

Sonderdruck
aus
Enzyklopädie der Psychologie

Themenbereich B
Methodologie und Methoden

Serie IV
Evaluation

Band 1
**Grundlagen und statistische Methoden
der Evaluationsforschung**

herausgegeben von
Prof. Dr. Heinz Holling

2009



Hogrefe · Verlag für Psychologie
Göttingen · Bern · Toronto · Seattle

4. Kapitel

Phasen der Evaluationsforschung

Nicola Döring

1 *Phasenmodelle in Wissenschaft und Praxis*

Die *Evaluationsforschung* hat das Ziel, die Wirksamkeit und Effizienz von Interventionen mit Hilfe wissenschaftlicher Methoden und Theorien zu bewerten. Weiterhin kann es Ziel von Evaluationsstudien sein, die Wirkungsweise von Interventionen kausal zu erklären und unerwartete Nebenwirkungen zu identifizieren (vgl. Holling in diesem Band). Evaluationsergebnisse sollen die Qualität einer Maßnahme bewerten und/oder Hinweise für Qualitätssteigerungen liefern. Sowohl bei Interventionen als auch bei den auf sie bezogenen Evaluationen handelt es sich um mehr oder minder langfristige und komplexe *Projekte*. Deswegen ist es sinnvoll, den Prozess der Evaluationsforschung zeitlich und sachlogisch zu strukturieren und in einzelne Phasen einzuteilen, in denen jeweils bestimmte Teilergebnisse (Meilensteine) erarbeitet werden. Die Gesamtheit der Phasen eines Projektes wird als *Lebenszyklus* (project life cycle) bezeichnet. *Phasenmodelle* dienen allgemein dazu, zeitlich ausgedehnte und komplexe Projekte sequenziell zu strukturieren und damit die Projektabwicklung und Projektdarstellung zu erleichtern. In der Evaluationsforschung sind unterschiedliche Phasenmodelle anwendbar: Phasenmodelle des Forschungsprozesses (vgl. Abschnitt 1.1) sowie Phasenmodelle für Projekte (vgl. Abschnitt 1.2), die hier in einem heuristischen Phasenmodell für Evaluationsforschungsprojekte zusammengefasst werden (vgl. Abschnitt 1.3).

1.1 Phasenmodelle des Forschungsprozesses

Im Grunde verlangt jede empirische Studie einen geordneten Ablauf, um das Ziel des wissenschaftlichen Erkenntnisgewinns zu erreichen. Dabei hat sich für den empirischen Forschungsprozess in den Sozialwissenschaften ein allgemeines

Phasenmodell eingebürgert, das sowohl die praktische Vorgehensweise als auch die Ergebnisdarstellung (z. B. Gliederung von Artikeln in Fachzeitschriften) strukturiert.

Zwar gibt es in der Ausdifferenzierung und Benennung der einzelnen Phasen leichte Abweichungen, doch im Kern besteht über die inhaltlichen *Phasen empirischer Forschung* in der Fachliteratur Konsens (vgl. z. B. Bortz, 2005, S. 2 ff.: sechs Phasen; Neuman, 2003, S. 13: sieben Phasen; Schnell, Hill & Esser, 1993, S. 118 ff.: neun Phasen; Kromrey, 2000, S. 72 ff.: elf Phasen). Der idealtypische lineare Phasenablauf empirischer Forschung wird lediglich im Bereich der theoriebildenden qualitativen Forschung teilweise durch einen zirkulären Ablauf ersetzt (Witt, 2001).

Jede Forschungsphase endet mit einem Arbeitsergebnis, das Voraussetzung für die nächste Phase ist (vgl. Tab. 1).

Evaluationsstudien lassen sich als sozialwissenschaftliche Studien kennzeichnen (vgl. Westmeyer in diesem Band). Dementsprechend müssen sich in einem Phasenmodell der Evaluationsforschung die Phasen des empirischen Forschungsprozesses wiederfinden.

Tabelle 1:
Phasenmodell des empirischen Forschungsprozesses

Phasen des Forschungsprozesses	Ergebnisse der Phasen
1. Problemwahl	Formulierung des Forschungsproblems bzw. der Forschungsziele
2. Gegenstandsexploration	Gegenstandsbeschreibung, Zusammenfassung des Forschungsstandes
3. Theoriebildung	Theoretisches Modell mit zugehörigen Forschungsfragen und Hypothesen
4. Untersuchungsplanung	Untersuchungsdesign, Stichprobenplan, Instrumente
5. Datenerhebung	Quantitative und/oder qualitative empirische Daten
6. Datenanalyse	Quantitative und/oder qualitative Ergebnisse
7. Interpretation	Lösung des Forschungsproblems, Aussagen zur Gültigkeit der Theorie, Schlussfolgerungen für Forschung und Anwendungspraxis
8. Publikation	Vorträge, Aufsätze etc.

1.2 Phasenmodelle für Projekte

Während Grundlagenforscher ihre Forschungsprojekte oft relativ autark und flexibel abwickeln können, sind Evaluationsforscher aus zwei Gründen viel stärker in wissenschaftsexterne Abhängigkeitsverhältnisse eingebunden:

- Ein Evaluationsforschungsprojekt ist immer mit einem Interventionsprojekt zeitlich und organisatorisch verzahnt, wodurch sich die Komplexität deutlich erhöht (Abstimmung zeitlicher Abläufe, Verständigung mit verschiedenen Interessengruppen etc.).
- Ein Evaluationsforschungsprojekt ist in der Regel eine Dienstleistung für einen Auftraggeber, weshalb die Studienabwicklung transparent und vertragskonform erfolgen muss (z. B. Finanzplanung, termingenaue Abgabe von Zwischen- und Endberichten).

Beide Anforderungen legen es nahe, Evaluationsforschungsprojekte ab einer gewissen Mindestgröße mit Methoden des systematischen *Projektmanagements* zu bearbeiten, einem Ansatz, der aus der Betriebswirtschaft stammt und beispielsweise Methoden der Zeit-, Finanz- und Personalplanung sowie des Kommunikations- und Risikomanagements umfasst. Systematisches Projektmanagement wird durch entsprechende Software unterstützt, wobei ein sehr breites Spektrum an nicht kommerziellen und kommerziellen Lösungen zur Verfügung steht (z. B. Microsoft „Project“).

Im Rahmen des Projektmanagements arbeitet man üblicherweise mit Phasenmodellen; es existieren jedoch auch Vorgehensmodelle, die sequenziell vorgeplante Phasen ablehnen. Die Anzahl der Projektphasen und auch der Formalismus, mit dem sie abgewickelt werden, sind von Art, Umfang, Risiko und Bedeutung eines Projektes sowie der gewünschten Einflussnahme des Auftraggebers abhängig (Kuster et al., 2006, S. 16). Je nachdem, ob es sich beispielsweise um Bauprojekte, um Produktentwicklungsprojekte oder um Organisations- und Informatikprojekte handelt, werden teilweise branchenspezifische Phasenbezeichnungen verwendet (Kessler & Winkelhofer, 2004, S. 124; Kuster et al., 2006, S. 16).

Im Kern gehen die meisten Phasenmodelle im Projektmanagement jedoch von fünf Phasen aus (vgl. Tab. 2). Diese laufen in der Praxis – auch in ihren Subphasen (z. B. einzelne Schritte der Realisierung) – nicht strikt sequenziell ab, sondern sind teilweise parallel und iterativ angelegt.

Generell werden Projektmanagementmethoden in der Forschungspraxis eher selten eingesetzt, allenfalls in großen Verbundprojekten sind Projektleiter tätig (vgl. Baumann, Escher & Witschi, 2005). Das mag einerseits daran liegen, dass Wissenschaftlerinnen und Wissenschaftler ihre Tätigkeit oft gar nicht explizit

Tabelle 2:
Phasenmodell für Projekte

Projektphasen	Ergebnisse der Phasen
1. Initiierung/Definition	Formulierung der Projektidee und Projektziele, Projektauftrag
2. Konzeption	Projektkonzept
3. Planung/Gestaltung/Spezifikation	Detailbeschreibung ausgewählter Arbeitsweisen und Lösungsvorschläge
4. Realisierung/Entwicklung/Umsetzung	Umgesetzte und qualitätsgeprüfte Lösungsvorschläge
5. Abschluss/Einführung/ Implementierung/Routine	Routineeinsatz der Lösung, Abschlussbericht, Verwertung der Ergebnisse

unter ökonomischen bzw. Management-Gesichtspunkten betrachten. Andererseits liegen bislang auch nur wenige Ansätze vor, die Methoden des systematischen Projektmanagements ausdrücklich auf Forschungsprojekte beziehen und deren Besonderheiten beachten.

Im Zuge der Sparmaßnahmen im öffentlichen Hochschulwesen und der wachsenden Bedeutung von Drittmittel- und Auftragsforschung (darunter eben auch Evaluationsforschung) wächst in den letzten Jahren jedoch das Bewusstsein für systematisches *Forschungsprojektmanagement* (research project management). Eine Vielzahl von Weiterbildungsangeboten für Wissenschaftlerinnen und Wissenschaftler vermittelt Projektmanagementkompetenz. Allerdings existieren nur wenige Fachbeiträge (z. B. Bode, 2000), Lehrbücher (z. B. Sharp, Peters & Howard, 2002) oder Handbücher (z. B. Badiru, 1996; Usherwood, 1996; Tarling, 2006) zum Thema Forschungsprojektmanagement. Ebenso selten ist auf den Forschungsbereich spezialisierte Projektmanagementsoftware zu finden (z. B. EFFCOR – Projektcontrolling für Forschungseinrichtungen und Universitäten, www.effcor.com).

Phasenmodelle für Projekte sind für die Evaluationsforschung nicht nur relevant, weil große Evaluationsprojekte systematisches Projektmanagement erfordern, sondern auch, weil die untersuchten Interventionen ihrerseits einen *Projekt- bzw. Produktlebenszyklus* durchlaufen. Es ist für die Evaluationsforschung von entscheidender Bedeutung, in welcher Projektphase sich das *Interventionsprojekt* gerade befindet: Soll z. B. eine neue Therapiemethode entwickelt (Konzeptphase der Intervention) oder eine etablierte Therapiemethode überprüft werden (Routinephase der Intervention)?

1.3 Phasenmodelle für Evaluationsforschungsprojekte

Phasenmodelle für Evaluationsforschungsprojekte müssen Prozessabläufe auf drei Ebenen integrieren:

1. Aus wissenschaftstheoretischer Perspektive müssen bei einer Evaluation die Phasen des empirischen Forschungsprozesses durchlaufen werden.
2. Aus Projektmanagementperspektive müssen die Projektphasen des Evaluationsprojektes systematisch organisiert werden.
3. Aus Entwurfsperspektive ist zu beachten, in welcher Phase oder welchen Phasen des Interventionslebenszyklus Evaluationsmaßnahmen greifen sollen und wie jeweils eine nützliche Rückkopplung der Evaluationsergebnisse in den Prozess der Interventionsentwicklung möglich ist.

Innerhalb der Evaluationsforschung haben sich unterschiedliche Ansätze zur Beschreibung des Vorgehens herausgebildet, etwa in Abhängigkeit von der bevorzugten Forschungsmethodologie oder der Art der Beteiligung der Interessengruppen (vgl. Wittmann in diesem Band). Einige dieser Evaluationsmodelle beinhalten Sequenzierungen und haben damit den Charakter von Phasenmodellen. Dies trifft unter anderem auf das bekannte CIPP-Modell (Context, Input, Process & Product Evaluation) zu, das in seiner ersten Fassung bereits 1966 von Daniel Stufflebeam publiziert wurde (Stufflebeam, 2007). Das Modell bildet die verschiedenen Phasen des Interventionslebenszyklus ab und beschreibt die in jeder Phase möglichen bzw. notwendigen Evaluationsschritte (vgl. Tab. 3).

Tabelle 3:
Verzahnung von Intervention und Evaluation (eigene Darstellung)

Projektphasen der Intervention	CIPP-Modell der Evaluation
1. Initiierung	Context Evaluation: What needs to be done?
2. Konzeption	
3. Planung	Input Evaluation: How should it be done?
4. Realisierung	Process Evaluation: Is it being done?
5. Abschluss	Product Evaluation: Did it succeed? – Impact – Effectiveness – Sustainability – Transportability

Je nach Status der Intervention kann gemäß dem CIPP-Modell eine Evaluationsstudie eine oder mehrere Teilevaluationen enthalten. Generell wird in vielen Praxisfeldern betont, dass Evaluationsmaßnahmen bereits in den frühen Projektphasen sinnvoll sind. Denn je weiter das Interventionsprojekt fortgeschritten ist, umso schwieriger und aufwendiger wird es, identifizierte Probleme nachträglich zu korrigieren. Somit wird die Kopplung von Interventionsentwicklungsphasen einerseits und Evaluationsforschungsphasen andererseits zu einem wichtigen Aspekt von Evaluationsprojekten (vgl. Gediga & Hamborg, 2002, zur Integration von Softwareentwicklung und Softwareevaluation).

Unabhängig davon, ob beispielsweise eine Prozessevaluation und/oder eine Produktevaluation durchgeführt wird, durchläuft das Evaluationsprojekt einen Projektlebenszyklus. Den Projektphasen lassen sich dabei jeweils bestimmte Projektmanagementaufgaben sowie bestimmte Schritte des empirischen Forschungsprozesses zuordnen, zudem sind evaluationsspezifische Arbeitsschritte integriert (vgl. Wottawa & Thierau, 1998; Stufflebeam & Shinkfield, 2007, S. 94).

Tabelle 4:
Ein Phasenmodell für Evaluationsforschungsprojekte

Projektphasen der Evaluation	Arbeitsschritte
1. Initiierung	1.1 Identifikation von Evaluationsmöglichkeiten 1.2 Prüfung von Evaluationszielen
2. Konzeption	2.1 Exploration der Evaluationsbedingungen 2.2 Entwicklung des Evaluationskonzepts 2.4 Einbeziehung der Evaluationsstandards 2.3 Vorbereitung des Projektmanagements 2.5 Formulierung des Evaluationsvertrags
3. Planung	3.1 Formulierung der Evaluationsfragen und -hypothesen 3.2 Untersuchungsplanung (Untersuchungsdesign, Stichprobenplan Instrumente) 3.3 Projektplanung
4. Realisierung	4.1 Erhebung von Evaluationsdaten 4.2 Analyse von Evaluationsdaten 4.3 Interpretation von Evaluationsdaten 4.4 Projektcontrolling
5. Abschluss	5.1 Abgabe der Evaluationsberichte 5.2 Organisation der Evaluationsnutzung 5.3 Durchführung einer Metaevaluation 5.4 Publikation der Evaluationsstudie

Die systematische Entwicklung von Phasenmodellen der Evaluationsforschung ist Aufgabe der *Evaluationstheorie*. Eine empirische Prüfung vorgeschlagener Phasenmodelle ist möglich, indem die Arbeitsabläufe von Evaluationsprojekten erhoben (z. B. Dokumentenanalyse, Interviews mit Evaluatoren) und mit dem Modell verglichen sowie mit Erfolgskriterien verknüpft werden. Entsprechende Forschungsergebnisse zu Phasenmodellen der Evaluationsforschung stehen bislang aus, sodass das hier in Tabelle 4 und im folgenden Text vorgestellte Modell als heuristisch einzustufen ist. Heuristischen Charakter haben mangels empirischer Prüfung im Übrigen die meisten präskriptiven Evaluationsmodelle (vgl. Wittmann in diesem Band): Es ist kaum bekannt, ob die klassischen Evaluationsmodelle überhaupt systematisch und modellkonform in die Praxis umgesetzt werden und ob die modellkonforme Umsetzung mit Projekterfolg zusammenhängt (Stufflebeam & Shinkfield, 2007, S. 68 f.; zur Kluft zwischen Evaluationsmodellen und Evaluationspraxis siehe auch Kromrey, 2003).

Dass Strukturierungshilfen und Phasenmodelle in der Praxis benötigt werden, steht jedoch außer Zweifel. Zur Erleichterung und Professionalisierung der praktischen Arbeit in der Evaluationsforschung hat sich die *Checklistenmethode* etabliert (für eine Sammlung von Evaluationschecklisten siehe die Website des Evaluationszentrums der Universität Michigan: www.wmich.edu/evalctr/checklists/). Als Gedächtnisstütze liefert die Checkliste eine geordnete – und oft auch chronologische – Auflistung wichtiger Fragen, Kriterien oder Arbeitsschritte der Evaluationspraxis, implizit enthält sie dabei auch jeweils ein theoretisches Modell, bei chronologischen Listen ein Phasenmodell der Evaluationsarbeit (vgl. Scriven, 2005).

2 Initiierungsphase der Evaluation

In der Initiierungsphase geht es darum, potenzielle Evaluationsprojekte zu akquirieren, indem Evaluationsmöglichkeiten identifiziert (vgl. Abschnitt 2.1) und Evaluationsziele kritisch beleuchtet werden (vgl. Abschnitt 2.2).

2.1 Identifikation von Evaluationsmöglichkeiten

Es existieren im Wesentlichen vier Wege für Evaluatoren, um an mögliche Evaluationsprojekte zu kommen (vgl. Stufflebeam & Shinkfield, 2007, S. 453 ff.):

1. *Projektausschreibungen*: Öffentliche sowie private/gewerbliche Ausschreibungen (RFP: request for proposal) für Evaluationsprojekte skizzieren in der Regel das Evaluationsobjekt, Evaluationsziele sowie Merkmale der Evaluationsstudie (z. B. Zeitrahmen, Budget, bevorzugte Forschungsmethoden; vgl. Bussmann, 1995, S. 25). Bieter sind eingeladen, Angebote für Evaluationsstudien

einzureichen (Evaluationskonzept, Zeitplan, Kostenplan etc.) sowie ihre Expertise nachzuweisen (z. B. Belege früherer Studien, Firmenprofil, Lebenslauf etc.). Üblicherweise wird über die Angebote in Abhängigkeit von der inhaltlichen Qualität und dem Kostenvoranschlag entschieden. Öffentliche Evaluationsausschreibungen finden sich unter anderem im Deutschen Ausschreibungsblatt (online unter www.deutsches-ausschreibungsblatt.de). Ausschreibungen unterscheiden sich hinsichtlich ihres Umfangs und ihres Detaillierungsgrades erheblich. Oftmals ist ausdrücklich vorgesehen, dass das eingereichte Evaluationskonzept im Falle eines Zuschlags später gemeinsam mit dem Auftraggeber verfeinert oder modifiziert werden kann.

2. *Projektaufträge*: Viele Evaluatoren arbeiten nicht selbstständig, sondern in Unternehmen oder Evaluationsinstituten. Dabei sind sie teilweise in die Projektaukquisition involviert, teilweise bekommen sie Evaluationsaufträge zugeteilt. Der Evaluationsauftrag kann sich dabei auf einen unternehmensinternen Gegenstand beziehen (interne Evaluation: z. B. Evaluation des unternehmensinternen Weiterbildungssystems) oder beim Kunden angesiedelt sein (externe Evaluation).
3. *Projektanfragen*: Manchmal suchen sich Auftraggeber auch gezielt Evaluatoren aus und bieten ihnen Projekte an. Dabei handelt es sich typischerweise um besonders namhafte Evaluatoren oder um Evaluatoren, mit denen das Unternehmen bereits früher zusammengearbeitet hat. Die Chancen, Projektanfragen zu erhalten, erhöhen sich für Evaluatoren in Abhängigkeit von Art und Anzahl erfolgreich durchgeführter früherer Evaluationsprojekte sowie entsprechender Öffentlichkeitsarbeit (z. B. öffentliche Vorträge, Seminare, Internetangebote).
4. *Evaluator-initiierte Projekte*: Auf der Basis ihrer Fachexpertise und Praxiserfahrung in einem bestimmten Evaluationsbereich können Evaluatoren auch selbst Projektideen entwickeln. Dabei kann es sich einerseits um Projekte handeln, die aus Evaluatorperspektive vor allem von wissenschaftlichem Interesse sind, publizierbare Ergebnisse versprechen und der fachlichen Profilierung dienen. Derartige Evaluationsforschungsprojekte können in kleinem Rahmen teilweise auf Selbstkostenbasis umgesetzt werden (z. B. informelle Kooperation mit Praxispartnern). Geht es jedoch um umfangreichere Projekte und finanzierte Projektaufträge, so müssen finanzkräftige potenzielle Auftraggeber von der Evaluationsidee überzeugt werden.

2.2 Prüfung der Evaluationsziele

Nicht jede Evaluationsmöglichkeit ist es wert, weiterverfolgt zu werden. Erfährt man etwa sehr spät von einer Projektausschreibung, so ist realistischerweise oft nicht genug Zeit vorhanden, um ein konkurrenzfähiges Angebot auszuarbeiten. Vorsicht ist auch geboten bei Ausschreibungen, deren Evaluationsbereich rela-

tiv weit von der eigenen Fachexpertise entfernt ist. Denn dies verlangt einerseits deutlichen Mehraufwand bei der Einarbeitung und außerdem wird der Zuschlag vom Qualifikationsprofil des Bieters abhängig gemacht (zu Evaluationsbereichen innerhalb der Psychologie vgl. Moosbrugger & Schweizer, 2002). Fachliche Nähe der Evaluatoren zum Evaluationsobjekt erleichtert und verbessert nicht nur das theoretische und methodische Arbeiten, sondern geht in der Regel auch mit höherer Glaubwürdigkeit und Autorität gegenüber den Stakeholdern einher sowie mit dem Zugriff auf einschlägiges soziales Kapital (z. B. Expertenkontakte, Kollegen-Netzwerke). Nicht nur Projektausschreibungen, sondern auch externe Projektaufträge, die man erhält, müssen kritisch überprüft werden, etwa dahingehend, ob auf Evaluatorsseite genügend personelle Ressourcen für ein unvorhergesehen angebotenes Projekt bereitgestellt werden können.

Neben den operativen Restriktionen (Zeit, Personal, Fachexpertise) ist bei der Entscheidung für ein Evaluationsprojekt die *Problemdefinition bzw. Zielsetzung* von herausragender Bedeutung für den Projekterfolg. Man unterscheidet zwischen den *Zielen einer Intervention* und den *Zielen, Zwecken oder Funktionen der Evaluation*. Es lassen sich fünf Hauptfunktionen bzw. Ziele der Evaluation differenzieren (vgl. Stockmann, 2000):

1. *Erkenntnisfunktion*: Evaluationsforschung trägt definitionsgemäß dazu bei, wissenschaftliche Erkenntnisse über die Eigenschaften und Wirkungen von Interventionen zu sammeln.
2. *Optimierungsfunktion*: Wo liegen Stärken der Intervention im Hinblick auf die Interventionsziele und wie lassen sie sich ausbauen? Wo liegen Schwächen der Intervention und wie lassen sie sich beseitigen?
3. *Kontrollfunktion*: Wird das Projekt korrekt umgesetzt? In welchem Maße (Effektivität) und mit welcher Effizienz (Kosten-Nutzen-Bilanz) werden die intendierten Wirkungen der Maßnahme (Interventionsziele) erreicht? Welche nicht intendierten positiven und negativen Nebenwirkungen treten auf?
4. *Entscheidungsfunktion*: Soll eine bestimmte Intervention gefördert, umgesetzt, weiterentwickelt, genutzt etc. werden oder nicht? Welche von mehreren vergleichbaren Interventionen soll gefördert, umgesetzt, weiterentwickelt, genutzt etc. werden?
5. *Legitimationsfunktion*: Sowohl die Durchführung von Evaluationsforschung als auch ihre Befunde sollen dazu beitragen, die Entwicklung und Durchführung einer Intervention nach außen zu legitimieren und vor allem über die Verwendung öffentlicher Gelder Rechenschaft abzulegen.

Um an diesen Evaluationszielen arbeiten und insbesondere Optimierungs-, Kontroll- und Entscheidungsziele erreichen zu können, ist es notwendig, dass die Merkmale und Ziele der Intervention konkretisiert und messbar gemacht

werden. Nicht selten zeigt sich, dass Auftraggeber die fragliche Intervention und ihre Ziele nur vage charakterisieren können. Wenn es laut Auftraggeber im Sinne der Kontrollfunktion z. B. Ziel der Evaluationsstudie sein soll zu prüfen, „ob Interneteinsatz in der Schule das Lernen verbessert“, so muss das Evaluationsteam vermitteln, dass eine derart pauschale Kausalaussage im Rahmen der Evaluationsstudie realistischerweise gar nicht zu beantworten ist. Stattdessen müssen bei der Transformation der *Grobziele* in *Feinziele* die Erwartungen hinsichtlich Kausalitätsnachweisen und Generalisierbarkeit der Befunde beispielsweise oft zugunsten von Zusammenhangsaussagen für ausgewählte Nutzungs- und Lernvariablen heruntergeschraubt werden.

Im Zuge der Modifikation von Grobzielen zu Feinzielen können die unterschiedlichen Perspektiven der Stakeholder stärker und kontroverser zutage treten als bei den allgemein konsensfähigen Grobzielen. So mögen beispielsweise Ärzte und Patienten darin übereinstimmen, dass mit Behandlungsfehlern professionell umzugehen ist und Evaluationsstudien zur Qualität des Fehlermanagements in Arztpraxen sinnvoll sind. Doch hinsichtlich der konkreten Definition von „Fehlern“ sowie der Erwartungen an „professionelles Verhalten“ zeigen sich im Detail starke Differenzen zwischen beiden Interessengruppen (vgl. Gallagher et al., 2003), obwohl natürlich auch Ärzte zuweilen die Rolle von Patienten einnehmen.

Für potenzielle Evaluatoren einer Intervention gilt es im Vorfeld einzuschätzen, wie Erfolg versprechend eine Transformation von Grobzielen (goals) zu praktikablen Feinzielen (objectives) in Kooperation mit Auftraggebern und weiteren Stakeholdern erscheint. Denn ohne eine wissenschaftlich operationalisierbare Problemdefinition kann Evaluationsforschung nicht erfolgreich betrieben werden.

Neben den oben genannten fünf *rationalen Evaluationszielen*, die mit dem wissenschaftlichen Selbstverständnis harmonieren und Ergebnisoffenheit beinhalten, wird Evaluation teilweise auch für fragwürdige Zwecke instrumentalisiert (vgl. Suchman, 1967; Wottawa & Thierau, 1998, S. 29 ff.). Es resultieren dann sogenannte *Pseudo-Evaluationen* (Stufflebeam & Shinkfield, 2007). So nutzen manche Auftraggeber Evaluation nur zur *Selbstdarstellung*, indem sie Mitarbeiter- oder Kundenbefragungen durchführen und somit Mitarbeiter- bzw. Kundenorientierung demonstrieren, ohne jedoch die dabei gewonnenen Daten jemals systematisch auszuwerten oder in die Praxis umzusetzen. Manche Auftraggeber versuchen, sich positive Evaluationsergebnisse durch großzügige Evaluationsaufträge zu erkaufen. Bei umstrittenen Interventionen kann die Evaluation auch für eine *Verantwortungsdelegation* der Entscheidungsträger an die Evaluierenden genutzt werden oder auch eine Methode sein, um praktisches Handeln zu *verzögern*. Evaluation kann zudem als politische *Durchsetzungshilfe* eingesetzt werden,

sei es von Befürwortern oder Gegnern einer Intervention (die dann z. B. Einfluss darauf nehmen, welche Fragen gestellt, welche Gruppen befragt und welche Ergebnisse publiziert werden).

Ein authentisches Fallbeispiel soll die Umstände einer Pseudoevaluation illustrieren (Stufflebeam & Shinkfield, 2007, S. 152 f.): Ein Forschungszentrum wird seit Jahren regelmäßig mit sehr positiven Ergebnissen hinsichtlich der Bedeutung, Qualität und Menge der produzierten Forschungsergebnisse extern evaluiert. Der neue Direktor des Zentrums bespricht mit dem Evaluationsteam, dass die vorliegenden positiven Evaluationen zwar die Moral der Beschäftigten steigerten, aber keinerlei Optimierungsideen lieferten. Deswegen sei es im Sinne der Qualitätssteigerung erforderlich, dass die Evaluation sich nun verstärkt darauf konzentriere, Schwächen aufzudecken, aus denen sich dann Verbesserungsmaßnahmen ableiten lassen. Dementsprechend erarbeitet das Evaluationsteam eine umfassende Aufstellung von Schwächen der Einrichtung und der dort abgewickelten Forschungsprogramme. Als das Evaluationsteam seine Ergebnisse präsentiert, stellt sich heraus, dass nicht nur die Beschäftigten des Forschungszentrums, sondern auch die Geldgeber anwesend und über die Negativbilanz entsetzt sind. Das Evaluationsteam muss erkennen, dass es durch den einseitigen Evaluationsauftrag von dem neuen Direktor instrumentalisiert wurde, dessen Interesse darin bestand, die Arbeit seines Vorgängers zu diskreditieren und laufende Forschungsprogramme durch andere seiner Wahl zu ersetzen.

Auftraggeber mit dominanten Instrumentalisierungsabsichten sind – wenn möglich – im Vorfeld zu identifizieren und zu meiden. Für derartige Einschätzungen sind Hintergrundinformationen aus dem Praxisfeld und Kollegenkontakte hilfreich.

Doch strategische Ziele verfolgen nicht nur Auftraggeber, sondern teilweise natürlich auch Evaluatoren, etwa wenn sie überzogene Erwartungen von Auftraggebern (etwa hinsichtlich globaler Wirksamkeitsnachweise) nicht als empirisch undurchführbar zurückweisen, um einen lukrativen Evaluationsauftrag nicht zu verlieren (vgl. Uhl, 1997). Oder wenn sie mit den Interventionsbeteiligten sympathisieren und negative Evaluationsbefunde zurückhalten, um eine weitere Förderung der Intervention nicht zu gefährden.

Abgesehen von *Fachkompetenz* und *Interesse* am Inhaltsbereich sind in der Evaluationsforschung *Wertfragen* virulent, und zwar aufgrund ihrer thematischen Vorgaben und ihres direkten Praxisbezuges stärker als in der thematisch selbst bestimmten und praxisferneren Grundlagenforschung. Wenn Evaluatoren beispielsweise daran mitarbeiten, Methoden des militärischen Schießtrainings oder Kampagnen radikaler Parteien hinsichtlich ihrer Stärken und Schwächen zu bewerten, damit sie dann verbessert werden können, nehmen sie damit Einfluss

auf die Gestaltung unserer Lebensbedingungen und müssen dies mit ihren Werten vereinbaren können. Wertfragen im Bereich *ethisch kontroverser Praxisfelder und Interventionen* werden teilweise durch Institutionen beantwortet, in denen Evaluatoren tätig sind (z. B. Leitbilder von Forschungseinrichtungen, die z. B. nur eine friedliche, zivile Nutzung von Forschungsergebnissen vorsehen), meist sind sie jedoch eine Frage der persönlichen Haltung.

3 Konzeptionsphase der Evaluation

Wurden mögliche Evaluationsprojekte identifiziert, deren Zielsetzungen machbar und verantwortbar erscheinen, so müssen im nächsten Schritt die Evaluationsbedingungen genauer exploriert (vgl. Abschnitt 3.1) und das Evaluationskonzept bzw. Projektangebot entwickelt werden (vgl. Abschnitt 3.2). Dabei ist es notwendig sicherzustellen, dass die gängigen Evaluationsstandards eingehalten (vgl. Abschnitt 3.3), das Projektmanagement vorbereitet (vgl. Abschnitt 3.4), und verbindliche Vereinbarungen zwischen Evaluatoren und Auftraggebern getroffen werden (vgl. Abschnitt 3.5).

3.1 Exploration der Evaluationsbedingungen

Evaluationsteam. Entscheidend für den Erfolg eines Evaluationsprojektes ist das Evaluationsteam, dessen Größe und Zusammensetzung von Art und Umfang des avisierten Projektes abhängt. Das Spektrum reicht von *Single-Evaluator-Projekten* bis zu *arbeitsteiligen Teams* mit einem Dutzend oder mehr Mitgliedern (vgl. Stufflebeam & Shinkfield, 2007, S. 42 ff.). Bei größeren Projekten gibt es insbesondere eine Arbeitsteilung zwischen Projektmanagement (Projektleiter) und Evaluationsforschung, wobei auf Seiten der Forschung erfahrene Evaluationsexperten insbesondere für die Untersuchungsplanung und Interpretation der Ergebnisse notwendig sind, während einzelne Schritte der Datenerhebung und Datenauswertung auch von geschulten Hilfskräften übernommen werden können (z. B. Tätigkeiten als Beobachter, Interviewer, Codierer etc.). Zur Qualitätssicherung sollte von vornherein eine *Metaevaluation* in das Evaluationsprojekt integriert werden, das heißt, eine Überprüfung der Evaluationsarbeit durch externe Evaluationsexperten. Große Evaluationsprojekte kommen heute kaum ohne Informations- und Kommunikationstechnologie aus (z. B. Aufbau von Datenbanken und Intranets, Beschaffung und Wartung von Spezialsoftware etc.), weshalb teilweise auch IT-Experten in einem Evaluationsteam mitarbeiten. Frühzeitige Überlegungen zur Zusammensetzung und Rekrutierbarkeit des Evaluationsteams sind notwendig, da Angaben zum Team für das Projektangebot benötigt werden und potenzielle Teammitglieder zur Vorbereitung des Projektes beitragen können. Je nachdem, ob potenzielle Teammitglieder extern

rekrutiert werden (z. B. Evaluationsexperten einer fremden Universität für die Metaevaluation) oder von der eigenen Organisation stammen (z. B. Mitarbeiter einer Evaluationsagentur) sind Vorabsprachen mit den Betroffenen sowie mit der Leitungsebene geboten.

Evaluationsgegenstand. Ein Evaluationskonzept kann nur dann adäquat vorgeschlagen werden, wenn die Merkmale der Intervention und ihrer Nutzer, die Bedingungen am Evaluationsort sowie die Erwartungen des Auftraggebers möglichst genau bekannt sind. Insbesondere wenn ein konkurrenzfähiges Angebot für eine Projektausschreibung erarbeitet werden soll, ist es angeraten, Hintergrundinformationen zu recherchieren. Abgesehen von publizierten Materialien sind vor allem informelle Gespräche wertvoll. Inhaltliche Vorklärungen sollten vor allem folgende Aspekte betreffen: Identifikation der relevanten Stakeholder (z. B. Entwickler, Umsetzer und Nutzer der Intervention) und ihrer jeweiligen Fragen, Bedürfnisse und Vorbehalte sowie die Möglichkeiten der Datenerhebung im Feld (z. B. Orte für Beobachtungen, Zeitpunkte für Befragungen, Zugriff auf Dokumente etc.). Hilfreich kann es auch sein, sich im Vorfeld genau zu überlegen, welche Personen, Gruppen oder Institutionen durch die Evaluation möglicherweise beeinträchtigt werden, um hier entsprechend Rücksprache zu halten oder Vorkehrungen zu treffen.

3.2 Entwicklung des Evaluationskonzepts

Das Evaluationskonzept (evaluation design) muss darlegen, wie die Evaluationsziele des Auftraggebers im vorgegebenen zeitlichen und finanziellen Rahmen und mit hoher wissenschaftlicher Qualität erreicht werden sollen. Das Evaluationskonzept ist das Kernstück eines Projektangebotes. Bei der Ausarbeitung des Evaluationskonzeptes kann man sich an erfolgreichen Konzepten aus anderen Projekten des fraglichen Evaluationsbereiches orientieren (Zugriff über Publikationen, Internet, Kollegen-Netzwerke). Zudem existieren Checklisten, die die einzelnen Arbeitsschritte der Evaluationskonzeptentwicklung beschreiben, etwa die „Evaluation Design Checklist“ (Stufflebeam & Shinkfield, 2007, S. 493 ff.).

Das Evaluationskonzept muss den Auftraggeber überzeugen. Darüber hinaus muss es im Falle eines Projektzuschlages die Projektabwicklung begleiten und die Projektdarstellung unterstützen. Wichtig ist es deswegen, das Konzept sachgerecht und anschaulich zu charakterisieren. Hierfür bietet sich ein Rückgriff auf die etablierten Evaluationsmodelle an. Die verschiedenen in der Literatur entwickelten Formen, Ansätze und Modelle der Evaluation sind auf ganz unterschiedlichen Dimensionen angesiedelt (vgl. Wittmann in diesem Band; Westermann, 2002; Stufflebeam & Shinkfield, 2007):

- *Nähe oder Distanz der Evaluatoren* zur Intervention (Selbstevaluation versus Fremdevaluation; interne Evaluation versus externe Evaluation),
- *Anzahl der zu untersuchenden Interventionen* (Einzelevaluation versus vergleichende Evaluation),
- *Entwicklungsphase der Intervention* (Ex-ante-Evaluation versus On-going-Evaluation versus Ex-post-Evaluation oder formative Evaluation versus summative Evaluation oder CIPP-Modell),
- *Forschungsfokussierung* auf Methode, Theorie oder Praxis (methodenbasierte Evaluation versus theoriebasierte Evaluation versus praxisbasierte Evaluation),
- *Grad der Partizipation* unterschiedlicher Stakeholder (partizipative Evaluation versus nicht partizipative Evaluation)

Ein Evaluationsprojekt kann auf verschiedenen Ebenen unterschiedlichen Modellen gleichzeitig zugeordnet werden (z. B. externe *und* partizipative Evaluation). Welches Modell oder welche Modelle von mehreren passenden dann prominent zur Charakterisierung und als Titel des Konzepts gewählt werden, sollte auf die Projektziele und den Auftraggeber zugeschnitten sein.

Das Evaluationskonzept muss gemäß dem Phasenmodell der Evaluation Aussagen zu folgenden Aspekten treffen (vgl. Tab. 4; vgl. Stufflebeam & Shinkfield, 2007, S. 493 ff.):

1. Evaluationsziele (Grobziele und Feinziele),
2. Forschungsfragen und Hypothesen, ggf. auch Theoriebildung,
3. Untersuchungsplanung (mit Bezug zu Evaluationsmodellen und Evaluationsstandards),
4. Projektmanagement (Zeitplan, Kostenplan, Projektorganisation),
5. Datenerhebung und Datenanalyse,
6. Art und Umfang der Evaluationsberichte,
7. Maßnahmen zur Evaluationsnutzung,
8. Metaevaluation.

Da das Evaluationskonzept vor Vertragsabschluss erstellt wird, muss auf Seiten des Evaluators hier eine Vorarbeit geleistet werden, die sich möglicherweise nicht auszahlt. Dementsprechend können Angaben im Evaluationskonzept – etwa zur Untersuchungsplanung – zunächst nur kursorisch erfolgen. Eine detaillierte Ausarbeitung erfolgt dann erst in der Planungsphase nach Vertragsabschluss.

Einzelne Evaluatoren sowie Evaluationsagenturen/-institute haben sich in der Regel auf die Umsetzung eines bestimmten Evaluationsmodells spezialisiert, sodass sie dieses Modell dann auf die jeweilige Konzeptentwicklung anwenden und somit ihren Arbeitsaufwand reduzieren (Selbstvorstellungen von Evaluationseinrichtungen und den von ihnen präferierten Evaluationsmodellen finden sich regelmäßig in der Zeitschrift für Evaluation).

3.3 Einbeziehung der Evaluationsstandards

Im Sinne der Qualitätssicherung soll Evaluation sich an spezifischen Standards orientieren. Evaluationsstandards sichern nicht nur die wissenschaftliche Qualität der Evaluationsforschung, sondern auch ihren gesellschaftlichen Nutzen. Zudem sollen Evaluationsstandards dabei helfen, in der Praxis sowohl die Interessen der Auftraggeber und anderen Stakeholder als auch die Interessen von Fremd- bzw. Selbstevaluatoren zu wahren.

Gemäß den *Evaluationsstandards der Deutschen Gesellschaft für Evaluation*, die aus dem einschlägigen amerikanischen Standard-Set (Sanders, 1999) abgeleitet sind und denen der Schweizerischen Evaluationsgesellschaft ähneln (seval, 2000), haben Evaluationen vier grundlegende Eigenschaften aufzuweisen (DeGEval, 2002b):

1. Nützlichkeit,
2. Durchführbarkeit,
3. Fairness und
4. Genauigkeit.

Diesen vier Qualitätsdimensionen sind insgesamt 25 Einzelstandards zugeordnet. Nicht in jeder Evaluationsstudie lassen sich alle Einzelstandards realisieren, da diese teilweise auch in Konkurrenz zueinander stehen. Die Evaluationsstandards umfassen neben den *Besonderheiten der Auftragsforschung* sowohl *wissenschaftliche Gütekriterien* (so umfasst der Genauigkeitsstandard beispielsweise Fragen der Objektivität, Reliabilität und Validität von Erhebungsinstrumenten) als auch *forschungsethische Richtlinien* (so werden Aspekte der Anonymität und des Datenschutzes beispielsweise unter dem Fairnesskriterium abgehandelt, das auch die Fairness von Testverfahren anspricht). Evaluationsstandards der Nützlichkeit und Durchführbarkeit sind für die Analysephase einer Evaluation besonders einschlägig.

Zudem ist bereits im Vorfeld zu analysieren, wie die Einhaltung der Evaluationsstandards während des Evaluationsprojektes sichergestellt und bei Bedarf nach Abschluss der Evaluation auch zusammenfassend bewertet werden soll (z. B. in Form einer formativen und/oder summativen Metaevaluation, vgl. Abschnitt 6.3). Im Falle von *Selbstevaluationen* sind zusätzlich zu den vier allgemeinen Qualitätsnormen zusätzliche *Muss-Standards* sicherzustellen, die unter anderem für gesicherte Rahmenbedingungen und einen Schutz der – typischerweise abhängig beschäftigten – Selbstevaluatoren sorgen sollen (Müller-Kohlenberg & Beywl, 2002).

Während sich die Evaluationsstandards auf die Qualität von Evaluationsstudien beziehen, sind die von der American Evaluation Association herausgegebenen

Guiding Principles for Evaluators (Shadish, Newman, Scheirer & Wye, 1995) auf die Evaluatoren und ihr Selbstverständnis zugeschnitten; sie fordern:

- die systematische, datenbasierte Untersuchung des zu evaluierenden Gegenstands,
- den Beleg ausreichender Fachkompetenz zur Durchführung von Evaluationsstudien,
- die Gewährleistung eines fairen und integren Evaluationsprozesses,
- den angemessenen Respekt gegenüber den Persönlichkeitsrechten aller Beteiligten sowie
- ein allgemeines Verantwortungsgefühl für die durch Programme beeinflusste öffentliche Wohlfahrt.

Auftraggebern wird empfohlen, nur Evaluatoren zu engagieren, die sich diesen Prinzipien verpflichtet haben. Dementsprechend ist es umgekehrt von Seiten der Evaluatoren wichtig, ihre Forschungsleitlinien gegenüber potenziellen Auftraggebern transparent zu machen.

3.4 Vorbereitung des Projektmanagements

Im Zuge der Konzeption einer Evaluationsstudie geht es nicht nur um eine Grobplanung des wissenschaftlichen Vorgehens, sondern auch um eine Grobplanung der operativen Umsetzung im Rahmen des Projektmanagements. Das sogenannte „magische Dreieck“ des Projektmanagements besteht aus

- Zeit (der Projektbearbeitung bis zum Schlusstermin),
- Kosten (für alle benötigten Ressourcen) und
- Leistungen (Quantität und Qualität gemäß Projektzielen).

Alle drei Komponenten stehen in einem wechselseitigen Abhängigkeitsverhältnis. Soll etwa die Projektzeit verkürzt werden, so werden sich die Leistungen verringern und/oder die Kosten (z. B. durch zusätzliches Personal) steigen. In der Konzeptionsphase erfolgt eine Projektgrobplanung vor allem hinsichtlich Zeit und Kosten, die dann in der Planungsphase detailliert wird.

Bei ausgeschriebenen Evaluationsprojekten ist die *Projektlaufzeit* in der Regel vorgegeben, während die Projektkosten in der Konzeptionsphase abzuschätzen sind. Zur Vorbereitung des *Zeitmanagements* im Evaluationsprojekt wird in der Regel in der Konzeptionsphase bereits ein *grober Zeitplan* mit den wichtigsten Meilensteinen entwickelt, der dem Projektangebot beigelegt wird. Bezüglich des Zeitmanagements ist insbesondere auch eine Verzahnung des Interventionsprozesses mit dem Evaluationsforschungsprozess notwendig, etwa für die Terminie-

rung von Vorher-Messungen sowie zur Sicherstellung rechtzeitiger Ergebnismeldungen.

Aus Sicht des Auftraggebers sollte der Aufwand für die Evaluation in vernünftiger Relation zum Aufwand der Intervention stehen. Üblicherweise werden bei der Entwicklung von Interventionen ca. 5 % des Projektbudgets für die Selbstevaluation veranschlagt; für externe Evaluationen ist in der Regel mit Ausgaben in Höhe von rund 10 bis 15 % der Kosten für die Intervention zu rechnen (Bundesamt für Gesundheit der Schweiz, 1997, S. 24). Das *Evaluationsbudget* umfasst dabei Personalkosten, Betriebskosten (z. B. Spesen, Sekretariatskosten), Sachkosten (z. B. Druck- und Versandkosten, Übersetzungskosten) sowie – wenn notwendig – Investitionskosten (z. B. Apparate, EDV). Typischerweise müssen Bieter bei öffentlichen Ausschreibungen Fixkostenangebote unterbreiten.

Um die für die Forschungsleistung notwendigen Kosten abzuschätzen, ist anhand der Evaluationsziele und des Evaluationskonzeptes ein Ressourcen- und *Kostenplan* zu erstellen. Dazu werden zunächst die im Projekt anfallenden Aufgaben identifiziert (Projektstrukturplan). Für die einzelnen Aufgaben wird dann angegeben, welche Ressourcen sie erfordern (Personal, Geräte etc.), woraus dann die Kosten abgeleitet werden. Das Projektmanagement ist bei größeren Evaluationsprojekten neben den Forschungsaufgaben als eigene Aufgabe einzubeziehen (z. B. Kosten für die Beschäftigung eines Projektleiters, Arbeitsplatz des Projektleiters, Reisekosten etc.). Für die Erstellung von Kostenplänen existieren im Projektmanagement entsprechende Methoden und Tools (Kuster et al., 2006, S. 124), zudem liegen spezialisierte Checklisten zur Erstellung von Kostenplänen für Evaluationsprojekte vor. Dabei wird dann auch nach unterschiedlichen Finanzierungsmodellen differenziert (Stufflebeam & Shinkfield, 2007, S. 522 ff.).

In das Konzept eines Evaluationsforschungsprojektes sind auch Angaben zur *Projektorganisation* aufzunehmen. Dies betrifft die Zusammenstellung des Evaluationsteams, die Kompetenzen der Teammitglieder sowie besondere Rollen (z. B. Projektleitung, Teilprojektleiter) und Gremien (z. B. Stakeholder-Beirat).

Für das *Kommunikationsmanagement* ist zu klären, wann, wie oft und über welche Kommunikationskanäle mit welchen Zuständigkeiten sich die einzelnen Stakeholder im Verlauf der Evaluation untereinander und mit dem Evaluationsteam austauschen und aktiv am Evaluationsprozess beteiligen.

3.5 Formulierung des Evaluationsvertrags

Erhält ein Projektangebot den Zuschlag oder nimmt man einen Projektauftrag an, so folgt der Abschluss des *Evaluationsvertrags* (evaluation contract), der die Rechte und Pflichten von Auftraggeber und Auftragnehmer fixiert. Insbesondere

bei umfangreichen Projekten sind die rechtlichen Bedingungen genau zu prüfen, etwa die Zahlungsmodalitäten, Möglichkeiten des Vertragsrücktritts oder Konsequenzen bei Nichteinhaltung einzelner vertraglicher Regelungen (für eine Checkliste zur Vorbereitung von Projektverträgen siehe z. B. Kuster et al., 2006, S. 86 f.). Sofern sich während eines Evaluationsprojektes bedeutsame Änderungen im Evaluationsmandat ergeben, so sollten diese ebenfalls vertraglich festgehalten werden. Bei internen Evaluationen oder selbst finanzierten Evaluationskooperationen wird in der Regel kein Vertrag abgeschlossen. In diesen Fällen ist es empfehlenswert, eine *Evaluationsvereinbarung* (memorandum of understanding) abzuschließen, um Verbindlichkeit und Transparenz für alle Beteiligten sicherzustellen.

Aus Sicht der Evaluatoren ist es besonders wichtig zu prüfen, dass sie ihre vertraglich vereinbarten Leistungen im vorgesehenen Umfang termingerecht erbringen können. Umgekehrt müssen sie auch darauf achten, dass die Pflichten des Auftraggebers spezifiziert sind, dies betrifft etwa Zugang zu Informationen, Dokumenten und Erhebungsorten, Unterstützung bei der Rekrutierung von Befragungspersonen, ggf. Bereitstellung von Arbeitsplätzen im Feld etc. (für eine Checkliste zur Erstellung von Evaluationsverträgen und Evaluationsvereinbarungen vgl. Stufflebeam & Shinkfield, 2007, S. 547 f.).

4 *Planungsphase der Evaluation*

Nachdem Evaluationsvertrag oder -vereinbarung abgeschlossen wurde, stehen die Detailplanung und die konkrete Vorbereitung der Evaluationsstudie an. Dies umfasst die Formulierung von Evaluationsfragen bzw. -hypothesen (vgl. Abschnitt 4.1), die Untersuchungsplanung inklusive Untersuchungsdesign, Stichprobenplänen und Instrumenten (vgl. Abschnitt 4.2) sowie die Projektplanung (vgl. Abschnitt 4.3). Im Sinne der Nützlichkeit und Durchführbarkeit der Evaluationsstudie ist in der Planungsphase wiederum ein enger Kontakt zu den Stakeholdern notwendig, ebenso eine enge Rückkopplung mit den Abläufen im Praxisfeld. Andernfalls besteht die Gefahr, dass „am grünen Tisch“ ein Studienentwurf ausgearbeitet wird, der rein organisatorisch nicht umsetzbar ist und/oder am Widerstand derjenigen Stakeholder scheitert, die nicht von Anfang an genügend einbezogen wurden.

4.1 Formulierung der Evaluationsfragen und -hypothesen

Evaluationsfragen können sich – in Abhängigkeit von der Entwicklungsphase der evaluierten Intervention – auf alle Komponenten des CIPP-Modells von Stufflebeam (Stufflebeam & Shinkfield, 2007) beziehen und beispielsweise folgendermaßen formuliert sein:

Kasten 1:
Beispiele für Evaluationsfragen

Context

- Welche Gruppen haben Bedarf an einer geplanten (oder bereits realisierten) Intervention?
- Welche vergleichbaren Interventionen existieren bereits?
- Welche Bedürfnisse und Besonderheiten haben die Adressaten der Intervention?
- Wie lassen sich Mittel zur kurz- und langfristigen Finanzierung der Intervention beschaffen?

Input

- Wie gut entspricht das Konzept der Intervention den Bedürfnissen und Besonderheiten der Zielgruppen?
- Welche Alleinstellungsmerkmale hat das Konzept im Vergleich zu alternativen Interventionen?
- Wie kostenintensiv wird die Umsetzung des Konzeptes?
- Wie berücksichtigt das Konzept mögliche Widerstände gegen seine Umsetzung?

Process

- Welche Komponenten der Intervention werden in welchem Maße tatsächlich realisiert?
- In welchen Situationen/an welchen Orten wird die Intervention wie intensiv genutzt?
- Welche Probleme treten bei der Umsetzung der Intervention auf?
- Wie kann die Intervention verbessert werden?

Product

- Wie groß ist die Nutzergruppe der Intervention? In welchem Verhältnis steht die aktuelle Nutzergruppe zur Zielgruppe (*Impact*)? Spezieller: Ist die Intervention Menschen mit körperlichen Beeinträchtigungen zugänglich (*Barrierefreiheit*)? Ist die Intervention von Menschen mit unterschiedlichem kulturellem Hintergrund gleichermaßen nutzbar (*Internationalisierung*)?
- Erreicht die Intervention die intendierten Wirkungen? In welchem Umfang werden die intendierten Wirkungen erzielt (*Effektivität*)? Spezieller: Gibt es differenzielle Wirkungen in unterschiedlichen Gruppen? Profitieren Frauen und Männer gleichermaßen von der Intervention, oder wird eine Gruppe benachteiligt (*Gender Mainstreaming*)?
- In welchem Verhältnis stehen Kosten und Nutzen der Intervention? Erreicht die Intervention erwünschte Ziele kostengünstiger als andere Interventionen (*Effizienz*)?
- Welche Vorkehrungen werden getroffen, um die Intervention an anderen Orten/für andere Zielgruppen anzubieten (*Portabilität*)?
- Welche Vorkehrungen werden getroffen, um die Intervention langfristig oder dauerhaft zu erhalten? Welchen Status hat die Intervention z. B. zwei Jahre nach Abschluss des ersten Evaluationsprojektes (*Nachhaltigkeit*)?

Die Formulierung der Evaluationsfragen sollte auf die vom Auftraggeber gestellten Anforderungen zugeschnitten sein, wobei Evaluatoren jedoch im Sinne der Evaluationsstandards auch dafür Verantwortung tragen, dass vom Auftraggeber ggf. vernachlässigte Aspekte dennoch in die Studie aufgenommen werden (z. B. Kosten-Nutzen-Analyse), damit eine ausgewogene und fundierte Qualitätsbewertung möglich ist.

Ein Teil der Evaluationsfragen lässt sich auch als *gerichtete Hypothesen* fassen, da implizit von der Wirkung der Intervention ausgegangen wird (z. B. statt der offenen

Frage „Erreicht die Intervention die intendierten Wirkungen?“ die gerichtete Hypothese „Die intendierte Wirkung X wird mit der Intervention in stärkerem Maße erreicht als ohne Intervention bzw. als mit einer alternativen Intervention“).

In der Grundlagenforschung werden die untersuchten Fragestellungen und Hypothesen in der Regel aus einer umfassenden Theoriebildung abgeleitet. In der Evaluationsforschung werden Evaluationsfragen und -hypothesen dagegen oft ohne umfangreichen Theoriehintergrund eher pragmatisch formuliert. Es existieren jedoch auch Evaluationsansätze mit stärkerem Theoriebezug: Bei der sogenannten *theoriebasierten Evaluation* wird der Frage nachgegangen, *welche* Interventionskomponenten für *welche* Zielgruppe unter *welchen* Umständen *wie* funktionieren und *warum*. Anstelle der Output-Orientierung geht es um eine Prozess-Orientierung mit dem Anspruch, die Wirkprozesse theoretisch zu modellieren. Bei der theoriebasierten Evaluation wird also die Erkenntnisfunktion der Evaluation besonders stark betont und Evaluationsforschung sowohl als theoriebildend als auch theorieprüfend angelegt, wobei Handlungs- und Veränderungstheorien besonders einschlägig sind. Wissenschaftstheoretisch handelt es sich dabei um technologische Theorien (vgl. Westmeyer in diesem Band). Gemäß der These „Es gibt nichts Praktischeres als eine gute Theorie“ („nothing as practical as good theory“, Lewin, 1951, S. 196) soll theoriebasierte Evaluation ihren pragmatischen Nutzen darin entfalten, dass sie effektivere Veränderungsmöglichkeiten aufzeigt. Innerhalb des theoriebasierten Ansatzes sind unterschiedliche Modelle zu unterscheiden, etwa die theoriebasierte Evaluation im engeren Sinne (*theory-based evaluation*, Birckmayer & Weiss, 2000), die theoriegeleitete Evaluation (*theory-driven evaluation*, Chen, 1990) sowie die sogenannte realistische Evaluation (*realistic evaluation*, Pawson & Tilley, 1997). Im Rahmen der realistischen Evaluation werden Hypothesen in der Weise aufgestellt, dass damit kleine Kausalmodelle zu Einzelkomponenten der Intervention prüfbar sind, in denen die Verursachungsprozesse zwischen Kontext, Wirkmechanismus und Ergebnis modelliert sind (*CMO-Modelle*: Context, Mechanism, Outcome; vgl. Pawson & Tilley, 1997).

4.2 Untersuchungsplanung

Die Hauptelemente der Untersuchungsplanung sind das Untersuchungsdesign, der Stichprobenplan und die Auswahl oder Entwicklung der Instrumente.

Untersuchungsdesign

Zur Kennzeichnung eines Untersuchungsdesigns sind eine Reihe von Kriterien heranzuziehen. Anhand des Untersuchungsortes sind *Labor- und Feldstudien* zu differenzieren, wobei die meisten Evaluationsstudien Feldstudien im Praxisfeld

sind (z. B. in Unternehmen, in Bildungseinrichtungen, in Krankenhäusern etc.). Gemäß der Art der zu erhebenden Daten sowie der Datenauswertung wird zwischen einem *quantitativ-statistischen* und einem *qualitativ-interpretativen* Ansatz unterschieden, wobei in der Praxis eine Kombination von quantitativen und qualitativen Methoden gängig ist (*Methoden-Mix*, *Mehrmethoden-Ansatz*, *Methoden-Triangulation*).

Es ist zudem zwischen *Querschnittsstudien* (ohne Messwiederholung) und *Längsschnittstudien* (mit Messwiederholung) zu unterscheiden. In Evaluationsstudien sind Querschnittsstudien als *Treatmentgruppe-Kontrollgruppe/n-Vergleiche* einschlägig, etwa um herauszufinden, wie sich eine mit der Intervention behandelte Gruppe von einer nicht behandelten Gruppe und/oder einer alternativ (bzw. herkömmlich) behandelten Gruppe unterscheidet. Längsschnittstudien sind dagegen geeignet, um Veränderungen nachzuvollziehen. Sie sind als *Vorher-nachher-Vergleiche* in vielen Evaluationsstudien realisiert. Besonders aussagekräftig ist eine Kombination beider Ansätze, indem sowohl Treatment- als auch Kontrollgruppe/n vorher und nachher untersucht werden. Ein solches Design kann zusätzlich aufgewertet werden, indem mit zeitlichem Abstand noch ein dritter Messzeitpunkt einbezogen wird (Follow-up-Erhebung), der mittel- oder langfristige Wirkungen erfasst. Ein *Vortest-Nachtest-Follow-up-Vergleichsgruppen-Plan* ist somit im Sinne hoher Erklärungskraft besonders empfehlenswert (Hager, 2000, S. 183).

Während in *echten experimentellen Studien* die Untersuchungsteilnehmer per Zufall den Untersuchungs- und Kontrollbedingungen zugeordnet werden (Randomisierung), wird in *quasiexperimentellen Studien* mit vorgefundenen oder willkürlich eingeteilten Gruppen gearbeitet, wobei diese zumindest hinsichtlich zentraler sozialstatistischer sowie interventionsrelevanter Variablen vergleichbar sein sollten. Will man beispielsweise den Effekt des Notebook-Einsatzes in Schulen evaluieren (vgl. Schaumburg, 2001), so wird man typischerweise eine Klasse mit Notebooks ausstatten (Treatmentgruppe) und eine andere nicht (Kontrollgruppe), um dann Unterschiede zu erfassen. Hier müssen also die *vorgefundenen Klassenverbände* als Untersuchungsgruppen dienen, weil zum Zweck der Evaluation kaum eine im Sinne des Erkenntnisgewinns optimale zufallsgesteuerte Neubildung von Schulklassen praktikabel wäre. Ein weiteres Evaluationsdesign zur Überprüfung der Wirksamkeit einer Intervention ist der *Ist-Soll-Vergleich*, dabei wird die behandelte Gruppe anhand vorher festgelegter Soll-Werte eingeschätzt.

Um die Wirksamkeit einer Intervention überzeugend nachweisen zu können (Kriterium der *internen Validität*), ist es wichtig, dass die im Vorher-nachher-, Treatment-Kontrollgruppen- oder Ist-Soll-Vergleich gefundenen Effekte (Ausprägungen der *abhängigen Variablen*) mit hoher Sicherheit auf die Intervention (*unabhängige Variable*) zurückzuführen sind. Die interne Validität ist gering,

wenn die Effekte durch andere Faktoren wie z. B. Umweltbedingungen oder bereits vorliegende Unterschiede zwischen den Untersuchungsgruppen erklärbar sind. Da Evaluationsstudien oft als quasiexperimentelle Feldstudien realisiert werden, lassen sich im Unterschied zu experimentellen Laborstudien weder die personengebundenen noch die umweltbezogenen Störvariablen exakt kontrollieren. Dementsprechend müssen gute Evaluationsstudien diesen Problemen schon bei der Designentwicklung mit entsprechenden Methoden entgegenwirken (z. B. Berücksichtigung von Kontrollvariablen wie z. B. Computererfahrung, Schulleistung, Lehrermotivation etc. im Notebook-Klassen-Beispiel; vgl. Cook & Campbell, 1979). Außerdem müssen die Erkenntnisgrenzen im Ergebnisbericht klar benannt werden, um ungesicherten Pauschal- oder Kausalaussagen entgegenzuwirken.

Ein wichtiges Evaluationsziel besteht häufig darin, nicht nur die intendierten Wirkungen der Intervention nachzuweisen, sondern auch nicht intendierte Nebenwirkungen. Dementsprechend breit muss das Spektrum der erfassten abhängigen Variablen sein. Insbesondere beim Einsatz standardisierter Methoden mit geschlossenen Fragen besteht die Gefahr, dass nicht antizipierte Nebenwirkungen übersehen werden.

Das Evaluationsdesign sollte mit allen relevanten Stakeholdern abgestimmt werden (z. B. im Rahmen einer Sitzung des Stakeholder-Beirates). Zudem sollte antizipiert werden, wie das Design ggf. zu modifizieren ist, falls sich auf Seiten der Intervention Veränderungen ergeben. Ein gutes Evaluationsdesign sollte sich nicht nur durch hohe interne Validität, sondern auch durch *Robustheit* auszeichnen, d. h. so ausgelegt sein, dass selbst unter suboptimalen Datenerhebungsbedingungen (z. B. wegen unvorhergesehener Veränderungen der Feldbedingungen) noch nützliche Resultate erzielt werden können.

Stichprobenplan

Die Datenerhebung im Rahmen einer Evaluationsstudie kann sich auf unterschiedliche Merkmalsträger beziehen (z. B. Personen, Situationen, Objekte, Ereignisse etc.). Personen sind bei psychologischen Interventionen die zentralen Merkmalsträger. Die Gesamtheit aller Personen, an die sich eine Intervention richtet, ist die Interventionspopulation. In der Regel wird eine Intervention nicht von der gesamten Zielgruppe, sondern nur von einer Teilgruppe genutzt, die als *Interventionsstichprobe* bezeichnet wird (vgl. Bortz & Döring, 2006, S. 128). Die Höhe der *Ausschöpfungsquote* (Größe der Interventionsstichprobe relativiert am Umfang der Zielgruppe) kann dabei ein Qualitätskriterium der Maßnahme sein. Ein weiterer Erfolgsindikator einer Maßnahme ist die *Ausschöpfungsqualität* (AQ) mit einem Idealwert von 100 (Rossi & Freeman, 1993):

$$AQ = 100 \cdot \left[\frac{\text{Anzahl erreichter Zielobjekte}}{\text{Anzahl aller Zielobjekte}} - \frac{\text{Anzahl unbefugter Teilnehmer}}{\text{Anzahl aller Teilnehmer}} \right]$$

Je größer die Anzahl der „unbefugten“ Teilnehmer, umso geringer die Ausschöpfungsqualität. „Unbefugte“ Teilnehmer können Personen sein, die eine Maßnahme nutzen oder nutzen müssen, ohne sie wirklich zu benötigen (z. B. an einer unternehmensinternen Trainingsmaßnahme nehmen Mitarbeiter teil, die die Inhalte bereits beherrschen oder die die Inhalte in ihrem Tätigkeitsfeld gar nicht benötigen).

Selten wird man die regulären Maßnahmenteilnehmer vollständig in die Evaluation einbeziehen (Vollerhebung), sondern sich auf Auswahlen (Stichproben) beschränken.

Für die Zusammenstellung von *Evaluationsstichproben* ist es unerlässlich, Personen aus der Interventionsstichprobe zu rekrutieren. Eine Evaluationsstichprobe lässt sich nur in Ausnahmefällen als *echte Zufallsstichprobe* aus der Grundgesamtheit aller interessierenden Objekte realisieren, stattdessen sind nicht probabilistische bzw. *willkürliche Stichproben* gängig (z. B. Ad-hoc-Stichprobe bei einer Veranstaltung, Teilnehmerrekrutierung im Schneeballprinzip, Zusammenstellung einer Quotenstichprobe von Freiwilligen). Willkürliche Stichproben sind leichter realisierbar, haben aber den Nachteil, dass aufgrund von Verzerrungen ihre statistische Repräsentativität fraglich bzw. unklar ist. Bei der Planung von Evaluationsstichproben ist jeweils zu berücksichtigen, dass die Zielgruppe der Intervention möglichst ohne systematische Verzerrungen (z. B. Ausschluss ganzer Stakeholder-Gruppen, Vernachlässigung von Wenig-Nutzern etc.) erfasst wird.

Um Nebenwirkungen der Intervention zu erfassen, kann es sinnvoll sein, auch die Gruppe der unbefugten Teilnehmer genauer zu beleuchten. Zudem werden für Evaluationsstichproben auch Mitglieder der Interventionszielgruppe benötigt, die nicht (non-user) oder nicht mehr (drop out) an der Maßnahme teilnehmen, da diese Gruppen Hinweise auf Zugangsbarrieren oder Nutzungsprobleme liefern können. Mitglieder der Zielgruppe, die nicht an der Maßnahme teilnehmen, werden zudem für Kontrollgruppendesigns benötigt. Abgesehen von den Maßnahmenteilnehmern sind insbesondere für Evaluationen in frühen Phasen der Interventionsentwicklung auch diejenigen Personen relevante Informanten, die an der Konzeption und Umsetzung der Intervention mitwirken (z. B. Lehrer, Trainer, Psychotherapeuten).

Für aussagekräftige inferenzstatistische Analysen ist jeweils von minimalen bzw. *optimalen Stichprobengrößen* auszugehen, die umso größer angelegt sein müssen, je kleiner der erwartete Effekt ist (vgl. Bortz, 2004). Um etwa die Wirksamkeit

einer auf Jugendliche zugeschnittenen Primärpräventionsmaßnahme im Bereich illegaler Drogen statistisch nachzuweisen, wären – aufgrund der geringen Verbreitung des Konsums illegaler Drogen – Stichprobengrößen im sechsstelligen Bereich notwendig (vgl. Uhl, 1997).

In qualitativen Evaluationsstudien werden verhältnismäßig wenige Informanten sehr ausführlich untersucht (z. B. durch Feldbeobachtung oder Leitfadeninterview). Anstelle von zufälligen oder zufallsähnlichen Auswahlverfahren sind hier *theoretische Stichprobenkonstruktionen* einschlägig, bei denen bewusst besonders interessante oder kontrastierende Fälle herausgesucht werden.

Instrumente

Die für die Evaluationsfragen und -hypothesen relevanten Variablen (Merkmale der Intervention und der Interventionsumsetzung, Kosten und Nutzen der Maßnahme, positive und negative Nebenwirkungen etc.) gilt es durch entsprechende Operationalisierungen messbar zu machen. Als Datenerhebungsmethoden kommen in der Evaluationsforschung sämtliche *sozialwissenschaftlichen Methoden* zum Einsatz (vgl. Bortz & Döring, 2006), das heißt vor allem Beobachtung, mündliche und schriftliche Befragung, Testung, Dokumentenanalyse sowie die Analyse von Verhaltensspuren (nonreaktive Methoden). Für diese Datenerhebungsmethoden sind jeweils konkrete *Instrumente* notwendig, seien es Beobachtungssysteme, Interviewleitfäden, Fragebögen, Tests, Checklisten oder Kategoriensysteme.

Sofern möglich sollten bereits erprobte und hinsichtlich ihrer wissenschaftlichen Gütemerkmale (Objektivität, Reliabilität, Validität) geprüfte Datenerhebungsinstrumente als Evaluationsinstrumente *ausgewählt* werden. Liegen diese nicht vor, so ist eine *Modifikation* vorhandener Instrumente angeraten. Die *Neukonstruktion* von Instrumenten erlaubt eine optimale Anpassung an die Evaluationsfragen, ist jedoch recht aufwendig – insbesondere, da die Qualität der neu entwickelten Instrumente erst im Rahmen von Pretests ermittelt und gesichert werden muss (zu den *Standards von psychologischen und pädagogischen Testverfahren* siehe American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999). Abgesehen von der Forschungsökonomie ist eine Verwendung etablierter Instrumente oft auch deswegen zu präferieren, weil Evaluationsbefunde dann mit bereits publizierten Ergebnissen vergleichbar sind. Damit die Datenerhebung im Feld nicht störend wirkt, ist auf eine *ökonomische Abwicklung* besonderen Wert zu legen (z. B. durch kurze Instrumente und komfortable Bearbeitung).

Neben den allgemeinen sozialwissenschaftlichen Datenerhebungsmethoden kommen in Evaluationsstudien teilweise auch spezifische Methoden und Instrumente

zum Einsatz, etwa Methoden der *Kosten-Nutzen-Rechnung* zur Effizienzprüfung (vgl. Holling in diesem Band) oder das *Benchmarking-Verfahren* für einen systematischen Vergleich der Merkmale verschiedener Interventionen und der Identifikation von Best Practice-Beispielen, an denen man sich orientieren kann (für den Einsatz von Benchmarking im Rahmen der Evaluation von Hochschulen vgl. Ziegele & Hener, 2004).

4.3 Projektplanung

Während bei der Untersuchungsplanung der angestrebte wissenschaftliche Erkenntnisgewinn bzw. die Evaluationsziele im Vordergrund stehen, konzentriert sich die Projektplanung auf die operative Umsetzung und vor allem auf die Zeit-, Kosten- und Arbeitsplanung. Die Grobpläne aus der Konzeptionsphase werden verfeinert und je nach Projektbedarf durch zusätzliche Pläne ergänzt (z. B. Qualifizierungsplan, Berichtswesenplan, vgl. Kessler & Winkelhofer, 2004).

5 Realisierungsphase der Evaluation

Auf der Basis der Ergebnisse der Planungsphase werden in der Realisierungsphase die für die Evaluationsziele notwendigen Evaluationsdaten erhoben (vgl. Abschnitt 5.1), analysiert (vgl. Abschnitt 5.2) und interpretiert (vgl. Abschnitt 5.3), zudem wird der operative Projektablauf durch ein Projektcontrolling begleitet (vgl. Abschnitt 5.4). Je nach Umfang und Dauer einer Intervention sowie der aktuellen Phase im Interventionslebenszyklus werden Erhebungs-, Analyse- und Interpretationsschritte mehrfach nacheinander oder auch parallel stattfinden, wobei es im Sinne eines iterativen Vorgehens oft Rücksprünge in die Planungsphase gibt.

Aus dem Satz der Evaluationsstandards sind bei der Datenerhebung vor allem Standards der Durchführbarkeit und Fairness relevant, während bei der Datenauswertung und Dateninterpretation der Schwerpunkt auf Standards der Genauigkeit liegt (vgl. DeGEval, 2002b, S. 39).

5.1 Erhebung von Evaluationsdaten

Im Vergleich zur Grundlagenforschung handelt es sich bei der Evaluationsforschung deutlich häufiger um Feldforschung, sodass *Störungen der Praxis* im Zuge der Datenerhebung zu vermeiden sind (z. B. durch ökonomische Instrumente, günstige Auswahl von Erhebungszeitpunkten, nonreaktive Datenerhebung etc.). Zudem sind Fragen des *Datenschutzes* besonders zu beachten, weil im Feld oftmals

die Anonymität nicht voll gewährleistet werden kann und zudem Beteiligte vor möglichen negativen Konsequenzen geschützt werden sollen. Nicht zuletzt sind die bereichsspezifischen *sozialen Normen* und *rechtlichen Regelungen* (z. B. Arbeitsrecht) beim Agieren des Evaluationsteams im Feld zu berücksichtigen. Die geringe Kontrolle der Feldbedingungen gegenüber Laborbedingungen erfordert es auch, dass Evaluationsforscher darauf vorbereitet sind, ihr Design und ihre Instrumente bei Bedarf kurzfristig zu modifizieren, um veränderten Praxisbedingungen gerecht zu werden. Entsprechende Entscheidungen sind dabei mit dem Auftraggeber und anderen Stakeholdern abzustimmen.

Während bei der summativen Evaluation typischerweise die Datenerhebung konzentriert erfolgt (z. B. am Beginn, in der Mitte und am Ende einer Intervention), genau vorgeplant und dementsprechend dokumentiert wird, besteht bei formativer Evaluation mit fortlaufender und oft auch explorativer Datenerhebung die Gefahr, dass die Informationssammlung ungenügend dokumentiert wird. Insbesondere bei formativen Evaluationen stoßen Phasenmodelle insofern an ihre Grenzen, als strikte Vorplanungen es verhindern oder erschweren können, sich flexibel auf neu entstehende Fragen einzustellen. Wichtig ist zu beachten, dass sich die Datensammlung nicht nur auf Wirkungen der Intervention bezieht, sondern auch auf die Konzeption und Umsetzung der Intervention. Zur Unterstützung der Erhebung von Evaluationsdaten liegen Handreichungen und Checklisten vor, die verschiedene Datenquellen systematisieren (vgl. Stufflebeam & Shinkfield, 2007, S. 569 ff.).

5.2 Analyse von Evaluationsdaten

Je nach Art der erhobenen quantitativen und/oder qualitativen Daten sowie der interessierenden Evaluationsfragen sind jeweils *adäquate quantitative und/oder qualitative Datenauswertungsmethoden* zu wählen und methodologisch korrekt anzuwenden (für ein breites Spektrum von entsprechenden Methoden vgl. Holling, in Vorbereitung). Im Falle einer formativen Evaluation sind häufige Zwischenauswertungen zu erstellen.

Evaluationsspezifische Besonderheiten im Vergleich zur Grundlagenforschung ergeben sich im Zusammenhang mit einer sachgerechten Datenanalyse anhand der formulierten Fragestellungen und Hypothesen kaum (Stufflebeam & Shinkfield, 2007). Besonders wichtig ist es allerdings, den *praktischen Nutzwert* im Hinblick auf die Evaluationsziele bei der Datenanalyse immer in den Vordergrund zu rücken. Für umfassende und komplexe inferenzstatistische Datenanalysen werden sich Auftraggeber und andere Evaluationsnutzer allenfalls dann interessieren, wenn diese Antworten auf die drängenden Evaluationsfragen liefern. Um den Nutzwert der Datenanalyse zu steigern empfiehlt es sich, Zwi-

schenergebnisse mit den Evaluationsnutzern zu diskutieren und von ihnen ggf. noch Anregungen für zusätzliche Datenanalysen entgegenzunehmen.

Angesichts der Ausgangsfragen und der Datenlage ist bei Evaluationsstudien jeweils kritisch zu überdenken, ob überhaupt *inferenzstatistische Methoden* indiziert sind. Dies ist nur dann der Fall, wenn a priori theoriebasierte Hypothesen formuliert wurden, während zur Beantwortung von vielen Evaluationsfragen deskriptive Befunde ausreichen. Ebenso erübrigen sich Signifikanztests, wenn einerseits alle Teilnehmer einer Maßnahme untersucht wurden (Vollerhebung) oder wenn selektiv nur sehr wenige Untersuchungsteilnehmer gewonnen werden konnten und somit die statistischen Voraussetzungen für entsprechende Analysen nicht erfüllt sind.

Zu bedenken ist auch, dass sich in vielen Evaluationsstudien vermutlich nur schwache Effekte nachweisen lassen, die bei geringer Stichprobengröße gar nicht signifikant werden können. Um hier zu gesicherten Befunden zu gelangen, empfiehlt es sich, anstelle der statistischen Signifikanz z. B. die *Musterläufigkeit* der Ergebnisse zu beachten: Angenommen, eine psychologische Online-Beratung wird in einem Vorher-nachher-Design überprüft, und es werden 15 Variablen einbezogen, wobei sich keine einzige Differenz als statistisch signifikant erweist. Die Schlussfolgerung, dass die Maßnahme wirkungslos war, könnte voreilig sein, wenn die Evaluationsstichprobe z. B. $n = 300$ Personen umfasst. Denn ein schwacher Effekt hätte für einen Signifikanznachweis die drei- oder vierfache Zahl an Probanden erfordert. Betrachtet man jedoch das Ergebnismuster und stellt fest, dass sich deskriptiv bei 13 von 15 Variablen Differenzen im Sinne einer Verbesserung gezeigt haben, so könnte dies eine ermutigendere Bilanz rechtfertigen. Erst recht, wenn diese quantitativen Befunde mit anderen Befunden übereinstimmen (z. B. Ergebnisse von Interviews mit Beratern und Ratsuchenden).

5.3 Interpretation von Evaluationsdaten

Bei der formativen Evaluation zielt die Interpretation der Zwischenergebnisse darauf ab, konkrete Empfehlungen für die Veränderung der Maßnahme oder ihrer Umsetzung abzuleiten. Bei der summativen Evaluation entspricht die Interpretation des empirischen Gesamtergebnisses der abschließenden Bewertung der Maßnahme hinsichtlich ihrer Nützlichkeit.

Während man in der Grundlagenforschung eine Studie durchaus mit divergierenden Befunden und offenen weiteren Forschungsfragen beenden kann, erfüllt eine Evaluationsstudie nur dann ihren Zweck, wenn sie die Fülle der Daten letztlich zu einem konsistenten Gesamtergebnis zusammenfasst. Auf dem Weg

zu diesem Ergebnis mögen zunächst divergierende Befunde offengelegt, reflektiert und mit den Stakeholdern diskutiert werden – aber letztlich mit dem Ziel einer wissenschaftlich abgesicherten Synthese durch die Evaluationsexperten. Da es bei der Evaluation um Bewertung geht, sollten Evaluatoreninnen und Evaluatoren am Ende auch einer solchen bewertenden Gesamtinterpretation nicht aus dem Weg gehen, so der auffallend vehemente Appell von Stufflebeam und Shinkfield (2007, S. 588):

„Evaluators who seek and present firm conclusions often face opposition and criticism and sometimes are wrong. However, if they ground their conclusions in systematic analysis and synthesis of a sufficient set of appropriate evidence and report appropriate caveats, we think they will be correct far more often than those who fail to practice systematic, conclusion-oriented evaluation and will be instrumental in helping their audiences make sound decisions and improve programs. Systematic evaluation is not and never will be an exact science, but it is an invaluable guide to progress. Evaluators who are steadfastly afraid of being wrong and consequently equivocate or exercise undue caution probably should seek other worthwhile work, such as bookkeeping, proofreading, window washing, watch repair, or house painting.“

5.4 Projektcontrolling

Während der gesamten Durchführung des Evaluationsprojektes, vor allem aber während der Realisierungsphase, muss die Projektabwicklung im Hinblick auf die Projektziele und Projektpläne fortlaufend und im Detail kontrolliert und gesteuert werden. Auf diese Weise sollen Probleme bei der Projektabwicklung frühzeitig erkannt und gelöst werden. Zudem sichert das Projektcontrolling die Transparenz gegenüber dem Auftraggeber und anderen Stakeholdern: Die Projektleitung kann zu jedem Projektzeitpunkt auf der Basis systematischer Controlling-Methoden (Soll-ist-Vergleiche, Kennzahlen) beispielsweise den Arbeitsstand oder Kostenstand berichten (vgl. Kuster et al., 2006, S. 143 ff.). Der Einsatz von Projektcontrolling-Methoden muss in Abhängigkeit vom jeweiligen Projekt skaliert werden und darf nicht zu einer unnötigen Bürokratisierung der Abläufe führen.

6 Abschlussphase der Evaluation

Die in der Realisierungsphase der Evaluation gewonnenen Ergebnisse gilt es in der Abschlussphase an unterschiedliche Stakeholder der Intervention zu *kommunizieren* (vgl. Abschnitt 6.1) und für Entscheidungen in der *Praxis zu nutzen*.

(vgl. Abschnitt 6.2). Auch eine wissenschaftliche Bewertung der Evaluation in Form von systematischer *Metaevaluation* kann sinnvoll oder notwendig sein (vgl. Abschnitt 6.3). Schließlich ist es für die Wissenschaftlichkeit der Evaluationsforschung wünschenswert, Evaluationsstudien auch zu publizieren (vgl. Abschnitt 6.4).

6.1 Abgabe der Evaluationsberichte

Bei einer *formativen Evaluation* erfolgt typischerweise eine fortlaufende, eher *informelle Rückkopplung* von Evaluationsergebnissen an den Auftraggeber bzw. die Interventionsentwickler etwa in Meetings, Telefonaten, E-Mails, Newslettern usw. In der Abschlussphase des Evaluationsprojektes wird üblicherweise erwartet, dass der Ertrag der formativen Evaluation in einem *Endbericht* dokumentiert wird, zum einen zur Ergebnissicherung, zum anderen als Beleg des Arbeitsaufwands sowie als Basis für eine mögliche *Metaevaluation*. Bei einer *summativen Evaluation* erfolgt die Kommunikation der Ergebnisse ohnehin stärker gebündelt und formalisiert in Form von *Präsentationen und Berichten*, wobei Zwischenberichte in einen abschließenden Endbericht integriert werden.

Die Darstellung von Evaluationsergebnissen sollte ausgewählten *Evaluationsstandards* der Nützlichkeit (z. B. Auswahl und Umfang der Informationen), der Fairness (z. B. Schutz individueller Rechte wie z. B. Datenschutz, Vertraulichkeit) und Genauigkeit (z. B. Angabe von Informationsquellen, begründete Schlussfolgerungen) genügen (vgl. DeGEval, 2002b, S. 39 f.). Die Kommunikation von Evaluationsbefunden muss zudem – gerade mit Blick auf die Nützlichkeit – *zielgruppenorientiert und verständlich* erfolgen, das heißt beispielsweise gut strukturiert und ohne überflüssigen Wissenschaftsjargon. Ablaufdiagramme, Tabellen, Grafiken, Fotografien und andere Varianten der Veranschaulichung sind empfehlenswert. Für alle Fragen und Hypothesen der Evaluation sollten eindeutige und wissenschaftlich fundierte Antworten formuliert werden.

Sofern dies im Evaluationsvertrag vorgesehen ist, muss der Evaluationsbericht dem Auftraggeber auch konkrete *Entscheidungs- und Handlungsempfehlungen* an die Hand geben. Dabei sind mögliche Fehl- und Überinterpretationen der Befunde durch Evaluationsnutzer möglichst explizit zu thematisieren. Schließlich muss ein umfangreicher Evaluationsbericht die zentralen Inhalte stets in einer *aussagekräftigen Zusammenfassung* (executive summary) bündeln, die über Ausgangslage, Ziel, Methodologie, wichtigste Ergebnisse, Schlussfolgerungen und Empfehlungen informiert. Bei Bedarf enthält der Evaluationsbericht auch ein *Glossar* mit Abkürzungen und Fachbegriffen sowie einen *Materialanhang* mit einschlägigen Dokumenten (z. B. Evaluationsauftrag, Fragebögen, Interviewleitfäden etc.).

6.2 Organisation der Evaluationsnutzung

Der Evaluationsbericht ist eine wichtige Basis für die praktische Umsetzung und Nutzung der Befunde, allerdings ist er allein in der Regel nicht ausreichend. Vielmehr müssen aus dem Bericht die zentralen Botschaften für unterschiedliche Zielgruppen extrahiert und in geeigneter Form (z. B. Merkblatt, Checkliste, Workshop, Seminar etc.) zum passenden Zeitpunkt an die jeweiligen Adressaten weitergegeben werden. Es kann sinnvoll sein, dass Auftraggeber und Evaluationsteam zusammenarbeiten, wenn es um die Evaluationsnutzung geht. Die folgende Checkliste fasst für die Evaluationsnutzung zentrale Fragen zusammen (Bundesamt für Gesundheit der Schweiz, 1997, S. 61, Checkliste 5.1):

- Haben wir die für unsere Planung und Entwicklung wichtigsten Aussagen der Studie identifiziert?
- Ist klar, was aus der Studie für die Umsetzung von Projekten folgt?
- Haben wir mögliche Folgen für die Aktivitäten verschiedener beteiligter Gruppen abgeschätzt?
- Haben wir überlegt, was nun zu tun ist?
- Ist klar, welche Rahmenbedingungen dafür geschaffen werden müssen?
- Wissen wir, welcher Aufwand mit der Durchführung solcher Folgearbeiten verbunden ist (personelle und finanzielle Ressourcen)?
- Haben wir die Personen oder Gruppen identifiziert, die am ehesten geeignet sind, die nächsten Schritte zu unternehmen?
- Haben wir überlegt, welche Personen/Gruppen Priorität haben? (Praktikabilität versus Ideal!)
- Ist uns klar, wie wir am besten an diese Gruppe(n) herantreten?
- Wer ist am ehesten geeignet, an diese Gruppe(n) heranzutreten?

6.3 Durchführung einer Metaevaluation

Mit *Metaevaluation* ist die Evaluation der Evaluation gemeint, es geht also um eine Qualitätsbewertung einer einzelnen Evaluationsstudie, nicht um eine Metaanalyse. Ebenso wie jede Grundlagenforschung eine kritische Reflexion von Methodenproblemen enthalten und die Aussagekraft ihrer Befunde kritisch diskutieren soll, ist es im Sinne der Qualitätssicherung und Transparenz wünschenswert, dass in Evaluationsprojekten reflektiert wird, welche Evaluationsstandards in welcher Weise umgesetzt werden konnten und wo und warum ggf. Probleme aufgetreten sind. Diese systematische Selbstbewertung (*interne Metaevaluation*) kann für das Evaluationsteam Katalysator von Verständigungs-, Professionalisierungs- und Lernprozessen sein und zudem als Qualitätsmerkmal in der Außenkommunikation dienen. Eine Metaevaluation kann am Ende eines Evaluationsprojektes summativ erfolgen oder auch formativ die Evaluations-

forschung von Anfang an begleiten (Stufflebeam, 2001; Stufflebeam & Shinkfield, 2007).

Nach Abschluss einer Evaluation kann es zwischen Auftraggeber, den für die Intervention Verantwortlichen und dem Evaluationsteam manchmal zu Auseinandersetzungen über die Qualität der Evaluationsstudie kommen. Auftraggeber könnten beispielsweise zu der Ansicht gelangen, die Studie sei ihr Geld nicht wert gewesen oder Interventionsentwickler könnten sich in ihrer Arbeit unfair behandelt fühlen. Schließlich können Urheberrechts- oder Datenschutzfragen virulent werden. Ist informell keine Konfliktlösung zu erzielen, so kann es wünschenswert sein, die Evaluationsstudie von externen Evaluationsexperten evaluieren und begutachten zu lassen, wobei typischerweise die gängigen Evaluationsstandards heranzuziehen und anhand von empirischen Daten (z. B. Dokumentenanalyse, Befragung usw.) zu prüfen sind. Zur Anfertigung von Gutachten im Rahmen *externer Metaevaluationen* liegen von der Deutschen Gesellschaft für Evaluation einschlägige Empfehlungen vor (DeGEval, 2002a).

Schließlich kann externe Metaevaluation auch eingesetzt werden, um die – z. B. in der Routinephase von Interventionen etablierte – *Selbstevaluation* der Beteiligten (z. B. Lehr- und Forschungsevaluation) periodisch oder inzidenziell zu prüfen.

6.4 Publikation der Evaluationsstudie

Deutlich seltener als in der Grundlagenforschung werden die Ergebnisse der Evaluationsforschung der wissenschaftlichen Fachöffentlichkeit vorgestellt. Dies liegt unter anderem an der Vertraulichkeit der gewonnenen Daten, an der Projektspezifität von Erkenntnissen (speziell bei formativer Evaluation), am Imageschutz für Auftraggeber bzw. Interventionsverantwortliche (speziell im Fall ungünstiger Evaluationsbefunde) sowie nicht selten auch am Zeitmangel der Evaluatoren, für die die praktische Evaluationsarbeit in der Regel eine höhere Priorität besitzt als das akademische Publizieren. Doch Evaluationsforschung könnte sich nicht als wissenschaftliches Feld behaupten, wenn eine kritische Auseinandersetzung in der Scientific Community und auch in der breiten Öffentlichkeit fehlen würde. Dafür ist die Publikation von Evaluationen sowie von Metaevaluationen notwendig.

Die Publikation von Evaluationsforschung auf *Tagungen* der Fachgesellschaften sowie auf *fachbezogenen Online-Portalen*, in *Fachbüchern* und insbesondere in *Fachzeitschriften* dient der akademischen Erkenntnisweitergabe und im Rahmen des Peer-Review-Verfahrens auch der wissenschaftlichen Qualitätssicherung. So können aus Evaluationsforschungsprojekten theoretische, methodische und

inhaltliche Erkenntnisse gewonnen werden, die dazu dienen, das Feld der Evaluationsforschung weiterzuentwickeln.

Sobald zu einer bestimmten Intervention eine kritische Menge von publizierten Evaluationsstudien vorliegt, sind *Metaanalysen* im Sinne einer systematischen Zusammenfassung der Befunde über das Evaluationsobjekt möglich. So legten beispielsweise Grawe, Donati und Bernauer (2001) eine sehr umfassende und seit der Erstveröffentlichung im Jahr 2001 sehr kontrovers diskutierte Metaanalyse zur Wirksamkeit der verschiedenen Psychotherapiemethoden vor. Während eine Metaanalyse von Evaluationsstudien zusammenfassend Auskunft über die Qualität von Interventionen geben soll, ist es das Ziel von Metaevaluationen die Qualität der Evaluationsforschung zu bewerten. Leider werden irreführenderweise auch Metaanalysen über Evaluationsstudien teilweise als Metaevaluationen bezeichnet (z. B. benennen Ashworth, Cebulla, Greenberg & Walker, 2004, ihre Metaanalyse von Evaluationsstudien über die Wirksamkeit von Arbeitslosenprogrammen in den USA als „Metaevaluation“).

In dem Maße, in dem sich Evaluation und Evaluationsforschung in nahezu allen Praxisfeldern etablieren, wird die Frage nach Qualität und Nutzen von Evaluationsstätigkeit – jenseits der Bewertung von einzelnen Evaluationsprojekten – zu einer wissenschaftlich und gesellschaftlich vordringlichen Frage. So wird beispielsweise die vom Hochschulrahmengesetz geforderte Hochschulevaluation in ihren Möglichkeiten, Grenzen und Konsequenzen äußerst kontrovers diskutiert. Kritiker befürchten beispielsweise, dass übermäßige Evaluationsforderungen und -aktivitäten („Evalutitis“) zu einer wachsenden Bürokratisierung führen, dass gängige quantitative Evaluationskriterien (z. B. Anzahl von Publikationen) keine validen Indikatoren für Forschungsqualität seien usw. (Berlin-Brandenburgische Akademie der Wissenschaften, 2002; Döring, 2005). Neben der kritischen Reflexion der Evaluationspraxis in einzelnen Evaluationsbereichen lässt sich auch grundsätzlicher fragen, inwiefern Evaluationsforschung als wissenschaftliche Tätigkeit überhaupt mit den Ansprüchen von Auftraggebern (z. B. politischen Entscheidungsträgern) kompatibel ist (vgl. Kromrey, 2003).

Neben *theoretischen Metadiskursen* über Evaluationsforschung liegen inzwischen eine Reihe *empirischer Metaevaluationsstudien* vor, die beispielsweise mit Methoden der Dokumentenanalyse oder mit Experten- und Teilnehmerbefragungen operieren (für eine publizierte Metaevaluation von 56 Evaluationsstudien im Bildungsbereich in der Schweiz siehe Stamm, 2003; für eine Metaevaluation von 14 Evaluationsstudien im HIV-/AIDS-Bereich in Afrika siehe Patel, 2002; für eine Metaevaluation von 11 Evaluationsstudien im Gender-Mainstreaming-Bereich in Spanien siehe Bustelo, 2003). Derartige Metaevaluationen befassen sich beispielsweise mit der Frage, wie nützlich die gewonnenen Evaluationsbefunde sind und in welchem Maße sie erfolgreich in die Praxis umgesetzt werden kann-

ten. Metaevaluationen dienen auch dazu, die Qualität von unterschiedlichen Evaluationsformen oder Evaluationsmethoden zu bewerten (für eine Evaluation der Methode der Kurzevaluation siehe Widmer, Rothmayr & Serdült, 1996).

7 Schlussfolgerung

Evaluationsforschung wird in Form von Projekten abgewickelt. Jedes Evaluationsforschungsprojekt durchläuft in der Praxis verschiedene Arbeitsphasen, mit jeweils spezifischen Ergebnissen, die in die nächste Phase eingehen. Phasenmodelle für Evaluationsforschungsprojekte sind bislang kaum theoretisch elaboriert und empirisch geprüft worden. Aus der Logik der Projektabwicklung sowie des empirischen Forschungsprozesses wurde ein heuristisches Phasenmodell abgeleitet, das aus fünf Projektphasen besteht, in denen jeweils spezifische Arbeitsschritte stattfinden. Dieses Modell geht insofern über die gängigen Modelle des empirischen Forschungsprozesses hinaus, als Aspekte des Projektmanagements sowie evaluationsspezifische Aufgaben integriert sind. Da Evaluationsforschungsprojekte stets mit Interventionsprojekten gekoppelt sind, ist bei der Weiterentwicklung von Phasenmodellen insbesondere die Abstimmung zwischen Interventions- und Evaluationsphasen genauer auszuarbeiten. Auf einer allgemeinen Ebene sind Phasenmodelle des Forschungsprozesses, der Interventionsentwicklung und der Projektabwicklung jeweils auch als Phasen in Problemlösungsprozessen zu interpretieren. Denn auf dem Weg vom unerwünschten Anfangszustand zum erwünschten Endzustand sind in der Regel immer wieder Barrieren und Hindernisse zu überwinden.

Literatur

- American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME) (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Ashworth, K., Cebulla, A., Greenberg, D. & Walker, R. (2004). Meta-Evaluation: Discovering what works best in welfare provision. *Evaluation*, 10 (2), 193–216.
- Badiru, A. B. (1996). *Project Management for Research: A guide for engineering and science*. New York, NY: Chapman & Hall.
- Baumann, D., Escher, O. P. & Witschi, U. (2005). Projektmanagement in der Forschung. *Projektmagazin.de – Das Fachmagazin im Internet für erfolgreiches Projektmanagement* 12/2005: Zugriff am 5.7.2006, <http://www.spm.ch/Dokumente/Projektmanagement%20in%20der%20Forschung.pdf>
- Berlin-Brandenburgische Akademie der Wissenschaften (Hrsg.). (2002). *Gegenworte. Hefte für den Disput über Wissen*. 5. Heft Frühjahr 2000: Gütesiegel für die Wissenschaft? Zur Diskussion über Qualität, Evaluierung und Standards. Berlin: Hrsg.

- Birckmayer, J. D. & Weiss, C. H. (2000). Theory-based evaluation in practice. *Evaluation Review*, 24 (4), 407–431.
- Bode, P. (2000). Quality and project management for scientific research in INAA. *Journal of Radioanalytical and Nuclear Chemistry*, 245 (1), 133–135.
- Bortz, J. (2004). *Statistik für Human- und Sozialwissenschaftler* (6. Aufl.). Heidelberg: Springer.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation* (4. Aufl.). Berlin: Springer.
- Bundesamt für Gesundheit (BAG). (1997). *Leitfaden für die Planung von Projekt- und Programm-evaluation*. CH-3003 Bern. Zugriff am 13. 10. 2008, <http://www.bag.admin.ch/evaluation/02357/02362/>
- Bussmann, W. (1995). *Evaluationen staatlicher Maßnahmen erfolgreich begleiten und nutzen. Ein Leitfaden*. Chur/Zürich: Ruegger. Auch als Online-Dokument: Zugriff am 13. 10. 2008, <http://www.crus.ch/dms.php?id=5266>
- Bustelo, M. (2003). Evaluation of Gender Mainstreaming: Ideas from a Meta-Evaluation Study. *Evaluation*, 9 (4), 383–403.
- Chen, H.-T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- DeGEval (Deutsche Gesellschaft für Evaluation) (2002a). *Angefragte Gutachten/Meta-Evaluationen auf Basis der DeGEval-Standards*. Zugriff am 13. 10. 2008, <http://www.degeval.de/calimero/tools/proxy.php?id=63>
- DeGEval (Deutsche Gesellschaft für Evaluation) (2002b). *Standards für Evaluation. Online-Dokument*. Zugriff am 13. 10. 2008, http://www.degeval.de/index.php?class=Calimero_Webpage&id=9023
- Döring, N. (2005). Für Evaluation und gegen Evaluitis. Warum und wie Lehrevaluation an deutschen Hochschulen verbessert werden sollte. In B. Berendt, H.-P. Voss & J. Wildt (2005), *Neues Handbuch Hochschullehre* (Ergänzungslieferung Juli 2005). Berlin: Raabe.
- Gallagher, T., Waterman, A., Ebers, A., Fraser, V. & Levinson, W. (2003). Patients' and Physicians' Attitudes Regarding the Disclosure of Medical Errors. *The Journal of the American Medical Association JAMA*, 289, 1001–1007.
- Gediga, G. & Hamborg, K.-C. (2002). Ergonomische Evaluation von Software: Methoden und Modelle im Software-Entwicklungsprozess. *Zeitschrift für Psychologie*, 210, 40–57.
- Grawe, K., Donati, R. & Bernauer, F. (2001). *Psychotherapie im Wandel – Von der Konfession zur Profession* (5. Aufl.). Göttingen: Hogrefe.
- Gregory, A. (2000). Problematizing Participation. A Critical Review of Approaches to Participation in Evaluation Theory. *Evaluation*, 6 (2), 179–199.
- Hager, W. (2000). Wirksamkeits- und Wirksamkeitsunterschiedshypothesen, Evaluationsparadigmen, Vergleichsgruppen und Kontrolle. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 180–201). Bern: Huber.
- Holling, H. (Hrsg.). (2008). *Anwendungen der Evaluation* (Enzyklopädie der Psychologie, Serie Methodologie und Methoden, Bd. 2). Göttingen: Hogrefe. (Manuskript in Vorbereitung)
- Kromrey, H. (2000). *Empirische Sozialforschung* (9. Aufl.). Opladen: Leske + Budrich.
- Kromrey, H. (2003). Evaluation in Wissenschaft und Gesellschaft. *Zeitschrift für Evaluation*, 2, 93–116.

- Lewin, K. (1951). *Field theory in social science*. New York: Harper & Row.
- Mittag, W. & Hager, W. (2000). Ein Rahmenkonzept zur Evaluation psychologischer Interventionsmaßnahmen. In W. Hager, J.-L. Patry & H. Brenzing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 102–128). Bern: Huber.
- Moosbrugger, H. & Schweizer, K. (2002). Evaluationsforschung in der Psychologie. *Zeitschrift für Evaluation*, 1, 19–37.
- Müller-Kohlenberg, H. & Beywl, W. (2002). *Standards der Selbstevaluation*. Online-Dokument. Zugriff am 15. 12. 2004, http://www.degeval.de/ak_soz/Selbstevaluation.rtf
- Neuman, W. L. (2003). *Social Research Methods* (5th ed.). Boston: Allyn and Bacon.
- Owen, J. M. & Rogers, P. J. (1999). *Program Evaluation. Forms and Approaches*. London: Sage.
- Patel, M. (2002). A meta-evaluation, or quality assessment, of the evaluations in this issue, based on the African Evaluation Guidelines: 2002. *Evaluation and Program Planning*, 25 (4), 329–332.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks: Sage.
- Pawson, R. & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Rebien, C. C. (1996). Participatory Evaluation of Development Assistance: Dealing with Power and Facilitative Learning. *Evaluation* 2 (2), 151–172.
- Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004). *Evaluation. A Systematic Approach*. London: Sage.
- Sanders, J. R. (Hrsg.). (1999). *Handbuch der Evaluationsstandards. Die Standards des „Joint Committee on Standards for Educational Evaluations“*. Opladen: Leske + Budrich.
- Schaumburg, H. (2001). Neues Lernen mit Laptops? Ein Überblick über Forschungsergebnisse zur Nutzung mobiler Computer in der Schule. *Zeitschrift für Medienpsychologie*, 13 (1), 11–21.
- Schnell, R., Hill, P. & Esser, E. (2003). *Methoden der Empirischen Sozialforschung* (6. Aufl.). München: Oldenbourg.
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, 17 (2), 151–162.
- Scriven, M. (2000). *The Logic and Methodology of Checklists*. Online Document. Zugriff am 5. 7. 2006, http://www.wmich.edu/evalctr/checklists/papers/logic&methodology_oct05.pdf.
- seval (Schweizerische Evaluationsgesellschaft) (2000). *Evaluationsstandards*. Zugriff am 5. 7. 2006, http://www.seval.ch/de/documents/seval_Standards_2001_dt.pdf.
- Shadish, W. R., Newman, D., Scheirer, M. A. & Wye, C. (Eds.). (1995). *The American Evaluation Association's Guiding Principles*. San Francisco: Jossey-Bass.
- Sharp, J. A., Peters, J. & Howard, K. (2002). *The Management of a Student Research Projekt* (3rd ed.). Aldershot: Gower.
- Stamm, M. (2003). *Evaluation und ihre Folgen für die Bildung – eine unterschätzte pädagogische Herausforderung*. Münster: Waxmann.
- Stockmann, R. (2000). *Evaluationsforschung*. Opladen: Leske + Budrich.

- Stufflebeam, D. L. (2001). The metaevaluation imperative. *The American Journal of Evaluation*, 22 (2), 183–209.
- Stufflebeam, D. L. (2007). *CIPP Evaluation Model Checklist* (2nd ed.). Zugriff am 5. 7. 2006, http://www.wmich.edu/evalctr/checklists/cippchecklist_mar07.pdf.
- Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. San Francisco, CA: Wiley.
- Suchman, E. A. (1967). *Evaluative Research. Principles and Practice in Public Service and Social Action Programs*. New York: Russel Sage.
- Tarling, R. (2006). *Managing Social Research: A Practical Guide*. Oxford: Routledge.
- Taut, S. & Brauns, D. (2003). Resistance to Evaluation. A Psychological Perspective. *Evaluation*, 9 (3), 247–264.
- Uhl, A. (1997). Probleme bei der Evaluation von Präventionsmaßnahmen im Suchtbereich. *Wiener Zeitschrift für Suchtforschung*, 20 (3/4), 93–109.
- Usherwood, T. (1996). *Introduction to Project Management in Health Research: A Guide for New Researchers*. Buckingham: Open University Press.
- Westermann, R. (2002). Merkmale und Varianten von Evaluationen: Überblick und Klassifikation. *Zeitschrift für Psychologie*, 210 (1), 4–26.
- Widmer, T., Rothmayr, C. & Serdült, U. (1996). *Kurz und gut? Qualität und Effizienz von Kurzevaluationen*. Chur Zürich: Rüegger.
- Witt, H. (2001). *Forschungsstrategien bei quantitativer und qualitativer Forschung*. *Forum Qualitative Sozialforschung*, 2 (1, Artikel 8). Zugriff am 5. 7. 2006, <http://www.qualitative-research.net/fqs-texte/1-01/1-01witt-d.pdf>.
- Wittmann, W. W. (1990). Brunswik-Symmetrie und die Konzeption der Fünf-Datenboxen. Ein Rahmenkonzept für umfassende Evaluationsforschung. *Zeitschrift für Pädagogische Psychologie*, 4, 241–251.
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation* (2. Aufl.). Göttingen: Huber.
- Ziegele, F. & Hener, Y. (2004). Benchmarking in der Hochschulpraxis – Effizienzsteigerung und Prozessoptimierung durch kooperatives Handeln. In W. Benz, J. Kohler & K. Landfried (Hrsg.), *Handbuch Qualität in Studium und Lehre* (E 7.2: S. 1–120). Berlin: Raabe. Auch als Online-Dokument: Zugriff am 13. 10. 2008, http://www.che.de/downloads/E_7_2_Ziegele_Hener.pdf