

Ziele von Teil I (Kapitel 1–3)

Dieser erste Teil soll in das Thema »Evaluation« einführen, Grundkenntnisse vermitteln und insbesondere Ihren Mut stärken, sich mit diesem Thema zu beschäftigen – nicht nur auf dem Papier dieses Buches, sondern auch in der Praxis.

Wichtigstes Ziel dieses ersten Teils ist, Sie »bei der Stange zu halten«. Das Schlimmste, was passieren könnte, wäre, dass Sie abgeschreckt dieses Buch und Thema zur Seite legen. Das Ziel »bei der Stange halten« würde in Idealform bedeuten, dass Sie nach dem Bearbeiten dieses Kapitels gleich Lust haben, in einem Ihrer Praxisfelder Evaluation zu betreiben, mehr noch: dass Sie nicht nur Lust haben, sondern es auch tun. Ein eher laues Ergebnis wäre es, wenn Sie alles hier Aufgeschriebene wiederholen können, es aber nicht in Praxishandeln umsetzen.

Als Ergebnis dieses ersten Teils sollen Sie

- ▶ definieren, was »Evaluation« meint, _____
- ▶ Evaluation abgrenzen einerseits gegen Forschung, andererseits gegen didaktische Rückmeldeformen wie Seminarkritik, Rückblick, Nachbereitung, _____
- ▶ zentrale Funktionen von Evaluation nennen: Wozu dient Evaluation? _____
- ▶ ein halbes Dutzend Indikatoren aufzählen, die als Kriterien zur Evaluation herangezogen werden können, _____
- ▶ erklären, was mit »Indikatorenschaukel« gemeint ist, _____
- ▶ drei Klein-Evaluationsformen kennen, _____
- ▶ einen Kursbeurteilungsbogen kennen, _____
- ▶ die Methode »Breitenevaluation« einsetzen, _____
- ▶ erklären, wie unterschiedliche Formen von Evaluation mit unterschiedlichen Vorstellungen vom Lernen zusammenhängen. _____

Eine ganze Menge für 70 Seiten? Überlegen Sie vor dem Weiterlesen jetzt zunächst, was Sie zu jedem dieser Punkte bereits wissen – versuchen Sie eine Ist-Stand-Schätzung aufzustellen. Und tragen Sie auf der ersten Linie nach jedem Punkt eine Prozentzahl ein: Bei wie viel Prozent, glauben Sie, liegt Ihr Ist-Wert?

1. Evaluation – was das (nicht) ist und wozu sie nützt

1.1 Zum Begriff

Wie bei jedem vieldiskutierten Begriff finden sich auch bei »Evaluation« unterschiedliche Definitionen. So wird bei Durchsicht der Definitionen der letzten dreißig Jahre die »schillernde Vielfalt der mit dem Begriff ‚Evaluation‘ assoziierten Vorstellungen« (*Wottawa/Thierau* 1998, S. 13) konstatiert, es wird festgestellt: »Aktuelle Evaluationsbegriffe sind vielschichtig« (*Balzer/Frey/Nenniger* 1999, S. 393), und zusammengefasst: »Evaluation umfasst ... eine Vielzahl von Bedeutungsfacetten und Vorgehensweisen, und in der Literatur finden sich zahlreiche, zum Teil sehr unterschiedliche Vorstellungen und Beschreibungen« (*Mittag/Hager* 2000, S. 103). Zu dieser schillernden Vielfalt beigetragen haben neben anderem die unterschiedlichen Wurzeln der Evaluationsforschung; *Rost* (2000, S. 129) nennt die Curriculumevaluation, die Evaluation sozialer Programme, die Evaluation therapeutischer Maßnahmen und die Evaluation im Kontext von Betrieben und Organisationen. Insbesondere beim Bezug auf englischsprachige Quellen trägt zur Unklarheit auch bei, dass »Evaluation« bzw. »to evaluate« in der englischen Umgangssprache nicht nur Fachtermini sind, sondern zunächst und vor allem schlicht und ohne besonderen Anspruch »Bewertung« bzw. »bewerten« meinen; dies hat bei deutschen Übersetzungen manchmal Missverständnisse hervorgerufen.

Betrachtet man weniger die Worte und Formulierungen der Evaluationsdefinitionen, sondern die darin angesprochenen Elemente, Aufgaben und Funktionen von Evaluation, dann lassen sich jedoch durchaus eine Reihe von Gemeinsamkeiten identifizieren. Als pädagogischer/andragogischer Fachbegriff im Deutschen wie im Englischen enthält der Begriff »Evaluation« im Wesentlichen drei Elemente, die sich in den meisten Evaluationsdefinitionen finden; diese werden in folgender Definition zusammengefasst.

Definition

Evaluation meint

1. das methodische Erfassen und
2. das begründete Bewerten von Prozessen und Ergebnissen zum
3. besseren Verstehen und Gestalten einer Praxis-Maßnahme im Bildungsbereich durch Wirkungskontrolle, Steuerung und Reflexion.

Diese Definition soll diesem Lehrbuch zu Grunde liegen.

1.1.1 »Erfassen« = methodisch organisiert + dokumentiert (Definitionselement 1a + b)

Der Fachbegriff »Evaluation« enthält zunächst das **Erfassen von Prozessen und Ergebnissen**. »Erfassen« bedeutet dabei eine explizite Handlung, die (a) **methodisch organisiert** und damit überprüfbar ist und deren Ergebnis (b) festgehalten – **dokumentiert** – wird (»schwarz auf weiß«).

- a) Für das, was man erfasst, wird die Bezeichnung »Daten« verwendet. Dies können Zahlen oder Worte sein; Bilder, Fotos, Artefakte u.ä. werden, wenn sie als Datengrundlage dienen, vor der Weiterverarbeitung in Zahlen und Worte »übersetzt«. Zum Erfassen der Daten verwendet man Instrumente: Fragebögen, Interviews, Tests, Checklisten, Beobachtungslisten. Die Spanne der Methoden der Datenerhebung reicht von strengen Forschungsdesigns mit Kontrollgruppen, Randomisierung, standardisierten Tests und Signifikanztests in der Tradition quantitativer Unterrichtsforschung bis zu qualitativ-naturalistischen Fallstudien in der Tradition responsiver Evaluation (vgl. *Beywl* 1998). *Widmer* (2000, S. 80) fasst die Diskussion mit dem Stichwort »Methodenpluralismus« zusammen. Charakteristisch für das methodisch organisierte Erfassen ist, dass man hinterher weiß, wie die Daten zustande kamen. »Einige Teilnehmer wurden befragt ...« erfüllt dieses Kriterium nicht. »Jeder fünfte Teilnehmer gemäß alphabetischer Reihenfolge wurde befragt; unmittelbar im Anschluss wurde jeweils ein Gedächtnisprotokoll angefertigt ...« erfüllt dieses Kriterium.

Die Bedingung, dass Evaluation methodisch erhobene Daten voraussetzt, findet sich übereinstimmend sowohl in der älteren als auch in der jüngsten Diskussion. Gelegentlich wird zwischen Evaluation (methodisch weicher) und Evaluationsforschung (methodisch strenger) unterschieden (z. B. *Suchmann* 1967, *Mittag/Hager* 2000), gelegentlich werden die Begriffe synonym verwendet (z. B. *Stockmann* 2000). Da es im Gegensatz zur englischen Sprache keinen umgangssprachlichen Konkurrenzbegriff gibt, soll hier für den deutschen Fachbegriff »Evaluation« die strengere Position vertreten werden, dass die methodische Erhebung von Daten unabdingbares Element von Evaluation ist.

»Es wurde mit Betroffenen gesprochen ...« oder »Beim Besuch der Einrichtung konnte der Eindruck gewonnen werden ...« – solche vagen Formulierungen zeigen, dass überhaupt keine Daten (= methodisch organisiert) erhoben wurden; alles, was nachfolgt, kann nicht mehr

1. Evaluation – was das (nicht) ist und wozu sie nützt

beanspruchen, Evaluation zu sein. Für allgemeine wertende Aussagen ohne systematische Datenbasis sollte nach diesem Verständnis der Begriff »Evaluation« nicht verwendet werden (sondern eben »Meinung«, »Bewertung« ..., oder »Visitation«, »Hospitation«, »Kontrollbesuch«, »Mitarbeiterbeurteilung«, »Rückmeldung« ...).

- b) Seltener wird explizit auf die Notwendigkeit der Dokumentation, Aufbereitung, Darstellung und Verfügbarkeit der Instrumente und Daten hingewiesen (zumeist unter dem Stichwort »Berichtlegung«, z. B. *Wottawa/Thierau* 2000, S. 159, *Balzer/Frei/Nenniger* 1999, S. 399), wohl weil dies Forschern selbstverständlich erscheint. Für Evaluationen, die zumeist nah bei und mit den Betroffenen und Beteiligten erarbeitet werden, ist eine solche »at-hand«-Verfügbarkeit der Instrumente und Daten unabdingbar, um das Gefühl von Willkür oder subjektiver Interpretation zu vermeiden und um immer klarzumachen: Evaluation ist ein datenbasiertes Verfahren!

Ausschlusskriterien

Wenn keine Dokumentation vorliegt, wenn Aussagen nicht mit zugänglichen Daten begründet werden können, dann sollte – so wird hier vorgeschlagen – nicht von Evaluation gesprochen werden.

Übungsaufgabe 1*

Kreuzen Sie an, ob bei folgenden Maßnahmen die Kriterien »methodisch organisiert« und/oder »Ergebnis dokumentiert« erfüllt ist:



| | methodisch organisiert | Ergebnis dokumentiert |
|--|------------------------|-----------------------|
| a) Am Ende eines Kurses redet man miteinander über den Kurs. | ☆ | ☆ |
| b) Man tauscht Meinungen über die Stärken und Schwächen eines Volkshochschulprogramms aus. | ☆ | ☆ |
| c) Ein »Fragebogen zur Kursbewertung« wurde ausgefüllt und liegt unausgewertet im Schrank. | ☆ | ☆ |

* Die Lösungen der Übungsaufgaben finden Sie im Anhang.

1.1 Zum Begriff

| | methodisch organisiert | Ergebnis dokumentiert |
|--|---------------------------|--------------------------|
| d) Am Kursende schreiben alle Teilnehmer Moderationskärtchen zur Frage »Was hat der Kurs mir gebracht (grüne Kärtchen), was hat mir gefehlt (rote Kärtchen)?« Diese werden an der Pinwand systematisch geordnet und mit den Teilnehmern, dem Kursleiter und dem Leiter des Tagungshauses besprochen. Danach wandern sie in den Papierkorb. | ☆ | ☆ |

Anstelle der hier gewählten Bezeichnung »Erfassen« wird in anderen Definitionen manchmal die Bezeichnung »Messung« verwendet. »Messung« bedeutet, dass Beobachtungen auf einer »mehr – weniger«-Skala angeordnet werden. Dies wäre eine strenge Auslegung des Evaluationsbegriffes, denn es gibt auch andere hilfreiche Formen, mit denen Prozesse und Ergebnisse erfasst und dokumentiert werden können.

1.1.2 »Bewerten« = Vergleich Ist-/ Soll-Wert + Begründung Soll-Wert (Definitionselement 2a + b)

»Evaluation« enthält als Zweites das **Bewerten von Prozessen und Ergebnissen** auf der Grundlage der erhobenen Daten. Wird »Evaluation« als Fachbegriff verwendet, dann ist dieses »Bewerten« etwas anderes als »seine subjektive Meinung äußern«: Man vergleicht die erhobenen Daten (»Ist-Wert«) mit dem, was man sich gewünscht hat (»Soll-Wert«). Dieser Soll-Wert sollte möglichst vorher, spätestens im Verlauf der Evaluationsuntersuchung durch Beschreibungen, Erläuterungen und Begründungen so nachvollziehbar gemacht werden, dass auch für einen Außenstehenden verständlich wird, warum welche Kriterien zu einem bestimmten Urteil führen. Der Maßstab muss also explizit begründet werden und kann damit überprüft, kritisiert und korrigiert werden.

Dieses »Bewerten« als zweites Element der Evaluationsdefinition wird als eine der problematischsten Anforderungen an Evaluation diskutiert. *Wottawa/Thierau* warnen: »Leider sind die Fragen der Zielsetzung in vielen Evaluationsprojekten die größte Schwachstelle« (1999, S. 83). Einhellig ist die Position, dass die Sammlung von Daten allein noch keine Evaluation ist. So formulieren *Thorndike/Hagen* bereits in einem in mehreren Auflagen erschienenen frühen Standardwerk: »It can not be too much emphasised that measurement at best provides only information, not judgment« (*Thorndike/Hagen* 1964, S. 13) Und: »The term ‚evaluation‘ ... includes more definitely the aspect of valuing – of saying what is desirable and good« (a. a. O.

S. 27). Um »bewerten« zu können, muss unumgänglich entschieden werden, was als Kriterium für »Erfolg/Nichterfolg« angesehen wird. Sind die generellen Zielsetzungen und Beurteilungsdimensionen formuliert, dann müssen Bewertungskriterien expliziert werden; *Wottawa/Thierau* (1998, S. 83ff.) unterscheiden zwischen Zielexplication, Bewertungsprozess und Bewertungs- und Entscheidungshilfen. *Mittag/Hager* (2000, S. 109) bezeichnen die »Formulierung und Präzisierung solcher Ist-Soll- oder Wird-Soll-Diskrepanzen« als Arbeitsschritt der Zielbestimmung. Zur Formulierung solcher Ist- und Soll-Werte werden zumeist Operationalisierungsstrategien nach *Mager* (1969, 1973 – siehe Kap. 9) eingesetzt (z. B. *Wottawa/Thierau* 1998, S. 92, *Mittag/Hager* 2000, S. 109, *Mittag/Jerusalem* 1997, S. 600).

Ein zentrales Evaluationsproblem ist die Frage: Wer definiert Kriterien und Maßstäbe? Zum Teil wird auf den Auftraggeber verwiesen, der solche Soll-Werte vorgebe (»Wenn nicht 80% der Teilnehmer die Prüfung bestehen, wechseln wir den Weiterbildungs-Anbieter«), zum Teil auf die Kompetenz des Evaluators (»Nur der Evaluator hat die Fachkompetenz und Neutralität« – z. B. *Scriven* 1972, S. 72, *Balzer/Frey/Nenniger* 1999, S. 398), zum Teil gesteht man das Urteil den Befragten selbst zu (»Wenn alle Studenten die Vorlesungen unterhaltend, abwechslungsreich und leicht einschätzen, dann haben wir eine gute Hochschule(!?)«). Manchmal ist der Ist-Soll-Wert-Vergleich banal-einfach (»Wer ist der billigste Anbieter?«), manchmal erfordert er eine aufwendige Zielanalyse (»Wer ist der beste Anbieter?«); und manchmal geschieht der für traditionelle Forschung undenkbarer Horror, dass sich im Evaluationsprozess das Objekt und seine Bewertung verändert: »Ein zunächst gewähltes Thema kann auch einfach verschwinden und an seiner Stelle kann ein anderes erscheinen. Oder die Breite und Weite seiner Bezüge verändern sich. Es kann sich als Teilproblem eines viel umfassenderen Zusammenhangs herausstellen ...« (*Kleining* 1982, S. 233). Das »Aufdecken von Anliegen und Konfliktthemen« (vgl. *Beywl* 1998, S. 230, S. 152), das Formulieren der Ziele und das Herausfinden der »eigentlichen Ziele« wird zunehmend als spezifische, oft schwierige und konfliktträchtige Aufgabe im Evaluationsprozess gesehen. Aber: Ohne solch explizite, dokumentierte Bewertungskriterien und Bewertungsmaßstäbe (kurz: Soll-Wert) ist eine »Evaluation« keine Evaluation!

Nach dem hier vertretenen Verständnis von Evaluation lässt sich damit ein weiteres **Ausschlusskriterium** formulieren: Exploration im Sinne von »mal nachschauen, was es im Untersuchungsfeld so alles gibt« – ist (noch) nicht Evaluation – auch wenn dabei riesige, netzbasierte Daten-

1.1 Zum Begriff

erhebungssysteme benutzt werden (»Jeder Student/Mitarbeiter kann jederzeit über unser Intranet seine Meinung zu ... abgeben«). Wenn nur beschrieben wird (»So ist es ...«) ohne dass Evaluatoren versuchen, begründet zu bewerten (»Was wir gefunden haben, ist gemessen an den explizierten Erwartungen/Vorgaben ... gut/schlecht!«), dann sollte nicht von Evaluation gesprochen werden (sondern eben von »Exploration«, »Untersuchung«, »Studie« o. ä.). Dies trifft auch auf den Fall zu, dass die gesammelten Daten selbst schon Wertungen enthalten: Eine Aussage wie »Ein Drittel der Befragten kreuzt die beiden linken Kästchen einer Likert-Skala an, beurteilt also die Vorlesung von Prof. X negativ« ist lediglich die Zusammenstellung von Daten, noch keine Evaluation. Nach diesem Verständnis muss der Evaluator auf Grund dieser Daten und der explizierten Bewertungskriterien und -maßstäbe dann selbst eine Bewertung vornehmen.

Dieses strenge Ausschlusskriterium findet sicherlich nicht allgemeine Zustimmung. In der Praxis hat es sich zu sehr eingebürgert, auch die Zusammenstellung wertender, erfolg- oder problemaufzählender Aussagen als »Evaluation« zu bezeichnen. Deshalb wurde oben auch im Konjunktiv formuliert: In diesen Fällen »... ‚sollte‘ nicht von Evaluation gesprochen werden«. Dennoch sei dieses Kriterium allen Evaluatoren dringend ans Herz gelegt: Es sollte **ihre** Aufgabe sein, auf Grund expliziter Kriterien wertende Aussagen zu machen! Die auf unklaren Kriterien beruhenden wertenden Aussagen von Befragten sind eine Informationsquelle, aber bedürfen dann der evaluatorischen Reflexion.

Übungsaufgabe 2:

Kreuzen Sie an, ob bei folgenden Maßnahmen das Kriterium »Vergleich Ist-/Soll-Wert« und/oder »Soll-Wert begründet« erfüllt ist:



| | Vergleich Ist-/Soll-Wert | Soll-Wert begründet |
|---|--------------------------|---------------------|
| a) Jemand blättert ein Programm betrieblicher Weiterbildung durch und meint, »es sei so nicht ausreichend.« | ☆ | ☆ |
| b) In einem von allen Teilnehmern am Kursende ausgefüllten Fragebogen wird auf einer Skala von 1 bis 5 der Lernerfolg mit dem Durchschnittswert 2,2 angegeben, das Kursklima mit 2,8. | ☆ | ☆ |

1. Evaluation – was das (nicht) ist und wozu sie nützt

| | Vergleich Ist-/Soll-Wert | Soll-Wert begründet |
|--|-----------------------------|------------------------|
| c) Anhand der Teilnehmerstatistik des Bildungswerks seines Kreises meint ein Pfarrer, 15% »unter 25-Jährige« seien viel zu wenig, das müssten mindestens 30% sein – und erklärt seiner verdutzten Bildungsworkleiterin, sie hätte erheblich das Ziel verfehlt. | ☆ | ☆ |

1.1.3 »Didaktische Funktion Praxisbezug« = Kontrolle + Steuerung + Reflexion (Definitionselement 3a + b + c)

Evaluation ist in der Pädagogik/Andragogik (und auch anderswo) kein Selbstzweck, sondern hat eine **didaktische Funktion**: Praxis soll eingeschätzt, verstanden und verbessert werden. Evaluation soll

- zeigen, was abgelaufene Maßnahmen bewirkt haben: Das ist der »Blick zurück« und die **Kontrollfunktion**. Und Evaluation soll
- helfen, kommende Maßnahmen besser zu gestalten: Das ist der »Blick nach vorn« und die didaktische **Steuerungsfunktion** von Evaluation. Außerdem hilft Evaluation,
- Bildungsmaßnahmen zu reflektieren (**Reflexionsfunktion**) und zu verstehen.

Aber: Alles gleichzeitig ist meist nicht möglich. Deshalb ist es wichtig, sich bei Evaluationsmaßnahmen von vornherein klar zu machen, was man erreichen will: Rechtfertigung nach außen oder Ergebnisdokumentation oder Korrektur von Bildungsmaßnahmen oder Schwachstellenanalyse oder Verstehen von Problemen oder ... Nur vor diesem didaktischen Hintergrund kann hilfreich entschieden werden, welche Daten man sammeln muss, wie diese Daten ausgewertet, aufbereitet und dokumentiert werden, auf Grund welcher Überlegungen und Erwartungen welche Soll-Werte formuliert werden und anhand welcher Indikatoren wie bewertet wird. Tut man dies nicht vorab, dann landet man leicht bei Daten, von denen man nicht weiß, wozu sie gut sind, und bei Bewertungen, die weh tun, aber nicht weiterhelfen.

Auch dieses dritte Element der vorgeschlagenen Definition – Evaluation erfüllt in einer konkreten Situation eine konkrete Funktion zum Verstehen und Verbessern einer konkreten Bildungsmaßnahme – findet sich vielfältig in der Literatur. In der frühen Diskussion formuliert *Stake* (1972, S. 112):

1.1 Zum Begriff

»Evaluation sollte ihre Aufgabe darin sehen, Daten für Entscheidungen zu sammeln ...«, *Cronbach* (1972, S. 41) sieht als »Ziel, Entscheidungen über ein Curriculum zu fällen«, und nach *Stufflebeam* (1972, S. 125) sollen damit »Informationen für Entscheidungsprozesse zur Verfügung gestellt werden«. *Scriven* (1986, S. 21) fordert: »Evaluation should be benefit-orientated, not explanation-orientated«. Auch in jüngster Zeit wird dieser Gegenstandsbezug betont, etwa wenn das *Joint Committee* (1999, S. 25) Evaluation als die »systematische Untersuchung der Verwendbarkeit oder Güte eines Gegenstandes« definiert, oder wenn nach *Balzer/Frey/Nenniger* (1999, S. 399) zu Evaluation eine »Nutzenbezogenheit« in Hinblick auf einen bestimmten Gegenstand gehört.

Dieser Bezug auf ein konkretes Evaluationsobjekt hat für die Arbeit eines Evaluators eine selten beschriebene, doch schnell spürbare Praxiskonsequenz: Alle Betroffenen und Beteiligten wollen möglichst schnell (Teil-, Vor-) Ergebnisse sehen. Während bei Forschungsprojekten schon einmal ein paar Jährchen ins Land gehen können, bevor die Ergebnisse vorliegen, will man in einer Evaluation am liebsten gestern wissen, wie bestimmte Dinge ankommen und was man besser tun oder lassen sollte. Auf diese an sich erfreuliche Verwertungs-Ungeduld sollten sich Evaluatoren bei ihrer Zeit- und Methodenplanung einstellen. Eine Evaluation, deren Ergebnisse erst vorliegen, wenn niemand mehr am Ergebnis interessiert ist, verfehlt ihr eigentliches Ziel: das bessere Verstehen und Gestalten einer Praxismaßnahme.

Die Forderung, dass Evaluation praxisnützlich sein soll, und die seit ca. dreißig Jahren anwachsende Erfahrung mit Evaluationen machen zunehmend bewusst, dass dieser Praxisbezug den Nutz- und Erkenntniswert von Evaluationen begrenzt. So fasst *Lee* in ihren »conclusions« über »Theories of Evaluation« zusammen: »One of the intriguing recent changes is the increasing recognition that definitive evaluation studies will often be quite limited in scope, and that specific evaluation findings are always time-limited in their usefulness, and not likely not to be easily generalized to other settings. As Scriven puts it (1991), »... yours is not the task of determining truth for the ages, but the best possible advice at the time it is needed.« (Lee 2000, S. 161). In der Abhängigkeit vom individuellen Projekt- und Zeit-Kontext liegt zugleich Stärke und Schwäche von Evaluation.

Die Orientierung auf ein konkretes Praxisobjekt hin bedeutet nicht notwendig, dass das untersuchte Objekt selbst verbessert wird: Auch eine abschließende, »summative« Evaluation, in der Entscheidungen fallen

wie »Programm wird unverändert fortgesetzt bzw. eingestellt«, ist Evaluation. Entscheidend ist, dass man eine nutzenbezogene Erkenntnis über ein bestimmtes konkretes Objekt gewinnen will; man will wissen, was das konkrete Programm X bewirkt, nicht, was alle Programme mit Elementen vom Typus X ganz allgemein bewirken.

Ausschlusskriterium

Damit lässt sich auch hier ausschließend abgrenzen: Geht es in einer Untersuchung um allgemeine Erkenntnisse ohne direkte Konsequenzen für oder über einen bestimmten untersuchten Gegenstand, dann handelt es sich nicht um Evaluation.

Übungsaufgabe 3

Prüfen Sie, welche der drei Elemente »methodisch erfasst (= ME)«, »begründet bewertet (= BB)« und »Praxisbezug (= PB)« Sie in folgenden Evaluationsdefinitionen identifizieren können. Schreiben Sie die Kürzel ME, BB und PB an die entsprechenden Textstellen.



- a) »Im Allgemeinen bedeutet Evaluation die Gewinnung von Informationen durch formale Mittel wie Kriterien, Messungen und statistische Verfahren mit dem Ziel, eine rationale Grundlage für das Fällern von Urteilen in Entscheidungssituationen zu erhalten« (*Stufflebeam 1972, S. 124*).
- b) Evaluation ist »die systematische Anwendung von Verfahren der Sozialforschung mit dem Ziel, die Konzeptualisierung und das Design, die Implementation und die Nützlichkeit sozialer Interventionsprogramme einzuschätzen« (*Rossi/Freedman 1982, S. 20*).
- c) »We define evaluation as a process for describing an evaluand and judging its merit and worth« (*Guba/Lincoln 1981, S. 303*).
- d) »Evaluation zielt auf die Verbesserung gesellschaftlicher Praxis, indem sie Programme, Maßnahmenbündel oder Materialien systematisch, d.h. methodisch angeleitet und an Gütekriterien überprüfbar, beschreibt und bewertet« (*Beywl 1998, S. 1*).

Auf zwei Besonderheiten der hier vorgeschlagenen Definition sei hingewiesen:

1. Evaluieren kann man Prozesse und Ergebnisse. Bei Prozessen lautet die Frage: »So wie das abläuft, ist das gut?« Bei Ergebnissen lautet die Frage: »Ist das, was am Ende herauskam, gut?«

2. In der Definition wird mit der Formulierung »Prozesse und Ergebnisse im Bildungsbereich« bewusst ein umfassenderer Anspruch formuliert; »Lern- oder Bildungsprozesse« wäre zu eng. Denn Evaluation bezieht sich nicht nur auf die mikrodidaktische Ebene, also das Erfassen, Bewerten und Gestalten dessen, was und wie Teilnehmer in Erwachsenenbildungsveranstaltungen lernen (was für diejenigen unter unseren Lesern von Interesse ist, die ihre Lehrveranstaltungen evaluieren wollen). Evaluieren kann man auch auf makrodidaktischer Ebene, also auf Weiterbildung bezogene Programme, Ziele, Zeit- und Kostenpläne, politische Entscheidungen, Standorte, Ausstattung ..., also auch das ganze »Drumherum«, das über Erfolg und Misserfolg von Weiterbildungs-Angeboten entscheidet (was für Leser von Interesse ist, die für organisatorisch-planerische Entscheidungen zuständig sind).

Evaluation als »politische« Tätigkeit

Die Festlegung, dass Evaluation sich auf konkrete Bildungspraxis richtet, hat Konsequenzen, die manchen Evaluator und manches Evaluationsprojekt in erhebliche Schwierigkeiten stürzte. Will Evaluation Entscheidungen vorbereiten, Aussagen über Verwendbarkeit oder Güte eines Gegenstandes machen, nutzenbezogene Informationen liefern, die für die evaluierte Bildungsmaßnahme praktische Konsequenzen haben, dann sieht sich Evaluation unausweichlich mit den Interessen der Betroffenen und Beteiligten, mit deren Machtstrukturen und deren Einflussnahme konfrontiert. Evaluation mischt sich ein, denn sie will/soll Folgen haben. Begründungszusammenhang und Verwendungszusammenhang können in diesem Verständnis nicht voneinander isoliert betrachtet werden. Dies mag erklären, warum Evaluatoren, die sich als »neutrale Untersucher«, »objektive Forscher« oder »selbstlose Helfer« verstanden und die parteiische Strukturierung ihres Untersuchungsfelds nicht hinreichend berücksichtigten, sich plötzlich mit einem Geflecht aus Anfeindungen, Boykott, Nicht-Unterstützung, Einflusversuchen und Folgenlosigkeit konfrontiert sahen, welche die Evaluation und ihre Ergebnisse obsolet machten.

Mit der Praxis- und Nutzenbezogenheit greift Evaluation einerseits in vorhandene Systeme ein, wird andererseits von den vorhandenen Akteuren und Strukturen beeinflusst. Evaluation spielt sich deshalb immer in einem »politischen« Kontext ab (vgl. dazu ausführlich *Beywl* 1998, S. 134ff., S. 203ff., *Patton* 1997, S. 341ff.). *Patton*, der konsequent eine »Utilization-Focused Evaluation« fordert und entwickelt, bietet eine freche Beschreibung der Unausweichlichkeit der politischen Dimension im Evaluationsprozess:

»Evaluation is NOT political under the following conditions:

- No one cares about the program.
- No one knows about the program.
- No money is at stake.
- No power or authority is at stake.
- And, no one in the program, making decisions about the program, or otherwise involved in, knowledgeable about, or attached to the program, is sexually active.« (Patton 1997, S. 352)

Gegenüber Forschung, die in höherem Maße zweckfrei sein kann, steht Evaluation immer in einem Spannungsfeld von Interessen; Beywl (2001, S. 151) spricht bezeichnend von »Konflikt als Regelfall der Evaluationspraxis«. »Deshalb kommt es sehr darauf an, wer eine Evaluation in Auftrag gibt, wer sie durchführt, welche Ziele von wem damit verfolgt werden, welche Vorgehensweisen gewählt und welche Methoden angewendet werden« (Stockmann 2000, S. 12). Es sei Evaluatoren dringend empfohlen, diese »Verflechtungen und Konflikte zwischen Auftraggebern, Evaluatoren, Entscheidern, Programmbetroffenen und Programmmitarbeitern – seien dies nun Individuen oder Organisationen« (Beywl 1998, S. 134) – bereits am Ausgangspunkt ihre Planung in Betracht zu ziehen – und zwar nicht als lästige Störvariable, sondern als zentrale Herausforderung einer Evaluation, die Wirkungen hervorbringen will.

1.1.4 Ausschlusskriterien und Abgrenzungen

Ein Begriff hat nur dann Trennschärfe, wenn auch klar ist, was er nicht einschließt. Eine Reihe von Ausschlusskriterien wurden bereits genannt:

- ▶ Evaluation setzt die methodische Sammlung und zugängliche Dokumentation von Daten voraus.
- ▶ Evaluation macht wertende Aussagen auf der Grundlage begründeter Soll-Wert- oder Zielbeschreibungen.
- ▶ Evaluation zielt auf das Verstehen, Bewerten und Verbessern konkreter Praxismaßnahmen.

Fehlt eines dieser Elemente, dann sollte nach dem hier vorgeschlagenen Verständnis nicht von Evaluation gesprochen werden.

Evaluation lässt sich damit auch **abgrenzen gegen Forschung**: Forschung sucht nach allgemeinen Erkenntnissen (»Welche Farben haben bei Weiterbildungsprogrammen den besten Aufmerksamkeitswert?«), Evaluation

untersucht und bewertet einen speziellen und realen Fall (»Lohnt sich der höhere Druckpreis für einen Vierfarbdruck unseres Programms?«). Forschungsfragen sind zumeist allgemeiner gestellt, sind theorieorientiert und sollen über den untersuchten Fall hinaus generelle Aussagen ermöglichen (»Wie gut ist die Lehre an deutschen Hochschulen?«). Forschung hat damit immer das Problem der Repräsentativität (»Lässt sich von der untersuchten Stichprobe auf die Grundgesamtheit schließen?«). Mit der Untersuchung eines einzigen Fachbereiches würde sich ein Forscher, der allgemein eine Aussage über Hochschullehre machen will, lächerlich machen. Anders in der Evaluation. Evaluation ist praxisorientiert-bescheidener: In einer Evaluation sind wir zufrieden, wenn wir die Lehre an dem einen, uns interessierenden Fachbereich einschätzen können. Die Forderung nach praxisbezogenem Nutzen stellt damit einen gewichtigen Unterschied zwischen Evaluation und Forschung dar: »The primary and essential difference between evaluation and fundamental research concerns utilization ... An evaluation is supposed to be used in immediate or future debate and decision-making on the problem or intervention at hand. This is not all the case with fundamental research. Fundamental researchers have no specific application in mind« (*Vedung* 2000, S. 110).

Ein weiterer klarer Unterschied zwischen Forschung und Evaluation wird darin gesehen, dass Evaluation unumgänglich (begründete) Wertungen, Beurteilungen beinhalten muss: »Während sich Grundlagenforschung weitgehend normativer Urteile enthalten kann, wird bei einer Evaluation immer eine Bewertung verlangt, sie ist Teil des Forschungsauftrags« (*Stockmann* 2000, S. 12). Forschung dagegen kann ein Ergebnis darstellen (»So ist es: ...«), und die Entscheidung über Art und Ausmaß der Nutzung dem überlassen, der etwas damit anfangen möchte.

Evaluation und Forschung gemeinsam ist das Bemühen um intersubjektive Überprüfbarkeit (»methodisch«): Auch Außenstehende sollen den Bewertungsprozess nachvollziehen können. Für beide gelten (zusammenfassend *Stockmann* 2000, S. 12) »die grundlegenden Regeln für die Sammlung valider und relevanter Daten«.

Bereits in der Frühdiskussion um Evaluation wird auf die notwendige Unterscheidung zwischen Evaluation und Forschung und auf die Probleme beim Ineinanderverfließen hingewiesen. So warnt *Glass*: »So wird manches Projekt der pädagogischen Forschung unzulänglich durchgeführt, weil man es Evaluation nennt; doch weit mehr Evaluationsuntersuchungen sind nutzlos, weil sie als pädagogische Grundlagenforschung behandelt werden« (*Glass* 1972, S. 168). Nach *Glass* besteht ein zentraler Unterschied darin,

»dass pädagogische Evaluation den **Wert**, pädagogische Forschung dagegen die wissenschaftliche **Wahrheit** einer Sache einzuschätzen versucht« (ebd. S. 169, Hervorhebung im Original). Auch neuere Ansätze betonen diese Unterscheidung. Ausführlich diskutiert *Beywl* (1998, S. 127–134) die Entwicklung »von der Logik der Forschung hin zu einer Logik der Evaluation« und arbeitet dabei sowohl Unterschiede als auch Gemeinsamkeiten heraus. *Rost* (2000, S. 131) vertritt die Position, »dass sich die wesentlichen Unterschiede zwischen Grundlagen- und Evaluationsforschung aus der Unterschiedlichkeit ihres jeweiligen Untersuchungsgegenstandes ergeben«. Für die angemessene Planung und den angemessenen Ablauf einer Untersuchung ist es damit unabdingbar, dass sich Untersucher klar sind, ob sie Evaluation oder Forschung betreiben wollen. Eine (unklare) Mischung schadet zumeist in beide Richtungen.

Abgrenzung gegen »Rückblick«

Evaluation lässt sich auf der anderen Seite abgrenzen gegen didaktische Rückmeldeformen wie »Seminarrückblick«, »Seminkritik«, »Nachbereitung«, »Feed-Back«, »Manöverkritik« u.ä.: Diese Formen stehen zumeist am Seminar-Ende. *Jagenlauf* beschreibt dies folgendermaßen:

»In der klassischen Manöverkritik steht der Aussprache nur wenig Zeit zur Verfügung, im Normalfalle die letzte Stunde oder die letzte halbe Stunde vor Ende der Veranstaltung: Der Kursleiter eröffnet die Aussprache, bittet um schonungslose Kritik. Nur wenige Teilnehmer melden sich; ehe auch die zurückhaltenderen Teilnehmer ihre Bemerkungen einbringen können, wird die Aussprache abgeschlossen. Irgendwelche Konsequenzen können nicht gezogen werden, da die Veranstaltung in der bisherigen Form nicht weitergeführt wird. Der Kursleiter bedankt sich für die Anregungen, obwohl sie – scheinbar – nutzlos sind« (Jagenlauf 1982, S. 4).

Diese didaktische Rückmeldeform unterscheidet sich von Evaluation sowohl dadurch, dass der Zugriff nicht methodisch erfolgt als auch keine »Daten« dokumentiert werden. Man weiß schon bald nicht mehr, was eigentlich gefragt war und was gesagt wurde und ob dies lediglich durch einen Wortführer erfolgte oder die Meinung aller war. Solche »rückblickende Beurteilungen« sind also weniger explizit-methodisch als dies für Evaluation verlangt wird, und ihnen fehlt das Bemühen um eine intersubjektive Überprüfbarkeit.

Allerdings: Didaktisch bedeutsam sind auch solche einfachen Rückmeldeformen, wenn sie didaktische Kontroll- und Steuerungsfunktion erfüllen.

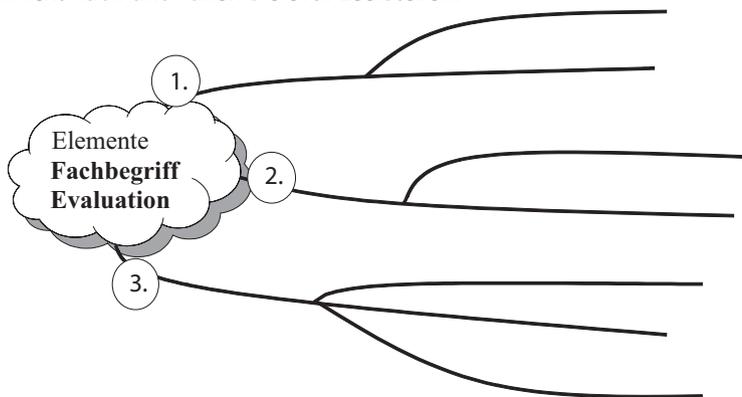
1.1 Zum Begriff

Auch wenn die am Ende einer Sitzung gestellte Fragen »Was kam heute nicht rüber? Was muss morgen nochmals bearbeitet werden?« keine Evaluation darstellen (weil sie nicht explizit-methodisch angelegt sind und weder die Daten noch die Bewertungskriterien dokumentieren), so sind sie dennoch wichtig für die Steuerung des Lernprozesses. Aufwendige Evaluationsverfahren wie Fragebogen- oder Interviewtechniken wären hier überzogen.

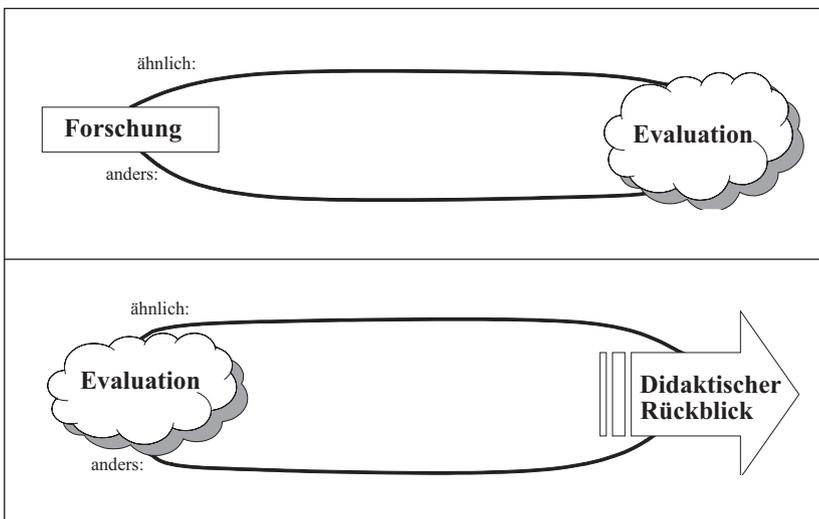
Solche rückblickenden Beurteilungen ohne Evaluationsanspruch sind auch bei makrodidaktisch-bildungsorganisatorischen Maßnahmen wichtig, z. B.: Am Halbjahresende äußert in der Kursleiterbesprechung jeder seine Meinung darüber, ob die Raumverteilung angemessen war. Das ist noch keine Evaluation, selbst wenn daraus wichtige Konsequenzen für das nächste Halbjahr gezogen werden: Es fehlt sowohl die nachvollziehbare Datenerfassung (»methodisch«) als auch die später nachlesbare Dokumentation; auch die Kriterien für die Entscheidung bleiben spontan-subjektiv. Dies ist keine Abwertung: Für die Steuerung ständig verbesserter Bildungsarbeit sind auch und gerade solche mit leichter Hand durchgeführten rück- und vorausblickenden Kleinbewertungen unabdingbar.

Übungsaufgabe 4:

Versuchen Sie, die in diesem Teil 1.1 genannten Definitionselemente in folgende Mind-Map einzutragen (Mind-Maps sind gezeichnete Strukturen, in die man die entscheidenden Stichworte einträgt; sie erlauben einen schnellen Überblick und wirken merkfördernd). Tragen Sie zunächst ein, was Ihnen spontan im Gedächtnis geblieben ist bzw. was Sie bereits wussten; lesen Sie dann im Text nach und füllen Sie evtl. Leerstellen.



1. Evaluation – was das (nicht) ist und wozu sie nützt



1.2 Wozu Evaluation?

Übungsaufgabe 5

Schreiben Sie drei Gründe auf, warum Sie keine Lust hätten, dass Ihre Arbeit in der Weiterbildung evaluiert würde:



1.
2.
3.

Übungsaufgabe 6

Schreiben Sie drei Vorteile auf, die Sie sich von einer Evaluation Ihrer Arbeit in der Weiterbildung erhoffen würden:



1.
2.
3.

Stichworte wie »Kontrolle«, »Bewertung« oder »Erfolgsmessung« lösen auch bei Erwachsenen oft alte Schulängste mit der Erinnerung an Fremdbeurteilung, Zensierung, Zeugnisse und Versagen, erlittene Ungerechtigkeit, demütigende Strafen und Konflikte aus. Niemand (Teilnehmer/Kursleiter/Programm-Macher ...) mag es, wenn seine eigene Leistung evaluiert wird. Dies mag ein Grund dafür sein, warum solche Ergebniskontrollen in der Erwachsenenbildung freiwillig so selten durchgeführt werden.

Bei einer Befragung von Kursleitern mit der obigen Frage, warum jemand keine Lust hätte, dass seine/ihre Weiterbildungsarbeit evaluiert würde, wurden spontan folgende **Gründe** genannt:

*Der am meisten genannte Grund ist **Angst vor einem schlechten Ergebnis**, dicht gefolgt von **Angst vor Kritik** (z. B. »Ich werde hinterfragt«, »Angst vor Zerpflücktwerden«). Und eine dritte Angst wird genannt: **Angst vor Konsequenzen** (z. B. »Der Mist wird öffentlich«, »mehr Vorbereitung«, »keine Weiterbeschäftigung«). Befürchtet wird ein **fremder Maßstab** (z. B. »Evaluierer hat andere Ansprüche als ich«, »Wer bekommt das Ergebnis zu sehen?«) und **zu viel Arbeitsaufwand**. Es wurde auch darauf verwiesen, dass Evaluation im Kurs stört (z. B. »Evaluation beeinträchtigt Teilnehmer/Kursgeschehen«). Schließlich gab es grundsätzliche Zweifel: Evaluation **nutzt nichts und ist unnötig** (z. B. »Ich brauch das nicht! Ich bin gut!!«). Auch wurden Zweifel angemeldet: »Geht das denn? (Seidenmalkurs)«*

Vergleichen Sie mit den von Ihnen in Übungsaufgabe 5 genannten Gründen – haben Sie ähnlich oder anders argumentiert?

Information über Erreichtes und Nichterreichtes muss nicht Sortier- und Disziplinierungsfunktion haben. Bereits Anfang der 70er Jahre wurde in der Schulpädagogik darauf hingewiesen, dass Information über Erreichtes und Nichterreichtes Voraussetzung ist, pädagogische Ideen auf ihre Tragfähigkeit zu überprüfen und Maßnahmen gezielt zu verbessern (z. B. *Ingenkamp* 1971, *Reischmann* 1974). Mit dem 1972 erschienenen Sammelband »Evaluation« führte *Christoph Wulf* diesen Fachbegriff für die breite Diskussion weiter, insbesondere indem er die auch heute noch grundlegende und lesenswerte amerikanische Diskussion zugänglich machte. Eine höchst lesenswerte Zusammenfassung der amerikanischen Entwicklungsgeschichte der Evaluation findet sich bei *Beywl* (1998, S. 14–43). Auf dem Prüfstand steht bei der Evaluation die pädagogisch/ andragogische Maßnahme (»erfolgreicher Kurs«, »zielgruppengemäße Werbung«, »bedarfsgerechte Planung« ...), nicht der Teilnehmer, Kursleiter, Planer.

1. Evaluation – was das (nicht) ist und wozu sie nützt

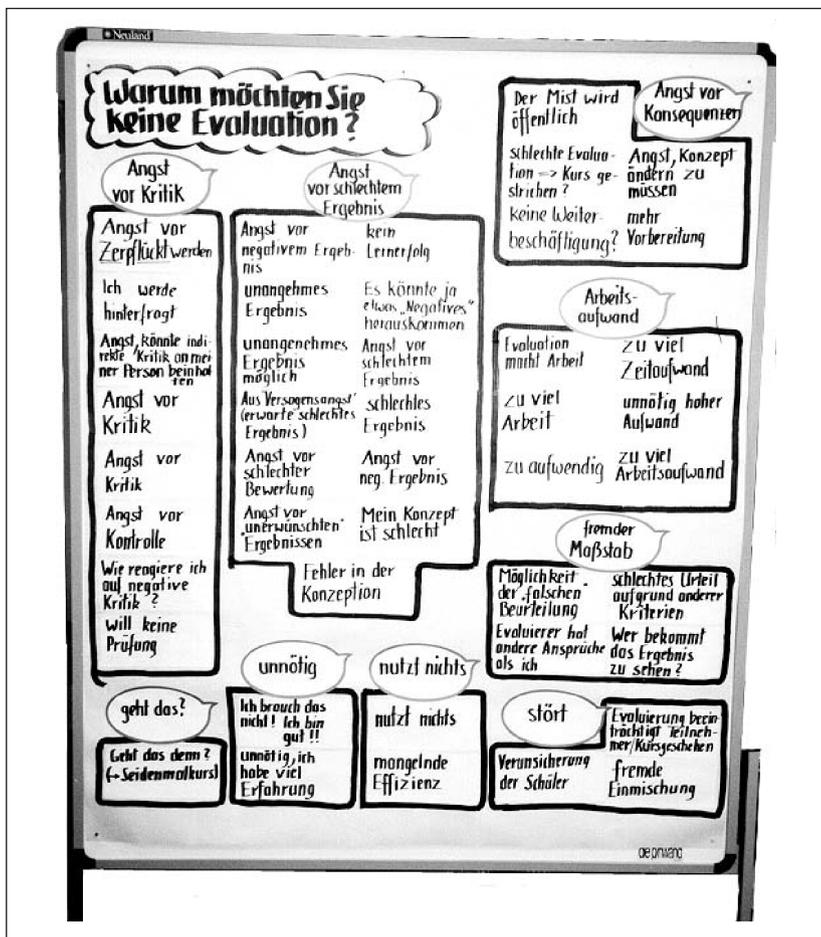


Abb. 1: Einwände gegenüber Evaluationsvorhaben

Dahinter steht ein Regelkreismodell: Eine Maßnahme wird geplant und durchgeführt, die Ergebnisse werden kontrolliert und mit den daraus gewonnenen Einsichten die verbesserte Fortsetzung oder nächste Maßnahme geplant. Ohne eine Ergebniskontrolle bliebe offen, ob die Maßnahme die gewünschten Wirkungen bringt, und ob/wie/wo Korrekturen nötig sind.

Bei einer Befragung von Kursleitern, welche Vorteile sie sich von einer Evaluation der Arbeit in der Weiterbildung erhoffen würden, wurden folgende Gründe genannt:

Am meisten erwartet werden **Verbesserungsvorschläge** (z. B. »besserer Lernerfolg«, »Verbesserung des Kurses«, »Anhaltspunkte, um den Kurs besser zu gestalten«). Die formative Chance (formativ/summativ wird gleich erklärt) der Evaluation steht hierbei im Vordergrund. Auf die summative Funktion verweisen Begründungen, die sich von Evaluation **Lob oder Anerkennung** versprechen (z. B. »Alle sehen, wie gut ich bin«, »Auf diesem Wege positives Feed-back erhalten«, »meine gute Arbeit öffentlich machen«), verbunden mit dem Vorteil, dass eine **höhere Objektivität** der Beurteilung gegeben ist (»objektivere Ergebnisse bei Fremdevaluation«, »dokumentierte Ergebnissicherung«). Zu diesen summativen Argumenten gehört auch die **Rückmeldefunktion** (»Wissen darüber, ob mein Kurs ankommt!«, »Wissen um Stärken und Schwächen des Programms«). Schließlich wird auch genannt, dass Evaluation **didaktische Reflexionshilfe** ist (z. B. »Möglichkeit, meine Arbeit zu hinterfragen«, »Erkennen eingeschliffener Fehler«) und damit »neue Anregungen, neue Ideen« bringen kann.

Vergleichen Sie mit den von Ihnen in Übungsaufgabe 6 genannten Gründen – überschneiden sich Ihre Erwartungen mit den hier genannten?

Evaluation kann aus zwei bzw. drei grundsätzlich unterschiedlichen Motiven durchgeführt werden (vgl. *Scriven* 1972, S. 62):

Summative Funktion

- ▶ Entweder wird damit am Ende einer Maßnahme der Erfolg aufsummiert, gewissermaßen als Ergebnis, Schluss-Strich und Endurteil, auf das nichts mehr folgt: »Das wurde (nicht) geleistet. Und deshalb war der Kurs/der Teilnehmer/der Kursleiter (nicht) gut.« Dies wird als summative Evaluation bezeichnet; sie dient »zum besseren **Verständnis** und zur besseren **Verwendung** des Programms« (*Scriven* 1972, S. 62. Hervorhebung im Original).

Formative Funktion

- ▶ Oder die Messung dient im Lernprozess als Informationsquelle für die effektive Formung des Weiterlernens: »Hier stehen wir jetzt. Das bedeutet, dass folgende Maßnahmen ergriffen werden müssen, um möglichst viele Teilnehmer möglichst weit zu fördern ...«. Dies wird als formative Evaluation bezeichnet.

Reflexive Funktion

- ▶ Neben diesen beiden direkten Absichten gibt es eine eher indirekte: Evaluation ist in didaktischer Hinsicht auch eine Denk- oder Reflexi-

onshilfe. Unabhängig davon, ob man formativ oder summativ zu mehr oder weniger bemerkenswerten Ergebnissen kommt, hilft Evaluation, didaktische Maßnahmen aus einem ungewöhnlichen Winkel zu betrachten. Dies öffnet die Augen für didaktische Zusammenhänge, die man anders kaum in den Blick bekommt.

Die Überprüfung, ob Maßnahmen in der Weiterbildung den ihnen zgedachten Erfolg haben, ist sowohl für den Lerner, den Kursleiter und Organisator als auch für die Weiterbildungs-Institutionen und die Öffentlichkeit von Interesse:

- ▶ Den **Lernern** wird der Erfolg rückgemeldet bzw. es werden Hinweise zur Weiterarbeit gegeben.
- ▶ Der **Kursleiter** bzw. der **Organisator/Veranstalter** kann didaktische Entscheidungen begründet fällen.
- ▶ Der **Institution** und **Öffentlichkeit** gegenüber kann gerechtfertigt werden, wozu bestimmte Investitionen gemacht werden.
- ▶ Außerdem werden für **Wissenschaft** und **Forschung** Datengrundlagen und Instrumente geschaffen.

Verzichtet man auf Erfolgskontrollen, dann ergeben sich sowohl praktische als auch theoretische Gefahren:

- ▶ Für die **Praxis** fehlt ein Maßstab, an dem und mit dem Bildungsangebote verbessert und weiterentwickelt werden; außerdem wird, wer nicht nachweist, was seine Tätigkeit bringt, in weniger bildungsfreundlichen Zeiten Schwierigkeiten haben, sein Angebot zu rechtfertigen.
- ▶ **Theorien** bleiben ohne empirische Kontrolle kaum mehr als wohlklingende und wohlmeinende Überzeugungen; Erwachsenenpädagogik bzw. Andragogik als Wissenschaft von der Bildung Erwachsener tut sich damit schwer, den Geruch der intuitiven Kunst loszuwerden.

Übungsaufgabe 7

Welche Gründe erscheinen Ihnen besonders einleuchtend für die Durchführung von Evaluation in der Weiterbildung?



.....

.....

.....

1.3 Was evaluieren? – Vermeiden Sie die »Indikatorenschaukel«!

Vergleichen Sie mit Ihrer Auflistung oben (warum keine Lust/was erhoffen). Hat sich etwas geändert? Ja? Dann herzlichen Glückwunsch! Nein? Schön, dann sind Sie schon mit gutem Vorwissen eingestiegen!

1.3 Was evaluieren? – Vermeiden Sie die »Indikatorenschaukel«!

Evaluieren kann man alles mögliche, z. B.:

- ▶ Programme: Ist unser Programm ansprechend? Haben wir genügend und die richtigen Computerkurse?
- ▶ Teilnahme: Wie viele TeilnehmerInnen kommen, wie viele sollten kommen?
- ▶ Kurserfolg: Sind unsere TeilnehmerInnen zufrieden? Lernen sie etwas in der Veranstaltung?
- ▶ Zielgruppenerreichung: Kommen die »richtigen« Leute in unser Bildungsprogramm?
- ▶ Inhalte: Bieten wir das an, was wir anbieten wollen/sollen/müssen?
- ▶ Kosten: Welche direkten/indirekten Kosten entstehen durch Angebot X? Sind diese vertretbar?
- ▶ Nutzen: Senkt die Schulung X den Ausschuss in der Produktion? Wie viele Arbeitslose werden nach unserer Qualifizierungsmaßnahme vermittelt? Wie ist diese Vermittlungsquote zu bewerten?
- ▶ Schwachstellen: Welche Probleme gefährden die Qualität unserer Angebote?

Gewarnt werden muss vor allumfassenden Fragestellungen. »Ist unsere Mensa gut?« ist keine evaluierbare Fragestellung! Denn was heißt »gut« für wen? Ist damit schmackhaftes Essen gemeint (nach wessen Geschmack)? Oder große Portionen? Oder möglichst billig – und meint das billig im Einkauf für das Studentenwerk oder billig im Verkauf für den Studenten? Oder Rauchverbot (was halten die Raucher davon)? Oder vitaminreich? Oder Parkplätze? Oder ein Mix aus alledem? Und wer fängt mit dieser Mix-Aussage dann was an?

Evaluatoren sollten sich nicht auf vage und globale Evaluationsaufgaben einlassen, die alles und jedes erfragen möchten! »Ist unser Programm gut?« – »Sind unsere Kursleiter Ihr Geld wert?« – »Sind unsere Teilnehmer hinterher bessere Demokraten?« – dies sind Anfangsanfragen, die erheblich enger gefasst werden müssen, ehe sie einer Evaluation zugeführt werden können. Solch vage und globale Anfangsanfragen sind notwendige Start-

punkte und damit ein erster Schritt zur Evaluation. Man muss aber noch eine ganze Weile am Schreibtisch weiterfragen und weiterdenken, bevor daraus eine evaluierbare Fragestellung wird.

Eine ausführliche Darstellung, wie man gewünschte Ergebnisse möglichst präzise und überprüfbar formuliert, wird Kapitel 9 anbieten. Hier soll schon vorab darauf hingewiesen werden, dass es entscheidend wichtig ist, **vorher und schriftlich festzulegen, was evaluiert werden soll**, welche Wirkung bzw. welches Ergebnis der evaluierten Maßnahme interessiert und welche Indikatoren als Kriterium herangezogen werden sollen. Insbesondere wenn man einen Auftraggeber hat, dem man über die Evaluation berichten möchte/ muss, dann empfiehlt es sich, vorher und schriftlich (!) gemeinsam abzuklären, welche erwarteten Wirkungen bzw. Ergebnisse sich hinter den Anfangsfragen verstecken (dies wird zumeist unter dem Stichwort »Kontrakt« diskutiert – z. B. zusammenfassend *Vedung* 2000, S. 113f.). Dabei sollte man sich auch einigen, ob die ausgewählten Kriterien/Indikatoren auch wirklich das beschreiben, was als Erfolg akzeptiert wird. Bei diesem Abklären der Ziele und Kriterien werden zu Beginn eines Evaluationsprozesses in einem ersten Schritt zumeist anfangs eher vage Zielbeschreibungen formuliert. Zur Vorbereitung (und dann natürlich erst recht zur Durchführung) einer Evaluation braucht man dann aber auch den zweiten Schritt der Präzisierung der erwarteten Wirkungen und den dritten Schritt der Einigung auf Indikatoren, aus denen man auf den Erfolg schließt.

Drei Schritte zur Präzisierung

In dieser Hinsicht ist Evaluation der Forschung ähnlich: Bevor die eigentliche Untersuchung beginnt, legt man fest,

1. was man untersuchen will (Zielbeschreibung),
2. was man als erwünschte oder befürchtete Wirkungen feststellen will und
3. welches die Kriterien und Indikatoren sind, die man erfassen kann und an denen man den Erfolg messen will (wobei man sowohl bei Forschung als auch bei Evaluation durchaus ein offenes Auge und Ohr dafür hat, wenn sich im Verlauf der Arbeit neue Erkenntnisse ergeben, die eine Änderung/Korrektur/Ergänzung der vorgefertigten Untersuchung erfordern).

Kriterien und Indikatoren

Zu dieser Vorab-Klärung gehört also auch die Festlegung, woran man erkennt, ob das gewünschte Ergebnis erreicht wurde. Kriterien und Indikatoren können z. B. Teilnehmerzahlen sein oder Kursthemen, Leistungs- oder Einstellungsergebnisse, Klima-Fragebogen, Zufriedenheitsschätzungen, produzierte Werkstücke oder Ausschusszahlen, Krankheitstage, Durchführungskosten, geschätzter Schaden bei Nicht-Schulung ...

Diese drei Schritte zunehmender Präzisierung der Evaluationsabsicht könnten im konkreten Fall so aussehen:

Beispiel

Ein Kursleiter fragt sich: »Ist mein Kurs gut?« Diese im ersten Schritt formulierte Beschreibung des Untersuchungsziels (»Ziel ist, etwas über meinen Kurs zu erfahren«) ist noch vage-allumfassend. Im zweiten Schritt präzisiert er deshalb, indem er als erwünschte Wirkung beschreibt: »Möglichst viele Teilnehmer sollen aktiv sein!« Als Indikator, ob diese erreicht ist oder nicht, formuliert er (dritter Schritt): »Der Kursleiter beansprucht weniger als 50% der Redezeit, zwei Drittel der Teilnehmer sagen etwas pro Stunde.« An diesen Indikatoren kann er nun erfassen und bewerten (erkennen Sie – siehe 1.1 – die Definitionselemente von »Evaluation«?), ob sein Kurs »gut« ist.

Nun kann man natürlich einwenden, dass die Redezeit nicht der entscheidende Indikator für einen »guten Kurs« ist! Genau darum geht es: Mit der Aufgabe, vorher und schriftlich festzulegen, was als Erfolg oder »Soll-Wert« akzeptiert würde, zwingt man die an der Evaluation Beteiligten zur Reflexion und Offenlegung der eigentlichen Ziele und Erwartungen. Zu diesem Zeitpunkt kann man in der Planung einer Evaluation durchaus noch andere Kriterien und Indikatoren vereinbaren – oder sich entscheiden, dass für dieses Mal die Ziele, Kriterien und Indikatoren akzeptabel sind und bei einer späteren Evaluation dann eben anderes untersucht werden soll.

Warnung: Indikatorenschaukel!

Wenn vorher kein Konsens über die als Erfolg geltenden Wirkungen und Indikatoren geschaffen wird, dann landet man hinterher schnell in der »Indikatorenschaukel«: Der Evaluator hat A untersucht, aber – so wird ihm vorgehalten – B wäre doch viel wichtiger gewesen. Denn »Wirkung von Weiterbildung« kann vieles sein:

1. Evaluation – was das (nicht) ist und wozu sie nützt

- ▶ dass ein Programm angeboten und auch besucht wird, (Aber kommen auch die »richtigen« Leute? Und genug? Und was wäre »genug«?)
- ▶ dass die Teilnehmer ein Angebot gut und hilfreich finden und wiederkommen, (Aber haben sie wirklich etwas gelernt, oder pflegen sie nur ihre sozialen Beziehung?)
- ▶ dass der Ausschuss in der Produktion sinkt, (Aber das ist doch keine Menschenbildung!)
- ▶ dass jemand sich in seinem Leben bereichert fühlt, (Aber dafür lassen sich doch keine öffentlichen Subventionen rechtfertigen!)

Werden Weiterbildungs-Wirkungen festgestellt, dann ist es ein leichtes Spiel, zu reklamieren, dass man doch ganz andere Indikatoren hätte heranziehen müssen. Mit dieser Indikatorenschaukel kann jede Wirkung ad absurdum geführt werden. Sichern Sie sich ab, indem Sie vorher und schriftlich festhalten: Was wird als Erfolgskriterien für den untersuchten Gegenstandsbereich akzeptiert?

Spätestens an dieser Stelle werden Evaluatoren ausführlich Erfahrung sammeln mit dem Power-Play der Betroffenen und Beteiligten, das oben unter dem Stichwort »Evaluation als politische Tätigkeit« angekündigt wurde. Das »Joint Committee ...« empfiehlt in seinen Standards (D2 Politische Tragfähigkeit): »Evaluationen sollten mit Voraussicht auf die unterschiedlichen Positionen der verschiedenen Interessengruppen geplant und durchgeführt werden, um deren Kooperation zu erreichen und um mögliche Versuche irgendeiner dieser Gruppen zu vermeiden, die Evaluationsaktivitäten einzuschränken oder die Ergebnisse zu verzerren respektive zu missbrauchen« (*Joint Committee ... 1999, S. 95*). Geduldiges Zuhören, sensibles Agieren, Bereitschaft, in mehreren Überarbeitungen Ziele aufzustellen, zu revidieren, zu erklären und erklären zu lassen und zwischen den Betroffenen zu vermitteln entscheiden bereits an dieser Stelle über die Durchführbarkeit und Nützlichkeit der Evaluation.

Übungsaufgabe 8

Dieser Text zeigte an einem Beispiel drei Schritte, durch die Sie vorher und schriftlich präzisieren, was Sie evaluieren wollen. Suchen Sie je ein Beispiel aus Ihrem eigenen Weiterbildungsbereich!



1. Ebene »Zielbeschreibung«: Was/welcher Bereich soll evaluiert werden?
Textbeispiel: Ist mein Kurs gut?

Ihr Beispiel Zielbeschreibung:

1.3 Was evaluieren? – Vermeiden Sie die »Indikatorenschaukel«!

2. Ebene »erwartete Ergebnisse, Wirkungen«: Welche Wirkungen oder Ergebnisse würde man akzeptieren, um den in 1. beschriebenen Sachverhalt angemessen zu überprüfen?

Textbeispiel: Teilnehmer sind aktiv.

Als Erfolg akzeptierte Wirkung(en) Ihrer Zielbeschreibung:

.....

3. Ebene »Indikatoren«: Auf welche Indikatoren, an denen man die Wirkung ablesen kann, kann man sich einigen?

Textbeispiel: Kursleiter beansprucht weniger als 50% Redezeit, zwei Drittel der Teilnehmer sagen etwas pro Stunde ...

Indikatoren für Ihr Beispiel:

.....

Meistens stellt man nach dieser den Evaluationsrahmen klärenden Vorarbeit zwei Dinge fest:

1. Eine Reihe der Evaluationsfragen sind bereits durch den Versuch einer treffenden Formulierung und genaueren Eingrenzung klarer geworden. Man erkennt, dass manche Fragen unlösbar, andere längst beantwortet sind, wieder andere zu komplex sind und deshalb in eine Reihe einfacherer Fragen zerlegt werden müssen. Das Bemühen um Evaluation nutzt also schon ehe man anfängt zu evaluieren. Dies ist die in 1.2 beschriebene Denk- oder Reflexionshilfe der Evaluation.

Empfehlung:

Halten Sie diesen Klärungsprozess in Protokollen wenigstens in Stichworten fest. Sie verfügen damit bereits über Unterlagen für den späteren Evaluationsbericht.

Sie können auf diese Weise zeigen, von wo man ausgegangen ist und was der Startpunkt der Überlegungen war (»Vorher-Nachher-Vergleich«). Meist hat man das nämlich schnell vergessen und merkt damit nicht mehr, welchen Erkenntnisgewinn man von diesem Nullpunkt aus bereits erarbeitet hat.

2. Im Prozess der Erarbeitung der Evaluationsfrage wird man in der Regel deutlich bescheidener in der Fragestellung. Global-vage Fragen werden konkretisiert und präzisiert und damit beantwortbar gemacht – um den Preis, dass sie enger geworden sind.

1.4 Zusammenfassung

Versuchen Sie, dieses erste Kapitel anhand folgender Fragen zusammenzufassen. Die Lösungen finden Sie auf den angegebenen Seiten dieses Buches:

Übungsaufgabe 9:



- a. Definieren Sie »Evaluation« (S. 18):

Evaluation meint

1.....
und 2.....
zum 3.

- b. Was ist bei »Forschung« und »Evaluation« ähnlich, was ist unterschiedlich? (S. 28f.)

ähnlich:
anders:

- c. Worin ähneln sich »Evaluation« und »Seminarrückblick«, worin unterscheiden sie sich? (S. 30f.)

ähnlich:
anders:

- d. Wozu dient Evaluation? (S. 35)

1.:
2.:
3.:

- e. Zählen Sie ein halbes Dutzend Indikatoren auf, die man als Kriterien zur Evaluation heranziehen kann! (S. 37)

.....
.....
.....
.....
.....
.....

f. Was meint »Indikatorenschaukel?« (S. 37)

.....
.....

Damit sind wir am Ende von Kapitel 1 angelangt. Gehen Sie jetzt noch einmal zu den »Zielen von Teil I« auf S. 17. Tragen Sie bei den ersten fünf Zielen (mit diesen befasste sich Kapitel 1) auf der zweiten Linie eine geschätzte Prozentzahl ein, wo Sie jetzt stehen.

Und als Schlussüberlegung: War das Erarbeiten der Zusammenfassungsaufgaben und die Schätzung Ihrer Punktzahl eine »Evaluation Ihres Lernerfolgs?«

2. Wie man es macht – probieren Sie selbst!

Machen lernt man durch Machen. Über eine Handlungskompetenz, wie zum Beispiel »eine Evaluation durchführen« kann man lange lesen, reflektieren, viel erklären und noch mehr kritisch hinterfragen: Aber ob Sie diese Kompetenz beherrschen, werden Sie erst dann wissen, wenn Sie es ausprobiert haben. Und auch was Evaluation bringt und was sie nicht bringt, lässt sich dann erheblich besser einschätzen, als wenn man nur darüber nachdenkt. Aus diesem Grunde gehen wir jetzt von der theoretischen Vorrede weg zum Ausprobieren, zum »Machen«.

Mit diesem Voranstellen der konkreten Handlung und des praktischen Erprobens vor die vertiefte theoretische Auseinandersetzung folgen wir alten und neuen pädagogischen Konzepten: Schon *Comenius* verweist in seiner *Didactica Magna* 1645 darauf, dass man Denkwege konkret gegangen sein muss, bevor man abstrakte und formale abkürzende Wege gehen kann: »Sofort als möglich ziehe man die sinnliche Wahrnehmung zu, damit alles sich leichter einprägt«. *Pestalozzi* formuliert, dass »Anschauung das Fundament jeder Erkenntnis« ist. Der Symbolische Interaktionismus (*George Herbert Mead*) verweist darauf, dass die Fähigkeit der Interaktion mittels Symbolen die konkrete Interaktionserfahrung voraussetzt. Und *Piaget* hat gezeigt, welche Bedeutung die konkrete Operation für den späteren Aufbau von formalen Operationen hat. Deshalb werden in diesem Kapitel vor weiteren theoretischen Überlegungen ausgewählte Beispiele für Evaluationshandeln vorgestellt. Die Beispiele wurden einerseits als Aufforderung zur Umsetzung in die Praxis ausgewählt, andererseits kann man besser über etwas nachdenken, was man vorher auch konkret erprobt hat.

Bei der konkreten Durchführung werden Sie bald merken, dass der Teufel im Detail steckt. Sie werden auch merken, dass Sie sich oft für das weniger Schlechte entscheiden müssen, weil Sie, wenn Sie alles optimal machen wollen, nie zum Handeln kommen. Nur der Theoretiker kann alles richtig und optimal, weil er es nicht machen, sondern nur mit schönen Worten beschreiben muss. Sie werden bei diesen Erprobungen auch merken, dass man sich bei beschränkten Mitteln und begrenzter Zeit auf einen Hauptzweck konzentrieren muss; darin liegen auch die Grenzen der Verwertbarkeit jeder Evaluation.

All dies kann man nur durch praktische Erfahrung erwerben. Es genügt also nicht, die folgenden Beispiele durchzulesen und durchzudenken; versuchen Sie, den einen oder anderen der folgenden Vorschläge in der

Praxis **auszuprobieren**. Auf diesem Weg nähern wir uns dem Ziel dieses Lehrbuchs: Dass Sie Evaluation »machen«. Und dass Sie beschreiben können, wo die Schwächen dieses Machens liegen, und inwieweit diese Schwächen die Aussagekraft der Ergebnisse einschränken.

Auf eine fast durchgehende Schwäche der folgenden Beispiele soll vorab hingewiesen werden: Zwar werden Daten einigermaßen methodisch gesammelt und dokumentiert, aber die Bewertung erfolgt oft leichtgewichtigsituativ-undokumentiert. Hier scheint der größte Lernbedarf zu liegen.

Die Beispiele 2.1 bis 2.3 beschreiben Evaluationsvorhaben auf mikrodidaktischer Ebene, also in Bezug auf die unmittelbare Kursdurchführung, die Beispiele 2.4 bis 2.6 auf makrodidaktischer Ebene, also der organisatorischen Planung und Kontrolle.

2.1 Drei Kurzformen zur Kursevaluation

2.1.1 Qualitative Bewertung von Kurselementen durch Kartenabfrage

Intention: Sie wollen am Kursende oder einem Zwischenabschnitt schnell Rückmeldung über die Beurteilung bestimmter Kurselemente (z. B. Dozent, Inhalt, Klima).

Adressaten der Evaluation: Teilnehmer, Kursleiter, Auftraggeber.

Zeitaufwand: 10 Minuten zur Durchführung, 20–30 Minuten zur Besprechung.

Voraussetzungen: keine. Geeignet bis ca. 12 Teilnehmer.

Material: Pinnwand, Kärtchen, dicke Filzschreiber.

Durchführung: Am Kursende (ca. 40 Minuten vor Schluss) teilen Sie rote und grüne Kärtchen aus und stellen eine Pinnwand zur Verfügung, auf der ein großer Kreis in drei Segmente aufgeteilt ist. Überschrift: »Wie fanden Sie ...«. In den drei Kreissegmenten steht »Dozent«, »Inhalt«, »Klima« (oder auch »Praxisbezug«, »Methode« – was Sie eben erfassen möchten). Bitten Sie die Teilnehmer, jeder möge einzeln auf grüne Kärtchen positive, auf rote Kärtchen kritische Aussagen zu jedem der drei Segmente schreiben (pro Karte nur eine inhaltliche Aussage, groß schreiben, dass von allen lesbar) und auf dem jeweiligen Segment anpinnen. Zeit zum Schreiben und Anpinnen 10 Minuten.

Lassen Sie die Teilnehmer dann interpretieren, was herausgekommen ist (z. B. rot/grün-Verteilung, wo häufen sich Aussagen ...). Gemeinsam

2. Wie man es macht – probieren Sie selbst!

überlegen, was dies für jeden (nicht nur den Kursleiter) bedeutet, für Kursfortsetzung (formativ) oder zukünftige Kurse. Nicht verteidigen, erläutern, abwehren, sondern als Feedback einfach stehen lassen. Und nicht nur das Negative anschauen, sondern sich auch über das Positive gemeinsam freuen!

Didaktische Reflexion: Hat auch die Funktion eines Abschlussrituals.

Dokumentation: Foto der Pinwand oder (insbesondere wenn zu klein geschrieben) Karten abschreiben.

Kommentar: Ein eher leichtgewichtiges Verfahren, aber schnell und ohne Vorbereitung anwendbar. Problem ist die Theorielosigkeit: Man weiß eigentlich nicht, warum man etwas wissen möchte und was man überhaupt herausbekommen will – es fehlt ein Soll-Wert, an dem man sich orientieren könnte. Lediglich bei der Auswahl der drei »Segmenttitel« müsste man sich überlegen, welche man auswählt, aber noch nicht einmal das ist notwendig. Deshalb bleiben solche Verfahren in der Regel blass: Wenn man sich vorher keine Gedanken darüber macht, was man wissen will, weiß man hinterher meist nicht, was man eigentlich weiß, und wozu das nützt. Dennoch: Erproben Sie diese Form; mit den dokumentierten Aussagen können Sie später mit einiger Plausibilität zeigen, dass bestimmte Ziele (nicht) erreicht wurden und dass bestimmte Kritik (nicht) geäußert wurde.

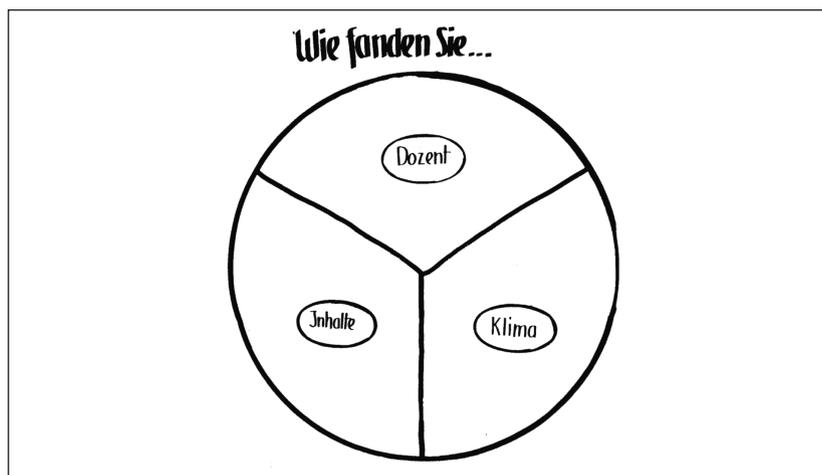


Abb.2: Kartenabfrage als qualitative Kursevaluation

Allerdings: Das Foto von der Pinwand ist lediglich eine Sammlung von Aussagen, eine Dokumentation von Daten, noch keine Kursevaluation. Für eine Evaluation müsste auch noch eine wertende Beurteilung darüber geschrieben werden, was dieses Foto bzw. diese Daten bedeuten: Mit welcher Plausibilität zeigen sie, dass bestimmte Ziele (nicht) erreicht wurden und dass bestimmte Kritik (nicht) geäußert wurde (richtig: das ist die gleiche Formulierung wie oben. Aber: jetzt handelt es sich nicht um eine spontan-flüchtige Meinungsäußerung, sondern um eine begründete und dokumentierte Bewertung, bei der auch argumentiert wurde, warum ein bestimmter Soll-Wert als Maßstab genommen wurde). »Hier sind die Daten, mach Dir selbst Deinen Reim!« ist keine Evaluation.

2.1.2 Quantitative summative »Thermometer«-Rückmeldung

Intention: Sie wollen am Ende eines Seminars eine schnelle und einfache Rückmeldung darüber, welche Ziele die Teilnehmer in welchem Ausmaß glauben erreicht zu haben.

Adressaten der Evaluation: Teilnehmer, Kursleiter, Institution.

Zeitaufwand: Durchführung am Kursende 5 Minuten, dazu kommt das Auswertungsgespräch.

Voraussetzung: Sie haben am Kursanfang die Ziele Ihres Kurses definiert.

Material: Die Ziele werden auf Karten geschrieben, unter jede Karte ein »Thermometer« (mit dickem Filzstift auf Papierstreifen gemalt, siehe Abbildung 3) geklebt. Für jeden Teilnehmer so viele Klebepunkte wie Zielkarten.

Durchführung: Die Ziele samt Thermometer werden an der Wand befestigt, jeder Teilnehmer erhält Klebepunkte entsprechend der Anzahl der Ziele. Anweisung: *»Wir sind am Ende dieses Kurses angelangt. Die Ziele, die in diesem Kurs erreicht werden sollten, sehen Sie an der Wand. Sie sollen mit einem Klebepunkt markieren, inwieweit Sie glauben, jedes dieser Teilziele erreicht zu haben: Wenn Sie glauben, es voll zu beherrschen, dann kleben Sie Ihren Punkt ganz oben, wenn Sie glauben, es halb zu beherrschen, in der Mitte usw.«*

Auswertung: Unmittelbar mit den Teilnehmern, die interpretieren, was ihnen die gepunkteten Thermometer sagen.

Dokumentation: Da die Thermometer zu sperrig sind, um sie direkt zu dokumentieren, ist es am einfachsten, sie zu fotografieren. Am besten

2. Wie man es macht – probieren Sie selbst!

verwendet man helle Karten/Streifen und dunkle Klebpunkte. Diese Fotos lassen sich – beispielsweise für ein Fotoprotokoll – am Kopierer für die Teilnehmer vervielfältigen. Oder verkleinern Sie die Karten/Streifen über den Kopierer. Man kann die Klebpunkte auch in Prozentzahlen umrechnen und damit Mittelwerte bilden; dabei geht aber die unmittelbare Anschaulichkeit vor allem der Streuung verloren.

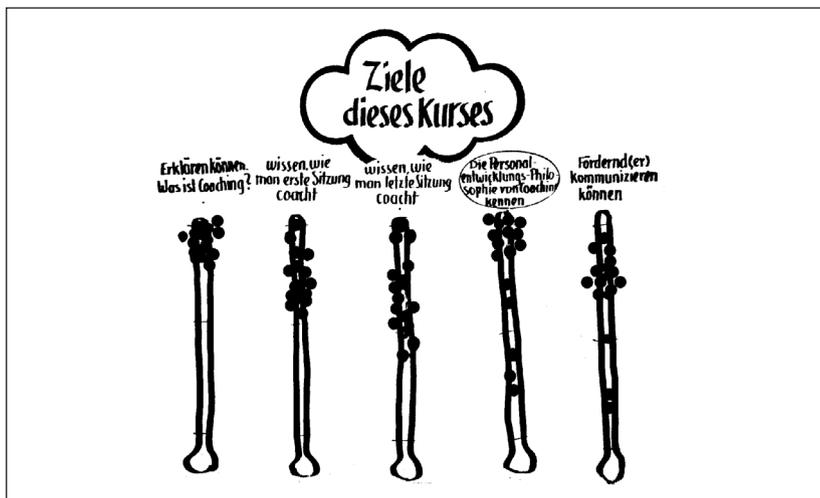


Abb.3: Thermometerauswertung

Kommentar: Hauptschwäche dieses Verfahrens ist, dass man nicht weiß, was die Teilnehmer tatsächlich können, sondern lediglich, was die Teilnehmer glauben, zu können. Geprüft wird bei diesem Verfahren also nicht Leistung, sondern Meinung. Auch der Soll-Wert (oberes Ende des Thermometers) bleibt ungewiss: Ist der Soll-Anspruch die Kompetenz, über die man nach höchster Perfektion verfügt, oder das, was man nach einer zeitlich begrenzten Schulung realistischerweise erwarten kann? Für die Teilnehmer und den Kursleiter wird diese Einschätzung um so treffender, je eindeutiger die Ziele formuliert sind (»Erklärt XY« ist eindeutiger einzuschätzen als »Beherrscht XY«).

Ein Vorteil dieses Verfahrens ist die wenig aufwändige und schnelle Durchführung sowie die sofortige Verfügbarkeit der Auswertung. Die Vorgabe mehrerer Teilziele erlaubt eine differenziertere Diskussion sowohl im Kurs als auch gegenüber der Institution; hier können sich in der Besprechung auch formative Aspekte ergeben (»Wer macht was, um noch Lücken zu

füllen?«). Dient diese Auswertung der summativen Darstellung gegenüber der Institution, dann gelten die genannten Einwände ebenfalls. Allerdings: Bevor man diese Ergebnisse als wertlos abtut, weil sie nur Teilnehmermeinung und nicht nachweisbare Ergebnisse erfassen, muss man plausible Gründe anführen, warum man den Teilnehmern die Fähigkeit zu einer angemessenen Einschätzung abspricht. Sind die Zielformulierungen für den Teilnehmerkreis eindeutig beurteilbar, dann fängt dies eine Menge der Einwände gegen diese Verfahren ab.

Dieses Verfahren fordert die **didaktische Reflexion** über die Formulierung der Teilziele heraus: Unabhängig vom Ergebnis der Evaluation zwingt dieses Verfahren dazu, sich vorher zu überlegen, was die Teilziele im Einzelnen sind. Dies Wissen um die zu erreichenden Ziele fördert in der Regel eine zielorientierte Arbeit im Kurs. Und wenn diese ganzen Überlegungen dann auch noch schriftlich dokumentiert sind (wieder: die Thermometer und Klebepunkte sind lediglich Daten, nicht Evaluation), dann nähern wir uns auch mit diesem Verfahren einer »Evaluation«.

2.1.3 Leistungsprüfung mit »Tests«: Was wird gekonnt?

Intention: In einem Kurs sollte ein bestimmtes Wissen oder Können vermittelt werden; in der Evaluation soll nun festgestellt werden, ob die Teilnehmer dieses Wissen oder Können auch beherrschen.

Adressaten: Teilnehmer, Kursleiter, Institution, Öffentlichkeit.

Zeitaufwand: je nach Umfang und Komplexität der Aufgaben hoch.

Voraussetzungen: Sie können Aufgaben formulieren, welche die entscheidenden Fähigkeiten erkennen lassen (dies wird in Kapitel 10 ausführlich behandelt werden).

Material: je nach geprüfter Leistung Aufgabenblätter oder reales Material.

Durchführung: formativ nach Stoffabschnitt, summativ am Kursende.

Beispiel: Sie führen eine »Einführung in den Computer« durch. Sie listen vor Beginn für sich auf, was zu einem bestimmten Zeitpunkt beherrscht werden muss, z. B. eine Diskette formatieren, Subdirectories anlegen, Dateien kopieren ... Damit formulieren Sie ein Aufgabenblatt, das jeder Teilnehmer gegen Kursende in prüfungsähnlicher Form abarbeiten muss, z. B.:

Aufgabenblatt

Abschlusstest »Einführung in den Computer«

1. Sie legen eine fabrikneue Diskette ins Laufwerk; sie lässt sich nicht beschreiben, Fehler wird angezeigt. Was ist die naheliegendste Ursache?
.....
2. Formatieren Sie die beiliegende Diskette!
3. Legen Sie darauf ein Unterverzeichnis »SPIELE« an!
4. Suchen Sie auf der Festplatte die Datei »SCHACH.EXE«! Wie lautet der komplette Pfadname?
.....
5. Kopieren Sie die Datei »SCHACH.EXE« in das Unterverzeichnis »SPIELE«.

Übungsaufgabe 10

Erkennen Sie die zwei Konstruktionsfehler in diesem Aufgabenblatt (obwohl das erst Thema von Kapitel 10.3.1 sein wird)?



Jeder Teilnehmer markiert auf seinem Aufgabenblatt, ob er die Aufgabe ohne fremde Hilfe sicher gelöst hat; ist das der Fall, dann hakt er diese Aufgabe auch an der Wandtafel ab. Auf diese Weise wird sichtbar, was beherrscht wird, und was im Kurs Nacharbeit benötigt.

Didaktische Reflexion: Diese Evaluationsform hat in der Regel den höchsten didaktischen Wert, trägt also am meisten dazu bei, Lernerfolg zu steigern. Wenn Sie solche zu erreichenden beschreibbaren Fähigkeiten vor Kursbeginn bei Ihrer Vorbereitung ausformulieren, zielbewusst darauf hinarbeiten und auch den Teilnehmern zu Beginn diese »Prüfungs«-Kriterien mitteilen (sie also nicht etwa zum Kursende mit einer »Klassenarbeit« überfallen), dann führt dieses »zielorientierte Lernen« zumeist zu hohen Erfolgsquoten, insbesondere, wenn Sie abschnittsweise (also formativ) vorgehen. Die »Prüfung« ist dabei kein an das Lernen unnütz angehängtes Ritual, sondern ist ein Lern- und Übungsangebot für die Teilnehmer, sich noch einmal des Lernerfolges zu vergewissern und an die Anforderungen »nachher« zu denken.

Dokumentation: Auf einem Aufgabenblatt wird eingetragen, wie viel Teilnehmer pro Aufgabe erfolgreich waren. Natürlich werden auch die Aufgaben dokumentiert.

Probleme:

- ▶ Aufgabenformulierung ist arbeitsaufwändig.
- ▶ Nur bestimmte Inhalte lassen sich in solche Aufgabenformen bringen.
- ▶ Prüfungsängste der Teilnehmer (ernstnehmen!), beispielsweise wenn der Prüfungsstoff so umfangreich oder vielfältig ist, dass eine für die Teilnehmer vorhersehbare Prüfung nicht möglich ist.
- ▶ Muss für jeden Inhalt neu gemacht werden.

Kommentar: Überzeugend, um erreichte Fertigkeiten nach außen darzustellen, aber auch, weil mit den Aufgaben und einem evtl. Maßstab ein Soll-Wert definiert ist (»Diese Aufgaben müssen zu 80% gelöst werden«). Oft aber scheut man sich, solche »einfachen« Fertigkeiten als Soll-Wert zu formulieren und abzuprüfen. Auch bleibt der Einwand, von den vielfältig möglichen Aufgaben seien nur wenige ausgewählt, und man müsste eigentlich viel mehr verschiedene Inhalte prüfen. Lassen Sie sich dadurch nicht entmutigen! Fangen Sie mit einigen wenigen zu beherrschenden Fertigkeiten an: Die Erfahrung wächst mit der Übung.

2.2 Der »Kursbeurteilungsbogen KBB« – ein quantitatives Standardinstrument

2.2.1 Kurzübersicht

Intention: Sie wollen am Ende eines Seminars eine schnelle und einfache Rückmeldung darüber, wie die Teilnehmer den Kurs in verschiedenen Dimensionen einschätzen. Diese Einschätzung soll methodisch möglichst zuverlässig sein, evtl. auch vergleichbar mit anderen Kursen.

Adressaten der Evaluation: Teilnehmer, Kursleiter, Institution.

Zeitaufwand: Durchführung am Kursende ca. 15 Minuten, ca. 1 Minute pro Teilnehmer für die Auswertung.

Voraussetzung: Sie müssen einen MS-DOS-Computer haben und in der Lage sein, ein Programm zu starten. Dann können Sie dieses Instrument in fast jedem Kurs und ohne didaktische Vorbereitung einsetzen. Der Kursbeurteilungsbogen wurde für Kurse konstruiert, die mindestens einen Tag oder drei 90-Minuten-Sitzungen lang sind. Für Einzelveranstaltungen, Vorträge oder Informationsveranstaltungen ist er nicht geeignet.

Material: Der Fragebogen als Kopiervorlage, eine Demo-Version des Computerauswertungsprogramms (maximal 9 Personen lassen sich mit dieser Demo-Version auswerten).

Hinweise zur Benutzung sowie eine Beispieldatei zum Üben liegen zum Download für Sie unter: www.ziel-verlag/gdw bereit.

Hinweis für Ungeduldige und Neugierige

Falls Sie gerade einen Kurs abschließen (oder ein Freund oder Kollege), dann nutzen Sie gleich die Gelegenheit zum Ausprobieren – hier über die Feinheiten und Hintergründe weiterlesen können Sie auch noch später: Kopieren Sie die entsprechende Anzahl Fragebogen, lassen Sie sie in einem Kurs einsetzen, laden Sie sich anschließend das Auswertungsprogramm unter www.ziel-verlag.de/gdw herunter und starten Sie das Programm »KBB-Auswertung-demo.exe«. Lassen Sie sich vom Programm führen, es ist weitgehend selbsterklärend. Wenn Sie mehr als 9 Personen auswerten wollen (das ist die Grenze der beigelegten Demo-Version), können Sie die Teilnehmer unterteilen (z. B. Männer/Frauen) und in zwei Arbeitsgängen auswerten.

Didaktische Reflexion: Mit seinen Dimensionen und Fragen weist der Kursbeurteilungsbogen – noch bevor er überhaupt eingesetzt ist – die Kursleiter auf Aspekte hin, an denen sie ihre praktische Arbeit, ihr Selbstverständnis und ihre eigenen Intentionen prüfen können. Zum Beispiel: In diesem Kursbeurteilungsbogen wird davon ausgegangen, dass der vermittelte Stoff nicht das Einzige ist, was zählt. Sondern es wird auch das Kursklima als eine Subdimension von Kursqualität verstanden. Dies fordert den Kursleiter heraus, sich zu überlegen, ob er dies auch so sehen mag oder nicht. Ebenso können einzelne Fragen aus dem Fragebogen den Kursleiter zur Reflexion anregen: Ist »In der Gruppe gibt es jetzt ein herzliches Gefühl der Zusammengehörigkeit« oder »Die Teilnehmer wurden ermutigt, eigene Fragen und Erfahrungen in den Kurs einzubringen«, etwas, was ich für diesen Kurs oder für zukünftige Kurse beabsichtige? Ist dies für mich als Kursleiter etwas, was angestrebt wurde, angestrebt werden soll?

Der Kursbeurteilungsbogen will in diesem Sinne gleichermaßen zur Reflexion (besser verstehen) als auch zu Handlungskonsequenzen (besser machen) anregen und Neugierde und Bereitschaft fördern, die eigene Arbeit zu überprüfen und daraus zu lernen. Der Kursbeurteilungsbogen will also nicht nur zu einer nachträglichen Kontrolle beitragen, sondern will auch anregen, den vorausgehenden Bildungsprozess zu verändern.

Der Kursbeurteilungsbogen KBB erfasst die Dimensionen:

- ▶ Stoffbeherrschung,
- ▶ Lernbereitschaft,
- ▶ Lernunterstützung und
- ▶ Klima.

Da die Fragen sich nicht auf bestimmte Kursinhalte beziehen, kann der Kursbeurteilungsbogen vielseitig eingesetzt werden; er erlaubt damit Vergleiche auch zwischen inhaltlich verschiedenen Kursen.

Der Kursbeurteilungsbogen besteht aus 39 Schätzfragen zum Ankreuzen. Die Durchführung dauert ca. 10–15 Minuten. Die Auswertung erfolgt über ein Computerprogramm. Die Messqualität wurde in ca. 40 Kursen geprüft und gesichert.

2.2.2 Wozu der Kursbeurteilungsbogen konstruiert wurde – Einsatzmöglichkeiten

Der Kursbeurteilungsbogen wurde mit der Absicht konstruiert, Kurse zu überprüfen (summativ) und zu verbessern (formativ).

1. Der Kursbeurteilungsbogen ist ein Angebot für Kursleiter, die sich selbst Rückmeldung über den eigenen Kurs verschaffen wollen. Kursleiter können den Kursbeurteilungsbogen zur Kontrolle und Selbstreflexion ihrer Arbeit einsetzen und als Anlass zum Gespräch mit den Teilnehmern über eine bestmögliche gemeinsame Kursgestaltung (Selbstevaluation) nehmen. Absicht des Kursbeurteilungsbogens in diesem Sinne ist: Nachdenken – Miteinander-Sprechen – Bessermachen.
2. Institutionen können dieses Instrument im Rahmen von Verbesserungsmaßnahmen ihres Kursangebotes einsetzen. Voraussetzung dieser Fremdevaluation durch Bereichsleiter, Bildungsreferenten oder pädagogische Mitarbeiter ist, dass bei dieser Anwendung auch das interpretierende Gespräch mit den Kursleitern stattfindet. Wird der Kursbeurteilungsbogen nicht vom Kursleiter selbst eingesetzt, sondern von Vorgesetzten, Bildungsreferenten oder Untersuchern, dann wäre es schlechter Stil, wenn der Kursleiter nicht über Art, Absicht und Ergebnis informiert würde. Und vor allem: Keine Befragung hinter dem Rücken des Kursleiters!
3. Schließlich kann der Kursbeurteilungsbogen als wissenschaftliches Instrument in Vergleichsgruppen/Kontrolldesigns in der Forschung eingesetzt werden.

2.2.3 Die Inhalte des Kursbeurteilungsbogens

Im Kursbeurteilungsbogen wurde ein »guter Kurs« mit vier Dimensionen umschrieben. Zwei Dimensionen beschreiben »Produkte« (»**Outcomes**«), also **Ergebnisse**, die sich durch den Kurs einstellen:

- 1. Stoffbeherrschung:** Selbstverständliche Erwartung an einen guten Kurs ist, dass die Teilnehmer die im Kurs zu vermittelnden Inhalte am Ende beherrschen. Stoffbeherrschung wird aus drei unterschiedlichen Subdimensionen zusammengesetzt:
 - ▶ Aus der **subjektiven Einschätzung des Erfolgs**, d.h. ob der Teilnehmer nach seinem eigenen Maßstab mit dem Gelernten zufrieden ist (Frage 13, 21, 25).
 - ▶ Fünf weitere Aussagen fragen unterschiedlich scharf nach der Einschätzung des **objektiven Lernerfolgs**, d.h. Erfolg wird gemessen am äußeren Maßstab, also »was eigentlich hätte gelernt werden müssen« (Frage 1, 9, 17, 29, 37).
 - ▶ Zwei Aussagen fragen, ob die Stoffbeherrschung auch bis zur **praktischen Anwendung** reicht (Frage 5, 33).
- 2. Lernbereitschaft, Lernfähigkeit:** Ein guter Kurs muss die Bereitschaft und die Fähigkeit steigern, auch später wieder gern und erfolgreich an Weiterbildung teilzunehmen. Mit dieser Dimension wird danach gefragt, ob der Teilnehmer über den unmittelbaren Stoff des Kurses hinaus Wirkungen, »Outcomes«, aus dem Kurs mitnimmt, die sich positiv auf sein zukünftiges Lernen auswirken.
 - ▶ Dazu gehört zunächst ein **allgemeines positives Gefühl gegenüber Weiterbildung** (Frage 27, 11, 19, 35).
 - ▶ Gefragt wird auch nach **Konsequenzen** aus der Kurserfahrung (Frage 7, 31).
 - ▶ Schließlich enthält diese Dimension noch Fragen zur Förderung des »**Lernens des Lernen**« (»Methodenkompetenz« – Frage 23, 3, 15):

Zwei weitere Dimensionen beschreiben Qualitäten der im Kurs ablaufenden Prozesse:
- 3. Lernunterstützung:** In einem guten Kurs wird der Lernprozess von den Teilnehmern als leicht, schnell, angenehm und vom Kursleiter unterstützt wahrgenommen.
 - ▶ Zunächst wird gefragt, ob der Kurs **allgemein als lernunterstützend** erlebt wurde (Frage 14, 10).

- ▶ Sodann wird nach verschiedenen **methodischen Gestaltungselementen** und damit indirekt nach der methodisch-didaktischen Gestaltungskompetenz des Kursleiters gefragt (Frage 18, 38, 34, 2).
 - ▶ Als lernunterstützend werden weiterhin Maßnahmen zur Teilnehmerorientierung und das Herstellen des **Bezugs zum Lerner** abgefragt (Fragen 6, 22, 26, 30).
4. **Klima:** In einem guten Kurs herrscht ein Klima gegenseitiger Achtung, partnerschaftlichen Helfens, entspannten Humors und persönlicher Echtheit.
- ▶ Ein Teil der Klima-Aussagen fragt allgemein nach der **emotionalen Getöntheit** des Lernprozesses (Frage 16, 24, 32, 36).
 - ▶ Ein zweiter Teil fragt nach Art und **Qualität der sozialen Interaktion** im Kurs. Dieser Perspektive wird besondere Bedeutung für das Kursklima zugemessen (Frage 4, 8, 12, 20, 28, 39).

2.2.4 Grenzen

Der Kursbeurteilungsbogen ist eine Schätzskala: Teilnehmer geben ihren subjektiven Eindruck vom Kurs wieder; erhoben wird Meinung, nicht Leistung. Diese subjektive Einschätzung wird durch vielerlei Faktoren beeinflusst. Der Kursbeurteilungsbogen misst deshalb nicht auf einer absoluten Skala; eine Durchschnittsbeurteilung von 3,0 bedeutet nicht unbedingt, dass dieser Kurs schlechter war als ein Kurs mit der Durchschnittsbeurteilung 2,5 – vielleicht hatten die Teilnehmer nur strengere Erwartungen. Der Gesamtdurchschnitt, das Ergebnis in den vier Dimensionen oder bei jeder einzelnen Aufgabe bedarf grundsätzlich der Interpretation.

Der Kursbeurteilungsbogen kann auch missverstanden werden als Instrument, Kursleiter zu be- oder zu verurteilen. Ob ein Kurs gelingt, hängt sicherlich auch vom Kursleiter ab. Deshalb gibt es im Kursbeurteilungsbogen Aussagen, die stärker dem Handeln des Kursleiters zugeordnet werden können (insbesondere in der Dimension »Lern-Unterstützung«). Aber andere Variablen, die der Kursleiter nur bedingt oder gar nicht beeinflussen kann, spielen für das Gelingen eines Kurses genauso eine Rolle: der äußere Rahmen (wie Ort, Zeit Raum), der Stoff (Schwierigkeit, Umfang) und insbesondere die Teilnehmer, die mit ihrem Vorwissen und Mitmachen, ihrem Lerninteresse und ihren Launen mitentscheiden, ob ein Kurs ein Erfolg wird. All dies geht in die Schätzung im Fragebogen ein. Deshalb heißt er auch »Kursbeurteilungsbogen« und nicht »Kursleiterbeurteilungsbogen«.

