

Karin Haubrich/Christian Lüders

Evaluation – mehr als ein Modewort?¹

1 Evaluation zwischen hoher Relevanz und begrifflicher Unschärfe

Seit etwa Mitte der neunziger Jahre des letzten Jahrhunderts hat in nahezu allen Gesellschaftsbereichen, Politik- und Praxisfeldern das fachliche und öffentliche Interesse an Evaluation sprunghaft zugenommen. Vor allem im Bereich der Kindheits-, Jugend- und Familienpolitik bzw. in den jeweiligen Praxisfeldern lassen sich seit kurzem allerorten entsprechende Nachfragen beobachten. Über erhöhte Aufmerksamkeit dürfen sich schließlich auch jene Politik- und Praxisfelder erfreuen, für die Evaluation in unterschiedlichen Formen traditionell ein selbstverständlicher Bestandteil der politischen und fachlichen Steuerung war. Dies gilt z.B. für die Arbeitsmarktpolitik und die Bildungspolitik.

Ein nur cursorischer Blick z.B. in die englische und amerikanische Diskussion belegt, dass dort die Beschäftigung mit Evaluation auf eine sehr viel ältere, breitere und selbstverständlichere Tradition zurück blicken kann (vgl. z.B. *Beywl* 1988, S. 14 ff.; *Mertens* 2000; *Worthen/Sanders/Fitzpatrick* 1997, S. 26 ff.), sodass es wert wäre, der Frage nachzugehen, was eigentlich das neue Interesse an Evaluation hierzulande motiviert hat. Es ist hier aber nicht der Ort, dieser Frage im Detail nachzugehen. Festgehalten werden kann jedoch, dass knapper werdende öffentliche Mittel, die sichtbar werdenden Grenzen des Sozialstaates und seiner Verteilungsmechanismen in der bislang in Deutschland vertrauten Form sowie die fachliche Notwendigkeit, über die Effekte der eigenen Praxis begründet Auskunft geben zu wollen und zu müssen, eine wesentliche Rolle dabei gespielt haben. Nicht zuletzt dürften auch die systemimmanenten Steuerungsprobleme moderner Gesellschaften, die sichtbar werdenden Grenzen nationaler Politiken und wachsender Entscheidungsbedarf angesichts konkurrierender Optionen bei enger werdenden Handlungsspielräumen Gründe dafür sein, verstärkt Evaluationen im Sinne von empirisch begründeten Aussagen über Qualität und Nutzen als eine Grundlage für Entscheidungen einzufordern.²

Jenseits dieser Frage nach den Anlässen für das aktuelle Interesse an Evaluation, ist zu konstatieren, dass damit – wie fast immer in solchen Fällen – zwei Probleme verbunden sind: Erstens führt die allseitige Beliebtheit und fast inflationäre Verwendung des Begriffes dazu, dass seine Bedeutung zunehmend unscharf wird, was wiederum die Notwendigkeit mit sich bringt, jedes Mal erläutern zu müssen, was man denn eigentlich meint, wenn man von Evaluation spricht. Da mutieren schon mal normale Teambesprechungen, Jahresberichte und Ein-

1 Wir bedanken uns herzlich bei Gerlinde Struhkamp, die die intensiven Diskussionen zu diesem Artikel durch engagierte Stellungnahmen und hilfreiche Anregungen sehr bereichert hat.

2 Jüngst hat *T.A. Schwandt* von Evaluation als einer Institution »with distinct cultural authority« (2002, S. 175) gesprochen, um darauf hinzuweisen, in welcher weitreichender Weise Evaluationen in modernen Industriegesellschaften die Beziehungen der Menschen zu ihrer sozialen Umwelt und ihren Institutionen beeinflussen. *Schwandt* begreift Evaluation als eine der »ruling apparatuses of society« (ebd.) bzw. »as a social practice of power« (S. 172). Auch wenn man derartigen Globaldiagnosen skeptisch gegenüber steht, weil man befürchten muss, dass als Nächstes jemand die Evaluations-Gesellschaft ausruft, verweisen sie doch immerhin auf einen diskussionswürdigen Aspekt: Angesichts der Steuerungsprobleme moderner Gesellschaften avanciert Evaluation zu einer Strategie, die insofern eine Lösung zu versprechen scheint, als sie dafür steht, begründetes Wissen über Qualitäten, Effekte und Kosten-Nutzen-Beziehungen erzeugen und damit einen Beitrag zur Reduktion von Komplexität und von Entscheidungsalternativen leisten zu können.

richtungsdokumentationen, der Einsatz einfacher Fragebögen zum Ende eines Workshops oder Zufriedenheitsabfragen bei Besucherinnen und Besuchern zu vermeintlich aussagekräftigen Evaluationen der eigenen Praxis. Es ist deshalb nicht weiter verwunderlich, dass selbst in der Fachdebatte über Evaluation die Definitionen des Begriffs in einem so erheblichen Maße variieren, dass man schon geneigt ist *J. L. Franklin* und *J. H. Trasher* (1976, S. 20 zitiert nach *Wottawa/Thierau* 1998, S. 13) Glauben zu schenken: »To say that there are as many definitions as there are evaluators is not too far from accurate«.

Zusätzlich erschwert wird die Sache jedoch nicht nur dadurch, dass sich unter dem Begriff Evaluation sehr heterogene, im Detail keineswegs kompatible Dinge verbergen, sondern auch weil es eine breite Forschungspraxis gibt, die zwar nicht unter dem Label Evaluation firmiert, der Sache nach aber als solche gelesen werden kann und meistens in der Fachdiskussion und Öffentlichkeit entsprechend verhandelt wird. Die derzeit bekanntesten Beispiele dafür sind Studien wie PISA, TIMSS, IGLU, LAU, MARKUS. Schon die mitunter aufkommende Debatte, ob und inwiefern es sich dabei um Schulleistungsstudien im strikten Sinn des Wortes handelt, indiziert das Problem. Verwendet man den Begriff Schulleistungsstudien, unterstellt man, dass die gemessenen Kompetenzunterschiede primär der Schule zuzuschreiben sind. Man nimmt also eine Ursache-Wirkungs-Zuschreibung vor. Begriffe wie Leistungsvergleiche, Kompetenzmessung, Grundbildungstest, Vergleichsstudie, Benchmarking u.ä. legen sich an dieser Stelle zunächst nicht fest. Unabhängig davon haben sowohl die Fachdiskussion als auch die öffentliche Debatte die PISA-Studie wie auch andere ähnlich gelagerte Studien als empirische Analysen verstanden und genutzt, die im Hinblick auf die untersuchten Aspekte Auskunft über die Leistungsfähigkeit des Schulsystems geben und die Daten zur Verfügung stellen, an Hand derer die Schule in Bezug auf ihren »outcome«, in diesem Fall gemessen an Hand der empirisch beschreibbaren Kompetenzniveaus der Schülerinnen und Schüler, bewertet werden kann. M.a.W.: Obwohl nicht selbst unter diesem Label firmierend, wurden diese Studien als Evaluationsstudien verwendet.

Zweitens werden mit Evaluation inzwischen derart unterschiedliche Aufgaben, überzogene Erwartungen und zu einem nicht geringen Teil »unheilige« Interessen verbunden, dass jeder, der sich darauf einlässt, bedroht ist daran zu scheitern. Da soll Evaluation einmal die zielgenaue Verwendung öffentlicher Mittel belegen, während vonseiten der politischen Opposition genau das Gegenteil erwartet wird. Für viele Träger fungiert sie als Legitimationsinstanz der eigenen Praxis, während man gelegentlich den Verdacht nicht los wird, dass von öffentlicher Seite immer wieder Sparzwänge das Interesse an Evaluation motivieren. Es wird gefordert, möglichst allgemeine Aussagen über die Wirkungen bestimmter Arbeitsformen zu gewinnen, während die Beteiligten eher an einer organisationsbezogenen Weiterentwicklung und Beratung interessiert sind. Evaluation soll ebenso einer weiteren Professionalisierung pädagogischer und sozialpädagogischer Fachpraxis den Weg ebnen, wie zugleich einer rationaleren, im Sinne von wirtschaftlicheren Steuerung des Praxisfeldes die wesentlichen Ansatzpunkte liefern.

Angesichts dieser nicht gerade übersichtlichen Diskussionslage ist es notwendig, ein wenig Licht auf die unterschiedlichen Formen und Verständnisse von Evaluation zu lenken und diese zumindest in einigen grundlegenden Aspekten zu sortieren, um dann im Einzelfall verstehen und klären zu können, was im jeweiligen Kontext Evaluation bedeutet.

2 Zwei Seiten einer Medaille: Evaluation in Alltag und Wissenschaft

Zunächst einmal bedeutet Evaluation in einem alltagsweltlichen Verständnis, dass irgendetwas durch irgendwen in irgendeiner Weise und an Hand von, nicht immer klar ausgewiesenen, Kriterien bewertet wird (vgl. *Kromrey* 1995, S. 313). Evaluation in diesem Sinne ist ein integraler Bestandteil alltäglicher Handlungspraxis. Kein Tag verläuft ohne eine Vielzahl impliziter oder expliziter Bewertungsprozesse: Das Abendessen wird ebenso bewertet wie der Fernsehfilm oder der Anzug und Schlips des Kollegen. Jeder Entscheidungsprozess impliziert ein Abwägen und Bewerten von bisherigen Erfahrungen, verfügbaren Wissensbeständen und möglichen Alternativen, von erwünschten Folgen und möglichen Nebenwirkungen an Hand von wie auch immer gearteten Kriterien. Im Kern geht es dabei immer wieder um die Bewertung von Prozessen und Zuständen, gelegentlich um die Frage, wie das zu bewertende Etwas ermöglicht bzw. verursacht wurde, und – wenn Aufwendungen und Kosten mit im Spiel sind – nicht selten auch um Mittel-Zweck-Relationen. Das alles kann als alltagsweltliche Evaluation bezeichnet werden, wenn damit ein bewusstes Abwägen und eine an Kriterien orientierte Prüfung verbunden ist. In diesem Sinne wären Bewertungen, die auf »Zufällen« beruhen, nicht als Evaluation zu bezeichnen – z.B. wenn man zur Erleichterung einer Bewertungsentscheidung eine Münze werfen oder auf göttliche Eingebung warten würde (vgl. *Lee* 2000, S. 130).

An dieses alltagssprachliche Verständnis von Evaluation schließt sich das professionelle und das wissenschaftliche Verständnis des Begriffes an. Evaluationen als professionelle Praxis und als eine Form sozialwissenschaftlicher Forschung sind charakterisiert durch ein systematisches und transparentes Vorgehen zur Bestimmung des Wertes eines Gegenstandes auf der Grundlage empirisch gewonnener Informationen. Sowohl die gewonnenen Daten als auch die daraus gezogenen Schlussfolgerungen und Bewertungen müssen nachvollziehbar und gültig bzw. valide sein. Diese Bestimmung von Evaluation stellt gewissermaßen einen minimalen Konsens darüber dar, was unter einer Evaluation im wissenschaftlichen Sinne zu verstehen ist.

Daher kann man – dem Vorschlag *H. Wottawas* und *H. Thieraus* (1998, S. 14) folgend – als zentrale Kennzeichen wissenschaftlicher Evaluation festhalten:

- Evaluationen haben immer etwas mit »Bewerten« bestimmter Untersuchungsgegenstände zu tun. Diese Evaluationsgegenstände – im englischen Sprachgebrauch auch »evaluand« (von deutschsprachigen Autoren Evaluandum genannt, vgl. *Widmer* 2000, S. 78 oder *Martinuzzi/Meyer* 2000, S. 457) – können sehr unterschiedlicher Art sein. Dazu zählen vor allem Politiken, Programme, Projekte, Maßnahmen, Leistungen, professionelle Handlungsansätze, Produkte, Personen, Organisationen, Technologien und die Forschung selbst (vgl. *Deutsche Gesellschaft für Evaluation* 2002, S. 13).
- Evaluationen basieren auf systematischer Informationsgewinnung. Sie verwenden professionelle Verfahren der Informationsbeschaffung bzw. empirisch-wissenschaftliche Methoden, insbesondere die qualitativen und quantitativen Methoden empirischer Sozialforschung (vgl. ebd.). Dabei besteht Einigkeit, dass die eingesetzten Methoden dem aktuellen Stand der professionellen Praxis bzw. der Forschung entsprechen sollten (vgl. *Wottawa/Thierau* 1998, S.14; *Kromrey* 2000, S. 22).

Strittig ist – wenn auch nicht bei vielen –, ob und inwiefern diese beiden Bestimmungsmomente mit dem Anspruch verbunden sein sollen und können, dass Evaluationen auf bestimmte

Verwendungszwecke oder Ziele hin systematisch anzulegen seien (vgl. dazu unten Abs. 4). Mit dieser Erweiterung der Definition von Evaluation wird ihr eine spezifische Funktion zugeschrieben. In diesem Sinne ergänzen *H. Wottawa* und *H. Thierau* ihre Definition um ein drittes Bestimmungsmoment:

- »Evaluation ist ziel- und zweckorientiert. Sie hat primär das Ziel, praktische Maßnahmen zu überprüfen, zu verbessern oder über sie zu entscheiden« (*ebd.* S. 14). Evaluationen verfolgen also die Bewertung eines Untersuchungsgegenstandes für konkrete und benennbare Verwendungszwecke. Diese sind neben der Erweiterung von Wissensbeständen (z.B. welche pädagogischen Ansätze in Bezug auf welche Klienten in spezifischen Kontexten welche Erfolge zeigen) vor allem die systematische Verbesserung oder Gestaltung eines Evaluationsgegenstandes (z.B. zur Optimierung von Konzepten, Prozessen oder Strukturbedingungen) sowie die bilanzierende Bewertung als Grundlage für Entscheidungen (beispielsweise die Fortsetzung oder Einstellung, die Einführung, Ausweitung oder Beschränkung eines Programms betreffend).³

Jedem dieser Merkmale wohnt unvermeidlich ein erheblicher Interpretationsspielraum inne. In der Fachdiskussion hat sich daraus eine so große Spannbreite möglicher Auslegungen jedes der drei genannten Kennzeichen ergeben, dass es fast schon überraschen mag, dass es für die Erstellung zentraler Regelwerke (wie die »Standards für Evaluation«⁴ der Deutschen Gesellschaft für Evaluation (2002) oder die »Guiding Principles for Evaluators« der American Evaluation Association, vgl. *Beywl* 1996) gelingen konnte, einen gemeinsamen Nenner zu finden.

Um im Folgenden nicht das gesamte, zum Teil auch verwirrende Spektrum⁵ aufrollen zu müssen, konzentrieren wir uns auf folgende Aspekte:

- Wir beschränken uns im Folgenden auf die Evaluation von Politiken, Programmen, Projekten, Praxisformen, Konzepten und Strategien. Unberücksichtigt bleiben dabei alle auf Personen bezogene Formen der Evaluation. Auch der spezifische Bereich der Evaluation im Wissenschaftssystem, also z.B. von Studiengängen und Fakultäten, wird hier nicht zum Thema gemacht.
- Zweitens fokussieren wir unsere Darstellung auf jene Formen von Evaluation, die man mit guten Gründen als eine Form der Forschung oder mindestens als forschungsbasierte professionelle Praxis bzw. bezeichnen kann, wobei wir ein größeres Gewicht auf den ersten Aspekt legen. Der Grund hierfür ist einfach: Evaluation als professionelle Praxis ist mittlerweile ein derart breites und unübersichtliches Feld, das von der Organisationsberatung bis zur Konsumforschung, von der Zufriedenheitsmessung von allem und jedem bis hin zu Zuschauerquotenanalysen reicht, sodass allein die Vermessung dieses Feldes Anstrengun-

3 Vgl. hierzu weiter unten Kapitel 6.2.

4 Die Standards für Evaluation formulieren Voraussetzungen für und Anforderungen an die Qualität von Evaluationen. Demnach sollen Evaluationen nützlich, durchführbar, fair und genau sein. Diese Standards wurden von der Deutschen Gesellschaft für Evaluation in Anlehnung an die Standards der Schweizerischen Evaluationsgesellschaft (*Widmer/Landert/Bachmann* 2000) und des Joint Committee on Standards for Educational Evaluation (1999) entwickelt. Sie wurden von der Mitgliederversammlung der Deutschen Gesellschaft für Evaluation im Jahr 2001 verabschiedet und befinden sich derzeit in einem Revisionsprozess.

5 Immerhin konnte *M.Q. Patton* (1997) allein 68 Typen von Evaluation an Hand des Kriteriums »ways of focusing evaluation« auflisten (*Patton* 1997, S. 192 ff.). Kombiniert mit anderen Merkmalen kommt man bei der Lektüre des Buches von *M.Q. Patton* auf weit über 300 verschiedene Typen, und es ist zu befürchten, dass dies keineswegs alle sind.

gen eigener Art notwendig machen würde, die den Rahmen dieser Überblicksdarstellung sprengen würden.

- Drittens werden wir uns nur am Rande mit den unterschiedlichen Strategien der Selbstevaluation befassen (vgl. von Spiegel 1993). Ohne diese Form der Bewertung von Fachpraxis abwerten zu wollen, kommt sie insofern schnell an systematische Grenzen, als die durch Selbstevaluationen gewonnenen Aussagen – z.B. durch die Fachkräfte – genau so weit reichen wie die Wissens- und Erfahrungshorizonte der beteiligten Personen. In diesem Sinne betrachten wir selbstevaluative Strategien als eine wichtige Erkenntnisquelle. In vielen Fällen erweisen sie sich als unverzichtbare Ergänzung zu externen Evaluationen. Sie können jedoch externe Evaluationen nicht ersetzen.

3 Bewerten als Kernaufgabe von Evaluation

Wie angedeutet, verbirgt sich unter dem Label Evaluation eine Vielzahl von Konzepten und Strategien, um Verfahren, Maßnahmen, Programme und Organisationen bzw. Institutionen etc. zu beschreiben und vor allem zu bewerten. Evaluationen können unterschiedliche Rollen oder Funktionen für die Wissenschaft und die Praxis übernehmen, eines bleibt jedoch immer das Ziel: Nämlich, dass Evaluationen den Wert eines Gegenstandes zu bestimmen haben.⁶ Evaluationen machen damit explizit, was andere Formen der Forschung eher zu vermeiden trachten. Denn spätestens seitdem das Prinzip der Wertfreiheit als allgemeines Gütekriterium für wissenschaftliche Forschung Anerkennung findet (vgl. Weber 1998; Orig. 1917), unternehmen Wissenschaftlerinnen und Wissenschaftler alles Erdenkliche, um dem Anspruch gerecht zu werden, ihre »(...) praktisch wertende, d.h. (die) Tatsachen (...) als erfreulich oder unerfreulich *beurteilende*, in diesem Sinne »bewertende« Stellungnahme (...)« (Weber 1988, S. 500; kursiv *im Orig.* gesperrt gedruckt) zu vermeiden. Weil es eine wertfreie Sprache aber nicht gibt, wissenschaftliche Ergebnisse dennoch nur sprachlich vermittelt werden können, stoßen sie dabei immer wieder an Grenzen⁷. Nichtsdestoweniger ist nach wie vor, z.B. bei der Erstellung von Forschungsberichten und der Präsentation von Forschungsergebnissen, die Unterscheidung zwischen und Trennung von beschreibender Rekonstruktion bzw. Deskription einerseits und fachlicher, d.h. ebenso sachlich wie auch normativ fundierter Bewertung andererseits für das wissenschaftliche Selbstverständnis leitend. Davon unabhängig ist es natürlich immer möglich, Beschreibungen von empirischen Sachverhalten als Evaluationen zu verwenden. So wird auch die normativ enthaltsamste Beschreibung, z.B. sozialer Ungleichheiten, unweigerlich dazu tendieren, von der ein oder anderen Seite als Evaluation z.B. einer gescheiterten Sozialpolitik genutzt zu werden. Umgekehrt – und dies wird bislang zu wenig gewürdigt – enthalten Evaluationsstudien, wenn sie handwerklich ordentlich durchgeführt worden sind, üblicherweise eine bemerkenswerte Menge an empirischem Wissen über Strukturen, Ablaufprozesse und Verfahren, das man auch ohne die daran anschließenden Bewertungen

6 Damit schließen wir uns *M. Scrivens* prominenter Definition von Evaluation an (1991a, S. 139). Allerdings kann man in der Literatur wesentlich engere Definitionen von Evaluation in dem Sinne finden, dass sie die Bewertungsdimension auszuschließen versuchen und Evaluation auf die Messung von Wirkungen oder Effekten einer Maßnahme, eines Programms etc. konzentrieren (z.B. *Wollmann* 2000) oder gar auf die ex post Beschreibung von Effekten und Wirkungen (vgl. European Commission 1999, S. 17). Derartige Konzeptionen streben eine »neutrale« Analyse von Ursache-Wirkungszusammenhängen an, ohne diese selbst zum Gegenstand fachlicher Bewertungen zu machen.

7 Vgl. hierzu die ausführlichen Auseinandersetzungen mit dieser Thematik bei *M. Scriven* (1983) sowie *E.R. House* und *K.R. Howe* (1999).

z.B. für die Theoriebildung oder die empirische Überprüfung theoretischer Annahmen nutzen könnte.

Wenn zuvor gesagt worden ist, dass Evaluation den Wert (»value«) eines Gegenstandes zu bestimmen habe, dann bedarf es an dieser Stelle der Einführung weiterer Unterscheidungen. Dieser kann nämlich danach bemessen werden, welchen Wert oder Nutzen (»worth«) er in bestimmten zeitlich und räumlich begrenzten Kontexten für bestimmte Zielgruppen bzw. Beteiligte und Betroffene⁸ (»stakeholder«) hat oder ob und inwiefern er in sich stimmig und fachwissenschaftlich abgesichert ist (»merit«, verstanden als Güte oder Qualität)⁹.

Allerdings gibt es unterschiedliche Auffassungen darüber, wer die Bewertung des Gegenstandes letztendlich vornehmen sollte und welche Kriterien dabei zugrunde zu legen sind. Während *M. Scriven* postuliert, dass es Aufgabe der Evaluatoreninnen und Evaluatoren ist, Bewertungen vorzunehmen (*Scriven 1991b*, zit. nach *Patton 1997*, S. 316), begrenzt *D. Stufflebeam* die Rolle der Evaluation darauf, eine allgemeine »audience« zu unterstützen, die Güte oder den Nutzen eines Gegenstandes zu bestimmen (2001, S. 11). Eine ähnliche Position findet sich bei Vertreterinnen und Vertretern responsiver und dialogorientierter Evaluationen, die auf die Informationsbedarfe der Beteiligten und Betroffenen reagieren. Diesem Verständnis zufolge besteht die Aufgabe der Evaluation darin, unterschiedliche Wertpositionen zu erheben und so transparent wie möglich zu beschreiben (*Stake 1995*; *Guba/Lincoln 1989*), damit die Leserinnen und Leser sich ein differenziertes eigenes Werturteil bilden bzw. in dialogischen Prozessen zu einem Urteil finden können.¹⁰

Wenn Evaluatoreninnen und Evaluatoren auf der Basis wissenschaftlicher Erhebungen Bewertungen vornehmen oder solche vorbereiten, so können sie grundsätzlich zwei alternative Wege beschreiten: Entweder sie erheben, beschreiben und analysieren die Wertpositionen der Untersuchungspopulation oder sie nehmen selbst auf der Grundlage transparenter und nachvollziehbarer Bewertungsverfahren eine Bestimmung des Wertes, des Nutzens, der Güte oder der Relevanz des Untersuchungsgegenstandes vor.¹¹ In beiden Fällen ist es erforderlich, dass die Bewertungskriterien soweit als möglich ausgewiesen werden. *E. Vedung* unterscheidet dabei mit Bezugnahme auf *W.R. Shadish*, *T.D. Cook* und *L.C. Leviton* (1991) zwischen deskriptiven und präskriptiven Kriterien. »Präskriptive Theorien der Bewertung gehen davon aus, dass bestimmte Werte anderen überlegen sind, während deskriptive Theorien Werte abbilden, die von anderen vertreten werden, ohne diese in Frage zu stellen oder zu behaupten, die eine Wertvorstellung sei die beste oder besser als einige der Alternativen« (*Vedung 1999*, S. 214). Deskriptive Kriterien wären u.a. Zielerfüllung, Qualitätsvorstellungen und Erwartungen von Klienten oder professionelle Auffassungen von Güte Merkmalen. Präskriptive Kriterien wären

8 In der deutschen Fachterminologie spricht man von Beteiligten, die eine aktive Rolle in Bezug auf den Evaluationsgegenstand haben (wie Auftraggeber oder Mitarbeiter bzw. Mitarbeiterinnen in einem Projekt) und Betroffenen, die wenig Einfluss auf den Untersuchungsgegenstand nehmen können, aber durch ihn entweder einen Vorteil oder einen Nachteil erfahren (z.B. als Zielgruppen des Programms oder als davon Ausgeschlossene).

9 *M. Scriven* hat die auf *E. Guba* und *Y. Lincoln* (1981) zurückgehenden klassischen Kategorien Güte (merit) und Wert/Nutzen (worth) in neueren Veröffentlichungen (vgl. *Scriven 2003*, S. 28) um die Dimension der Signifikanz bzw. Relevanz erweitert, die insbesondere für die Bestimmung der Qualität von Forschung von Bedeutung ist.

10 Siehe hierzu ausführlicher die Systematisierung unterschiedlicher Evaluationsmodelle nach ihrer Berücksichtigung von Werten *Beywl, W./Speer, S./Kehr, J.* (2004).

11 Damit wird jedoch die Frage aufgeworfen, wie wissenschaftliche Erhebungsmethoden, die in erster Linie für die Beschreibung der Wirklichkeit und nicht für deren Bewertung entwickelt und begründet wurden, für Evaluationszwecke weiterzuentwickeln sind (vgl. hierzu ausführlicher *Lüders/Haubrich 2003*).

demgegenüber präferierte gesellschaftliche Wertmaßstäbe, wie Bedürfnisse der Klienten, das Gemeinwohl oder Gerechtigkeit (vgl. *ebd.*).

Verwenden Evaluatoren und Evaluatorinnen deskriptive Kriterien, so legen sie die normative Position Dritter ihrer Evaluation zugrunde. Dies ist u.a. der Fall, wenn »wertedistanziert« (*Beywl/Speer/Kehr* 2004) die Zielerreichung einer Intervention untersucht oder der Frage nachgegangen wird, ob die erhofften Wirkungen einer Intervention, um diagnostizierte Defizite zu beheben, tatsächlich eintreten. Deskriptive Kriterien kommen auch zur Anwendung, wenn sich die Evaluation an fachlichen Standards oder in einem partizipativen Vorgehen an den Maßstäben der Zielgruppen eines Programms orientiert. Demgegenüber erfordern präskriptive Kriterien, dass die Evaluatoreninnen oder Evaluatoren selbst, indem sie sich bspw. auf eine ethische oder politische Theorie beziehen, bestimmte Wertpositionen bevorzugen und dies begründen.

So formal und akademisch diese Unterscheidungen zunächst auch anmuten mögen, für die Forschungspraxis und erst recht für den öffentlichen Umgang mit Evaluationsergebnissen sind sie von größter Bedeutung. Denn neben die Frage, ob und inwiefern die zu Grunde liegenden Daten den Standards wissenschaftlicher Forschung entsprechen, tritt nun in Evaluationen das Problem der in Anspruch genommenen normativen Grundlagen und ihrer Begründung. Abgesehen davon, dass die dafür notwendigen Bedingungen bekanntermaßen immer Gefahr laufen einem infiniten Begründungsregress ausgesetzt zu werden, der im Zweifelsfall nicht bei den christlichen Grundüberzeugungen des Abendlandes endet, steht jede Bewertung unter den Bedingungen einer weltanschaulich pluralistischen Gesellschaft vor dem unvermeidbaren Problem, dass man es ggf. auch »ganz anders« sehen könnte. Selbst wenn man nicht der These anhängt, dass Bewertungen beliebig sind, so ist doch zu konstatieren, dass die Evaluationsdiskussion es bislang versäumt hat, sich den damit zusammenhängenden Fragen systematisch zu stellen. Aus unserer Sicht kann die Lösung jedoch nicht im Suchen bzw. Postulieren essentialistischer Wertpositionen bestehen. So bleibt als Ausweg nur das Setzen auf demokratische Verfahren der Aushandlung und der diskursiven Rechtfertigung (vgl. *Habermas* 1992 sowie *House/Howe* 1999). In diesem Sinne ist aus unserer Sicht jede Evaluation in doppelter Hinsicht und unvermeidlich auf demokratische Strukturen verwiesen: Zum einen bedarf es des demokratisch legitimierten Auftrages und zum anderen müssen die Verfahren und vor allem die in Anspruch genommenen Kriterien über demokratische Verfahren legitimiert sein. So argumentieren z.B. auch *M.M. Mark*, *G.T. Henry* und *G. Julnes* (2000).

4 Evaluationen sind zielorientiert – müssen sie auch nützlich sein?

Auf den ersten Blick scheint es nahe zu liegen, diese Frage mit einem eindeutigen »Ja« zu beantworten. Denn aus welchem Grunde sollten Evaluationen denn in Auftrag gegeben werden, wenn nicht zur Gewinnung nützlicher Ergebnisse? Und in der Tat gewannen insbesondere in den 80er Jahren Definitionen von Evaluation an Bedeutung, die vor allem auf die Nützlichkeit und faktische Verwendung von Evaluationsergebnissen abzielten.¹² Diese Ausrichtung der Evaluationsfachdebatte, wie sie vor allem in Nordamerika geführt wurde, lässt sich zurückführen auf die angesichts hoher öffentlicher Erwartungen enttäuschende Erfahrung, dass Evaluationsergebnisse oft als wenig nützlich wahrgenommen wurden bzw. de facto nicht genutzt wurden – zumindest nicht in dem Sinne, wie es die Auftraggeber ursprünglich erwartet und

¹² Siehe hierzu die Debatte zwischen *M.Q. Patton* und *C.H. Weiss*, zusammengefasst in *Alkin M.C.* (Hrsg.): *Debates on Evaluation*. Newbury Park, London, 1990, S. 171ff.

die Evaluatorinnen und Evaluatoren erhofft hatten (vgl. *Weiss 1998*). *M.Q. Patton* ist der prominenteste Vertreter dieser Position: »I use the term *evaluation* quite broadly to include any effort to increase human effectiveness through systematic data based inquiry« (1990, S. 11). Betont wird in dieser Definition die Steigerung der Wirksamkeit menschlichen Handelns auf der Basis systematischer empirischer Analysen. Evaluation wird über ihre Funktion oder genauer gesagt, ihre angestrebte Verwendung im Sinne der Verbesserung der jeweiligen Konstellation definiert.¹³ In einem ähnlichen, in den Details etwas differenzierteren Sinne definiert auch *W. Beywl* Evaluation: Sie »(...) zielt auf die Verbesserung gesellschaftlicher Praxis, indem sie Programme, Maßnahmenbündel oder Materialien systematisch, d.h. methodisch angeleitet und an Gütekriterien überprüfbar, beschreibt und bewertet« (1988, S. 1). Damit wird das durch Evaluationen zur Verfügung gestellte Wissen darauf festgelegt, dass es in konkreten Verwendungskontexten genutzt werden kann: Entweder als Grundlage für die Verbesserung des Evaluationsgegenstandes (formative Evaluation) oder um eine Steuerungsentscheidung auf Basis einer bilanzierenden Bewertung (summative Evaluation) zu ermöglichen oder im Sinne einer längerfristigen Optimierung fachlicher Praxis (wissensgenerierende Evaluation) (*Beywl 2001*, S. 21).¹⁴

Gemeinsam ist diesen Definitionen, dass die Optimierung gesellschaftlicher Praxis in das Zentrum der Funktionsbestimmung von Evaluation gerückt wird. Auch wenn diese evaluations-theoretische Orientierung zumindest in ihren Grundzügen durch die »Nützlichkeitsstandards« der »Standards für Evaluation« eine deutliche Bestätigung erfährt (vgl. Joint Committee on Standards for Educational Evaluation/Sanders 1999; Deutsche Gesellschaft für Evaluation 2002), gibt es gute Gründe, sie auf ihre Plausibilität hin zu diskutieren. Dabei sind für uns vor allem zwei Argumente von Bedeutung:

- Aus einer normativen Perspektive stellt sich die Frage, ob das, was aus einer bestimmten Perspektive als nützlich eingeschätzt wird, denn wirklich hilfreich im Sinne einer Optimierung gesellschaftlicher Praxis ist. Oder noch allgemeiner formuliert: wer definiert wie den jeweiligen Nutzen und wie kann entschieden werden, was jeweils als nützlich zu gelten hat. Wenn die Auftraggeberin die Mafia ist – um ein scheinbar absurdes, in der Sache und in einigen Landstrichen Europas jedoch realistisches Beispiel zu nennen –, können für sie Ergebnisse nützlich sein, die für andere geradezu schädlich sind.¹⁵ Einen Ausweg aus diesem Problem suchen Evaluationstheoretiker, die Evaluationen entweder in den Dienst einer emanzipatorischen Praxis (»advocative evaluation«) stellen und von daher betonen, dass Evaluationen vor allem dazu beitragen sollten, gerade gesellschaftlich benachteiligten Gruppen eine Stimme zu verleihen (z.B. ethnischen Minderheiten), oder die Evaluation an demokratische Entscheidungsprozesse koppeln und sie dabei als ein Instrument im Dienst sozialer Verbesserungen (»social betterment« vgl. *Mark/Henry/Julnes 2000*) definieren.
- Systematisch noch folgenreicher ist eine zweite Überlegung. Sowohl die Ergebnisse der sozialwissenschaftlichen Verwendungsforschung (vgl. *Beck/Bonß 1989; Lüders 1993*) als auch die alltägliche Erfahrung (z.B. im Bereich der Politikberatung, der Verwendung von Forschungsergebnissen in der Öffentlichkeit etc.) machen unmissverständlich deutlich, dass die Praxis sozialwissenschaftliches Wissen *autonom* nutzt. Autonome Nutzung bedeu-

13 Der von ihm entwickelten Evaluationsansatz der »utilization-focused evaluation« (vgl. *Patton 1997*) spitzt diese Position noch zu, indem dort betont wird: »(...) evaluation is done for and with specific, intended primary users for specific, intended uses« (*ebd.*, S. 23).

14 Diese Aspekte werden weiter unten in Kap. 6.2. ausführlicher diskutiert.

15 Vgl. hierzu *House, E.R./Howe, K.R.* (1999, S. 16).

tet, dass der Umgang mit wissenschaftlichem, d.h. theoriegeleitet gewonnenem und methodisch geprüftem Wissen den Eigenlogiken der Praxis folgt und dass der praktische Umgang mit dem sozialwissenschaftlichen Wissen *nicht* von Seiten des Wissenschaftssystems plan-, vorhersag- und steuerbar ist. Dabei hängt die Verwendung sozialwissenschaftlichen Wissens offensichtlich nicht unbedingt von der Qualität dieses Wissens ab. Es gibt keinen Grund, warum dies nicht auch für die Evaluationsforschung gelten sollte. Dies hat zur Konsequenz, dass ein Verständnis von Evaluation z.B. im Sinne von *M.Q. Patton* oder *W. Beywl* unvermeidlich in grundlegende verwendungstheoretische Probleme führt (vgl. hierzu *Lüders/Haubrich* 2003).

Vor diesem Hintergrund plädieren wir dafür, dass deutlicher zwischen dem Fakt, dass Evaluationen in aller Regel einen Auftraggeber haben, der die Ergebnisse in einer bestimmten Art und Weise zu verwenden beabsichtigt, und dem Anspruch an Evaluationen, nutzbare und brauchbare Ergebnisse zu erzielen, unterschieden wird. Sicherlich ist es angesichts der besonderen institutionellen Rahmenbedingungen von Evaluation in aller Regel sinnvoll, diesen Nutzungsinteressen Rechnung zu tragen, indem Evaluationen so angelegt werden, dass die erwarteten Verwendungskontexte, aber auch die möglichen Nebenfolgen für andere Beteiligte und Betroffene berücksichtigt werden. Man könnte sagen, dass dies ein Teil der professionellen Praxis von Evaluation ist. Jedoch bleibt die Beurteilung der Nützlichkeit bzw. Nutzbarkeit¹⁶ von Evaluationen angesichts der Eigenlogik der Nutzung von Ergebnissen in der Praxis immer an den situativen Kontext gebunden. Dementsprechend stellt es aus unserer Sicht eine Überforderung evaluativer Forschungspraxis dar, wenn man ihr Nützlichkeit als wesentliches Qualitätskriterium abfordern würde. Angesichts der von Evaluatorinnen und Evaluatoren nicht beeinflussbaren oder gar steuerbaren Eigendynamik der Verwendungskontexte, z.B. in Politik und Praxis, würde dies auf ein Hase und Igel-Spiel hinauslaufen, das die Forschung nur verlieren kann. Anders formuliert: Der Anspruch auf Nützlichkeit setzt voraus, dass der Verwendungskontext für die jeweiligen Evaluationsergebnisse stabil und kalkulierbar bleibt oder zumindest ein Nachjustieren angesichts sich verändernder Rahmenbedingungen möglich ist. Dies ist nach aller Erfahrung eher selten und nur in begrenztem Maße der Fall.

5 Evaluation zwischen professioneller Praxis und Forschung

Die Diskussion, ob und inwiefern Evaluation nützliche Ergebnisse zu erzeugen habe, hat weitreichende Konsequenzen für das eigene Selbstverständnis. So dominiert in der deutschen Fachdebatte – zumindest in einschlägigen Lehr- und Handbüchern – ein Verständnis von Evaluation, das diese als eine Variante angewandter Sozialforschung für eine besondere Gruppe von Fragestellungen beschreibt (*Bortz/Döring* 2002; in ähnlicher Weise *Wottawa/Thierau* 1998 und *Kromrey* 2000). Dabei werden die Verwendungskontexte eher als restriktive Rahmenbedingungen verstanden: »Für sie (Evaluation) gilt jedoch als »hinderliche« Bedingung, dass im Mittelpunkt nicht die Logik einer auf Erkenntnisgewinnung, Verallgemeinerung und Übertragbarkeit ausgerichteten Wissenschaft steht, sondern die Handlungslogik eines auf »Erfolg« seines Tuns ausgerichteten Praxisprojektes, häufig eines staatlich-administrativen oder von sozialen Organisationen getragenen Interventionsprogramms« (*Kromrey* 1995, S. 315).

¹⁶ In Anlehnung an *M. Scriven* (gemäß seiner Ausführungen in dem Seminar »Hot Topics of Evaluation« auf der Jahrestagung der American Evaluation Association in Washington D.C., 2002) wird hier zwischen »nutzbaren« und »nützlichen« Ergebnissen unterschieden. Nutzbar meint in diesem Kontext nur, dass Ergebnisse in nicht planbaren Verwendungszusammenhängen potentiell genutzt werden können. Demgegenüber impliziert »nützlich« die tatsächliche Nützlichkeit der Ergebnisse für benennbare Personen in konkreten Kontexten.

Abgesehen davon, dass die Einsortierung als angewandte Sozialforschung u.E. mehr Probleme erzeugt als löst, weil alle wissenschaftstheoretischen Fragen unbeantwortet bleiben und nur ein diffuser Zwischenstatus zwischen Wissenschaft und Praxis postuliert wird, zeigt der Blick in den englischen Sprachraum, dass man auch zu ganz anderen Positionsbestimmungen kommen kann. Dort wird Evaluation gegenüber der Sozialforschung eher ein eigenständiger Charakter zugewiesen – zumindest im Sinne einer eigenständigen Profession mit einem breiten Praxisfeld, das aus unterschiedlichen wissenschaftlichen Disziplinen fachliches Wissen und Instrumente bezieht (vgl. bspw. *Worthen/Sanders/Fitzpatrick 1997*, S. 35 ff.). Daneben finden sich Positionen, die Evaluation als sich entwickelnde eigenständige Disziplin (z.B. *House, 1994*, zit. a.a.O., S. 49) oder als Transdisziplin (*Scriven 1991a, 2003*) beschreiben.

Aus unserer Sicht führen all diese Ansätze nicht viel weiter. Die Bestimmung von Evaluation als einer *professionellen Praxis*, die Verfahren und Wissen aus unterschiedlichen Disziplinen unter der eigenen Perspektive zusammenfasst und weiterentwickelt, beschreibt zutreffend einen großen und wichtigen Bereich. Nicht mehr enthalten in diesem Verständnis sind jedoch jene Formen von Evaluation, die vorrangig forschungsbasiert sind bzw. die sich selbst als *Forschung* verstehen. Der Begriff »Transdisziplin« (vgl. *Scriven 1991a*, S. 141 f.) wiederum ist u.E. zu diffus, weil er zwar darauf hinweist, dass Evaluation ein Verfahren ist, das nicht an die klassischen Disziplinen, also z.B. die Soziologie, Politologie, Erziehungswissenschaft, Psychologie etc. gebunden ist, aber mehr eben auch nicht. Völlig in die Irre führen die gelegentlich vorgetragenen Überlegungen, Evaluation als eigene »Disziplin« zu verstehen. Denn aus unserer Sicht lässt sich Evaluation nicht als Disziplin im Sinne der seit dem 19. Jahrhundert etablierten gegenstandsbezogenen Form der Organisation von Wissenschaft begreifen, so dass die Bedeutung des Disziplinbegriffes, wenn man ihn auf Evaluation beziehen würde, völlig entleert werden würde.

Demgegenüber etwas bescheidener schlagen wir vor, Evaluation als Dachbegriff für die Standards, Regeln und Verfahren der Bewertung von Gegenständen auf der Basis von Zustandsbeschreibungen zu verstehen. Gleichsam eine Ebene tiefer kann man dann zwischen der professionellen, weitgehend verberuflichten Evaluationspraxis auf der einen Seite und Evaluation als Teil des Wissenschaftssystems, i.e. Evaluationsforschung, auf der anderen Seite unterscheiden.¹⁷ In ähnlicher Weise unterscheiden auch andere Autoren und Autorinnen seit *E. Suchmans* 1967 erschienenen Buch »Evaluative Research« zwischen Evaluation und Evaluationsforschung (vgl. z.B. *Rossi/Freeman 1993*, S. 5 ff.; *Bortz/Döring 2002*, S. 101 ff.) – im Gegensatz etwa zu Ansätzen, die die Begriffe explizit oder de facto synonym verwenden (vgl. z.B. *Stockmann 2000 b*, S. 11; *Shaw 1999* oder *Clarke 2000*). In der jüngeren deutschsprachigen Diskussion wurde mit der Erstellung der Standards für Evaluation in Deutschland durch die Deutsche Gesellschaft für Evaluation (DeGEval) insofern eine etwas anders gelagerte Unterscheidung zwischen Evaluation und Evaluationsforschung getroffen, als – aus unserer Sicht in irritierender Weise – Evaluationsforschung mit Forschung *über* Evaluation gleichgesetzt wird (Deutsche Gesellschaft für Evaluation 2002, S. 36).

17 Diese Unterscheidung soll selbstverständlich nicht die Annahme nahe legen, dass Evaluationsforschung nicht professionell sei bzw. nicht als solche verstanden werden könne. Selbstverständlich sollte sie dies sein. Nur, wie man anderenorts zwischen Disziplin und Profession, also z.B. zwischen der Disziplin Erziehungswissenschaft und der Profession Pädagogik, sinnvoll unterscheidet, spricht viel dafür, auch in diesem Fall Theorie und Forschung auf der einen Seite und kompetente Berufspraxis auf der anderen Seite zu unterscheiden, ohne damit eine qualitative Hierarchie zu behaupten.

Wenn zuvor zwischen professioneller Evaluationspraxis einerseits und Evaluationsforschung andererseits unterschieden wurde, dann stellt sich die Frage, auf der Basis welcher Kriterien diese Unterscheidung zustande kommt, wenn doch beide auf die Feststellung des Wertes eines Gegenstandes abzielen. Aus unserer Sicht gibt es darauf nur eine Antwort: Der Unterschied zeigt sich in den eingesetzten Verfahren und den damit verbundenen Gütekriterien. Für professionelle Berufspraxis gilt dabei üblicherweise, dass sie auf pragmatische Verfahren der Informationsgewinnung und Bewertung setzt und dabei vorrangig an der Lösung praktischer Probleme in bestimmten Kontexten interessiert ist. Dementsprechend ist die Gültigkeit der Ergebnisse in diesen Fällen meistens mit der praktischen Bewährung gleichzusetzen (vgl. hierzu z.B. die Guidelines zur Einschätzung der Qualität von Befragungsmethoden bei *Mark/Henry/Julnes* 2000, S. 86 ff.). Dabei gilt es zu bedenken, dass die professionelle Evaluationspraxis mittlerweile ein breites Spektrum aufweist. Dies beginnt bei den zahlreichen und phantasievollen Varianten, z.B. die Zufriedenheit von Teilnehmerinnen und Teilnehmern zu erfassen oder ein Feedback derselben zu erhalten – z.B. in Form von Ampelfeedback (vgl. *Beywl/Schepp-Winter* 2000, S. 44). Daneben gibt es jene Konzepte, die auf dem Expertenstatus einer oder mehrerer Personen basieren. In dieser Tradition stehen beispielsweise Ansätze der »connoisseurship« Evaluation, Juryverfahren oder Wettbewerbe. Auch Bewertungen durch Fachkollegen und -kolleginnen innerhalb eines Berufsfeldes oder einer Profession, wie beispielsweise »Peer-Review«-Verfahren sind systematische Bewertungsverfahren, die vor allem auf dem Expertenstatus der evaluierenden Kolleginnen und Kollegen gründen. Die Glaubwürdigkeit und Gültigkeit der Ergebnisse hängt in diesen Fällen in hohem Maße davon ab, ob und von wem der Experte bzw. die Expertin als glaubwürdig und kompetent eingeschätzt wird. Andere Formen der professionellen Praxis verknüpfen externe Expertise mit internen Formen der Selbstevaluation (vgl. z.B. *von Spiegel* 1993; *Liebold* 1998). Zu nennen sind schließlich auch die ausgefeilten Strategien der Konsumforschung und der Warrentests – z.B. in Form von moderierten Fokusgruppen.

Demgegenüber ist die wissenschaftliche Evaluationsforschung an soweit wie möglich verallgemeinerbaren Aussagen interessiert.¹⁸ Die Verfahren orientieren sich an den Standards wissenschaftlich-systematischer Erkenntnisgewinnung; leitend sind für sie die Regeln und Gütekriterien sozialwissenschaftlicher Forschung. Dabei ist zu ergänzen, dass die Frage noch strittig ist, wie wissenschaftliche Verfahren, der Anspruch auf Systematik und Transparenz, das Prinzip der Wiederholbarkeit und der Validität der Ergebnisse mit der Aufgabe, Bewertungen vorzunehmen, in Einklang gebracht werden können (vgl. ausführlich *Lüders/Haubrich* 2003).¹⁹

18 Zu betonen ist, dass wir mit dieser Bestimmung nicht die sonst übliche Unterscheidung zwischen Evaluation und Grundlagenforschung übernehmen. Für Evaluation wird dabei üblicherweise postuliert, dass diese nützlich sein und genutzt werden soll (z.B. *Vedung* 2000, S. 110 ff.). Dem schließen sich auch andere an, die letztendlich den Anwendungsbezug zum zentralen Merkmal machen (wie *Bortz/Döring* 2002, *Wottawa/Thierau* 1998, *Kromrey* 2000). Aus unserer Sicht ist diese Unterscheidung nicht nur auf Grund der zuvor schon dargestellten verwendungstheoretischen Argumentation problematisch (vgl. oben Abs. 4), sondern auch angesichts der allerorten beobachtbaren Prozesse der Politisierung, Mediatisierung und Ökonomisierung von Wissenschaft nicht mehr haltbar (vgl. *Weingart* 2001; noch radikaler in dieser Hinsicht: *Gibbons/Limoges/Nowotny/Schwartzman/Scott/Trow*, 1994; *Nowotny/Scott/Gibbons* 2001).

19 So finden sich in der Literatur Positionen, die, zurückgehend auf das Wertfreiheitspostulat, darin eine grundlegende Schwierigkeit sehen, während andererseits auch schon mal das Gegenteil behauptet wird: »There is nothing in the philosophy of science forbidding basic research to be openly normative« (*Vedung* 2000, S. 110).

6 Typen der Evaluation

6.1 Selbst- und Fremdevaluation, interne und externe Evaluation

In der Fachliteratur findet sich mittlerweile eine Vielzahl von Vorschlägen, um das unübersichtliche Feld der Evaluation zu sortieren. Eine der grundlegenden Unterscheidungen ist dabei die Differenz zwischen Innen und Außen. So wird einmal zwischen interner und externer Evaluation unterschieden. Bei externen Evaluationen wird der fremde Blick von außen auf den Evaluationsgegenstand gerichtet. Sie werden von Personen vorgenommen, die nicht Mitglieder der Organisation sind, in der der Untersuchungsgegenstand angesiedelt ist. Es besteht also per Definition eine weitgehende Unabhängigkeit zwischen der Verantwortlichkeit für den Untersuchungsgegenstand und für die Evaluation. Interne Evaluationen werden dagegen von internen Evaluatorinnen und Evaluatoren gesteuert und durchgeführt, die gewissermaßen die Arbeit ihrer Kollegen und Kolleginnen bewerten (vgl. *Beywl/Heiner* 2000, S. 114 ff.).

Davon zu unterscheiden sind Formen der Selbstevaluation, die in Deutschland in einer Reihe von Praxisfeldern, wie z.B. in der Kinder- und Jugendhilfe (vgl. die Broschürenreihe »Qualitätssicherung in der Kinder- und Jugendhilfe« des BMFSFJ) und der beruflichen Weiterbildung (vgl. *Krekel/Raskopp*, 2003, S. 9) zu einer weit verbreiteten Praxis geworden sind: »Im Gegensatz zur Fremdevaluation durch Außenstehende, meist Sozialwissenschaftler, ist es bei der Selbstevaluation der Sozialarbeiter, der als »Forscher« in eigener Sache den Verlauf der Ergebnisse seines beruflichen Handelns untersucht« (*Heiner* 1988, S. 7). Entscheidend für die Bestimmung einer Evaluation als Selbstevaluation ist, dass »(...) die Personen, die für die Programmplanung und -umsetzung zuständig sind, (...) auch für die Evaluation verantwortlich (...)« sind (*Beywl/Schepp-Winter* 2000, S. 31). *W. Beywl und E. Schepp-Winter* betonen dabei, dass der Untersuchungsgegenstand von geringer Komplexität sein und zu 100% in der Zuständigkeit der Selbstevaluatorinnen bzw. -evaluatoren liegen sollte. Dies bedeutet, dass Programmsteuerung (Planung und Umsetzung des Programms) und Evaluationssteuerung »in einer Hand« vereint sind.

Bei anderen Autoren wird zur Unterscheidung zwischen Selbst- und Fremdevaluation, interner und externer Evaluation nochmals deutlicher zwischen der Evaluationsdurchführung und Evaluationssteuerung differenziert, woraus sich neue Kombinationsmöglichkeiten ergeben (vgl. *Widmer* 2000, S. 79; vgl. *Schratz* 1999, S. 219). Gerade bei internen Evaluationen gibt es allerdings häufig Mischformen in der Verteilung der Programm- und Evaluationsverantwortlichkeiten und damit graduelle Übergänge zur Selbstevaluation – was den abstrakten Systematiken eine gewisse Blutleere verleiht. Jedoch sensibilisiert die Unterscheidung zwischen Steuerung und Durchführung für die genaue Betrachtung der faktischen Verteilungen steuernder Kompetenzen auf die an der Evaluation Beteiligten sowie für die in aller Regel ungleiche Machtverteilung. In der Realität sind die Kompetenzen der Evaluationssteuerung nicht nur sehr heterogen verteilt, sondern häufig auch komplex: So gibt es nicht selten extern *und* intern besetzte »Steuerungsgruppen« für Evaluationen und selbst in der »reinen« Evaluationsdurchführung sind steuernde Potentiale enthalten (wie in der Festlegung der Indikatoren, der Erhebungs- und Auswertungsverfahren und insbesondere in den Verfahren der Bewertung). Daher bedarf es u. E. einer genauen und differenzierten Berücksichtigung dieses Aspektes. Dies erfordert aber gleichzeitig eine weitaus höhere Transparenz der Steuerungsarchitektur von Evaluationen als dies bislang in der Forschungspraxis vorzufinden ist.

6.2 Formative und summative Evaluation

Das Begriffspaar »formativ« und »summativ« geht zurück auf *M. Scriven* (1991a, 1996) und hat sich auch im deutschen Sprachraum zu einer Standardunterscheidung entwickelt. *M. Scriven* zitiert *R. Stake* mit dem Ausspruch: »Wenn der Koch die Suppe probiert, dann ist dies formativ, wenn der Gast sie probiert, dann ist dies summativ«. Allerdings ist dieses so anschauliche und oft verwendete Zitat auch irreführend, denn man könnte den Eindruck gewinnen, dass lediglich die Verwendung der Ergebnisse ausschlaggebend dafür ist, ob eine Evaluation summativ oder formativ ist. Diese Interpretation findet sich auch in der Fachliteratur wieder.²⁰ U.E. ist aber entscheidend, dass es für die Planung und Durchführung einer Evaluation einen Unterschied macht, ob sie summativ oder formativ sein soll. Davon hängen Entscheidungen über das Evaluationsdesign, die Fragestellungen, relevante Indikatoren, die Erhebungs- und Auswertungsverfahren, die Form der Ergebnisvermittlung und auch die Art der Berichterstattung ab. Das Kernelement der Unterscheidung zwischen summativen oder formativen Evaluationen liegt daher in der mit der Evaluation verbundenen Intention, für welche Kontexte und welche Adressaten Informationen gewonnen und bewertet werden sollen.

Üblicherweise werden formative Evaluationen während der Entwicklung eines Programms – manchmal auch preformativ genannt – oder während der Umsetzung eines Programms durchgeführt.²¹ Die Ergebnisse werden typischerweise für diejenigen bereitgestellt, die das Programm steuern (z.B. Projektleiter und -leiterinnen, Maßnahmenträger etc.) und umsetzen (Mitarbeiter und Mitarbeiterinnen in einem Projekt, einer Maßnahme o.ä.) mit der Absicht, dass sie genutzt werden können, um das Programm zu verbessern. Die Berichte bleiben meist intern (vgl. *Scriven* 1991a, S. 168f.).

Summative Evaluationen werden oftmals erst nach Abschluss des Programms durchgeführt, obwohl sie auch bereits begleitend zu einem Programm einsetzen können. Die Ergebnisse werden in der Regel für eine externe Öffentlichkeit (z.B. Fachkräfte des jeweiligen Praxisfeldes) oder Entscheidungsträger (Geldgeber, Auftraggeber, Politiker oder mögliche zukünftige Nutzer) produziert. Häufig dienen sie als Grundlage für Entscheidungen über das Programm (wie Fortführung, Ausweitung oder Einschränkung). Die Entscheidungen können aber auch in der Zukunft liegen und vorab nicht genau bekannt sein. Dies ist z.B. der Fall, wenn Politiker zu einem späteren Zeitpunkt aus den Erfahrungen und Bewertungen eines Programms für die Gestaltung einer neuen Maßnahme Informationen beziehen. Letzteres wird von anderen Autoren, insbesondere *M.Q. Patton*, als wissensgenerierende Verwendung (»knowledge-oriented use«) bezeichnet (*Patton* 1997, S. 64). Er argumentiert, dass formative und summative Evaluationen an einen instrumentellen Nutzen gebunden seien: entweder Verbesserung oder Entscheidung. Daher bilde die wissensgenerierende Evaluation, deren Verwendungszusammenhänge nicht konkret geplant werden (können), eine eigenständige Kategorie. Sie trage insbesondere zu einer Vermehrung konzeptionellen Wissens bei, das für die Fachpraxis und Politik

20 *B.R. Worthen, J.R. Sanders* und *J.L. Fitzpatrick* (1997) bezeichnen ebenso wie *M.Q. Patton* (1997) formativ und summativ als zwei unterschiedliche »uses« also Verwendungsformen von Evaluationen. Auch in der deutschsprachigen Literatur lassen sich solche Zuspitzungen auf die Verwendungsseite finden. So betont bspw. *T. Widmer* ebenfalls die Nutzungsdimension dieser beiden Begriffe (vgl. 2000, S. 80).

21 Allerdings wird formative Evaluation nicht durch die zeitliche Komponente allein definiert, weshalb sich eine Gleichsetzung mit Begleitforschung verbietet. Es ist prinzipiell ebenso möglich programmbegleitend eine summative Evaluation durchzuführen (*Scriven* 1996).

in »länger gezogenen Schleifen« (vgl. *Beywl* 2001, S. 21 als Wiederabdruck von *Beywl* 1999) hilfreich sein kann.²²

6.3 Fünf Aufgabentypen von Evaluation

Zu einer ganz anderen Typologie kommen *J.M. Owen* und *P.J. Rogers*, die Evaluationen nach den unterschiedlichen Aufgaben, die sie in Bezug auf Programme und Politiken übernehmen, sortieren. Sie unterscheiden zwischen fünf Typen und aufeinander aufbauender Formen der Evaluation (vgl. *Owen/Rogers* 1999, S. 39 ff.):

Proaktive Evaluation

Diese Evaluationsform wird vor Beginn eines Programms gewählt, um den Bedarf für ein solches Programm, die Rahmenbedingungen einer geplanten Umsetzung und/oder den aktuellen Forschungsstand dazu sowie gute Praxisbeispiele zu erheben. Zweck einer solchen Evaluation ist es, ein Programm in der Konzeptionsphase systematisch zu verbessern. Teilweise ist diese Zwecksetzung verbunden mit einer Überprüfung der Evaluierbarkeit eines solchen Programms oder der Erhebung von Daten der aktuellen Ausgangssituation.

Klärende Evaluation

Diese Form wird gewählt, wenn in der Anfangsphase der Implementation eines Programms das Konzept unter Berücksichtigung der konkreten Umsetzungs- und Kontextbedingungen verfeinert und konkretisiert werden soll. Typische Fragen sind hier: Welche Ergebnisse sind konkret zu erreichen? Welche Schritte sind im Programm angelegt, um diese zu erreichen? Ist das Programmkonzept schlüssig oder gibt es Konkretisierungs- oder Ergänzungsbedarf? Welche Programmelemente müssen modifiziert werden, um die angestrebten Ergebnisse zu maximieren?

Typische Formen solcher Evaluationen sind beispielsweise die Rekonstruktion der impliziten Theorie, wie bestimmte Programmelemente zu bestimmten Ergebnissen führen sollen, und deren Überprüfung und Weiterentwicklung im Hinblick auf Kontextbedingungen. Ein anderes Beispiel ist die Überprüfung der Evaluierbarkeit, sofern sie nicht schon vorab durchgeführt wurde.

Interaktive Evaluation

Interaktive Evaluationsformen zielen auf die Verbesserung eines Programms während der Laufzeit ab. Sie bieten Informationen darüber, mit Hilfe welcher Interventionen, in welchen Einzelschritten und unter welchen Rahmenbedingungen bestimmte Ergebnisse angestrebt werden. Geklärt werden dabei in der Regel Fragen, wie: Wie sehen die konkreten Umsetzungsformen aus? Sind die Umsetzungen konsistent mit dem Programm? Funktionieren die Umsetzungsformen? Wie können diese Implementationen weiterentwickelt werden, um das Programm wirkungsvoller zu gestalten? Resümierend können nach Ablauf des Programms

22 *M. Scriven* (1996) unterscheidet demgegenüber explizit zwischen der bilanzierenden Bewertung und einer möglicherweise folgenden Entscheidung, indem er darauf hinweist, dass beides einer anderen Logik folgt – die Bilanzierung ist eine wissenschaftliche Aufgabe, die Möglichkeit diese Bilanzierung als Entscheidungsgrundlage zu wählen, ist eine davon logisch unabhängige Handlung. Summative Evaluation kann in diesem Sinne sowohl als Entscheidungsgrundlage verwendet werden wie auch zur Mehrung konzeptionellen Wissens.

Aussagen getroffen werden, unter welchen Rahmenbedingungen welche Strategien mit Erfolg oder mit weniger Erfolg umgesetzt werden konnten.

Typische Evaluationsansätze dieser Art sind Responsive Evaluation, Qualitätsüberprüfung, Empowerment Evaluation, Cluster Evaluation, z.T. auch interaktive Sozialforschung bzw. Handlungsforschung.

Monitoring

Monitoring zielt typischerweise auf die Feinsteuerung eines Programms und die Legitimation/Rechtfertigung der Programmumsetzung. Typischerweise werden Fragen verfolgt, wie: Erreicht das Programm die Zielgruppen? Erreicht die Umsetzung bestimmte Benchmarks oder spezifizierte Kriterien? Wie unterscheidet sich die Implementation zwischen den Projekten? Gibt es einzelne Projekte, die Unterstützung benötigen, um das Programm effektiver umzusetzen? Wo stehen wir heute verglichen mit einem Monat zuvor? Wie können wir das Programm steuern, damit es effektiver/effizienter wird? Resümierend können nach Ablauf des Programms Detailaussagen über konkrete Umsetzungsformen und ihre jeweilige Zielerreichung oder Programmtreue formuliert werden.

Monitoring wird in der Regel eingesetzt, wenn schon ausreichend Klarheit über Umsetzungsformen eines Programms besteht und Informationen über erreichbare Ergebnisse vorliegen. Typischerweise werden Management-Informationen-Systeme oder Leistungsmessung eingesetzt.

Wirkungsevaluation

Wirkungsevaluationen werden am Ende oder nach Ablauf eines Programms durchgeführt bei Programmen, deren Umsetzung bereits ausreichend erprobt worden ist und klare Kenntnisse über die Konzeption und Implementation des Programms vorliegen. Sie dienen in der Regel der bilanzierenden Beurteilung, ob ein Programm seine angestrebten Ergebnisse erreicht hat oder nicht. Typische Fragen sind dabei, ob das Programm so wie geplant implementiert worden ist, ob die Ziele erreicht wurden, welche nicht-intendierten Wirkungen aufgetreten sind, ob und in welchem Ausmaß die Intervention zu den fest gestellten Ergebnissen geführt hat, welche Auswirkungen unterschiedliche Formen der Implementation auf die Ergebnisreichung haben oder ob das Programm kosteneffizient gearbeitet hat.

Typische Evaluationsansätze sind Prozess-Ergebnisstudien, Outcome-Evaluationen, Kosten-Nutzen-Studien, Evaluationen der Zielerreichung, Programmtheorieevaluationen oder zielfreie Evaluationen. Üblicherweise werden dabei vorher festgelegte Kriterien zum Maßstab genommen. Als Untersuchungsanlage werden, experimentelle oder quasi-experimentelle Designs mit Vergleichs- oder Kontrollgruppen bevorzugt.

Die kategoriale Rationalität dieser Typisierung von *J. M. Owen* und *P. J. Rogers* darf nicht zu der Erwartung verführen, dass der Durchgang vom Beginn eines Programms bis hin zur Wirkungsanalyse schnell in fünf Schritten zu erledigen sei. Unsere Erfahrungen aus laufenden Programmen belegen immer wieder, dass man gerade in vergleichsweise jungen Praxisfeldern zufrieden sein kann, wenn man tragfähige Antworten im Hinblick auf die zu klärenden Programmfragen und ansatzweise Antworten hinsichtlich der Beschreibung der konkreten Umsetzungsformen und der Zielgruppenerreichung erhält. Welche Form der Evaluation zum Einsatz kommen kann, hat also wesentlich auch damit zu tun, wie bewährt und entwickelt ein Pro-

gramm ist. Aus dieser Perspektive würde es z.B. keinen Sinn machen, Wirkungsüberprüfung im engen Sinne des Wortes bei Modellprogrammen durchzuführen. Denn deren Aufgabe ist es, die interne Programmstruktur erst zu entwickeln und zu erproben.

7 Evaluation in der Kinder- und Jugendhilfe

Zu Beginn dieses Beitrages wurde festgehalten, dass sich Evaluation und vor allem die von ihr erwarteten Ergebnisse gegenwärtig besonderer Aufmerksamkeit erfreuen. Zugleich dürfte der Durchgang durch die aktuelle Diskussion deutlich gemacht haben, dass eine ganze Reihe von konzeptionellen Fragen noch ungeklärt sind, dass die Fachdebatte nach wie vor in hohem Maße vor allem durch die amerikanische Diskussion und ihre eher pragmatische Tradition geprägt ist und dass ein klares Profil, was Evaluation ist, was sie leisten kann und an welchen Stellen sie mit überzogenen Erwartungen konfrontiert wird, im deutschsprachigen Raum bislang nur in Ansätzen erkennbar ist. M.a.W.: Man hat es mit einem Forschungs- und Praxisbereich zu tun, der hierzulande gerade erst sein eigenes Profil und Selbstverständnis entwickelt. Dazu gehört, dass die Deutsche Gesellschaft für Evaluation noch vergleichsweise jung ist (Gründung 1997), dass es erst seit zwei Jahren eine einschlägige Fachzeitschrift gibt, die »Zeitschrift für Evaluation«, und dass es bislang für viele evaluationsspezifische Aspekte nur eine Hand voll einschlägiger deutschsprachiger Methoden- und Handbücher gibt.

Dies alles gilt in besonderem Maße auch für die Kinder- und Jugendhilfe. So erklingt in diesem Bereich die Forderung nach Evaluationen seit einigen Jahren besonders eindringlich und mit den §§ 78 a ff. SGB VIII wurde die Verpflichtung zur Bewertung der Qualität der Leistungsangebote als Moment der Leistungsvereinbarung gesetzlich festgeschrieben (vgl. hierzu *J. Merchel* in diesem Heft). Wohl auch als Reaktion auf die gestiegenen Erwartungen wurden in den letzten Jahren – durch das BMFSFJ gefördert (vgl. dazu auch den Beitrag von *R. Wiesner* in diesem Heft) – zwei größere Studien im Bereich »Hilfen zu Erziehung« durchgeführt, die so genannte JULE-Studie (vgl. *Bundesministerium für Familie, Senioren, Frauen und Jugend* 1998) und die Jugendhilfe-Effekte-Studie (vgl. *Schmidt/Schneider/Hohm/Pickartz/Macsenaere/Petermann/Flosdorf/Hölzl/Knab* 2002). Diese Studien stellen insofern ein Novum dar, als damit seit langem zum ersten Mal Regelstrukturen in der Kinder- und Jugendhilfe auf ihre Effekte hin untersucht worden sind, während es bis dahin meistens Modellprogramme waren, die aus den Mittel des Bundesministeriums wissenschaftlich begleitet bzw. evaluiert wurden (vgl. als ein Beispiel den Beitrag von *S. Hoops* und *H. Permien* in diesem Heft). Daneben wurden in den letzten Jahren und werden gegenwärtig in der Kinder- und Jugendhilfe eine Vielzahl von kleineren und mittleren Studien realisiert (vgl. den Beitrag von *Chr. Erzberger* in diesem Heft). Soweit zu sehen, sind viele dieser Studien auf bestimmte Einrichtungsformen, lokale Trägerstrukturen oder Angebotsformen fokussiert und stehen in einem unmittelbaren Verwertungszusammenhang.²³ Es darf deshalb auch nicht überraschen, dass von dort aus bislang keine nennenswerten überregionalen Impulse zur konzeptionellen und methodischen Weiterentwicklung des Feldes ausgegangen sind.

23 Im Rahmen des durch das BMFSFJ geförderten Projektes »Strategien und Konzepte externer Evaluation in der Kinder- und Jugendhilfe« versucht das DJI die aktuelle Forschungslage zu sichten, zu sortieren und die konzeptionelle und inhaltliche Spreu vom Weizen zu trennen sowie auf der Basis der bislang vorliegenden Erfahrungen Vorschläge für eine konzeptionelle Weiterentwicklung der Forschung und Praxis in diesem Bereich zu machen. Mitarbeiterinnen in diesem Projekt sind Frau *Karin Haubrich* und Frau *Gerlinde Struhkamp*.

In diesem Sinne gilt *cum grano salis* noch heute die Problemanzeige des 11. Kinder- und Jugendberichtes, dass gerade im deutschsprachigen Raum die für eine systematische und ertragreiche Evaluationsforschung im Bereich der Kinder- und Jugendhilfe die »(...) notwendigen Instrumente, Verfahren und Standards derzeit nur ansatzweise zur Verfügung stehen. Schon ein kurzer Blick in die jüngere Literatur zur Evaluationsforschung in der Bundesrepublik Deutschland (vgl. z.B. Müller-Kohlenberg/Münstermann 2000; Stockmann 2000a) dokumentiert die bislang randständige und methodologisch wenig entwickelte Rolle der Kinder- und Jugendhilfe in diesem Bereich. Es bedarf deshalb der gezielten Entwicklung von praxistauglichen Standards, Verfahren und Methodologien und – darauf aufbauend – einer breiten Empirie. (...) Evaluationen (haben sich dabei) nicht nur auf die einzelnen Angebote, Arbeitsformen oder Einrichtungen zu beziehen (...), sondern (...) auch (auf) Politiken (policies), Programme und vor allem Infrastrukturen im lokalen Kontext (...)« (*Deutscher Bundestag* 2002, S. 256). Zu ergänzen wäre allein, dass es dabei nicht nur um die intendierten, sondern ebenso auch um die nicht-intendierten Auswirkungen auf die Rahmenbedingungen, Verfahren und die Beteiligten ginge.

Wenn Evaluation angesichts der in sie gesetzten Erwartungen ihre Rolle als eine seriöse und relevante Informationsquelle (neben anderen) für politische Entscheidungen auf einem hohen Niveau ausfüllen will, ist es notwendig ihre realistischen Möglichkeiten und Grenzen noch genauer als bisher zu klären. Vor allem sollte immer deutlich bleiben, dass Evaluationen zwar in systematischen und transparenten Verfahren gut begründete Informationen bereit stellen, jedoch niemals Entscheidungen abnehmen können.

Literatur:

- Alkin, Marvin C. (Ed.): *Debates on Evaluation*. Newbury Park, London, New Delhi, Sage 1990.
- Beck, Ulrich/Bonß, Wolfgang (Hrsg.): *Weder Sozialtechnologie noch Aufklärung? Analysen zur Verwendung sozialwissenschaftlichen Wissens*. Frankfurt/M., Suhrkamp 1989.
- Beywl, Wolfgang: *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation*. Frankfurt/M., Bern, New York, Paris, Peter Lang Verlag 1988.
- Beywl, Wolfgang: *Anerkannte Standards und Leitprinzipien der amerikanischen Evaluation*. In Heiner, Maja (Hrsg.): *Qualitätsentwicklung durch Evaluation*. Freiburg/Br., Lambertus 1996, S. 85–107.
- Beywl, Wolfgang: *Evaluation der Weiterbildung*. In: Künzel, Klaus (Hrsg.): *Internationales Jahrbuch der Erwachsenenbildung* 27, 1999, S. 29–48.
- Beywl, Wolfgang: *Programmevaluation in pädagogischen Praxisfeldern: begriffliche und konzeptionelle Grundlagen*. In: *Materialien zur Erwachsenenbildung (Konzepte der Qualität in der Erwachsenenbildung. Aufsätze und Protokolle im Rahmen der Werkstatt am Bundesinstitut für Erwachsenenbildung St. Wolfgang vom 2.-3. Okt. 2000; hrsg. vom österreichischen Bundesministerium für Bildung, Wissenschaft und Kultur)*. Heft Nr. 2, Wien & Strobl (2001), S. 18–34. Downloadbar unter: http://wwwapp.bmbwk.gv.at/medien/6048_PDFzuPubID88.pdf [15.07.2004]

- Beywl, Wolfgang/Heiner, Maja: Kapitel 3.1. Darstellung des Verfahrens. In: Merchel, Joachim (Hrsg.): Qualitätsentwicklung in Einrichtungen und Dienstleistungen der Erziehungshilfe. Methoden, Erfahrungen, Kritik, Perspektiven. IGFH, Frankfurt/M. 2000, S.111–132.
- Beywl, Wolfgang/Schepp-Winter, Ellen: Zielgeführte Evaluation von Programmen – ein Leitfaden. (Materialien zur Qualitätssicherung in der Kinder- und Jugendhilfe, Heft 29). Berlin 2000.
- Beywl, Wolfgang/Speer, Sandra/Kehr, Jochen: Wirkungsorientierte Evaluation im Rahmen der Armuts- und Reichtumsberichterstattung. Köln 2004. Downloadbar unter: <http://www.bmgs.bund.de/download/broschueren/A323.pdf> [15.07.2004]
- Bundesministerium für Familie, Senioren, Frauen und Jugend (Hrsg.): Leistungen und Grenzen der Heimerziehung. Ergebnisse einer Evaluationsstudie stationärer und teilstationärer Erziehungshilfen. Forschungsprojekt Jule (Schriftenreihe des Bundesministeriums für Familie, Senioren, Frauen und Jugend, Bd. 170). Stuttgart, Kohlhammer 1998.
- Bortz, Jürgen/Döring, Nicola: Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler. Berlin, Heidelberg, New York, Springer ³2002.
- Clarke, Alan/Dawson, Ruth: Evaluation Research. An Introduction to Principles, Methods and Practice. London, Thousand Oaks, New Delhi, Sage 2000 (Reprint); Orig. 1999.
- Deutsche Gesellschaft für Evaluation (Hrsg.): Standards für Evaluation. Köln 2002. Downloadbar unter: <http://www.degeval.de/standards/index.htm> [15.07.2004]
- Deutscher Bundestag: Bericht über die Lebenslagen junger Menschen und die Leistungen der Kinder- und Jugendhilfe in Deutschland. Elfter Kinder- und Jugendbericht. BT 14/8181 vom 04.02.2002. Berlin 2002.
- European Commission (Ed.): Evaluating Socio-Economic Programmes. Glossary of 300 Concepts and Technical Terms (MEANS Collection Vol. 6). Office for Official Publications of the European Communities, Luxembourg 1999.
- Franklin, Jack L./Trasher, Jean H.: An Introduction to Program Evaluation. New York, Wiley 1976.
- Gibbons, Michael/Limoges, Camille/Nowotny, Helga/Schwartzman, Simon/Scott, Peter/Trow, Martin: The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies. London, Thousand Oaks, New Dehli, Sage 1994.
- Guba, Egon G./Lincoln, Yvonna S.: Effective Evaluation. Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches. San Francisco, Jossey-Bass 1981.
- Guba, Egon G./Lincoln Yvonna S.: Fourth Generation Evaluation. Newbury Park, London, New Dehli, Sage 1989.
- Habermas, Jürgen: Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats. Frankfurt/M., Suhrkamp 1992.
- Heiner, Maja (Hrsg.): Selbstevaluation in der sozialen Arbeit. Fallbeispiele zur Dokumentation und Reflexion beruflichen Handelns. Freiburg/Br., Lambertus 1988.

- House, Ernest R. (Ed.): *Philosophy of Evaluation. New Directions for Program Evaluation*, no. 19, San Francisco, Washington, London, Jossey-Bass 1983.
- House, Ernest R.: *The Future Perfect of Evaluation. Evaluation Practice*, 15 (1994), S. 239–247.
- House, Ernest R./Howe, Kenneth R.: *Values in Evaluation and Social Research*. Thousand Oaks, London, New Dehli, Sage 1999.
- Joint Committee on Standards for Educational Evaluation/Sanders, James R. (Hrsg.): *Handbuch der Evaluationsstandards*. Opladen, Leske + Budrich 1999.
- Krekel, Elisabeth M./Raskopp, Kornelia: *Qualität und Evaluation in der beruflichen Weiterbildung*. In: *Berufsbildung in Wissenschaft und Praxis*. (Schwerpunktheft: Evaluation und Qualität in der Berufsbildung). *Zeitschrift des Bundesinstituts für Berufsbildung*, 32. Jg., Heft 6 (2003), S. 8–12.
- Kromrey, Helmut: *Evaluation. Empirische Konzepte zur Bewertung von Handlungsprogrammen und die Schwierigkeiten ihrer Realisierung*. In: *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie* 15 (1995), S. 313–336.
- Kromrey, Helmut: *Die Bewertung von Humandienstleistungen. Fallstricke bei der Implementations- und Wirkungsforschung sowie methodische Alternativen*. In: Müller-Kohlenberg, Hildegard/Münstermann, Klaus (Hrsg.): *Qualität von Humandienstleistungen. Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen*. Opladen, Leske + Budrich 2000, S. 19–57.
- Lee, Barbara: *Theories of Evaluation*. In: Stockmann, Reinhard (Hrsg.): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder (Sozialwissenschaftliche Evaluationsforschung, Bd. 1)*. Opladen, Leske + Budrich 2000a, S. 127–164.
- Liebold, Christiane: *Leitfaden für Selbstevaluation und Qualitätssicherung (Materialien zur Qualitätssicherung in der Kinder- und Jugendhilfe, Heft 19)*. Bonn 1998.
- Lüders, Christian: *Spurensuche. Ein Literaturbericht zur Verwendungsforschung*. In: Oelkers, Jürgen/Tenorth, Heinz-Elmar (Hrsg.): *Pädagogisches Wissen*. Weinheim, Basel, Beltz 1993, S. 415–437.
- Lüders, Christian/Haubrich, Karin: *Qualitative Evaluationsforschung*. In: Schweppe, Cornelia (Hrsg.): *Qualitative Forschung in der Sozialpädagogik*. Opladen, Leske + Budrich 2003, S. 305–330.
- Mark, Melvin M./Henry, Gary T./Julnes, George: *Evaluation. An Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs*. San Francisco, Jossey-Bass 2000.
- Martinuzzi, André/Meyer, Wolfgang: *Evaluationen im Umweltbereich. Ein Beitrag zum Nachhaltigen Wirtschaften?* In: *DIW Vierteljahreshefte zur Wirtschaftsforschung*, »Evaluation im Spannungsfeld zwischen Wissenschaft und Politik«, Jg. 69, Heft 3 (2000), S. 449 – 463.
- Mertens, Donna M.: *Institutionalizing Evaluation in the United States of America*. In: Stockmann, Reinhard (Hrsg.): *Evaluationsforschung. Grundlagen und ausgewählte For-*

- schungsfelder (Sozialwissenschaftliche Evaluationsforschung, Bd. 1). Opladen, Leske + Budrich 2000a, S. 41–56.
- Müller-Kohlenberg, Hildegard/Münstermann, Klaus (Hrsg.): Qualität von Humandienstleistungen. Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen. Opladen, Leske + Budrich 2000.
- Nowotny, Helga/Scott, Peter/Gibbons, Michael: Re-Thinking Science. Knowledge and the Public in an Age of Uncertainty. Cambridge & Oxford, Polity Press 2001.
- Owen, John M./Rogers, Patricia J.: Program Evaluation. Forms and Approaches. London, Thousand Oaks & New Dehli, Sage 1999.
- Patton, Michael, Q.: Utilization-Focused Evaluation. Newbury Park, London & New Dehli, Sage ³1997.
- Patton, Michael Q.: Qualitative Evaluation and Research Methods. Newbury Park, London & New Dehli, Sage ²1990.
- Rossi, Peter H./Freeman, Howard E.: Evaluation. A Systematic Approach. Newbury Park, London & New Dehli, Sage ⁵1993.
- Schmidt, Martin/Schneider, Karsten/Hohm, Erika/Pickartz, Andrea/Macsenaere, Michael/Petermann, Franz/Flosdorf, Peter/Hözl, Heinrich/Knab, Eckart: Effekte erzieherischer Hilfen und ihre Hintergründe. (Schriftenreihe des Bundesministerium für Familie, Senioren, Frauen und Jugend, Bd. 219). Stuttgart, Kohlhammer 2002.
- Schwandt, Thomas A.: Evaluation Practice Reconsidered. (Counterpoints – Studies in Postmodern Theory of Education, Vol. 211). New York, Washington u.a., Peter Lang Verlag 2002.
- Schratz, Michael: Selbstevaluation als Bemühen, Qualität zu verstehen und zu entwickeln. In: Pädagogisches Forum (1999), S. 219–222
- Scriven, Michael: The Evaluation Taboo. In: House, Ernest R. (ed.): Philosophy of Evaluation. New Directions for Program Evaluation, no. 19, San Francisco, Washington, London, Josey-Bass 1983, S. 75–82.
- Scriven, Michael: Evaluation Thesaurus. Newbury Park, London & New Dehli, Sage ⁴1991 (a).
- Scriven, Michael: Beyond Formative and Summative Evaluation. In: McLaughlin M.W./Phillips D.C.: Evaluation and Education: At Quarter Century. (90th Yearbook of the National Society for the Study of Education). Chicago, University of Chicago Press 1991(b), S. 18–64.
- Scriven, Michael: Types of Evaluation and Types of Evaluator. In: Evaluation Practice, Vol. 17, No. 2 (1996), S. 151–161
- Scriven, Michael: Evaluation in the New Millennium. The Transdisciplinary Vision. In: Donaldson, Steward I./Scriven, Michael (Eds.): Evaluating Social Programs and Problems. Visions for the New Millennium. London, Mahwah, New Jersey, Lawrence Erlbaum Associates Publishers 2003, S. 19–41.

- Shadish, William R. Jr./Cook, Thomas D./Leviton Laura C.: Foundations of Program Evaluation. Theories of Practice. Newbury Park, London, New Dehli, Sage 1991.
- Shaw, Ian F.: Qualitative Evaluation. London, Thousand Oaks & New Dehli, Sage 1999.
- Spiegel, Hiltrud, von: Aus Erfahrung lernen. Qualifizierung durch Selbstevaluation. Münster, Votum 1993.
- Stake, Robert: The Art of Case Study Research. Thousand Oaks, London, New Dehli, Sage 1995.
- Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder (Sozialwissenschaftliche Evaluationsforschung Bd. 1). Opladen, Leske + Budrich 2000 (a).
- Stockmann, Reinhard: Evaluation in Deutschland. In: ders. (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder (Sozialwissenschaftliche Evaluationsforschung, Bd. 1). Opladen, Leske + Budrich 2000 (b), S. 11–40.
- Stufflebeam, Daniel L.: Evaluation Models (New Directions For Evaluation Number 89, Spring 2001). San Francisco, Jossey-Bass 2001.
- Suchman, Edward A.: Evaluative Research. Principles and Practice in Public Service & Social Action Programs. New York, Sage 1976 (Reprint); Orig. 1967.
- Vedung, Evert: Evaluation im öffentlichen Sektor. (Studien zu Politik und Verwaltung; Band 64). Wien, Köln, Graz, Böhlau Verlag 1999.
- Vedung, Evert: Evaluation Research and Fundamental Research. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder (Sozialwissenschaftliche Evaluationsforschung, Bd. 1). Opladen, Leske + Budrich 2000, S. 103–126.
- Weber, Max: Der Sinn der »Wertfreiheit« der soziologischen und ökonomischen Wissenschaften (1917). In: Ders.: Gesammelte Aufsätze zur Wissenschaftslehre (hrsg. von Johannes Winkelmann). Tübingen, Mohr 1988, S. 489–540.
- Weingart, Peter: Die Stunde der Wahrheit? Zum Verständnis der Wissenschaft zu Politik, Wirtschaft und Medien in der Wissensgesellschaft. Weilerswist, Velbrück Wissenschaft 2001.
- Weiss, Carol H.: Have We Learnt Anything New About the Use of Evaluation? In: American Journal of Evaluation 19 (1998), S. 21–33.
- Widmer, Thomas: Qualität der Evaluation. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder (Sozialwissenschaftliche Evaluationsforschung, Bd. 1). Opladen, Leske + Budrich 2000, S. 77–102.
- Widmer, Thomas/Landert, Charles/Bachmann; Nicole: Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards), 2000
Downloadbar unter: http://www.seval.ch/de/documents/seval_Standards_2001_dt.pdf
[15.07.2004]
- Wollmann, Helmut: Evaluierungsforschung von Verwaltungspolitik und –modernisierung. In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte For-

schungsfelder (Sozialwissenschaftliche Evaluationsforschung, Bd. 1). Opladen, Leske + Budrich 2000, S. 195–231.

Worthen, Blaine R./Sanders, James R./Fitzpatrick, Jody L.: Program Evaluation. Alternative Approaches and Practical Guidelines. New York u.a., Logman ²1997.

Wottawa, Heinrich/Thierau, Heike: Lehrbuch Evaluation. Bern, Göttingen, Toronto, Seattle, Verlag Hans Huber ²1998.

*Verf.: Karin Haubrich/Dr. Christian Lüders, Deutsches Jugendinstitut,
Abteilung Jugend und Jugendhilfe, Nockherstraße 2, 81541 München,
haubrich@dij@de, lueders@dij.de*