatheorie ermöglicht.

⁴Wie bereits mehrmals angedeutet, sehe ich diese Besonderheit in der Wertfrage, im Übergang von der Datensammlung über einen Evaluanden zur Bewertung desselben.

2.6. Evaluation als Bewertung

In der Diskussion und Abgrenzung zu den unterschiedlichen Definitionsversuchen dürfte meine Position inzwischen deutlich geworden sein: Ich stimme mit jenen Auffassungen überein, die im Evaluationsbegriff vor allem den Prozeß der Bewertung hervorheben:

Evaluation is the determination of a thing's value. (Worthen und Sanders 1987:22)
Evaluation is the process of determining the merit, worth and value of things, and evaluations are the products of that process. (Scriven 1991c:1)

Nach diesem Verständnis hat der Begriff der Evaluation sowohl Prozeß- als auch Produktcharakter:

- Als *Prozeß* ist unter Evaluation eine Form von angewandter Forschung (Evaluationforschung) zu betrachten, die sowohl handlungsanleitend (Entscheidungshilfe, „Problem“lösung) als auch auf Erkenntnisgewinn (Generieren von Weltwissen) ausgerichtet ist. Der starke Praxisbezug (pragmatische Charakter) schließt weder die Möglichkeit von Grundlagenforschung und Theoriebildung (Evaluationstheorie) noch die selbstbezügliche Metaebene aus (Metaevaluation als eine Evaluation von Evaluationen).
- Als *Produkt* ist unter Evaluation sowohl das Ergebnis der forschenden Aktivitäten in einem bestimmten Evaluationsfeld zu verstehen, aber auch die sich aus diesen Ergebnissen seit den 60er-Jahren formierende Disziplin selbst.

Als zusammenfassende These für diesen Abschnitt, des Versuchs einer Definition des Evaluationsbegriffes, möchte ich daher vorschlagen:

These: Unter Evaluation sind alle Aktivitäten und/oder Ergebnisse zu verstehen, die die Bedeutung, Verwendbarkeit, (Geld-)Wert, Wichtigkeit, Zweckmäßigkeit,... einer Sache beurteilen bzw. bewerten. Nur dieses weit gefaßte Verständnis von Evaluation kann sowohl die Charakteristika besonderer Evaluationsfelder berücksichtigen als auch einen adäquaten Beitrag zur Theoriebildung leisten.

3. Grundfragen der Evaluationstheorie

These: Obwohl sich jede Evaluation mit ganz besonderen inhaltlichen Problemen auseinandersetzen muß, Probleme, deren Lösung jeweils besondere Methodologien, Verfahren und Modelle der Evaluation bevorzugen, gibt es sowohl gemeinsame grundlegende Inhalte und eine einheitliche Logik der Evaluation. Deshalb ist es durchaus richtig, wenn man Evaluation als eigene Disziplin ansieht.

Die inhaltliche Einheit möchte ich nachfolgend durch die Auflistung von grundlegenden Fragestellungen aufzeigen. Die in fünf Gruppen zusammengestellten Fragen sind für alle Evaluanden relevant. In jeder Fragengruppe gibt es außerdem eine metatheoretische Fragestellung, die über die Evaluation als Disziplin hinaus weist und eher einem übergeordnetem Gebiet (z.B. Philosophie, Erkenntnistheorie, Sozialtheorie etc.) zuzuordnen ist.

3.1. Inhalt/Funktion/Wirkungsweise:

Allgemeine Fragen: Welche inhaltliche Probleme sollen durch den Evaluanden gelöst werden? Kann der Evaluand verbessert werden? Macht es Sinn (ist es wert) ihn zu verbessern? Wenn nicht: Welche Alternativen gibt es?

Metafragen: Wie entwickeln sich (diese Typen von) Evaluanden (z.B. Programme, Produkte etc.)? Wie können sie verbessert, geändert, eingestellt werden? Entsprechend dem jeweiligen Evaluationsfeld handelt es sich bei den angeführten Metafragen um soziologische (Programm-Evaluationen), maschinenbautechnische (z.B. Autos in einer Produktevaluation) curriculare (z.B. Evaluation von Lehrveranstaltungen) usw. Fragen.

Beispiele: Wie ist der Evaluand intern strukturiert? Welche Funktion erfüllt er? Wie arbeitet er, wirkt er? Welche externen Rahmenbedingungen beeinflussen ihn? Wie erfolgt ein Wandel, eine Veränderung? Welche Folgen bedeutet eine Veränderung?

3.2. Wissenskonstruktion

Allgemeine Fragen: Wie komme ich zu Fakten? Was ist ein gültiges Datum? Was zählt als sichere Antwort? Was verursacht diese Sicherheit/Gewißheit?

Metafragen: Wie komme ich zur Erkenntnis sozialer Zusammenhänge? Welche Zusammenhänge gilt es zu erkennen? (Fragen zur Erkenntnistheorie)

Beispiele: Ontologische Fragestellungen – Was ist Realität? Epistemologische Fragestellungen – Wie wird gültiges Wissen begründet? Was ist das Wesen von Erkenntnis? Wie wird Erkenntnis gewonnen? Was sind die Grenzen der Erkenntnis? Methodologische Fragestellungen – Wie wird Wissen konstruiert/gewonnen? Welche Techniken der Wissenskonstruktion gibt es? Wie funktionieren sie? Wie werden sie angewendet?

3.3. Bewertung

Allgemeine Fragen: Kann der Evaluand als gut (wertvoll) bezeichnet werden? Wodurch wird "gut" bzw. „wertvoll“ spezifiziert? Was rechtfertigt bzw. wodurch wird meine wertende Schlußfolgerung gerechtfertigt?

Metafragen: Wie komme ich von Fakten zu Werten? Welche Werte sind zu bevorzugen? (Werttheorie, Ethik)

Beispiele: Metatheorie – Was sind Werte? Wie werden sie begründet? Präskriptive (normative) Theorie – Welche Werte sind von besonderem Wert? (Wertehierarchie) Deskriptive Theorie – Beschreibung der Werte ohne Hierarchie.

3.4. Anwendung des Wissens

Allgemeine Fragen: Wie können die Ergebnisse der Evaluation sinnvoll umgesetzt werden? Was ist zu tun, damit die Ergebnisse nicht verfälscht oder unberücksichtigt bleiben?

Metafragen: Wie werden die Evaluationsergebnisse benützt? Wie sollen sie benützt werden?

Beispiele: Beschreibung der möglichen Verwendungsarten bei Berücksichtigung der verschiedenen Interessensgruppierungen einschließlich möglicher Hindernisse. Beschreibung verschiedener zeitlicher Rahmenbedingungen für die Verwendung. Beschreibung der Möglichkeiten die der Evaluator besitzt, um die Verwertung der Ergebnisse zu erleichtern.

Gerade die Komplexität von Evaluationen, die sowohl dem Postulat der sachlichen Richtigkeit (Wahrheit) als auch den unterschiedlichen Interessenorientierungen (den sog. „Stakeholders“) unterworfen sind, verhindern sehr oft, daß ihre Ergebnisse auch

umgesetzt werden können. (Shadish, Cook und Leviton1991) führen insgesamt sechs Widerstände an:

One obstacle is that evaluative results often threaten entrenched interests. Evaluation always has an adversarial relationship with some parties to public debates; its findings do not enter debates as uncontested nuggets of truth. A second obstacle is that social programs as a whole (as opposed to projects or elements) rarely die or get replaced. When they do, it is mostly for political or economic reasons, not because they were evaluated negatively. Hence evaluative information about program effects does not affect policy quickly. A third obstacle is that service deliverers engage in practices for reasons beside efficacy, such as convenience, habit, and security. A fourth obstacle is that political decision making is a slow process, and change is almost/ always incremental. A fifth obstacle is that policymakers and managers use information in many ways. Although political decision makers appreciate information about program performance (Weiss & Weiss, 1991), they must attend to conflicting values, interests, and expediency; and they have less power to modify programs than outside observer might think. Constraints are imposed upon them by past decisions and current fiscal realities, and by the political realities of social programming. A sixth obstacle is that even if an innovation is highly successful, local personnel may still not adopt it. (Shadish, Cook und Leviton1991:55)

3.5. Pragmatische Praxis

Allgemeine Fragen: Wie komme ich zu einem durchführbaren Design, das meinen Limitationen (Fertigkeiten, Zeit, Kosten usw.) Rechnung trägt? Welche Methoden sind sinnvoll und versprechen das gewünschte Ergebnis mit einem optimalen Aufwand? (Frage nach der Strategie = Frage nach dem Design der Evaluation, Frage nach der Taktik = Frage nach der zu verwendeten Methode).

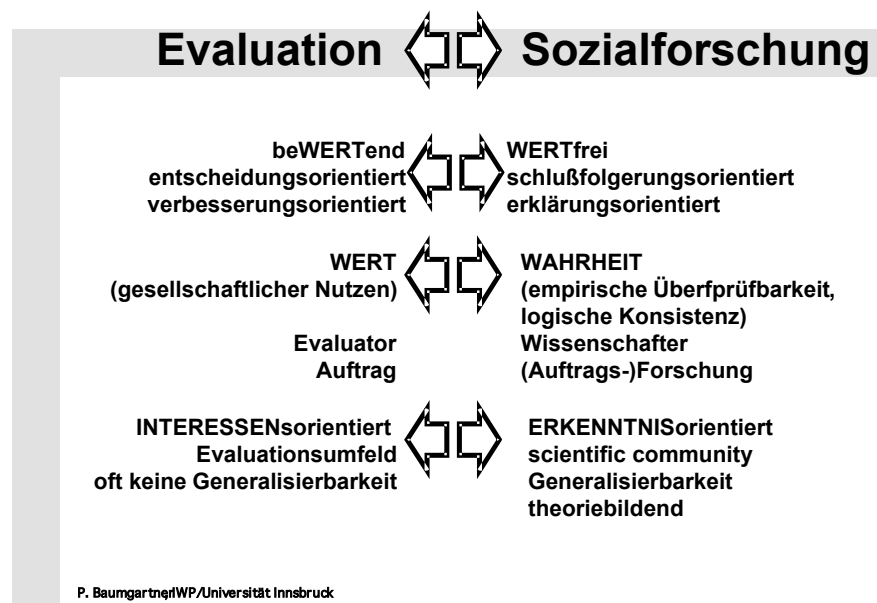
Metafragen: Welche Rollen als Evaluator kann ich einnehmen (Erzieher, Berater, methodologischer Experte, Richter?) Welche Rolle soll ich einnehmen? Welche Fragen soll ich stellen? Welche Methoden soll ich anwenden?

Beispiele: Soll überhaupt eine Evaluation durchgeführt werden? Was ist das Ziel/ der Zweck der Evaluation? Welche Rolle soll der Evaluator einnehmen? Welche Fragen sollen untersucht werden? Welches Design ist angebracht? Welche Aktivitäten werden gesetzt, um die Ergebnisse umzusetzen?

4. Ablauflogik von Evaluationen

4.1. Vergleich: Sozialforschung versus Evaluation

Eine Sichtweise, die in der Frage der Aufstellung und Zuweisung von Werten das zentrale Anliegen von Evaluationen sieht, ist nicht trivial, sondern hat ganz spezifische Auswirkungen auf Ablauf und Praxis von Evaluationen. Ich möchte dies exemplarisch in einem Vergleich Sozialforschung versus Evaluation demonstrieren:



Es ist für das Verständnis der Unterschiede dabei äußerst instruktiv, wenn wir uns kurz die unterschiedliche immanente Logik von sozialwissenschaftlichen Forschungsprozessen und Evaluationen vor Augen führen.

Der logische Ablauf sozialwissenschaftlicher Untersuchungen kann grob in drei Phasen eingeteilt werden (vgl. Friedrichs 1990:50-55):

- *Entdeckungszusammenhang (context of discovery)*: Darunter wird der Anlaß, (z.B. soziales Problem, Theoriebildung, Auftrag) der zu Entstehung eines Forschungsprojektes führt, verstanden. Er ist klar interessenorientiert und hebt aus der Unendlichkeit der Realität bestimmte für die Forscher „interessante“ Zusammenhänge heraus.
- *Begründungszusammenhang (context of justification)*: Darunter sind jene methodologischen Schritte zu verstehen, mit deren Hilfe das (herausgegriffene) Problem untersucht werden soll. Diese Phase stellt den eigentlichen wissenschaftlichen Bereich der Untersuchung dar, der möglichst exakt, reliabel, valid und intersubjektiv überprüfbar („objektivierbar“) sein soll.
- *Verwertungs- und Wirkungszusammenhang (context of utilization)*: Darunter sind alle Folgen der wissenschaftlichen Untersuchung zu rechnen, ob sie nun intendiert oder nicht intendiert sind. Streng genommen gehört hierzu auch bereits die Darstellung der Ergebnisse (wissenschaftlicher Artikel, Referate, Bericht etc.), da sie sowohl die Zugänglichkeit der Studie als auch deren Wirksamkeit beeinflusst. Wie die erste Phase ist auch dieser Abschnitt von Interessensorientierungen und Werturteilen geprägt.

Zur Logik des Ablaufs von Evaluationen wird nach wie vor ein heftiger wissenschaftlicher Disput geführt. Gerade die unterschiedlichen Sichtweisen zur Evaluation, wie sie sich in den oben erwähnten Definitionsversuchen äußern, sind eine Ursache für die wenig übereinstimmenden Positionen. Die nachfolgende Zusammenstellung folgt Scriven, dessen wissenschaftsphilosophische und -theoretische Arbeiten zur Evaluationslogik großen Einfluß und Verbreitung gefunden haben (vgl. (Scriven 1980 und 1991a):

- *Formulierung von Wertkriterien*: In der ersten Phase werden jene Kriterien ausgewählt und definiert, die der Evaluand erfüllen muß, um als gut, wertvoll etc. gelten zu können.

- *Formulierung von Leistungsstandards*: Für jedes einzelne Kriterium wird eine Norm definiert, die der Evaluand erreichen muß, damit das Kriterium als erfüllt angesehen werden kann (Operationalisierung).
- *Messung und Vergleich (Analyse)*: Nun wird jedes Kriterium beim Evaluanden untersucht, gemessen und mit den jeweils vorgegebenen Leistungsstandards verglichen.
- *Werturteil (Synthese)*: In dieser letzten und wohl schwierigsten Phase von Evaluationen müssen die verschiedenen Ergebnisse zu einem einheitlichen Werturteil integriert werden.

Die nachfolgende Tabelle versucht, diese Unterschiede zwischen Sozialforschung und Evaluation zusammenzufassen (in Anlehnung an Wottawa und Thierau 1990:48ff.).

These: Während die Sozialwissenschaften dem Postulat der Wertfreiheit verpflichtet sind, stellt die Zuweisung von Werten in Evaluationen gerade das zentrale Element dar. Während in den Sozialwissenschaften die Problemdefinition und die Verwertung aus dem eigentlichen Forschungszusammenhang ausgeklammert wird, werden Evaluationen (Inhalt, Methodologie und Design) gerade durch die Interessen und Ziele von Auftraggebern und Beteiligten charakterisiert.

4.2. Wertansprüche und Wertzuweisung

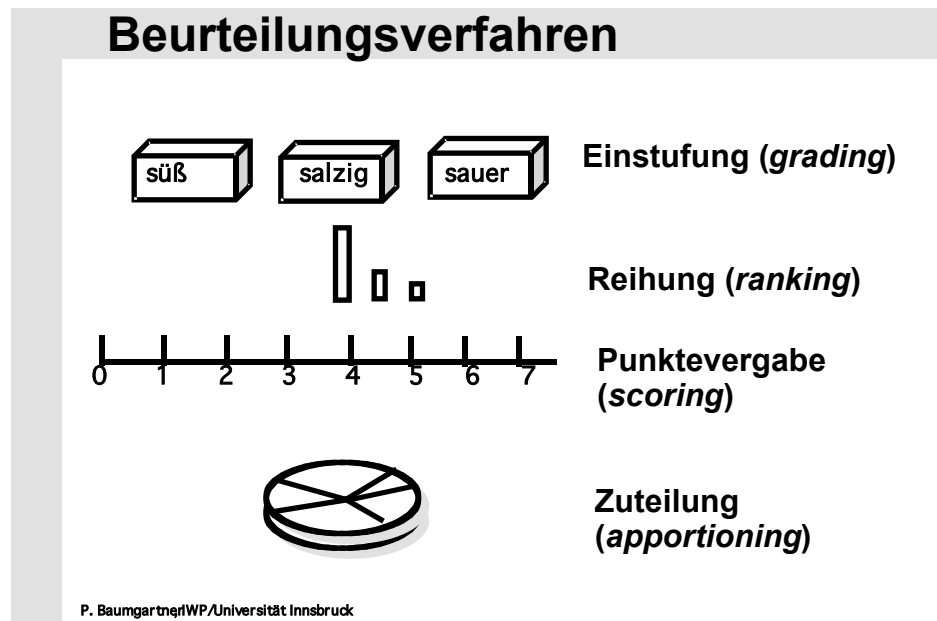
Als zentrales Mittel zur Definition von Wertansprüchen kann eine dreistufige *Bedürfnisanalyse* dienen:

- *Ebene 1 – Notwendigkeiten, Erfordernisse (necessitata, oft auch einfach als Bedürfnisse bezeichnet)*: Diese müssen unbedingt erfüllt sein, damit dem Evaluanden ein Wert zugesprochen werden kann. Sie sind in gewisser Weise ein Ausscheidungskriterium (k.o.-Kriterium). Der Einsatz einer Lernsoftware ist z.B. unmöglich, wenn sie auf der vorhandenen Hardware nicht lauffähig ist.
- *Ebene 2 – Vorzüge, weitergehende Ansprüche (desiderata, oft einfach auch als Wünsche bezeichnet)*: Hier handelt es sich um Funktionen und Eigenschaften, die über das absolute Minimum hinausgehen und den Wert des Evaluanden heben. Manchmal kann es der Fall sein, daß für ein bestimmtes Wertkriterium ein gewisses Minimum eine Notwendigkeit darstellt, darüber hinausgehend es jedoch bloß einen Vorzug darstellt. So stellt z.B. ein gewisse maximale Dauer der Systemantwort eine kognitionspsychologische Notwendigkeit zur Realisierung unmittelbaren Feedbacks bei interaktiver Lernsoftware dar. Obwohl eine weitere Unterschreitung dieser Grenze oft wünschenswert sein kann, bringt es keinen weiteren qualitativen Vorteil.
- *Ebene 3 – Ideale*: Obwohl sie nicht realisierbar bzw. kaum erreichbar sind, stellen sie doch eine wichtige Zielvorstellung dar. Sie geben an, in welche Richtung der Evaluand verbessert werden könnte bzw. sollte.

Eine gute Bedürfnisanalyse erfüllt nicht nur bereits wesentliche Momente der ersten zwei Phasen der Ablauflogik von Evaluationen, sondern ist auch für die Operationalisierung und Integration der Ergebnisse das entscheidende Instrument.

Für die Zuweisung von Werten (Beurteilungsverfahren) lassen sich grundsätzlich vier Methoden unterscheiden⁵:

⁵Die (leider häufige) Verwechslung bzw. Vermischung dieser unterschiedlichen Verfahren führt zu methodisch unsauberen Evaluationen.



- *Einstufung (grading)*: Die Beurteilung findet an Hand eines vorweg definierten Bewertungsmaßstabes statt. Dies ist z.B. dann der Fall, wenn bei einer Klausur die Beurteilung streng nach der Anzahl der beantworteten Fragen⁶ erfolgt. Häufig wird jedoch von den Lehrkräften aus optischen Gründen die Verteilung so „nachgebessert“, daß etwa eine Normalverteilung entsteht („grading on the curve“). In diesem Fall handelt es sich nicht mehr um Einstufung, sondern um Reihung.
- *Reihung (ranking)*: Hier werden die Evaluanden relativ zueinander beurteilt. Es entsteht eine Reihenfolge („gut - besser - am Besten“, „häufig - selten - nie“ usw.). Zu den Abständen der Evaluanden untereinander kann keine Aussage gemacht werden (Ordinalskala). Unsere Schulnoten z.B. stellen bloß eine Reihung dar: Operationen wie Addition, Division (wie sie z.B. für den Notendurchschnitt berechnet werden) sind – streng gesehen – eigentlich nicht zulässig.
- *Punktevergabe (scoring)*: Wenn Punkte vergeben werden, so ist unbedingt darauf zu achten, daß die Abstände zwischen den einzelnen Punkten bedeutungsvoll und gleich sind (Intervall- oder falls es einen Nullpunkt gibt: Ratio-Skala, auch metrische Skala genannt), ansonsten handelt es sich um bloße Reihung.
- *Aufteilung, Zuteilung (apportioning)*: Hierbei werden vorhandene Ressourcen entsprechend der Wertigkeit der Evaluanden aufgeteilt (z.B. Zuteilung von Budgetmitteln). Es ist eine häufig geübte Praxis die Begrenztheit von Ressourcen durch ein Ranking zu verschleiern und nur den oberen Plätzen eine Leistung zuzuteilen. Ein durchaus wertvoller Evaluand kann dadurch leer ausgehen: Obwohl seine Wertigkeit zu anderen – noch mit Ressourcen bedachten Evaluanden – gleich oder nicht zu argumentieren ist.

Selbstverständlich können die verschiedenen Methoden der Wertzuweisung auch kombiniert vorkommen: So können z.B. theoretisch in einer Evaluation alle Bewertungsverfahren von der Einstufung (Ausscheidung, Qualifikation) über eine Reihung und einer Vergabe von Punkten (z.B. bei Zeit-, Gewicht- oder Längenmessungen) bis zur Zuteilung (Preisverleihung) angewendet werden.

⁶Wir wollen hier der Einfachheit halber annehmen, daß alle Fragen gleiche Wertigkeit haben, d.h. gleich gewichtet sind.

Für eine genaue Analyse der (oft komplexen und vielschichtigen) Evaluanden müssen immer verschiedene Faktoren bzw. Komponenten gleichzeitig betrachtet werden. Das wichtigste Problem hierbei ist die Festlegung der relativen Wertigkeit (Gewichtung) der verschiedenen Dimensionen (vgl. Baumgartner 1995, Fricke 1995). Auf dieses Problem kann ich hier nicht eingehen, sondern verweise hier nur auf das dafür notwendige *qualitative* Gewichtungsverfahren (Baumgartner und Payr 1996).

5. Taxonomie von Evaluationsansätzen

These: Die definitorische Besonderheit von Evaluationen, wie sie in diesem Aufsatz vertreten wird (Zuweisungen von Werturteilen), ist auch ein ausgezeichnetes Kriterium für eine Einteilung der verschiedenen bei Evaluationen verwendeten Ansätze (Modelle). Gelingt es damit, eine stimmige Gliederung der mannigfachen Evaluationskonzepte durchzuführen, so wird damit gleichzeitig auch die Nützlichkeit der hier vorgeschlagenen Definition illustriert.

5.1. Ausgrenzungen

Je nachdem, wie Evaluationen mit der grundsätzlichen Frage der Zuweisung von Werturteilen umgehen, lassen sich bereits als erste grobe Annäherung bestimmte Studien als unechte Pseudo- und Quasi-Evaluationen stigmatisieren (Stufflebeam und Shinkfield 1985:45-57). In dieser Hinsicht stellt die Gleichung (Evaluation = Bewertung) bereits ein sehr scharfes Trennkriterium dar.

5.1.1. Pseudo-Evaluationen

Darunter sind alle Untersuchungen einzuordnen, die entweder politisch gesteuert sind oder ganz klar die Festigung (Bestätigung) einer vorgefaßten Meinung intendieren. Besonderes Kennzeichen dieser Art von Studien ist es, daß keine vollständige, umfassende und ausgewogene Analyse und Bewertung vorgenommen wird. Ausgangspunkt dieser Erhebungen sind:

- Gefährdete Interessen einer Adressatengruppe von echten Evaluationen (Interessenskonflikt). Die Pseudoevaluation soll Argumente für die gefährdete Position liefern und so die damit verbundene Interessensgruppierung stärken (= *politisch kontrollierte Studie*). Meistens werden diese Art von Untersuchungen verdeckt durchgeführt. Dadurch wird einerseits vermieden, daß die Öffentlichkeit vorzeitig ihre Aufmerksamkeit auf den Interessenskonflikt lenkt. Andererseits bleibt die Studie – falls ihre Ergebnisse den Auftraggebern nicht entsprechen – in der Schublade und wird nicht veröffentlicht.
- Der Versuch, durch gezielte Verbreitung bestimmter Informationen andere Interessensgruppierungen in ihrem Verhalten zu beeinflussen. Meistens dient sie dazu, ein bestimmtes Objekt (z.B. Konsumprodukt) in einem besonders vorteilhaften Licht erscheinen zu lassen (= *Public-Relation Studie*). Ein besonderes Kennzeichen dieser Studien ist ihre methodologische Fragwürdigkeit („quick and dirty“), die meistens zu einem (intendierten) systematischen Fehler führt.

Pseudo-Evaluationen geben nur vor, Evaluationen zu sein. Sie versuchen, die Autorität von echten Evaluationen für ihre eigenen (dubiosen) Interessen einzusetzen. Im weiteren Artikel werde ich daher diese Form der Untersuchungen nicht mehr weiter betrachten.

5.1.2. Quasi-Evaluationen

Hierbei handelt es sich um Untersuchungen, die zwar methodisch korrekt durchgeführt werden, jedoch bereits eine eingeschränkte – nicht mehr weiter zu hinterfragende – Ausgangsfragestellung haben. Besonderes Kennzeichen dieser Analysen ist es, daß sie eine Begründung, Diskussion und eventuelle Kritik der aufgestellten Wertansprüche vernachlässigen oder aber kritiklos zulassen. Sie nehmen die Aufgabenstellung unhinterfragt hin und beschäftigen sich sogleich mit der Auswahl einer adäquaten Methode zur Untersuchung der Problematik. Typische Beispiele für Quasi-Evaluationen sind:

- *Ziel-orientierte Evaluationsansätze* wie sie z.B. von Ralph Tyler in den 30-er Jahren entwickelt worden sind. Dazu ist auch die von Provus entwickelte Diskrepanz-Analyse zu zählen. Ausgehend von breit formulierten Zielen, die dann verfeinert und operationalisierbar gemacht werden, sollen Diskrepanzen zwischen Ziel und Realisierung festgestellt werden. Im Extremfall – wie z.B. bei gewissen Management-Informationen-Systemen (MIS) wird nur mehr beobachtet, ob der Evaluand gewisse Minimalkriterien überschreitet bzw. erfüllt (*Monitoring*). Obwohl zielbasierte Ansätze scheinbar objektiv sind, bedeuten sie immer eine Art von Tunnelvision, weil nur mehr vorgegebene Ziele untersucht werden. Damit ist die Legitimität der Untersuchung gefährdet, außerdem bleiben nicht intendierte Effekte unberücksichtigt. Scriven 1991a schlägt daher vor, diese mögliche Verzerrung (*bias*) durch eine ergänzende zielfreie Evaluation (*goal-free evaluation*) zu korrigieren. Dabei wird der Evaluand völlig unvoreingenommen untersucht. Offizielle Ziele, programmatische Papiere, Meinungen des Staffs und des Management etc. werden in dieser ersten Phase absichtlich nicht erhoben.
- *Experimentelle Untersuchungen* wie sie z.B. im quasi-experimentellen Forschungsdesign (Vergleichsgruppen) üblich sind (vgl. Thorndike et al. 1991; Wiersma 1991). So wird beispielsweise der Lernerfolg zweier Gruppen untersucht, die unterschiedlichen Maßnahmen (treatments) ausgesetzt worden sind (z.B. traditioneller Unterricht versus Verwendung von interaktiver Software). Die sorgfältige Beachtung methodologischer Forderungen (Vortest, Ähnlichkeit der beiden Gruppen in anderen als der untersuchten Variablen wie Alter, Geschlecht etc.) verhindert nicht, sondern im Gegenteil begünstigt die Mißachtung der ihnen implizit zugrunde liegenden Ziele. Die darin enthaltenen Werturteile werden unwidersprochen und z.T. sogar unbewußt akzeptiert. Was gilt z.B. als Kriterium für einen Lernerfolg? Ist es wirklich die bloße Erinnerung bei einem multiple-choice Test oder die richtige und vollständige verbale Reproduktion der vermittelten Inhalte bei offenen Fragen? Obwohl die komplexen Untersuchungsinstrumente (Fragebogen, multiple-choice Test) zwar methodisch einwandfrei konstruiert worden sind, messen sie immer nur das, was bereits als Ausgangspunkt ihrer Konstruktion unhinterfragt angenommen wurde („methodischer Zirkelschluß“). Und das kann oft auch völliger Unsinn sein („garbage in - garbage out“)

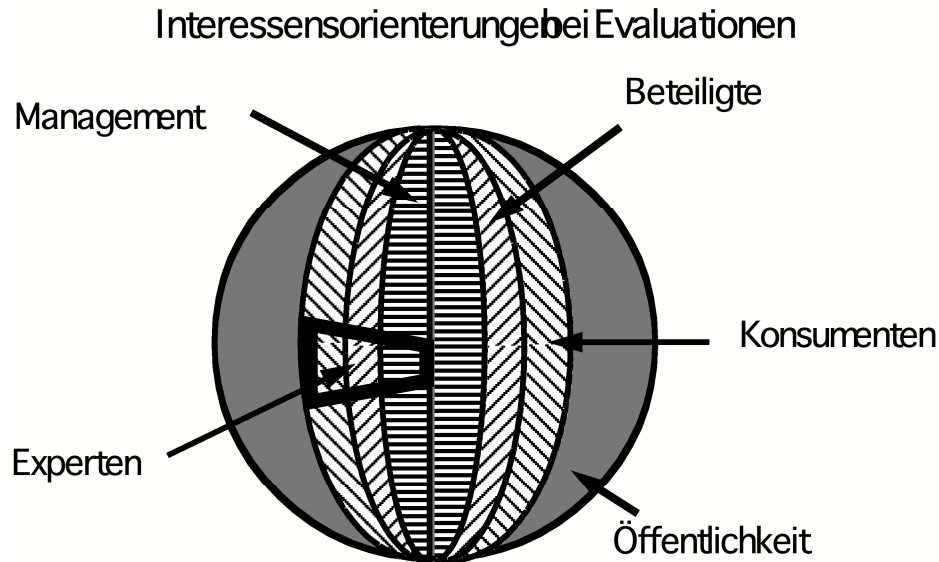
Im Gegensatz zu den Pseudo-Evaluationen können Quasi-Evaluationen durchaus ihre Berechtigung haben und im Einzelfall sogar sehr wertvoll sein. Sie klammern jedoch sowohl grundsätzliche Fragen zu den Zielsetzungen und den damit verbundenen Werturteilen als auch moralische Aspekte aus und sind oft interessensdominiert.

5.1.3. Echte Evaluationen

Hier handelt es sich um Untersuchungen, die sowohl die Definition als auch Begründung und Beurteilung von Wertansprüchen (Zieldefinition) in den Mittelpunkt der Analyse stellen. Diese Erhebungen zeichnen sich durch Methodenvielfalt, die Möglichkeit einer zielfreien Evaluation und den Versuch einer „gerechten“ Vorgangsweise und Darstellung aus (Interessenausgleich).

5.2. Interessensorientierungen von echten Evaluationen

Selbstverständlich werden auch echte Evaluationen nicht in einem interessensneutralen Raum durchgeführt. Gerade die definitorische Bedeutung von Wertzuweisungen bei Evaluationen ist interessensgeleitet. Je nach der unterschiedlichen Interessensorientierung können daher die echten Evaluationen weiter unterteilt werden. Die nachfolgende Grafik soll einen plastischen Eindruck dieser verschiedenen Gruppierungen vermitteln:



5.2.1. Management-orientierte Ansätze

Diese Evaluationsformen fungieren oft auch als Entscheidungsfindungs-Modelle. Von ganz besonderer Bedeutung ist dabei das von Stufflebeam ausgearbeitete *CIPP-Modell* (Stufflebeam und Shinkfield 1985), das einem systemtheoretischen Ansatz verpflichtet ist. Die Abkürzung steht für *Context-, Input- Process- und Product-Evaluation*, die jeweils verschiedene Abschnitte bzw. Phasen der Evaluation charakterisieren.

- Die *Kontextevaluation* soll Planungsentscheidungen evaluieren: Welche Bedürfnisse sollen durch einen Evaluanden abgedeckt werden? Mit der Kontextevaluation sollten die institutionellen Rahmenbedingungen definiert, Zielgruppe(n) und ihre Bedürfnisse eruiert und bewertet werden. Dabei sollen auch bereits mögliche Probleme (Konflikte), die diesen Bedürfnissen zugrunde liegen, diagnostiziert und beurteilt werden, ob die vorgeschlagenen Ziele die festgestellten Bedürfnisse überhaupt abdecken können.
- Die *Inputevaluation* dient der Bewertung von Strukturentscheidungen: Welche Ressourcen (Finanzen, Personal, Zeit) sind vorhanden? Welche alternativen Strategien (Evaluanden) kommen für die Befriedigung der festgestellten Bedürfnisse in Betracht? In dieser Phase werden auch mögliche Designs für die Implementierung ausgearbeitet und bewertet. Welche Pläne scheinen die besten Umsetzungsmöglichkeiten zu haben?
- Die *Prozessevaluation* widmet sich der Analyse von Implementierungsentscheidungen: Wie gut wurde der Evaluand implementiert? Welche Probleme gefährden seinen Erfolg? Es wird der Evaluand nicht nur in seiner statischen Erscheinung, sondern auch in seiner Entwicklung und funktionalen Aktivitäten beurteilt. Welche Korrekturen sind notwendig? Die Fragen in dieser Phase helfen, ein kontinuierliches Monitoring-System für das Management aufzubauen.

- Die *Produktevaluation* schließlich widmet sich der Beurteilung von Verwertungsentscheidungen: Welche Ergebnisse hat der Evaluand gebracht? Wie weit wurden Bedürfnisse auch tatsächlich befriedigt? Was soll mit dem Evaluanden weiter geschehen?

Die Beschreibung und Bewertung aller Phasen wird jeweils mit den anderen drei Evaluationsarten rückgekoppelt.

Das CIPP-Modell ist vor allem in betriebswirtschaftlichen Feldern sehr beliebt, weil es sich klar auf die Entscheidungsträger (Management) bezieht und mit ihnen eng zusammenarbeitet. Das Sammeln von überflüssigen Informationen wird meist vermieden, weil als Ausgangspunkt die zu treffenden Entscheidungen dienen und daher ein dementsprechender Fokus immer vorhanden ist. Als Nachteil wird allerdings oft gerade diese enge Konzentration auf die Entscheidungsträger moniert. Es wird ein unfaires und undemokratisches Vorgehen befürchtet, bei dem die Interessen der Evaluierten, aber auch der allgemeinen Öffentlichkeit zu wenig berücksichtigt werden.

Auch die Beachtung der Trennung von Bewertung und Entscheidung kann in diesem Ansatz leicht verwischt werden. Obwohl die Aufgabe des Evaluators ganz klar die Formulierung von Werturteilen und die Bewertung selbst ist, soll er/sie keine Entscheidungen vorschlagen oder gar treffen. Nach Cronbach (1983) sollen EvaluatorInnen die näheren Umstände beleuchten, aber nicht die Entscheidungen diktieren. Sie soll den AuftraggeberInnen helfen, die Komplexität der Probleme zu verstehen (d.h. sie sollen Entscheidungsgrundlagen liefern), aber keine einfachen Antworten auf zu eng definierte Fragestellungen geben.

5.2.2. Konsumenten-orientierte Ansätze

Diese Evaluationsmodelle werden häufig auch als Produkt-Evaluationen bezeichnet. Die Idee dabei ist, daß jeder Evaluand auch als Produkt betrachtet werden kann, das von bestimmten Adressatengruppen konsumiert wird. Dieser Ansatz beinhaltet zwar offensichtlich eine Marktorientierung, doch muß dies sich nicht in eng betriebswirtschaftlichen Kategorien niederschlagen. So kann unter einem *educational product* zwar alles verstanden werden, was am Bildungsmarkt erhältlich ist (Workshops, Lernsoftware, Lehrbücher, jede Form von Materialien, selbst Dienste und Methoden: Checklisten, Reviews und Reports), doch werden damit öffentliche (volkswirtschaftliche) Produkte (Schulen, Hochschulen etc.) nicht ausgeschlossen.

Der Konsumenten-orientierte Ansatz wird vor allem durch unabhängige Institutionen gefördert, finanziert bzw. durchgeführt (z.B. Konsumentenvereinigung). Die wichtigsten Ausarbeitungen für dieses Modell finden sich in den Arbeiten von Michael Scriven (1980; 1981; 1991a und b; 1996). Die Produktevaluation ist zwar meistens eine *summative* (= zusammenfassende, abschließende) Evaluation, allerdings kann sie in einem größeren Zeitrahmen auch als *formative* (= gestaltende, eingreifende) Evaluation dienen, indem neue oder gerade entwickelte Produkte sich in Zukunft ja auch den vorab definierten Kriterien stellen müssen.

Als Vorteil dieses Modells gilt, daß es große Publikumswirkung hat und hilft, bestimmte Standards aufzustellen und durchzusetzen (z.B. Ergonomie bei Büromöbeln). Produkt-Evaluationen sind daher ein gewisses Gegengewicht gegenüber den Versprechungen der Industrie geworden. Außerdem haben sie auch eine wichtige Bildungsfunktion, weil die Konsumenten durch ihre Berichte die Produkte in ihrer Funktionalität kritischer betrachten.

Allerdings sind auch eine Reihe von Nachteilen der Konsumenten-orientierten Ansätze anzuführen: So besteht z.B. die Gefahr, daß die Evaluationskosten den Produktpreis erhöhen können. Firmen versuchen, Zeit und Kosten von internen Evaluationen, die von den Forschungs- oder Prüfabteilungen durchzuführen sind, in die Preisgestaltung einzubeziehen. Andererseits fördert die Orientierung an definierte Standards nicht gerade die kreative Suche nach (neuen) Alternativen. Ein weiterer Nachteil besteht darin, daß bei Produkt-Evaluationen lokale Initiativen nicht gefördert werden. Aus Gründen der Professionalität und Kostenreduktion werden zentrale Stellen einge-

richtet, die dann die Evaluationen stellvertretend vornehmen und entsprechende Empfehlungen publizieren. Im Bildungsbereich heißt dies, daß die Auseinandersetzung über Stärken und Mängel eines bestimmten Produkts (z.B. Buch, Software) nicht von den Ausbildern selbst durchgeführt wird. Statt geprüftem Lehrmaterial wären jedoch materialprüfende Lehrer bildungspolitisch oft weit sinnvoller!

5.2.3. Experten-orientierte Ansätze

Sie gehören wahrscheinlich zu den ältesten und am weitesten verbreiteten Evaluationsansätzen. Obwohl in allen Modellen die Zuweisungen von Werten letztlich durch Subjekte erfolgen muß und daher in gewisser Weise subjektive Werturteile beinhalten, besteht die Besonderheit von Experten-orientierten Ansätzen gerade darin, daß sie damit kein Problem haben, sondern ganz offen und direkt auf dieser subjektiven professionellen Expertise aufbauen. Beispiele für die Expertenorientierung sind: Kommissionen zur Antragsprüfung (Forschungsantrag), Karriere (Habitationskommission), Beglaubigungskörperschaften (*Akkreditierung*) bzw. Vergabe von Qualitätssiegel (z.B. das Gütesiegel „EuroStudyCenter“ im Bereich der Fernstudien), Peer-Reviews bei wissenschaftlichen Zeitschriften, Preisverleihungen (z.B. der European Academic Software Award) etc.

Experten-orientierte Ansätze können nach dem Grad der Formalisierung (formales vs. informelles Begutachtungssystem) und der Anzahl der Experten (Komitee vs. individueller Bewertung) unterschieden werden (vgl. Worthen und Sanders 1987:99). Die für die Ausprägung der jeweiligen Dimension entscheidenden Fragen sind:

- Existiert eine Struktur für laufende Evaluationen?
- Gibt es publizierte Standards, die Teil dieser Evaluationen sind?
- Gehen in die Evaluation Meinungen von mehreren Experten ein?
- Haben die Ergebnisse der Evaluation Auswirkungen auf den Status des Evaluanden?

Von diesen traditionellen und bereits seit längerem geübten Verfahren ist der unkonventionelle Ansatz der Kennerschaft (engl. *connoisseurship* z.B. Weinkenner) von Elliot Eisner zu unterscheiden (Eisner 1979 und 1985): Dabei wird eine quantitative Analyse (und manchmal überhaupt ein formales Vorgehen) abgelehnt. Es wird für eine detaillierte narrative Beschreibung plädiert, die auch emphatische Momente nicht ausschließt. Künstlerische und humanistische Betrachtungen sind diesem Ansatz wichtiger als rein wissenschaftliche Kriterien (daher ist dieses Evaluationsmodell auch als *Kritizismus* z.B. in *Literaturkritik* bekannt).

Die Vorteile Experten-orientierter Ansätze liegen wie bei der Produktevaluation (die ja auch oft von Experten durchgeführt wird) in der laufenden Entwicklung von Qualitätsbewußtsein, Förderung guter Beispiele, Zurschaustellung und Kritik schlechter Fälle. Im allgemeinen führen selbst informelle Evaluierungen mit der Zeit zu einer gewissen Formalisierung (Richtlinien), die dann selbst wiederum zur formativen Evaluierung dienen können. Damit ist – trotz der bewußt subjektiven Komponente – auch in diesem Ansatz eine intersubjektive Überprüfbarkeit und Metaevaluierung gegeben. Schon allein das Bewußtsein, einer Evaluierung durch Experten ausgesetzt zu werden, führt zu verstärkten Anstrengungen (self-reviewing). Die durch das Kennerschaftsmodell erworbenen Perspektiven sind häufig nicht nur interessant, sondern eröffnen oft völlig neue, kreative und konstruktive alternative Sichtweisen.

Das wichtigste Problem, das mit diesem Ansatz jedoch untrennbar verbunden ist, ist die geringe Reliabilität des Verfahrens: Die Urteile der Experten sind oft schwer nachzuvollziehen, d.h. andere Experten kommen oft zu ganz anderen Resultaten. Außerdem vertreten die Experten häufig ganz andere Ansichten als die öffentliche Meinung, bzw. haben einen anderen „Geschmack“. Sie stellen daher keine zuverlässigen Führer für die Allgemeinheit dar. Zusätzlich birgt das Eisner-Modell auch noch die Gefahr der Irrelevanz (Validitätsproblem) in sich.

Weitere Schwierigkeiten in diesem Modell liegen im Umgang mit den Interessenskonflikten: Einerseits ist die Kritik manchmal zu wenig scharf, da die evaluieren-

den Experten selbst wiederum evaluiert werden („Eine Krähe hackt der anderen kein Auge aus!“); andererseits ist sie zu übertrieben, z.B. dann, wenn der jeweilige Experte sich selbst im evaluierten Gebiet bei knappen Ressourcen behaupten muß. Vor allem aber ist zu beachten, daß die Experten meistens aus inhaltlichen Gründen (Domain-Experts) ausgewählt wurden und wenig Erfahrung mit Evaluierungen haben können, d.h. keine Evaluationsexperten sind.

5.2.4. Öffentlichkeits-orientierte Ansätze

Darunter werden all jene Modelle verstanden, die nicht nur eine geplante Anstrengung unternehmen, unterschiedliche (meist Pro- und Kontra-) Sichtweisen in die Gesamtevaluation zu integrieren, sondern die konfligierende Bewertung und Zuweisung von Werten auch öffentlich austragen (*Hearing*). Dazu gehören die sogenannten Gegnerschafts-orientierten Ansätze (*adversary* oder auch *advocate-adversary*-approaches), wo zwei getrennte Teams jeweils die Vorzüge bzw. Nachteile des Evaluanden untersuchen. Eine besondere Erweiterung dieses Modells ist der *judicial-approach*, der den Prozeß vor Gericht als Metapher nimmt und weit über einfaches pro und kontra hinausgeht (Phase der Faktensammlung, Kreuzverhör von Zeugen, Beweiserhebung und andere gerichtliche Prozeduren wie Jury, Richter etc.)

Die Vorteile dieser Modelle sind vielfältig:

- Diese Evaluationen erzeugen großes öffentliches Interesse.
- Sie haben eine bildende Rolle sowohl für die Öffentlichkeit als auch für die „Geschworenen“. (Welche Argumente bzw. Fakten sind wichtig? Was soll bewertet werden?)
- Eine einseitige Sichtweise wird bereits durch das Design vermieden.
- Dieser Ansatz unterscheidet sich stark vom sozialwissenschaftlichen Modell der Evaluation, das oft nur für ein kleines, auserlesenes Publikum („scientific community“) verständlich ist (Scriven 1991a, 203).
- Durch die Einbeziehung unterschiedlicher Standpunkte wird eine größere Glaubwürdigkeit erreicht, die für nachfolgende Maßnahmen die Akzeptanzschwelle senkt.
- Öffentlichkeits-orientierte Ansätze erfordern immer eine intensive Planung und haben in sich bereits eine Art Meta-Evaluation eingebaut (Welche Daten sind relevant? Welche verschiedene Interpretationen/Sichtweisen/Kritiken gibt es?)

Allerdings stehen den zahlreichen Vorteilen auch eine Reihe ernster Kritikpunkte gegenüber:

- Die Öffentlichkeitsorientierung ist relativ teuer.
- Diese Ansätze streben vor allem nach Fairness. Gerechtigkeit ist jedoch nicht mit korrekter Bewertung gleichzusetzen. Das birgt die Gefahr in sich, daß der Interessensausgleich (Kompromiß) auf Kosten der Beurteilung stattfindet.
- Diese Ansätze schaffen durch die öffentliche Kontroverse oft auch „heißes Blut“, wodurch bereits bestehende Differenzen zu offenen Widersprüchen auswachsen können. Dies wird noch durch die Konkurrenzsituation von Pro- und Kontra-EvaluatorenInnen verstärkt. Damit werden Kompromisse und oft die weitere Arbeit erschwert („vergiftete Atmosphäre“).
- Die Anlehnung an die Gerichtsmetapher kann auch zu schweren Schäden (z.B. im Bildungsbereich) führen: Prozesse finden erst bei vermuteten Überschreitungen statt, Evaluationen sollten jedoch gerade diese krisenhaften Situationen meistern helfen. Bei Evaluationen geht es nicht um die Frage „schuldig“ oder „nicht schuldig“. Außerdem betont das Gerichtsmodell die Rolle der wirkungsvollen Rhetorik gegenüber der detaillierten Analyse:

Öffentlichkeits-orientierte Ansätze sind wegen der Betonung von Kritik und der im Verhältnis dazu geringeren Rolle der Ursachenanalyse vorwiegend für summative Evaluationen brauchbar. Die Anwendung dieses Evaluationsmodells ist daher besonders dann sinnvoll, wenn

- viele Menschen vom Evaluanden betroffen sind,

- über den Evaluanden schon öffentliche Debatten geführt worden sind und daher ein gewisses Vorverständnis im Bewußtsein der öffentlichen Meinung bereits vorhanden ist,
- summative Entscheidungen zu treffen sind,
- externe Evaluatoren die Untersuchung durchführen,
- klare Themen vorgegeben sind,
- Administratoren und Politiker über die Intensität (Zeit und Kosten) dieses Ansatzes informiert sind und sie bejahen,
- die notwendigen Ressourcen (Zeit, Kosten, Personal) für die Ausarbeitung unterschiedlicher (gegensätzlicher) Standpunkte vorhanden sind.

5.2.5. Teilnehmer-orientierte Ansätze

Darunter sind alle Ansätze zu verstehen, die nicht nach einem präskriptiven Muster vorgehen, sondern auf die Bedürfnisse und Interessen der Teilnehmer bzw. unmittelbar von der Evaluation Betroffenen eingehen und natürliche (vs. experimentelle, künstliche) Methoden anwenden. Daher werden diese Modelle auch als natürliche Evaluationen (*naturalistic evaluations*) bezeichnet. Bei den verwendeten Methoden handelt es sich meistens um qualitative Ansätze, obwohl dies kein unbedingtes Erfordernis darstellt. Häufig werden Methoden aus der Anthropologie verwendet, wie z.B. die von Geertz vorgeschlagene „dichte Beschreibung“ (1973 und 1983). Dadurch bekommen diese Ansätze große Ähnlichkeit mit ethnographischen Studien.

Teilnehmer-orientierte Modelle gehen nicht starr von vorgefaßten Zielen aus und sind daher sehr flexibel. Mit der Einbeziehung von subjektiven Präferenzen und Interessensorientierungen aller Beteiligten widersprechen sie dem traditionellen objektivistischen Wissenschaftsparadigma. Sie betonen induktives statt deduktives Schlußfolgern, sind auf Verstehen (Interpretationen) statt auf statistische Zusammenhänge ausgerichtet und bevorzugen Feld- bzw. Fallstudien. Sie gehen von einem Weltbild multipler Realitäten (Perspektiven) statt von einer einzigen objektive Realität aus und erheben dementsprechend auch die Daten von unterschiedlichen Gesichtspunkten und mit unterschiedlichen Methoden. In der Methodenliteratur wurde dieses Verfahren, das den Evaluanden von verschiedenen Perspektiven betrachtet als *Triangulation* (Daten-, Methoden-, Forscher- und Theorie-Triangulation) bekannt (vgl. Patton 1991:187ff.).

Teilnehmerorientierung ist derzeit in Mode, so daß ich hier nachfolgend nur eine Auflistung der vielen unterschiedlichen Ansätze geben kann:

- *Countenance-Modell* (Stake 1967): Beschreibung und Bewertung mit Hilfe einer Matrix
- *Transactional-Modell* (Riphey 1973): 5-Phasen-Konfliktmanagement, das va. notwendige Veränderungsprozesse begleiten soll.
- *Illuminative-Modell* (Parlett und Hamilton 1976): Dichte Beschreibung um das Verstehen komplexer Prozesse zu fördern. Es finden keine Wertungen statt, sondern es wird bloß eine progressive Fokussierung (Eingrenzung des Evaluanden, der Fragestellung...) und Triangulation (Perspektivwechsel) vorgenommen.
- *Democratic-Modell* (MacDonald 1974 und 1976): Unterscheidet bürokratische, autokratische und demokratische Evaluationen (je nachdem, wer über die Ergebnisse, verwendeten Methoden, Ziele etc. bestimmt). Es ist ein pluralistischer Ansatz, der jedoch jegliche Wertzuweisung vermeidet.
- *Responsive-Modell* (Stake 1975, Beywl 1988): Interagiert mit den InteressensvertreterInnen (*Stakeholder-Ansatz*) mit dem Ziel, deren Informationsprozeß untereinander bzw. deren Verständnis über den Evaluanden zu fördern.
- *Naturalistic-Modell* (Guba und Lincoln 1981, Lincoln und Guba 1985, House 1980): Ist eigentlich kein eigenes Modell, sondern eher eine Verallgemeinerung der genannten Verfahrensmodelle. Statt Wahrheit wird Glaubwürdigkeit, statt externe Validierung wird Anwendbarkeit (Passung, „fit“), statt Konsistenz oder Reliabilität wird Auditabilität (Prüfung durch eine nochmalige Evaluation, die durch

Dritte, d.h. externe Evaluatoren durchgeführt wird), statt Neutralität oder Objektivität wird Bestätigbarkeit forciert.

Die Vorteile von teilnehmerorientierten Ansätzen liegen darin, daß sie

- die betroffenen Menschen bei den Evaluationen in den Mittelpunkt rücken,
- die Bedeutung einer breiten Sichtweise mit multiplen Perspektiven betonen,
- auf neue Erkenntnisse ausgerichtet sind (ethnographische Forschung),
- methodisch und vom Design äußerst flexibel sind,
- daß durch die Einbeziehung aller Beteiligten dieser Ansatz durch die Evaluation selbst das Bewußtsein aller Beteiligten verändert, wodurch die spätere Implementierung der Ergebnisse erleichtert wird,

Die Nachteile einer Teilnehmerorientierung liegen,

- im relativ hohen Zeit- und Kostenaufwand,
- in der Neigung, extern sinnvolle Evaluationsstandards nicht zu erfüllen, da das aktuelle Feld mit seinen Bedürfnissen und Problemen Priorität hat,
- in den notwendig hohen sozialen Kompetenzen der EvaluatorsInnen, die flexible und in direkten Kompromissen mit den Betroffenen auszuhandelnde Vorgangsweisen erforderlich machen,
- in der improvisierten Erscheinungsform dieses Evaluationsmodelles, das gegenüber dem – als besonders wichtig empfundenen – Kriterium der Glaubwürdigkeit teilweise kontraproduktiv wirkt,
- in der häufig zu beobachtenden Enthaltensamkeit bei der Zuweisung von Werturteilen,
- der Konzentration auf nur wenige typische Fälle, die gut ausgesucht werden müssen, um systematische Fehler zu vermeiden,
- der Betonung der Prozeßaspekte der Evaluation, die gegenüber den Ergebnissen und der Bewertung Vorrang genießen.

6. Evaluation computerunterstützten Lernens

These: 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 wenig berücksichtigt.

Nicht nur um diese These zu untermauern, sondern damit sich auch bereits erste konkrete Anhaltspunkte für Evaluationsstrategien abzeichnen, werde ich nun versuchen, die Besonderheiten eines computerunterstützten Lernprozesses näher zu beschreiben. Aus analytischen Gründen und um die Darstellung etwas zu erleichtern, untersuche ich hier nur den individuellen mediengestützten Lernprozeß.

6.1. Interaktivität

Die zentrale neue Eigenschaft der sogenannten „Neuen Medien“ für den Lernprozeß besteht in der Möglichkeit des unmittelbaren individuellen Feedbacks, das durch die Interaktivität moderner multimedialer Software gegeben ist. Unter *Interaktivität* verstehe ich die Möglichkeit, daß Benutzer nicht bloß Rezipienten sind, sondern in den medial vermittelten Informations-, Kommunikations- und Lernprozeß gestaltend eingreifen. Das betrifft sowohl die Gestaltung der Inhalte, ihre Reihenfolge als auch die Zeitdauer, die mit einzelnen Phasen des Prozesses zugebracht werden. Die Interaktion kann die verschiedensten und vielfältigsten Formen annehmen: Texteingabe, Zeigen auf ein Grafikobjekt, selbständiges Aufsuchen von Daten (Hypertext), „entdeckendes“

Arbeiten in sogenannten Mikrowelten, zeitkritische Justierung komplex vernetzter Parameter in Simulationen, Manipulation von Eingabegeräten (Spaceball, Datenhandschuh,...) in virtuellen Umgebungen und so weiter.

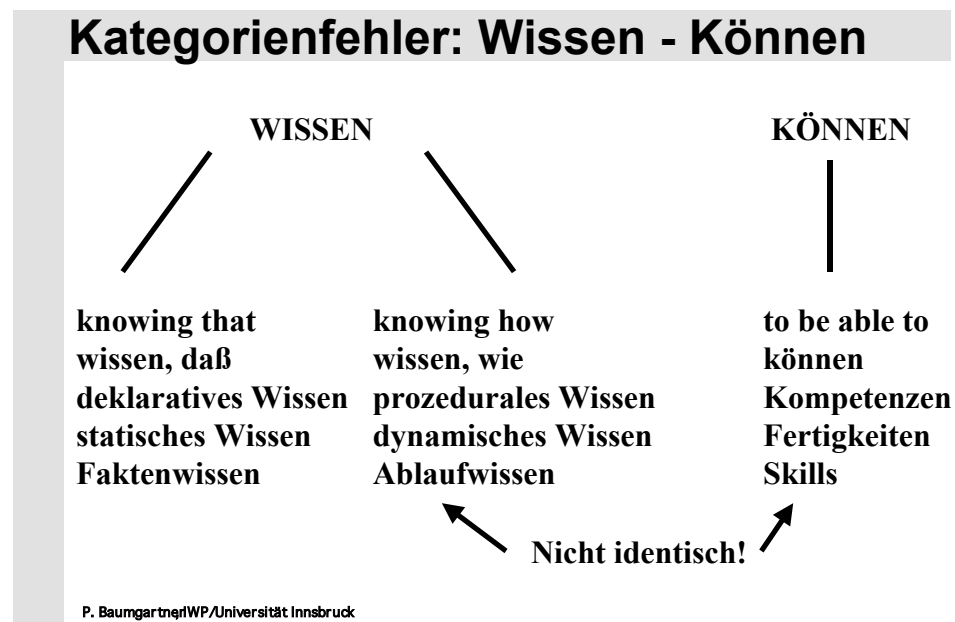
Bei den klassischen Print- oder audio-visuellen Medien ist eine Gestaltung des Prozeßablaufs dem Medium entweder nur äußerlich (z.B. indem ein bestimmtes Buch, Radio- oder Fernsehprogramm ausgewählt, weggelegt beziehungsweise abgeschaltet wird), oder bloß auf einseitige Aktionen (ohne „Inter“) des prinzipiell linear konzipierten Mediums reduziert (z.B. Wiederholen, Überspringen von Buchkapiteln, Videosequenzen).

Schon der Begriff „Benutzer“ drückt dieses aktivere Verhältnis zum Medium aus. Während die traditionellen Ausdrücke der klassischen audio-visuellen Medien, wie beispielsweise „Hörer“ oder „Seher“ eine eingeschränkte, nur auf die rezeptive Verarbeitung von speziellen Wahrnehmungskanälen abzielende Begrifflichkeit zur Grundlage haben, schlägt sich im Benutzerbegriff zumindest bereits ein instrumentelles Verhältnis zum Medium nieder.

Bereits aus dieser ersten – scheinbar trivialen – Feststellung ergeben sich weitgehende Konsequenzen für die Evaluation des Lernprozesses: Es ist nämlich keineswegs mehr klar und eindeutig, was unter *Lernerfolg* zu verstehen ist. Die bisherige traditionelle, jedoch oft vorherrschende Sichtweise, daß der Lernerfolg in der Reproduktion und/oder Anwendung von theoretischem Wissen (sei es nun Fakten- oder prozedurales Wissen) besteht, greift eindeutig zu kurz und ist noch überwiegend von den „alten“ Medien, d.h. der mehr oder weniger passiven Wissensvermittlung, geprägt.

Damit aber werden die heute immer noch sehr beliebten vergleichenden Untersuchungen des Lernerfolgs von Gruppen, die unterschiedliche Medien (z.B. Buch versus Software, oder allgemeiner: präsentierendes versus interaktives Medium) zum Lernen verwenden, vollkommen obsolet. Ausgehend von gleichen, jedoch unterschiedlich aufbereiteten Lerninhalten, wird hierbei nach *Effizienz* („Werden die Dinge richtig getan?“) statt *Effektivität* („Werden die richtigen Dinge getan?“) des Lernens gefragt. Meistens handelt es sich um eine Erhebung des kognitiven Lernerfolgs, der – aus Gründen der Vergleichbarkeit – zudem noch mit gleichen Prüfverfahren (z.B. schriftliche Tests für beide Gruppen) festgestellt werden soll. Dieses Evaluationsdesign ist aber in zweifacher Hinsicht absurd:

- Entweder handelt es sich bloß um einen Vergleich rezeptiven Lernens, wo es in erster Linie um die Präsentation (Vermittlung) von Inhalten mittels Text, Grafik oder Bewegung geht. In diesem Fall werden aber die neuen Eigenschaften und Vorteile der interaktiven Möglichkeiten gerade nicht genutzt! Es ist daher wenig überraschend, wenn diese Untersuchungen dann feststellen, daß interaktive Lernsoftware nicht besser als ein gutes Buch ist...
- Oder aber es wird in diesen Evaluationen gezeigt, daß bestimmte Zusammenhänge mit interaktiven Medien besser gelernt werden können. Diese vergleichenden Gruppenuntersuchungen sind dann nicht nur methodisch, sondern auch inhaltlich meist recht spitzfindig, weil – qua Evaluationsdesign – der Lernerfolg mit gleichen Inhalten und gleicher Prüfmethode festgestellt werden soll. Untersucht wird die Aneignung interaktiver Fertigkeiten, die aber bei der einen Methode (z.B. Buch) gar nicht gelehrt bzw. geübt werden. Dementsprechend sind auch die Ergebnisse („interaktive Medien sind beim Erwerb interaktiver Fertigkeiten besser“) trivial. Zurückzuführen sind diese unsinnigen Evaluationen auf die Gleichsetzung von prozeduralem Wissen („know how“ bzw. Wissen, *wie* etwas geht) mit der tatsächlichen Fertigkeit, dem *skill* selbst: Wenn ich weiß, wie ein Autoreifen gewechselt wird, heißt das lange noch nicht, daß ich es auch tatsächlich kann.



Ein weiterer Fehler in beiden Varianten besteht meiner Ansicht nach darin, daß das Erfolgskriterium – der Lernerfolg – nicht kritisch hinterfragt und neu definiert wird. Oder provozierender formuliert: Während die interaktiven Medien neue Lernparadigmen eröffnen, sind die traditionellen Evaluationen immer noch einer alten kognitivistisch-dominierten Sichtweise verpflichtet. Sie befinden sich in einer ähnlichen Situation, wie der Zuschauer, der die berühmte Tänzerin Isadore Duncan nach einer gelungenen Performance nach den Sinn ihres Tanzes fragte und zur Antwort bekam: „Welchen Sinn hätte es zu tanzen, wenn ich es auch in Worte ausdrücken könnte?“

Statt allgemein von Lernerfolg zu sprechen - und dabei meistens sprachlich formulierbares Wissen zu meinen, müssen bei der Evaluation interaktiver Medien vor allem die durch die Interaktivität angestoßenen Lernprozesse evaluiert werden. Diese lassen sich jedoch meistens gerade nicht sprachlich erfassen (vgl. dazu ausführlicher: Baumgartner 1993).

6.2. 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 hauptsächlich in der Form von privatem Studium – oft auch in den eigenen vier Wänden – unabhängig von Lehrkräften oder anderen Lernenden statt. Das ist ein großer Unterschied zu den traditionellen Lernformen, die durch den Besuch einer Serie von Unterrichtseinheiten direkt beobachtet und evaluiert werden können (Thorpe 1988:20). Viele Methoden der traditionellen Evaluation, (z.B. Beobachtung, Gruppendiskussion,...) 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 adäquat.

Andererseits eröffnen sich durch das neue Medium aber auch völlig neue Möglichkeiten, die für die Evaluation genutzt werden können. So kann durch eine geeignete Software der aktuelle Lernprozeß (wie navigiert wird, welche Aufgaben wie gelöst werden usw.) protokolliert werden. Das Netzwerk selbst kann für die Evaluation sowohl als Informationskanal (z.B. indem der Fragebogen als Webseite publiziert wird, sog. *Intrakommunikation*) und/oder auch als individueller Rückkanal verwendet

werden (z.B. wenn der ausgefüllte Fragebogen über e-mail an die EvaluatorInnen zurückgeschickt wird).

6.3. Heterogene Ausgangssituation und Bedürfnisstruktur

Die Möglichkeit des individuellen Lernens kann oft heterogene Ausgangssituationen auf der LernerInnenseite bedeuten. Während wir es beim Gruppenunterricht in der traditionellen Schulsituation meistens mit relativ homogenen demographischen und inhaltlichen Komponenten zu tun haben (Alter, Vorkenntnisse), ist dies entsprechend den Möglichkeiten des neuen Mediums tendenziell kaum mehr der Fall. Das betrifft nicht nur den bereits erwähnten individuellen Zugang, sondern va. auch die Möglichkeit, seine Lernprozesse individuell zu gestalten. Während der traditionelle Gruppenunterricht nur bei relativ homogenen Voraussetzungen funktioniert, ist computerunterstütztes Lernen nicht mehr dieser Restriktion unterworfen. Was, wann, wie lange und wie oft gelernt wird, liegt – falls es nicht explizit anders vorgesehen ist – in der Kontrolle der Lernenden.

Diese heterogene Ausgangssituation bringt automatisch auch eine unterschiedliche Bedürfnisstruktur mit sich. Potentiell ist mediengestütztes Lernen nicht nur in den Zugangs- und den Eingangsvoraussetzungen völlig offen, sondern auch in den Zielvorstellungen und den damit zusammenhängenden Motivationsstrukturen. Diese unterschiedlichen Bedürfnisse beeinflussen aber nicht nur den – wie immer definierten – Lernerfolg (Motivation), sondern sind auch für seine Definition („Was ist unter Lernerfolg zu verstehen?“) entscheidend. Während beim traditionellen Unterricht das Ziel meist vorgegeben ist (Abschluß, Prüfung, Zeugnis) braucht dies im mediengestützten Lernen nicht mehr der Fall zu sein. Vom sporadischen Aufsuchen individuell interessanter Informationen bis zur Absolvierung eines ganzen Kursprogrammes mit Prüfungen - *anything goes*.

Für Evaluationen heißt dies, daß sie komplexer werden und sowohl die unterschiedlichen Ausgangssituationen als auch die individuellen Lernstrategien reflektieren müssen.

6.4. Wahlfreiheit (individuelle Lernorganisation)

Mediengestützte Bildungsprozesse sind – so wird oft argumentiert – in einer gewissen Hinsicht sogar einfacher als personenorientierter Unterricht zu evaluieren. Durch das Medium findet nämlich scheinbar eine klarere Trennung zwischen Material und Lernprozeß statt. Aus diesem Grunde wird der selbständigen Evaluierung von Materialien oft größere Bedeutung zugesprochen (Produktevaluation z.B. von einem Lehrbuch).

Tatsächlich stellt dies für interaktive Software jedoch in mehrfacher Hinsicht einen Trugschluß dar:

- Erstens geht es gerade nicht um die inhaltliche *statische* Qualität des Materials, sondern darum, wie weit es in der Lage ist, Lernprozesse durch Interaktionen zu initiieren und/oder zu unterstützen. Die abstrakte inhaltliche Analyse von Software ist daher wenig sinnvoll. Dementsprechend sind auch die beliebten Check- und Prüflisten für Softwareevaluationen nur mit äußerster Vorsicht und eingeschränktem Wirkungsgrad zu genießen.
- Zweitens ist es beim mediengestützten Lernen gerade wichtig, das Augenmerk der Evaluation nicht nur auf den Lerneffekt einer einzelnen Kurseinheit (Wissens-transfer) zu legen. Das würde bedeuten, im alten Paradigma der (sequentiellen) Vermittlung theoretischen Wissens zu bleiben. Das Medium (die Software, das Internet) bietet praktisch unbegrenzt viele Lernangebote relativ unstrukturiert an. Auswahl, Reihenfolge und Bearbeitungsstrategie können von den Lernenden größtenteils selbst gesteuert werden. (Welcher Abschnitt bzw. Kurs wird wann und wie konsumiert?) Damit ist aber die Metaebene der individuellen Lernorgani-

sation („Lernen lernen“) angesprochen, die ebenfalls in das Design der Evaluation einbezogen werden muß.

Evaluationen mediengestützten Lernens werden aus den beiden oben genannten Gründen daher nicht einfacher sondern komplexer und verlangen infolge der großen individuellen Steuermöglichkeiten häufig qualitative Forschungsmethoden.

6.5. Lernsituation (soziale Lernorganisation)

Eine weitere wesentliche Eigenschaft mediengestützten Lernens, die es zu evaluieren gilt, besteht in der Verteilung des Materials, der Betreuung, der physikalischen (Hard- und Software, Leitungskapazitäten), sozialen und administrativen Organisation. Aus diesem Grunde greifen rein kognitionswissenschaftliche Konzepte in der inhaltlichen Gestaltung bzw. Interpretation der Evaluation zu kurz und müssen mit (organisations)soziologischen Ansätzen ergänzt werden.

Insbesondere ist es wichtig, zwischen den Effekten der Verteilung des Materials und den dynamischen Inhalten bzw. der interaktiven Didaktik des Materials zu unterscheiden. Da sowohl Inhalt als auch Verteilung mediengestützt umgesetzt werden kann (z.B. Transfer von Software über ftp-Server, Kurs im Internet über WWW-Server), müssen diese Effekte in der Evaluation analytisch zerlegt werden. Nur diese Trennung von *delivery technology* und *instruction technology* ist in der Lage, die jeweiligen gemessenen Effekte auch den entsprechenden Komponenten (dynamischer Inhalt, individuelle oder soziale Lernorganisation) kausal zuschreiben zu können.

Eine weitere Gefahr ist die Verwechslung bzw. Gleichsetzung von Interaktionen zur Steuerung der Software und Interaktionen, die aus pädagogischen Motiven verursacht werden (steuernde und didaktische Interaktionen). In vielen Fällen wird stolz gemeldet, daß „die Software gut angenommen wurde“. Dabei ist aber nicht klar, ob die unter Umständen einfache bzw. intuitive Bedienung der Software auch zu Lernerfolgen im zu vermittelnden Fachgebiet geführt hat.

6.6. Kostenanalyse

Weil es sich beim mediengestützten Lernen um alternative Lernverfahren mit neuartigen Lernorganisationen handelt, sind Kostenanalysen besonders wichtig. Bisher gibt es dazu relativ wenig Untersuchungen und Erfahrungen. Die nachfolgenden Gedanken sollen exemplarisch zwei Punkte veranschaulichen, die es bei der Kostenanalyse mediengestützter Lernprozesse zu beachten gilt (vgl. Levin 1981 und 1983):

Der *Kostenbegriff* ist nicht bloß mit den Ausgaben in der Buchhaltung gleichzusetzen, sondern es ist ein Kostenbegriff zu verwenden, der auch soziale Kosten beinhaltet. Kosten sind dabei als Opfer einer Alternative anzusehen:

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, 1981:30)

Dementsprechend sind z.B. auch Schenkungen von Hard- oder Software oder auch die Lernzeit der Studierenden einzubeziehen, weil diese Ressourcen gesellschaftlich auch anders verwendet werden hätten können.

Weiters ist es wichtig zu unterscheiden: *Kosten-Nutzen-Analyse (cost-benefit analysis)* und *Kosten-Wirksamkeits-Analyse (cost-effectiveness analysis)*. Im ersten Fall werden alle Kosten und Nutzen in Geldwert ausgedrückt und miteinander verglichen im zweiten Fall werden die Kosten zwar als Geldwert verrechnet, der Nutzen jedoch nicht, sondern z.B. in Testpunkten ausgedrückt.

Während mit der *Kosten-Nutzen-Analyse* inhaltlich völlig unterschiedliche Teilgebiete (z.B. Bildungs- vs. Transportwesen) miteinander verglichen werden können, ist dies bei der *Kosten-Wirksamkeits-Analyse* nicht möglich. So ist z.B. schon ein Ver-

gleich zwischen Lese- und Rechenfertigkeiten nicht mehr zulässig, weil sie verschiedene Tests erforderlich machen, deren Punktwerte nicht mehr dasselbe aussagen (keine metrische Skala mit gleichen Abständen und Nullpunkt).

Literatur

- Baumgartner, P. 1993. *Der Hintergrund des Wissens. Vorarbeiten zu einer Kritik der programmierbaren Vernunft*. Klagenfurt: Kärntner Druck- und Verlagsanstalt.
- Baumgartner, P. 1995. Didaktische Anforderungen an (multimediale) Lernsoftware. In: *Information und Lernen mit Multimedia*, Hg. von L. J. Issing und P. Klimsa. Weinheim: Psychologie-Verl.-Union. 241-252.
- Baumgartner, P. und S. Payr. 1990. Körper, Kontext und Kultur. Explorationen in den Hintergrund des Wissens. *Informatik Forum. Fachzeitschrift für Informatik*, Nr. 2: 62-74.
- Baumgartner, P. und S. Payr. 1996. Learning as action: A social science approach to the evaluation of interactive media. In: *Proceedings of ED-MEDIA 96 - World Conference on Educational Multimedia and Hypermedia*, Hg. von P. Carslon und F. Makedon. Charlottesville: AACE. 31-37.
- Beywl, W. 1988. *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation*. Frankfurt/M.: Peter Lang.
- Cronbach, L. J. 1983. *Designing Evaluation of Educational and Social Programs*. 2. Aufl. San Francisco: Jossey-Bass.
- Eisner, E. W. 1979. *The educational Imagination: On the design and evaluation of school programs*. New York: Macmillan.
- Eisner, E. W. 1985. *The art of educational Evaluation. A personal view*. London: The Falmer Press.
- Fricke, R. 1995. Evaluation von Multimedia. In: *Information und Lernen mit Multimedia*, Hg. von L. J. Issing und P. Klimsa. Weinheim: Psychologie-Verl.-Union. 401-413.
- Friedrichs, J. 1990. *Methoden empirischer Sozialforschung*. 14. Aufl. Opladen: Westdeutscher Verlag.
- Geertz, C. 1973. *The Interpretation of Cultures*. o.O.: Basic Books.
- Geertz, C. 1983. *Local Knowledge. Further Essays in Interpretive Anthropology*. New York: Basic Books.
- Götz, K. 1993. *Zur Evaluierung beruflicher Weiterbildung*. Weinheim: Deutscher Studien Verlag.
- Gronlund, N. E. 1976. *Measurement and Evaluation in Teaching*. New York: McGraw-Hill.
- Guba, E. G. und Y. S. Lincoln. 1981. *Effective evaluation*. San Francisco: Jossey-Bass.
- House, E. R. 1980. *Evaluating with validity*. Beverly Hills, CA: SAGE.
- Levin, H. M. 1981. Cost Analysis. In: *New techniques for evaluation*, Hg. von N. L. Smith. Beverly Hills, CA: SAGE. 13-70.
- Levin, H. M. 1983. *Cost-effectiveness: A primer*. Beverly Hills, CA: SAGE.
- Lincoln, Y. S. und E. G. Guba. 1985. *Naturalistic Inquiry*. Beverly Hills, CA: SAGE.
- MacDonald, J. B. 1974. An evaluation of evaluation. *Urban Review*. Bd. 7, Nr. 1: 3-14.
- MacDonald, J. B. 1976. Evaluation and the control of education. In: *Curriculum evaluation today: Trends and implications*, Hg. von D. Tawney. London: Macmillan.
- Madaus, G. F., M. Scriven und D. L. Stufflebeam, Hg. 1983. *Evaluation Models*. Boston: Kluwer.

- Parlett, M. und D. Hamilton. 1976. Evaluation as illumination: A new approach to the study of innovatory programs. In: *Evaluation studies review annual*, Hg. von G. V. Glass. Beverly Hills, CA: SAGE.
- Patton, M. Q. 1991. *Qualitative evaluation and research methods*. 2. Aufl. Newbury Park: SAGE.
- Rippey, R. M., Hg. 1973. *Studies in transactional evaluation*. Berkeley, CA: McCutchan.
- Rossi, P. H. und H. E. Freeman. 1989. *Evaluation. A Systematic Approach*. 4. Aufl. Newbury Park: SAGE.
- Rossi, P. H., H. E. Freeman und G. Hofmann. 1988. *Programm-Evaluation. Einführung in die Methoden angewandter Sozialforschung*. Stuttgart: Enke.
- Scriven, M. 1980. *The Logic of Evaluation*. Inverness, CA: Edgepress.
- Scriven, M. 1981. Product Evaluation. In: *New techniques for evaluation*, Hg. von N. L. Smith. Beverly Hills, CA: SAGE. 121-166.
- Scriven, M. 1991a. *Evaluation Thesaurus*. 4. Aufl. Newbury Park: SAGE.
- Scriven, M. 1991b. Evaluation: Logic's Last Frontier In: *Critical Reasoning in Contemporary Culture*, Hg. von R. A. Talaska: SUNY Press.
- Scriven, M. 1991c. Introduction: The Nature of Evaluation. In: *Evaluation Thesaurus*, Hg. von M. Scriven. 4. Aufl. Newbury Park: SAGE. 1-43.
- Shadish, W. R. J., T. D. Cook und L. C. Leviton. 1991. *Foundation of Program Evaluation. Theories of Practice*. Newbury Park: SAGE.
- Stake, R. E. 1967. The countenance of educational evaluation. *Teachers College Record*, Nr. 68: 523-540.
- Stake, R. E. 1975. *Evaluation the arts in education. A responsive approach*. Columbus, OH: Charles E. Merrill.
- Stufflebeam, D. L. und A. J. Shinkfield. 1985. *Systematic Evaluation*. Boston: Kluwer.
- Thorndike, R. M., G. K. Cunningham, R. L. Thorndike et al. 1991. *Measurement and Evaluation in Psychology and Education*. 5. Aufl. New York: Macmillan.
- Thorpe, M. 1988. *Evaluation Open & Distance Learning*. Essex: Longman.
- Weber, M. 1988a. Der Sinn der "Wertfreiheit" der soziologischen und ökonomischen Wissenschaften. In: *Gesammelte Aufsätze zur Wissenschaftslehre*, Hg. von M. Weber. 7. Aufl. Tübingen: UTB Mohr. 489-540.
- Weber, M. 1988b. Die "Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis. In: *Gesammelte Aufsätze zur Wissenschaftslehre*, Hg. von M. Weber. 7. Aufl. Tübingen: UTB Mohr. 146-214.
- Wiersma, W. 1991. *Research Methods in Education*. 5. Aufl. Needham Heights, MA: Simon & Schuster.
- Will, H., A. Winteler und A. Krapp, Hg. 1987. *Evaluation in der beruflichen Aus- und Weiterbildung*. Heidelberg: Sauer.
- Worthen, B. R. und J. R. Sanders. 1987. *Educational evaluation: Alternativ approaches and practical guidelines*. White Plains: Longman.
- Wottawa, H. und H. Thierau. 1990. *Lehrbuch Evaluation*. Bern: Huber.