

# Evaluation

Alfred Uhl

**Beitrag zu:**

**(1) Stimmer, F. (Hrsg.): Suchtlexikon.  
Oldenbourg, München, 1999**

**(2) Stimmer, F. (Hrsg.): Lexikon der Sozialpädagogik  
und der Sozialarbeit, 4. Auflage.  
Oldenbourg, München, 1999**

## 1. Der Begriff „Evaluation“ in Umgangssprache und Wissenschaft

Das aus dem Lateinischen kommende Wort „Evaluation“ bedeutet „Bewertung“. Dem umgangssprachlichen Verständnis entsprechend hat Evaluation eine funktionale (Sachurteil: „Funktioniert der Ansatz im Sinne der Zielvorstellungen?“) und eine ethische Komponente (Werturteil: „Ist der Ansatz mit grundlegenden Wertvorstellungen vereinbar?“). Diesem Begriffsverständnis entsprechend ist die reine Beschreibung von Vorgängen (Deskription) und die Entwicklung von Hypothesen, Modellen und Theorien (Exploration) keine Evaluation.

Im wissenschaftlichen Kontext hat sich allerdings ein gänzlich anderes Verständnis von Evaluation eingebürgert. Dem Postulat der Wertfreiheit entsprechend werden Werturteile gänzlich ausgeklammert, alle Schritte des wissenschaftlichen Handelns hingegen - von Deskription (Beschreibung) über Exploration (Hypothesenformulierung) bis zur Bestätigung (Hypothesenprüfung) – eingeschlossen. Wie man aus Formulierungen wie „formative Evaluation“ oder „Prozessevaluation“ deutlich erkennen kann, wird Evaluation im wissenschaftlichen Kontext als Überbegriff über eine sehr große und sehr heterogene Klasse von Tätigkeiten verstanden. Evaluation im wissenschaftlichen Sinn schließt alle Varianten wissenschaftlichen Handelns ein und ist damit keinesfalls auf Wirksamkeitsprüfung beschränkt.

Die Diskrepanz zwischen dem wissenschaftlichen und dem umgangssprachlichen Begriffsverständnis sowie der Umstand, dass der wissenschaftliche Evaluationsbegriff als Überbegriff über eine sehr heterogene Klasse ganz unterschiedlicher wissenschaftlicher Tätigkeiten steht, trägt regelmäßig zur Verwirrung aller Beteiligten bei und behindert den rationalen Diskurs über das Thema „Evaluation“ recht nachhaltig. Als Konsequenz aus dieser äußerst unbefriedigenden Situation kann man nur empfehlen, soweit wie möglich auf den vagen und mehrdeutigen Überbegriff „Evaluation“ zu verzichten und immer genau zu bezeichnen, auf welche Unterkategorie man sich im konkreten Fall bezieht.

## 2. Ein europäischer Versuch einer Synthese

Um einen konstruktiven Beitrag zur Präzisierung von evaluationsrelevanten Konzepten und Definitionen zu leisten, wurde 1994 im Rahmen der COST-A6 Aktion der Europäischen Kommission eine Arbeitsgruppe konstituiert. Diese umfasste 21 internationale Experten aus 14 Ländern. Aufgabe der Arbeitsgruppe war es, sich kritisch und konstruktiv mit Begriffen und Konzepten auseinanderzusetzen, die in Zusammenhang mit Prävention und Evaluation bedeutsam sind. Indem bei der Präzisierung und Vereinheitlichung auf traditionelle Konzepte aufgebaut wurde, sollte die begriffliche Kontinuität mit der wissenschaftlichen Tradition weitestgehend gewahrt bleiben. Im Zuge dieser, nach der Delphi-Methode durchgeführten Konsensusstudie wurde ein vierdimensionales Klassifikationsmodell „Daten-Zeit-Methodologie-Evaluator-Klassifikation“ (DZME-Klassifikation) entwickelt und eine Reihe von Evaluationskategorien nach inhaltlichen Gesichtspunkten aufgelistet und präzisiert (Uhl, 1998).

### 3. Wichtige traditionelle Klassifikationssysteme, auf die die DZME-Klassifikation und „inhaltliche Klassifikation“ der COST-A6-Arbeitsgruppe aufbauen

#### 3.1. PEI-Klassifikation (Prozessevaluation vs. Ergebnisevaluation vs. Impactevaluation; Clayton & Cattarello, 1991)

Die sehr gebräuchliche PEI-Klassifikation unterscheidet anhand der Art der berücksichtigten Daten nach drei Kategorien:

- Prozessevaluation. (Process Evaluation) steht für die systematische Erfassung des gesamten Prozesses (Prozessdaten) während der Durchführung einer Intervention, was den gesamten Interaktionsprozess zwischen Programmausführenden und Zielpersonen umfasst.
- Ergebnisevaluation (Outcome Evaluation) steht für die Untersuchung, ob erwartete Effekte (erwartete Ergebnisdaten) nach Abschluss einer Intervention eingetreten sind.
- Impact Evaluation steht für die Erfassung von Interventionseffekten, die über die vorgesehenen Zielgruppen und erwarteten Effekte (nicht erwartete Ergebnisdaten) hinausgehen. (Für diesen Begriff gibt es kein brauchbares deutsches Äquivalent.)

#### 3.2. SPE-Klassifikation („strukturelle Qualität“ vs. „Prozessqualität“ vs. „Ergebnisqualität“; Donabedian, 1980)

Auch die SPE-Klassifikation nimmt Bezug auf die Art der berücksichtigten Daten. Die der „struktureller Qualität“, „Prozessqualität“ und „Ergebnisqualität“ zugrundeliegenden Datenqualitäten lassen sich wie folgt beschreiben:

- strukturelle Daten sind Daten, die strukturelle Rahmenbedingungen beschreiben, wie „Ort der Intervention“, „Qualifikation der das Programm ausführenden Personen“, „Charakteristika der Zielpersonen“, usw.
- Prozessdaten sind Daten, die die Ausführung der Programme erfassen (das Verhalten der Programmausführenden = Programm-Input).
- Ergebnisdaten sind Daten, die die wünschenswerten Auswirkungen auf die Zielgruppe sowie Kosten, die das Programm verursacht hat, zum Inhalt haben (Programm-Output).

Das SPE-Konzept unterscheidet sich inhaltlich vom PEI-Konzept durch die Einführung des Begriffes „strukturelle Daten“, dadurch, dass nicht zwischen erwarteten und nicht erwarteten Ergebnissen unterschieden wird, sowie dadurch, dass der Begriff „Prozessdaten“ auf das Verhalten der Programmausführenden beschränkt wird.

#### 3.3. FS-Klassifikation (formative Evaluation vs. summative Evaluation; Scriven, 1967)

Die ebenfalls sehr gebräuchliche SE-Klassifikation nimmt Bezug auf eine zeitliche Dimension, nämlich darauf, ob das Präventionsprogramm zum Zeitpunkt der Evaluation noch entwickelt (geformt) wird oder schon abgeschlossen ist (und daher zusammenfassend beurteilt werden kann).

Es ist sinnvoll beide Phasen jeweils noch einmal zu unterteilen, wodurch sich 4 abgegrenzte Phasen ergeben (Vier-Phasen-Modell):

- In der präformativen Phase (Konzeptphase) wird auf rein reflexiver Basis ein Präventionskonzept entwickelt und bewertet. Präformative Evaluation kommt ohne praktische Erprobungsschritte - d.h. ohne prospektiv orientierte empirische Schritte - aus. Die präformative Phase schließt mit einem ersten vorläufigen Programmentwurf ab.
- In der formativen Phase (Entwicklungsphase) wird dann - aufbauend auf den in der präformativen Phase entwickelten vorläufigen Programmentwurf - durch wiederholte praktische Erprobung ein konkretes Präventionsprogramm geformt. Formative Evaluation zielt auf die rasche und flexible Erfassung von Schwachstellen mit dem Ziel, vorläufige Programmentwürfe kontinuierlich umzuformen und so lange zu verbessern, bis sich ein Programm ohne offensichtliche Schwachstellen ergibt.
- In der ersten summativen Phase (Erprobungsphase) finden Forschungsstrategien statt, die erst einsetzen, nachdem die Entwicklung eines neuen Präventionsprogramms abgeschlossen wurde. Im Zuge der formativen Evaluation in dieser Phase soll das fertige Programm nun zusammenfassend beurteilt werden.
- In der zweiten summativen Phase (Routinephase) sollte der Erfolgsnachweis bereits erbracht sein. Bei summativer Evaluation in dieser Phase geht es darum zu gewährleisten, dass die Qualität der Programmdurchführung gewährleistet bleibt und nach unerwarteten längerfristigen Effekten bzw. nach relevanten Veränderungen der Rahmenbedingungen Ausschau zu halten.

### **3.4. DEH-Klassifikation („deskriptive Evaluation“ vs. „explorative Evaluation“ vs. „hypothesenprüfende Evaluation“; z.B. Popper, 1976 oder Tukey, 1977)**

Eine weitere - und bezüglich der Aussagekraft von evaluierender Forschung ganz besonders wichtige - Klassifikation unterteilt nach den Kategorien „Deskription“, „explorative Forschung“ und „hypothesenprüfende Forschung“. Man könnte diese Dimension als „methodologische Dimension“ bezeichnen.

Deskription ist die unterste Stufe wissenschaftlichen Vorgehens. Deskriptive Evaluation ist die bloße Erfassung und Dokumentation von Phänomenen sowie deren Kategorisierung und Zusammenfassung, ohne daraus neue Hypothesen ableiten zu wollen.

Explorative Forschung geht über die reine Deskription hinaus und stellt damit die zweite Stufe wissenschaftlicher Vorgangsweise dar. Diese Form der Datenanalyse

- zielt auf die Entdeckung neuer Phänomene,
- liefert Impulse, um neue Hypothesen und Theorien zu entwickeln,
- ist grundsätzlich divergent orientiert,
- ist keinen strengen methodologischen Regeln unterworfen und
- alle Ergebnisse haben grundsätzlich nur vorläufigen Charakter.

Hypothesenprüfende Forschung versucht mit den Regeln der Wahrscheinlichkeitsrechnung und der schließenden Statistik Zufallseffekte von substantiellen Effekten abzugrenzen. Diese Form der Datenanalyse stellt die höchste Stufe wissenschaftlicher Vorgangsweise dar. Diese

- zielt auf die Prüfung von Hypothesen und Theorien,
- ist grundsätzlich konvergent orientiert,
- ist strengen methodologischen Regeln unterworfen und
- die Ergebnisse können in einem gewissen Sinn als wissenschaftlich gesichert gelten.

Da man Entscheidungen für bestimmte Theorien und Modelle sowie für die Anwendung bestimmter Techniken auf hypothesenprüfende Ansätze aufbauen sollte, werden hypothesenprüfende Forschungsansätze häufig „Entscheidungsstudien“ genannt. Hypothesenprüfende Ansätze, die die Wirksamkeit von Interventionen zum Ziel haben, werden auch als „Wirksamkeitsstudien“ bezeichnet.

### **3.5. IE-Klassifikation („interne Evaluation“ vs. „externe Evaluation“)**

Die IE-Klassifikation unterscheidet nach der Position jener Person, die die Hauptverantwortung für eine Evaluation trägt. Man könnte diesen Aspekt als „Evaluatordimension“ bezeichnen: Als interne Evaluation wird eine Evaluation bezeichnet, bei der die für die Evaluation hauptverantwortliche Person zum engeren Kreis der Programmentwickler und/oder Anwender gehört. Als externe Evaluation wird eine Evaluation bezeichnet, die primär in den Händen eines unabhängigen Evaluators liegt.

## 4. Synthese und Präzisierung klassischer Konzepte durch die COST-A6 Arbeitsgruppe (Uhl, 1998)

### 4.1. DZME-Klassifikation

Als Synthese und Weiterführung der im letzten Abschnitt angeführten traditionellen Klassifikationssysteme ergab sich in der erwähnten COST-A6 Konsensstudie die DZME-Klassifikation („Daten-Zeit-Methodologie-Evaluatordimension-Klassifikation“). Diese baut auf folgenden vier Dimensionen auf:

die Datendimension (D)

- Strukturdaten
- Prozessdaten
- erwartete Ergebnisdaten
- unerwartete Ergebnisdaten
- Kontextdaten

die Zeitdimension (Z)

- Konzeptphase = präformative Phase
- Entwicklungsphase = formative Phase
- Überprüfungsphase = erste summative Phase
- Routinephase = zweite summative Phase

die methodologische Dimension (M)

- deskriptiv
- explorativ
- hypothesenprüfend

die Evaluatordimension (E)

- interne Evaluation
- externe Evaluation

Der DZME-Ansatz zur Klassifikation von Evaluationsprojekten kann, sofern die Begriffe korrekt verwendet werden, einen wesentlichen Beitrag zur Präzisierung des Dialogs leisten. Er reicht alleine aber nicht aus, um die Komplexität der Aufgabenstellung vollständig abzubilden. Es erschien der COST-A6 Arbeitsgruppe daher zweckmäßig, diesen eher abstrakten Ansatz durch ein stärker inhaltlich orientiertes System zu ergänzen. Letzteres nimmt teilweise auch auf Begriffe Bezug, die in ersterem System eine Rolle spielen.

## 4.2. Inhaltliche Klassifikation von Evaluation

### 4.2.1. Ethische Evaluation

In der wissenschaftlichen Evaluation wird, dem Postulat der Wertfreiheit entsprechend, der wertorientiert ethische Aspekt häufig ignoriert, unterbetont bzw. über logisch/sachliche Argumentation verschleiert. Da die ethische Beurteilung von Präventionsansätzen aber auch dann, wenn diese nicht explizit erwähnt wird, implizit in die Forschungsstrategien einfließt, ist es zweckmäßig zu fordern, dass „ethische Evaluation“ ausdrücklich zum Thema gemacht wird. Werturteile sollten nicht bloß indirekt - und damit unreflektiert - in Forschungsdesigns und Schlussfolgerungen einfließen.

Besonders wichtig ist die ethische Evaluation natürlich in der Konzeptphase, aber selbstverständlich sollten ethische Überlegungen auch in allen späteren Phasen eine Rolle spielen.

### 4.2.2. Historische Evaluation

Als historische Evaluation kann man Expertise auf der Basis von eigener Erfahrung und/oder der vorhandenen wissenschaftlichen Literatur bezeichnen. Das in Zusammenhang mit Evaluation eher ungewöhnliche Attribut „historisch“ (historische Daten, historische Referenzwerte, historische Kontrollgruppe, usw.) ist in der Methodik für klinische Studien (Clinical Trials Methodology) etabliert, und es erscheint zweckmäßig, dieses sprachliche Konzept auch in der Evaluationsforschung zu etablieren.

### 4.2.3. Methodologische Evaluation

In engem Zusammenhang mit der auf wissenschaftlichen Erkenntnissen basierenden historischen Evaluation steht die Fragen, ob und wie weit die Schlussfolgerungen, die die jeweiligen Autoren der Forschungsberichte aus ihren empirischen Studien gezogen haben, aus statistisch-methodologischer Sicht korrekt sind (interne Validität), und ob die Ergebnisse sinnvollerweise auf die, für das Präventionsprogramm vorgesehene Anwendungssituation übertragbar sind (externe Validität).

### 4.2.4. Formative Evaluation

Der Begriff „formative Evaluation“ bezieht sich einerseits auf eine bestimmte Phase der Programmentwicklung - „das Formen des Programms in der Entwicklungsphase“ - und andererseits auf eine, dieser Phase angemessene ganz bestimmte explorative Vorgangsweise. Während ersterer Aspekt in der Zeitdimension des DZME-Ansatzes berücksichtigt ist, wird letzterer Aspekt hier nach inhaltlichen Gesichtspunkten gesondert angeführt und erörtert.

In der Programmentwicklungsphase steht der Programmentwickler vor der Aufgabe, Strukturelemente eines vorläufigen Programmentwurfes rasch und flexibel zu überprüfen und Schwachstellen zu beseitigen. Die Technik, um das möglich ökonomisch zu erreichen, besteht in der wiederholten Abfolge von Überprüfung und Anpassung, wobei das im Idealfall so lange weitergeht, bis sich das Konzept als praktisch durchführbar und im Sinne der Zielvorgaben erfolgversprechend erweist.

Da sich viele Schwachstellen vorläufiger Programmentwürfe in der Anwendung recht unmittelbar manifestieren, sind hypothesenprüfende Forschungsstrategien (komplexe Forschungsdesigns, große Stichproben und/oder strenge methodische Regeln) in diesem Zusammenhang kaum zweckmäßig. Primär sollte eine Serie von kleinen Erprobungen und Pilotstudien zum Einsatz kommen. Die Blickrichtung sollte möglichst offen in alle Richtungen gehen, um auch unerwartete Problemfelder zu erfassen (divergent, explorativ). Auch sollte die Datenerfassung überwiegend prozessorientiert erfolgen, da man nur durch die unmittelbare Beobachtung des Geschehens Zugang zu Phänomen erhält, die sich einer ergebnisorientierten Betrachtung verschließen.

In manchen Fällen ist auch in der Programmentwicklungsphase eine ergebnisorientierte Vorgangsweise angemessen, wobei allerdings in dieser Phase, die Einhaltung strenger methodologischer Prinzipien verzichtbar ist.

#### 4.2.5. Evaluation der Durchführbarkeit (Feasibility Evaluation)

Die Frage, ob ein bestimmtes Programm praktisch durchführbar ist, spielt vor allem in der Entwicklungsphase und in der Überprüfungsphase eine wichtige Rolle. Während man in der Entwicklungsphase allerdings divergent, explorativ vorgeht und vorläufige Programmentwürfe anhand von kleinen Erprobungen und Pilotstudien kontinuierlich anpasst (formative Evaluation), ist in der Überprüfungsphase eine hypothesengeleitete systematische Überprüfung des fertigen Programms an größeren Stichproben unter Alltagsbedingungen zu planen („Durchführbarkeitsstudie“ bzw. „Feasibility Study“).

#### 4.2.6. Monitoring unerwünschter Nebeneffekte

Da die Anzahl der möglichen Problemfelder fast unbegrenzt ist und sich in Abhängigkeit von den Rahmenbedingungen immer neue Probleme ergeben können, spielt auch die Erfassung von unerwünschten Nebeneffekten in allen Phasen eine wesentliche Rolle. Meist wird die Forschungsstrategie divergent, explorativ sein.

#### 4.2.7. Evaluation der Wirksamkeit (Wirksamkeitsnachweis)

Die zentrale Frage, nachdem ein Programm entwickelt wurde, ist zweifelsohne die Frage: „Funktioniert es im Sinne der Zielvorstellungen?“.

Die Evaluation der Wirksamkeit kann grundsätzlich auf drei Arten erfolgen: global, partiell und historisch.

##### 4.2.7.1. Globaler empirischer Wirksamkeitsnachweis

Als „globalen empirischen Wirksamkeitsnachweis“ kann man den experimentellen oder quasiexperimentellen Nachweis bezeichnen, dass ein Programm tatsächlich in der Lage ist, die gewünschten Effekte („primäre Zielvariablen“) in der Zielgruppe zu bewirken. Dieser Ansatz kommt forschungslogisch am nächsten an einen Beweis der Wirksamkeit heran und könnte als „empirischer Wirksamkeitsnachweis im engeren Sinne“ bezeichnet werden. Ein globaler empirischer Wirksamkeitsnachweis sollte, wenn immer das möglich ist, geführt werden. Es ist aber trotzdem keinesfalls zweckmäßig, diesen induktiven (statistischen) Ansatz als unbedingten Standard festzuschreiben, da dieser Zugang in der Praxis oft an ökonomischen und praktischen Erkenntnisgrenzen scheitert.

##### 4.2.7.2. Partieller empirischer Wirksamkeitsnachweis

Als „partiellen empirischen Wirksamkeitsnachweis“ kann man den experimentellen oder quasiexperimentellen Nachweis bezeichnen, dass wesentliche Teile des dem Programm zugrundeliegenden Wirkungsmodells zutreffen.

##### 4.2.7.3. Historischer Wirksamkeitsnachweis

Als historischen Wirksamkeitsnachweis kann man die Ableitung der Programmwirksamkeit aus vorhandenen (historischen) Daten bezeichnen; d.h. wenn die Wirksamkeit bereits deduktiv aus einer empirisch gut fundierten Theorie abgeleitet werden kann. Grundlage für einen historischen Wirksamkeitsnachweis ist eine umfassende historische und methodologische Evaluation, die auf empirisch gut belegte Zusammenhänge zurückgreifen kann.

#### 4.2.8. Qualitätssicherung (QS)

Es steht außer Frage, dass jede Form der Evaluation irgendeinen Aspekt der Qualität eines Programms oder einer Intervention zum Inhalt hat, und es ist daher semantisch möglich, Qualitätssicherung (QS) als Synonym für „Evaluation“ zu verstehen. Dieser inflationäre Gebrauch des Begriffes QS ist weit verbreitet aber nicht zweckmäßig. Nützlich wird der Begriff „QS“ allerdings, wenn man ihn auf die Qualität der Programmdurchführung bezieht. Zentrale Fragestellung der QS ist dann, ob ein Programm korrekt, d.h. den Instruktionen entsprechend angewendet wird und nicht, ob es wirkt. QS, im Sinne dieses Begriffsverständnisses, findet ausschließlich in der Routinephase statt. Programmwirksamkeit sollte man in dieser Phase bereits voraussetzen können. Wird QS intern organisiert, so ist es zweckmäßig, von Qualitätsmanagement (QM) zu sprechen, und wird QS von externen Evaluatoren durchgeführt, so ist der Begriff Qualitätskontrolle (QK) angemessener.

#### 4.2.9. Strukturelle Evaluation

Auch strukturelle Evaluation findet grundsätzlich in der Routinephase statt. Dieser Ansatz ist primär deskriptiv und zielt auf strukturelle Aspekte der Programmanwendung. Das ist z.B.: In wievielen Schulen, in wievielen Klassen und von wievielen Lehrern wird ein bestimmtes Präventionsprogramm verwendet?

#### 4.2.10. Kontextevaluation

Präventionsprogramme werden unter bestimmten Rahmenbedingungen (Kontext) für gewisse Situationen entwickelt. Auch wenn zunächst eindeutig nachgewiesen werden konnte, dass Programme unter bestimmten Rahmenbedingungen erfolgreich sind, so darf man das natürlich nicht einfach auf gänzlich andere Situationen oder geänderte Rahmenbedingungen übertragen. In diesem Sinne ist es nötig, in der Routinephase laufend zu erheben, ob sich der Kontext seit der Programmimplementierung entscheidend geändert hat, und ob sich daraus Veränderungen der erwarteten Effekte ableiten lassen.

#### 4.2.11. Impact Evaluation

Diese, auf ursprünglich nicht erwartete Effekte abzielende Evaluationsform, ist ebenfalls primär in der Routinephase anzusiedeln und grundsätzlich explorativ angelegt. Monitoring unerwünschter Nebeneffekte, ein Aspekt der bereits angesprochen wurde, findet in der Routinephase als Teil der Impact Evaluation statt.

#### 4.2.12. Evaluation der Wirtschaftlichkeit

Die Evaluation der Wirtschaftlichkeit beinhaltet nach (Yates, 1994) Kosten-Nutzenanalysen (Cost-Benefit Analyses - CBA) und Kosten-Effektivitätsanalysen (Cost-Effectiveness Analyses - CEA). Ziel der CBA ist es, Programmkosten zu rechtfertigen, und Ziel der CEA ist es, Grundlagen für die Entscheidung zwischen konkurrierenden Programmen zu liefern.

CBA vergleicht Kosten eines Programms mit den positiven Auswirkungen (Nutzen). Dabei ist es nötig, sowohl Kosten als auch Nutzen über eine gemeinsame Einheit - üblicherweise Geld - zu quantifizieren. Die zentrale Frage ist: „Zahlt sich das Programm aus?“

CEA dient dazu, gleichwertige Programme zu vergleichen. Die zentrale Frage ist: „Welches Programm ist überlegen?“ Mit CEA ist es möglich Programme zu vergleichen, deren Kosten und Nutzen nicht auf eine gemeinsame Dimension reduzierbar sind. Der Kunstgriff ist, dass man alle Dimensionen außer einer konstant hält. So kann man z.B. auf die Einbeziehung der monetären Kosten verzichten, wenn man zwei gleich teure Programme hinsichtlich nicht-monetärer Effekte vergleicht oder umgekehrt sich auf die monetären Kosten konzentrieren, wenn die Annahme gerechtfertigt ist, dass beide Programme identische nicht-monetäre Effekte erzielen.

**Literatur:**

- Clayton, R.R., Cattarello, A.: Prevention Intervention Research: Challenges and Opportunities. in: Leukefeld, C.G., Bukovsky, W.J. (ed.): Drug Abuse Prevention Intervention Research: Methodological Issues. NIDA Research Monograph 107, Rockville, 1991
- Donabedian, A.: Explorations in Quality Assessment and Monitoring. Vol.1. The Definition of Quality and Approaches to its Assessment. Health Administration Press. Ann Arbor, 1980
- Popper, K.R. (1976): Logik der Forschung, sechste verbesserte Auflage. J.C.B.Mohr, Tübingen
- Scriven, M.: The Methodology of Evaluation. in: Tyler, R.W., Gagne, R.M., Scriven, M.(ed.): Perspectives of Curriculum Evaluation. Chicago, Rand-Mc.Nally, 1967
- Tukey, J.W.: Exploratory Data Analysis. Addison-Wesley, Reading, 1977
- Uhl, A.: Evaluation of Primary Prevention in the Field of Illicit Drugs - Definitions - Concepts - Problems. Results of an International Consensus Study within the COST-A6 Action of the European Union. in Springer, A. & Uhl, A.(Ed.): Evaluation Research in Regard to Primary Prevention of Drug Abuse. COST-A6 publication, Commission of the European Communities, Brussels, 1998
- Yates, B.T.: Toward the Incorporation of Costs, Cost-Effectiveness Analysis, and Cost-Benefit Analysis Into Clinical Research. J-Consult-Clin-Psychol., 62, 4, 729-736, 1994.