



HANDBUCH DER PSYCHOLOGIE

Handbuch Statistik, Methoden und Evaluation

Heinz Holling
Bernhard Schmitz
(Hrsg.)



HOGREFE



HANDBUCH DER PSYCHOLOGIE

hrsg. von J. Bengel, H.-W. Bierhoff, V. Brandstätter, M. Eid, D. Frey, P. A. Frensch, J. Funke, S. Gauggel, M. Hasselhorn, M. Herrmann, H. Holling, M. Jerusalem, J. H. Otto, F. Petermann, T. Rammsayer, H. Reinecker, B. Schmitz, W. Schneider, H. Schuler, Kh. Sonntag, M. Steller, R. Volbert und H. Weber.

Band 13

Handbuch Statistik, Methoden und Evaluation

hrsg. von Heinz Holling und Bernhard Schmitz

weitere Bände:

Handbuch der Allgemeinen Psychologie: Kognition

hrsg. von Joachim Funke und Peter A. Frensch

Handbuch der Allgemeinen Psychologie: Motivation und Emotion

hrsg. von Veronika Brandstätter und Jürgen H. Otto

Handbuch der Entwicklungspsychologie

hrsg. von Marcus Hasselhorn und Wolfgang Schneider

Handbuch der Sozialpsychologie und Kommunikationspsychologie

hrsg. von Hans-Werner Bierhoff und Dieter Frey

Handbuch der Persönlichkeitspsychologie und Differentiellen Psychologie

hrsg. von Hannelore Weber und Thomas Rammsayer

Handbuch der Neuro- und Biopsychologie

hrsg. von Siegfried Gauggel und Manfred Herrmann

Handbuch der Psychologischen Diagnostik

hrsg. von Franz Petermann und Michael Eid

Handbuch der Klinischen Psychologie und Psychotherapie

hrsg. von Franz Petermann und Hans Reinecker

Handbuch der Arbeits- und Organisationspsychologie

hrsg. von Heinz Schuler und Karlheinz Sonntag

Handbuch der Pädagogischen Psychologie

hrsg. von Wolfgang Schneider und Marcus Hasselhorn

Handbuch der Gesundheitspsychologie und Medizinischen Psychologie

hrsg. von Jürgen Bengel und Matthias Jerusalem

Handbuch der Rechtspsychologie

hrsg. von Renate Volbert und Max Steller

HANDBUCH DER PSYCHOLOGIE

Handbuch Statistik, Methoden und Evaluation

herausgegeben von

Heinz Holling und Bernhard Schmitz

HOGREFE



GÖTTINGEN · BERN · WIEN · PARIS · OXFORD · PRAG · TORONTO
CAMBRIDGE, MA · AMSTERDAM · KOPENHAGEN · STOCKHOLM

Prof. Dr. Heinz Holling, geb. 1960. 1969–1976 Studium der Mathematik, Psychologie und Soziologie in Würzburg und Berlin. 1974–1987 Wissenschaftlicher Mitarbeiter an der FU Berlin und der Universität Osnabrück. Promotion 1980 (Dr. phil.) und 1987 (Dr. rer. nat.). 1987 Habilitation. 1987–1993 Vertretungsprofessor an den Universitäten Oldenburg, Münster und Mannheim. Seit 1993 Professor für Statistik und Quantitative Methoden am Psychologischen Institut der Universität Münster.

Prof. Dr. Bernhard Schmitz, geb. 1950. 1970–1977 Studium der Mathematik in Düsseldorf. 1976–1981 Studium der Psychologie in Düsseldorf und Berlin. 1978–1983 Assistent am Psychologischen Institut der FU Berlin. 1984 Promotion. 1984–1997 Wissenschaftlicher Mitarbeiter am Max-Planck-Institut für Bildungsforschung in Berlin. 1993 Vertretungsprofessur an der Universität Potsdam. 1995 Habilitation. 1995 und 1996 Gastprofessuren an der Universität Potsdam. Seit 1997 Professor für Pädagogische Psychologie an der TU Darmstadt.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© 2010 Hogrefe Verlag GmbH & Co. KG
Göttingen · Bern · Wien · Paris · Oxford · Prag · Toronto
Cambridge, MA · Amsterdam · Kopenhagen · Stockholm
Rohnsweg 25, 37085 Göttingen

<http://www.hogrefe.de>

Aktuelle Informationen · Weitere Titel zum Thema · Ergänzende Materialien



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Satz: ARThür Grafik-Design & Kunst, Weimar
Druck: AZ Druck und Datentechnik GmbH, Kempten
Printed in Germany
Auf säurefreiem Papier gedruckt

ISBN 978-3-8017-1848-0

Inhalt

Vorwort

<i>Heinz Holling & Bernhard Schmitz</i>	10
---	----

I Grundlagen der Methodenlehre

Grundkonzepte der Skalierung

<i>Thomas Staufenbiel & Ingwer Borg</i>	15
---	----

Stichprobenziehung

<i>Siegfried Gabler</i>	27
-------------------------------	----

Experimentelle Versuchspläne

<i>Wolfgang Ellermeier & Wolfgang Bösche</i>	37
--	----

Quasi-experimentelle Versuchspläne

<i>Silke Hertel, Julia Klug & Bernhard Schmitz</i>	49
--	----

Einzelfallpläne

<i>Michaela Schmidt</i>	63
-------------------------------	----

Gütekriterien

<i>Rainer Westermann & Jeanette Krohn</i>	71
---	----

Wissenschaftstheorie

<i>Hans Westmeyer</i>	87
-----------------------------	----

II Datenerhebung

Bildgebende Verfahren

<i>Lutz Jäncke</i>	101
--------------------------	-----

Biochemische Verfahren

<i>Andrea Catherina Gierens, Petra Pütz & Dirk Helmut Hellhammer</i>	112
--	-----

Elektrophysiologische Verfahren

<i>Patrick Khader & Frank Rösler</i>	120
--	-----

Befragung	
<i>Kirsten van de Loo</i>	131
Beobachtungsverfahren: Vom Datenmaterial zur Datenanalyse	
<i>Tina Seidel & Manfred Prenzel</i>	139
Interview	
<i>Heinz Reinders</i>	153
Tagebuch	
<i>Meike Landmann & Michaela Schmidt</i>	165
Nonreaktive Methoden: Vermeidung reaktiver Effekte in der psychologischen Forschung	
<i>Bastian Frithjof Benz</i>	173
Qualitativ orientierte Verfahren	
<i>Philipp Mayring</i>	179
Internetbasierte Methoden	
<i>Timo Gnambs & Bernad Batinic</i>	191
Ambulantes Assessment	
<i>Jochen Fahrenberg</i>	201
Event-Sampling	
<i>Sandra Ohly & Dieter Zapf</i>	213
 III Evaluation	
Grundlagen der Evaluationsforschung	
<i>Christiane Spiel, Petra Gradinger & Marko Lüftenegger</i>	223
Modelle der Evaluation	
<i>Renate Soellner</i>	233
Entscheidungs- und Nutzenanalyse	
<i>Katrin Borchering & Daniel Ullrich</i>	244
Zielexplication und Standards in der Evaluationsforschung	
<i>Christiane Spiel, Marko Lüftenegger, Petra Gradinger & Ralph Reimann</i> ...	252

Planung und Durchführung von Evaluationsstudien	
<i>Nicola Döring</i>	261

IV Grundlegende statistische Verfahren

Univariate deskriptive Statistik	
<i>Christof Nachtigall, Steffi Pohl & Sven Hartenstein</i>	275

Bivariate deskriptive Statistik	
<i>Markus Wirtz & Rolf Ulrich</i>	288

Wahrscheinlichkeitstheorie	
<i>Frank Lammers</i>	302

Wahrscheinlichkeitsverteilungen	
<i>Britta Colver</i>	314

Punktschätzung	
<i>Anna Doeblér</i>	326

Konfidenzintervalle	
<i>Otto B. Walter</i>	335

Hypothesentestung	
<i>Jörg-Tobias Kuhn</i>	350

Effektgröße und Teststärke	
<i>Edgar Erdfelder, Franz Faul, Axel Buchner & Lutz Cüpper</i>	358

Tests auf Anpassung und Assoziation	
<i>Ralf Schulze</i>	370

Ein-Stichproben-Tests	
<i>Philipp Alexander Freund</i>	382

Zwei-Stichproben-Tests	
<i>Jonas Pablo Bertling</i>	391

Einfache und multiple Regression	
<i>Wolfgang A. Rauch & Helfried Moosbrugger</i>	410

Einfaktorielle Varianzanalyse	
<i>Helfried Moosbrugger & Siegbert Reiß</i>	420

Mehrfaktorielle Varianzanalyse und Varianzanalyse mit Messwiederholung	
<i>Helfried Moosbrugger & Siegbert Reiß</i>	439

V Weiterführende statistische Verfahren

Generalisierte lineare Modelle	
<i>Gerhard Tutz & Carolin Strobl</i>	461

Allgemeines Lineares Modell (ALM)	
<i>Helfried Moosbrugger, Dorothea Mildner & Karl Schweizer</i>	472

Theorie und Analyse kausaler Effekte	
<i>Christiane Fiege, Ulf Kröhne & Rolf Steyer</i>	487

Logistische Regressionsverfahren	
<i>Martin Spieß</i>	496

Logistische Regressionsverfahren für mehrkategoriale Zielvariablen	
<i>Martin Spieß & Gerhard Tutz</i>	509

Loglineare Modelle	
<i>Thorsten Meiser</i>	518

Multivariate Varianzanalyse und Diskriminanzanalyse	
<i>Johannes Andres</i>	525

Mehrebenenanalyse	
<i>Johannes Hartig & Katrin Rakoczy</i>	538

Clusteranalyse	
<i>Christian Tarnai</i>	548

Latent-Class-Analyse	
<i>Anton K. Formann</i>	556

Mischverteilungsmodelle	
<i>Fridtjof W. Nussbeck, Michael Eid & Christian Geiser</i>	562

Korrespondenzanalyse	
<i>Jörg Blasius</i>	569
Multidimensionale Skalierung	
<i>Thomas Staufenbiel & Ingwer Borg</i>	575
Exploratorische Faktorenanalyse	
<i>Sabine Krolak-Schwerdt & Frank M. Spinath</i>	591
Strukturgleichungsmodelle	
<i>Norman Rose, Steffi Pohl, Hendryk Frank Böhme & Rolf Steyer</i>	600
Längsschnittstudien	
<i>Georg Rudinger</i>	612
Veränderungsmessung	
<i>Karl Christoph Klauer & Rainer Leonhart</i>	624
Zeitreihenanalyse	
<i>Franziska Perels</i>	632
Ereignisanalyse	
<i>Hans-Peter Blossfeld</i>	642
Messtheorie von Rasch	
<i>Jürgen Rost</i>	649
Item-Response-Modelle für kategoriale Daten	
<i>Claus H. Carstensen</i>	660
Item-Response-Modelle für ordinale Daten	
<i>Matthias von Davier</i>	667
Multitrait-Multimethod-Analyse	
<i>Christian Geiser, Michael Eid, Fridtjof W. Nussbeck,</i> <i>Tanja Lischetzke & David A. Cole</i>	679
Generalisierbarkeitstheorie	
<i>Heinz Holling</i>	686
Beurteilerübereinstimmung	
<i>Christof Schuster</i>	700

Modellvergleich <i>Andreas Klein</i>	708
Resampling-Verfahren <i>Günther Gediga</i>	714
Missing-Data-Analyse <i>Oliver Lüdtke & Alexander Robitzsch</i>	723
Bayes'sche Methoden in der Statistik <i>Wim J. van der Linden</i>	730
Meta-Analyse <i>Wolfgang Viechtbauer</i>	743
Neuronale Netzwerke <i>Fred H. Hamker & Marc Zirnsak</i>	757
Die Autorinnen und Autoren des Bandes	769
Autorenregister	781
Sachregister	795

Planung und Durchführung von Evaluationsstudien

Design and Implementation of Evaluation Studies

Nicola Döring

Evaluationsstudien haben das Ziel, den *Wert oder Nutzen einer Maßnahme, eines Produkts oder einer Methode* (allgemein: eines Evaluationsgegenstandes) mit Hilfe wissenschaftlicher Methoden zu beurteilen (Stufflebeam & Shinkfield, 2007, S. 7 ff.). Dabei kann es sich um Evaluationsgegenstände handeln, die bereits längere Zeit vorliegen, erstmals umgesetzt, gerade noch entwickelt oder erst für die Zukunft geplant werden. Psychologische Evaluationsstudien befassen sich primär mit *psychologischen Interventionen*, also mit Maßnahmen oder Produkten, die menschliches Erleben und Verhalten von außen zielgerichtet beeinflussen sollen (z. B. psychotherapeutische Behandlungsmethoden, computergestützte Lernprogramme, Marketingkampagnen). Im Rahmen von Evaluationsstudien zu bereits vorliegenden Evaluationsgegenständen spielen zwei Bewertungsaspekte eine herausgehobene Rolle (vgl. Gollwitzer & Jäger, 2007, S. 7):

- Die **Wirksamkeitsmessung** (Effektivitätsmessung): In welchem Umfang erreicht die Maßnahme die angestrebten Ziele – möglichst ohne unerwünschte Nebenwirkungen?
- Die **Effizienzmessung**: Steht der Nutzen der Maßnahme in vernünftigem Verhältnis zu Aufwand und Kosten?

Bei Evaluationsgegenständen, die sich noch in der Planungsphase befinden (z. B. Planung des neuen psychologischen Masterstudiengangs), geht es im Rahmen einer Evaluation beispielsweise darum, zunächst die *Machbarkeit* zu überprüfen oder eine *Bedarfsanalyse* durchzuführen.

Eine *formative Evaluation* untersucht die Intervention bereits während ihrer Entwicklung und liefert fortlaufend konkrete Verbesserungsvorschläge. Demgegenüber bietet eine *summative Evaluation* eine abschließende Qualitätsbewertung, z. B. zum Vergleich unterschiedlicher Therapie-, Trainings- oder Präventions-Maßnahmen (vgl. Scriven, 1967).

Im Sinne der Neutralität und Professionalität werden Evaluationsstudien oft als *Fremdevaluationen* bei Evaluationsexperten (sogenannte *Evaluatoren*) in Auftrag

gegeben. Für die alltägliche Qualitätssicherung ist dies jedoch zu aufwändig und kostspielig, weshalb zunehmend mit *Selbstevaluationen* gearbeitet wird.

Evaluationsstudien sind *Forschungsprojekte*, werden deswegen auch als *Evaluationsprojekte* bezeichnet, und müssen Prozessabläufe auf drei Ebenen integrieren:

1. Aus *wissenschaftstheoretischer Perspektive* müssen bei einer Evaluation die Phasen des empirischen Forschungsprozesses durchlaufen werden.
2. Aus *Projektmanagement-Perspektive* müssen die Projektphasen des Evaluationsprojektes systematisch organisiert (z. B. Zeit- und Kostenplanung) und kontrolliert werden (Projekt-Controlling). Eine besondere Herausforderung besteht darin, die verschiedenen Interessengruppen (Stakeholder) im Umfeld der zu evaluierenden Maßnahme (Interventionsentwickler, Interventionsnutzer, indirekt Betroffene, Entscheidungsträger in Wirtschaft und Politik etc.) angemessen einzubeziehen, etwa durch einen *Stakeholder-Beirat* in der Projektstruktur.
3. Aus *Perspektive der Designwissenschaft*, die sich mit Entwurf und Entwicklung neuer Methoden, Maßnahmen und Produkte befasst, ist zu beachten, wann im Lebenszyklus eines Evaluationsgegenstandes Evaluationsaktivitäten greifen sollen. In der Regel wird dazu geraten, eben nicht erst fertig entwickelte Produkte oder Maßnahmen zu evaluieren (z. B. marktreife Lern-Software), sondern bereits in ganz frühen Entwicklungsphasen mit empirischer Evaluation zu beginnen, damit Evaluationsergebnisse (z. B. mangelnde Akzeptanz bestimmter thematischer, inhaltlicher oder formaler Aspekte der Lern-Software in Teilen der Zielgruppe) gleich in den Gestaltungsprozess einfließen können (vgl. Döring & Ingerl, 2008).

Die Planung und Durchführung von Evaluationsstudien lässt sich in fünf Phasen gliedern (vgl. Tab. 1), deren Arbeitsschritte im Folgenden genauer dargestellt werden (vgl. Döring, 2009; Gollwitzer & Jäger, 2007, S. 37).

Tabelle 1: Fünf-Phasen-Modell für Evaluationsforschungsprojekte

Projektphasen der Evaluation	Wichtigste Arbeitsschritte
1. Initiierung	1.1 Akquise von Evaluationsprojekten 1.2 Bewertung von Evaluationsprojekten
2. Konzeption	2.1 Entwicklung des Evaluationskonzepts 2.2 Abschluss des Evaluationsvertrags
3. Planung	3.1 Formulierung von Evaluationsfragen und -hypothesen 3.2 Untersuchungsplanung
4. Realisation	4.1 Erhebung und Analyse von Evaluationsdaten 4.2 Interpretation der Evaluationsbefunde
5. Abschluss	5.1 Organisation der Evaluationsnutzung 5.2 Durchführung einer Meta-Evaluation

1 Initiierungsphase der Evaluation

In der Initiierungsphase geht es darum, potenzielle Evaluationsprojekte zu akquirieren und hinsichtlich ihrer Erfolgsaussichten kritisch zu beleuchten.

1.1 Akquise von Evaluationsprojekten

Es existieren für Evaluatoren im Wesentlichen vier Wege der Projektakquise (vgl. Stufflebeam & Shinkfield, 2007, S. 453 ff.):

1. *Projektausschreibungen*: Öffentliche sowie private/gewerbliche Ausschreibungen für Evaluationsprojekte sind entsprechenden Publikationen und Datenbanken zu entnehmen (z. B. www.deutsches-ausschreibungsblatt.de).
2. *Projektaufträge*: Viele Evaluatoren arbeiten nicht selbstständig, sondern in Unternehmen oder Evaluationsinstituten und erhalten in diesem Rahmen Evaluationsaufträge. Der Evaluationsauftrag kann sich auf einen unternehmensinternen Gegenstand beziehen (*interne Evaluation*: z. B. Evaluation des unternehmensinternen Weiterbildungssystems) oder bei einem Kunden angesiedelt sein (*externe Evaluation*).
3. *Projektanfragen*: Manchmal suchen sich Auftraggeber auch gezielt Evaluatoren aus und bieten ihnen Projekte an.
4. *Evaluator-initiierte Projekte*: Auf der Basis ihrer Fachexpertise und Praxiserfahrung in einem bestimmten Evaluationsbereich können Evaluatoren auch selbst Projektideen entwickeln und sich dafür passende Auftraggeber bzw. Finanzierungsmöglichkeiten suchen.

1.2 Bewertung von Evaluationsprojekten

Ob es im Einzelfall lohnens- und wünschenswert ist, ein potenzielles Evaluationsprojekt weiterzuverfolgen, hängt aus Sicht der Evaluatorin oder des Evaluators von ganz unterschiedlichen Rahmenbedingungen ab: Etwa von der Passung des Evaluationsbereiches bzw. Evaluationsgegenstandes zum eigenen Qualifikationsprofil, dem aktuellen fachlichen Interesse an der zu evaluierenden Maßnahme oder den Finanzierungsmöglichkeiten.

Aus wissenschaftstheoretischer Perspektive ist es besonders wichtig, zwei Erfolgsbedingungen der *Auftragsforschung* sicherzustellen:

1. Der Evaluationsauftrag muss in Abstimmung mit Auftraggeber und anderen Interessengruppen so *konkretisierbar* bzw. *operationalisierbar* sein, dass er im Rahmen einer sozialwissenschaftlichen Studie zu bearbeiten ist. Problematisch sind Fälle, in denen die Ziele der zu evaluierenden Intervention nicht eindeutig definierbar sind oder von der Evaluation sehr weitreichende Kausalitäts- und Wirkungsnachweise erwartet werden, obwohl im Rahmen der Maßnahme nur sehr begrenzt empirische Befunde erhoben werden können (vgl. Uhl, 1997).

2. Die Evaluationsstudie muss *ergebnisoffen* abgewickelt werden. Problematisch ist es, wenn Interessenvertreter durch ihren Einfluss dafür sorgen wollen, dass die Studie nur bestimmte wünschenswerte Ergebnisse hervorbringt. Unter derartigen Bedingungen werden die Evaluatoren instrumentalisiert und die Evaluation degeneriert zur Schein- oder *Pseudo-Evaluation* (Stufflebeam & Shinkfield, 2007, S. 145 ff.; Wottawa & Thierau, 1998, S. 29 ff.).

2 Konzeptionsphase der Evaluation

Erscheint die Übernahme eines bestimmten Evaluationsprojekts aus Sicht des Evaluators wünschenswert, so ist im nächsten Schritt das Evaluationskonzept bzw. Projektangebot zu entwickeln. Kommt daraufhin der Auftrag zustande, ist der Abschluss eines Evaluationsvertrags angezeigt.

2.1 Entwicklung des Evaluationskonzepts

Das Evaluationskonzept (evaluation design) ist das Kernstück eines Projektangebotes und muss darlegen, wie die Evaluationsziele des Auftraggebers im vorgegebenen zeitlichen und finanziellen Rahmen mit hoher wissenschaftlicher Qualität erreicht werden sollen. Bei der Ausarbeitung des Evaluationskonzeptes kann man sich an erfolgreichen Konzepten aus anderen Projekten des fraglichen Evaluationsbereiches orientieren sowie an Checklisten, die die einzelnen Arbeitsschritte der Evaluationskonzept-Entwicklung beschreiben, etwa die „Evaluation Design Checklist“ (Stufflebeam & Shinkfield, 2007, S. 493 ff.; zur Checklisten-Methode in der Evaluationsforschung siehe Scriven, 2000).

Das Evaluationskonzept sollte Aussagen zu folgenden Punkten treffen:

1. Welche *Evaluationsfragen und -hypothesen* sollen mit der Studie beantwortet werden? Geht es beispielsweise um die Beschreibung von Ist-Zuständen und/oder Veränderungen, um die Messung von Wirksamkeit und/oder Effizienz?
2. Welchem *Evaluationsmodell* (z. B. nutzungsorientierte, partizipative oder theoriebasierte Evaluation) soll die Studie folgen? Und daraus abgeleitet, welche Untersuchungsdesigns sollen umgesetzt werden (z. B. experimentelle und/oder nicht experimentelle Evaluation)?
3. Mit welchen quantitativen und/oder qualitativen *Verfahren der Datenerhebung und Datenanalyse* soll gearbeitet werden? In welchem Umfang sollen Daten erhoben werden (z. B. Anzahl, Art und Größe der verwendeten Stichproben)?
4. Wie wird sichergestellt, dass die Evaluationsstudie die gängigen *Evaluationsstandards* erfüllt? Gemäß der Gesellschaft für Evaluation e. V. (DeGEval, 2002b) sollen 25 Standards erfüllt werden, die vier Qualitätsdimensionen zugeordnet sind: Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit. Nicht in jeder

Evaluationsstudie lassen sich alle Einzelstandards realisieren, da diese teilweise in Konkurrenz zueinander stehen.

5. Welche Vorkehrungen werden für ein professionelles *Projektmanagement* getroffen (z. B. Instrumente der Kostenkalkulation, Beschäftigung eines Projektleiters)?
6. Wie wird die *Nutzung der Evaluationsergebnisse* für konkrete Veränderungsmaßnahmen in der Praxis gefördert?
7. Wird die Evaluationsstudie ihrerseits auf wissenschaftliche Qualität und Einhaltung der Evaluationsstandards überprüft? Wenn ja, wie wird diese *Meta-Evaluation* durchgeführt?
8. Welche *Ressourcen* werden für die Evaluationsstudie insgesamt benötigt: Personal, Räume, Technik, Finanzplan?
9. Wann ist laut *Zeitplan* mit aussagekräftigen Zwischen- und Endergebnissen zu rechnen?

2.2 Abschluss des Evaluationsvertrags

Erhält ein Projektangebot den Zuschlag oder nimmt man einen Projektauftrag an, so folgt der Abschluss des *Evaluationsvertrags* (evaluation contract). Er fixiert die Rechte und Pflichten des Evaluators (z. B. Abgabetermine) sowie des Auftraggebers (z. B. Bereitstellung von Untersuchungsteilnehmern und Arbeitsplätzen im Feld). Insbesondere bei umfangreichen Projekten sind die rechtlichen Bedingungen genau zu prüfen, etwa die Zahlungsmodalitäten oder Konsequenzen bei Nichteinhaltung einzelner vertraglicher Regelungen. Sofern sich während eines Evaluationsprojektes bedeutsame Änderungen im Evaluationsmandat ergeben, sollten diese ebenfalls vertraglich festgehalten werden.

Bei internen oder selbst finanzierten Evaluationen wird in der Regel kein Vertrag, sondern eine *Evaluationsvereinbarung* (memorandum of understanding) zwischen den Beteiligten unterzeichnet (für eine Checkliste zur Erstellung von Evaluationsverträgen und -vereinbarungen siehe Stufflebeam & Shinkfield, 2007, S. 547 f.).

3 Planungsphase der Evaluation

Nach dem Abschluss des Evaluationsvertrages stehen die Detailplanung und konkrete Vorbereitung der Evaluationsstudie an. Dies umfasst vor allem die Formulierung von Evaluationsfragen und -hypothesen sowie die Untersuchungsplanung.

3.1 Formulierung von Evaluationsfragen und -hypothesen

Da Evaluationsstudien gemäß den Evaluationsstandards für die Praxis nützlich sein sollen, haben viele Evaluationsfragen *pragmatischen Charakter*. Sie zielen darauf ab, Stärken und Schwächen des Evaluationsgegenstands zu identifizieren. Neben

Effektivität und Effizienz adressieren Evaluationsfragen auch oft Bedingungen der aktuellen und zukünftigen Bereitstellung und Nutzung einer Intervention. Ein Beispiel sind *Evaluationsfragen zur Nachhaltigkeit (sustainability)*: „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?“ Das *CIPP-Modell* (Context, Input, Process, Product) von Daniel Stufflebeam ist ein *Evaluationsmodell*, das die zentralen praxisbezogenen Evaluationsfragen für die verschiedenen Phasen der Interventionsentwicklung vorgibt (Stufflebeam & Shinkfield, 2007, S. 325 ff., vgl. Tab. 2).

Tabelle 2: CIPP-Modell und Evaluationsfragen (Stufflebeam, 2007)

CIPP-Modell	Zentrale Evaluationsfragen
Context	What needs to be done?
Input	How should it be done?
Process	Is it being done?
Product – Impact – Effectiveness – Sustainability – Transportability	Did it succeed? – Were the right beneficiaries reached? – Were their needs met? – Were the gains for the beneficiaries sustained? – Did the processes that produced the gains prove transportable and adaptable for effective use in other settings?

Das *nutzungsorientierte Evaluationsmodell* (utilization-focused evaluation) von Michael Patton (1997) empfiehlt, die Formulierung von Evaluationsfragen von vorne herein konsequent danach auszurichten, wer die Ergebnisse wofür verwenden will. Dementsprechend sind dann die zukünftigen *Evaluationsnutzer* (z. B. Klinikleiter, Schulleiter) bei der Formulierung von Evaluationsfragen maßgeblich einzubeziehen. Da Evaluationsnutzer oftmals keine Forscher sind, sind sie vom Evaluationsteam durch entsprechende Motivations-, Moderations- und Schulungsmaßnahmen zu unterstützen.

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

Traditionell wird in vielen *ziel-/wirkungsorientierten Evaluationsstudien* (goal-based evaluation) der Nachweis der Wirksamkeit in den Mittelpunkt gerückt, also

die Übereinstimmung zwischen den angezielten und tatsächlichen Wirkungen einer Maßnahme. Dafür wird beispielsweise ein Katalog der angestrebten Therapieziele einer neuen Therapiemethode erstellt und die Zielerreichung empirisch durch Befragung, Testverfahren und/oder Beobachtung geprüft. *Partizipative Evaluationsmodelle* betonen, dass die Zieldefinition der Maßnahme nicht den Interventionsentwicklern (z. B. Psychotherapeuten) allein überlassen bleiben sollte, sondern dass die Erwartungen und Bedarfe aller Stakeholder, insbesondere der potenziellen *Interventionsnutzer* (z. B. Patienten, aber auch Angehörige von Patienten) umfassend zu berücksichtigen sind (vgl. Ulrich & Wenzel, 2003; Stahl, Bromme, Pieschl, Hölzenbein & Kiffe, 2005). Dies gilt gemäß dem demokratischen Anspruch partizipativer Evaluation auch und gerade für unterprivilegierte Betroffenenengruppen (z. B. Psychiatriepatienten, Drogenkranke, Arbeitslose).

Evaluationsfragen lassen sich teilweise auch als *Evaluationshypothesen* fassen, da implizit von der Wirkung der Intervention ausgegangen wird. Beispielsweise kann statt der offenen Frage „Erreicht die Intervention die intendierten Wirkungen?“ eine gerichtete Hypothese formuliert werden: „Die intendierte Wirkung X wird mit der Intervention in stärkerem Maße erreicht als ohne Intervention bzw. als mit einer alternativen Intervention“. Diese Hypothese könnte z. B. in einem Kontrollgruppen-Design experimentell geprüft werden. Im Rahmen der sogenannten *theoriebasierten Evaluation* (theory-based/-driven evaluation) wird besonderer Wert darauf gelegt, dass die Evaluationshypothesen aus Theorien der Interventionsnutzung und -wirkung abgeleitet sind und Aussagen über die Wirkungsweise der Maßnahme treffen. Somit können die Evaluationsergebnisse dann nicht nur zur Verbesserung der Maßnahme, sondern auch zur Theorieentwicklung beitragen (Birckmayer & Weiss, 2000).

So nahe liegend es ist, die Qualität des Evaluationsgegenstandes an vorher festgelegten Zielsetzungen zu messen, so irreführend kann diese Vorgehensweise sein, etwa wenn sie den Blick der Evaluatoren und der anderen Stakeholder zu sehr verengt. Die *zielfreie Evaluation* (goal free evaluation) arbeitet deswegen gerade *nicht* mit detaillierten, zielbezogenen Evaluationshypothesen oder -fragen (z. B. Wie stark verbessert sich die Rechtschreibleistung durch Nutzung des neuen Rechtschreibprogramms?), sondern lässt externe Evaluatoren (die keinerlei Informationen über die Zielsetzung der Intervention erhalten) der *offenen Frage* nachgehen, wie der Evaluationsgegenstand genutzt wird und welche Konsequenzen seine Nutzung hat (Scriven, 1991). Dieses Vorgehen – das in einer Mehrmethoden-Studie mit anderen Evaluationsmodellen kombinierbar ist – soll die Wahrscheinlichkeit steigern, vorher nicht erwartete Effekte und Nebenwirkungen durch inhaltlich breit gestreute, unvoreingenommene Beobachtungen und Befragungen aufzudecken (z. B. könnte sich zeigen, dass das Rechtschreibprogramm regelmäßig genutzt wird und dadurch Interesse an und Nutzung von anderen Lernprogrammen steigen).

3.2 Untersuchungsplanung

Die Untersuchungsplanung muss so angelegt sein, dass die Evaluationsfragen beantwortet werden können. Wenn eine summative Wirksamkeitsüberprüfung im Mittelpunkt steht, dann ist der *Vortest-Nachtest-Follow-up-Vergleichsgruppen-Plan* im Sinne hoher Erklärungskraft besonders empfehlenswert (Hager, 2000). Dieser experimentelle Untersuchungsplan sieht vor, dass die Interventionsteilnehmer vor, unmittelbar nach und noch einmal zu einem späteren Zeitpunkt untersucht werden, so dass Erlebens- und Verhaltensänderungen über die Zeit hinweg messbar sind. Durch die parallel untersuchte Vergleichs- bzw. Kontrollgruppe kann zudem überprüft werden, ob die Veränderungen in der Interventionsgruppe auf die Maßnahme zurückzuführen sind, oder sich durch andere Einflüsse ergeben haben.

Experimentelle Evaluationen sind aus organisatorischen Gründen in der Praxis oft nicht durchführbar. Zudem liefern sie keine konkreten Hinweise zur Verbesserung der Maßnahme. Untersuchungspläne für formative Evaluationen sind deswegen meist nicht experimentell und sehen einen Mix verschiedener quantitativer und qualitativer Verfahren vor.

4 Realisationsphase der Evaluation

Auf der Basis der Ergebnisse der Planungsphase werden in der Realisationsphase die für die Evaluationsziele notwendigen Evaluationsdaten erhoben, analysiert und interpretiert. Zudem wird der operative Projektablauf durch ein Projektcontrolling begleitet. Je nach zugrunde liegendem Evaluationsmodell (z. B. zielorientierte und/oder zielfreie Evaluation) sowie in Abhängigkeit von Umfang/Dauer der Intervention (z. B. einmaliger Weiterbildungs-Workshop oder vierjährige Therapieausbildung) und der aktuellen Phase im Interventionslebenszyklus (z. B. Planung eines neuen versus Überprüfung eines etablierten Studiengangs) finden Erhebungs-, Analyse- und Interpretationsschritte einmal, mehrfach nacheinander oder auch parallel statt.

4.1 Erhebung und Analyse von Evaluationsdaten

Evaluationsspezifische Besonderheiten im Vergleich zur Grundlagenforschung ergeben sich im Zusammenhang mit einer an vorher festgelegten Evaluationsfragen/-hypothesen orientierten Datenerhebung und Datenanalyse kaum (für Checklisten siehe Stufflebeam & Shinkfield, 2007, S. 569 ff.). Evaluationsdaten sind dann aussagekräftig, wenn für die verschiedenen Qualitätsdimensionen der Maßnahme (z. B. einer Lehrveranstaltung) jeweils passende empirische Indikatoren/Kriterien gefunden und gemessen werden (z. B. studentische Zufriedenheit mit einer Lehrveranstaltung gemessen auf einer Schulnotenskala). Zudem müssen für die Qua-

litätskriterien auch *Qualitätsstandards* existieren, also Richtwerte, anhand derer zu entscheiden ist, welche Ausprägungen des Kriteriums für schlechte oder gute Qualität sprechen (z. B. Minimalstandard: welche Durchschnittsnote bei einer studentischen Lehrveranstaltungsbewertung muss mindestens erreicht werden, um von guter Lehre zu sprechen; vgl. Döring, 2005).

4.2 Interpretation der Evaluationsbefunde

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 Qualitätsbewertung der Maßnahme. Evaluatoren sollten wertenden Interpretationen nicht aus dem Weg gehen, selbst wenn sie teilweise unpopulär sind oder auf Kritik stoßen (Stufflebeam & Shinkfield, 2007, S. 588). Wenn Evaluatoren nur Einzelergebnisse berichten und die abschließende Gesamtbewertung des Evaluationsgegenstandes dann allein dem Auftraggeber oder den Evaluationsnutzern überlassen, würde dies der Zielsetzung der meisten Evaluationsmodelle widersprechen. Andererseits sehen partizipative Evaluationsmodelle eine aktive Beteiligung aller Interessengruppen auch am Prozess der Ergebnisinterpretation vor.

5 Abschlussphase der Evaluation

Die in der Realisationsphase der Evaluation gewonnenen Erkenntnisse gilt es in der Abschlussphase an unterschiedliche Stakeholder der Intervention zu kommunizieren und für Entscheidungen in der Praxis zu nutzen. Auch eine Qualitätsbewertung der Evaluation in Form einer Meta-Evaluation kann sinnvoll oder notwendig sein.

5.1 Organisation der Evaluationsnutzung

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. Bei einer summativen Evaluation erfolgt die Kommunikation der Ergebnisse ohnehin stärker gebündelt und formalisiert in Form von Präsentationen und Berichten. *Evaluationsberichte* sollen Wissenschafts-Jargon vermeiden und dafür anschaulich und nutzenorientiert gehalten sein.

Zur Förderung der *Evaluationsnutzung* müssen aus dem Bericht die zentralen Botschaften für unterschiedliche Zielgruppen extrahiert und in praxistauglicher Form

zum passenden Zeitpunkt an die jeweiligen Adressaten weitergegeben werden (z. B. Merkblatt, Checkliste, Workshop, Seminar etc.). Es muss ausdrücklich diskutiert und entschieden werden, welche praktischen Konsequenzen sich aus den Evaluationsergebnissen für Einsatz, Nutzung und Gestaltung der Intervention ergeben und wer hier wann und wie aktiv werden soll.

5.2 Durchführung einer Meta-Evaluation

Mit *Meta-Evaluation* ist die Evaluation der Evaluation gemeint, es geht also um eine Qualitätsbewertung einer einzelnen Evaluationsstudie. (Die Meta-Evaluation ist *nicht* zu verwechseln mit einer *Meta-Analyse*.) Eine Meta-Evaluation kann formativ durchgeführt werden, etwa indem externe Evaluationsexperten (typischerweise Fachkollegen) regelmäßig über den Ablauf der Evaluationsstudie informiert werden und korrigierende Hinweise geben. Dagegen bewertet eine summative Meta-Evaluation abschließend die Qualität der Evaluationsstudie. Externe Meta-Evaluationen haben gegenüber einer Selbstbewertung der Evaluatoren (interne Meta-Evaluation) den Vorteil größerer Glaubwürdigkeit.

Eine positive externe Meta-Evaluation ist einerseits ein Expertenachweis für die Evaluatoren und andererseits eine Rückversicherung für den Auftraggeber, dass die Ergebnisse der Studie stichhaltig sind. Eine Meta-Evaluation stützt sich typischerweise auf Dokumentenanalysen (Evaluationskonzept, Zeit- und Kostenplan, Zwischen- und Endberichte etc.) sowie Befragungen der Projektbeteiligten (DeGEval, 2002a).

6 Ausblick

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*, wie das hier beschriebene Fünf-Phasen-Modell, beruhen auf Praxiserfahrungen und sind bislang kaum wissenschaftlich geprüft worden. Es wäre eine Aufgabe der Methodenforschung zu untersuchen, ob und inwieweit Evaluationsprojekte tatsächlich den in der Literatur beschriebenen Abläufen und Checklisten folgen und inwiefern dies mit Projekterfolg zusammenhängt (vgl. Balzer, 2005).

Weiterführende Literatur

- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Gollwitzer, M. & Jäger, R. (2007). *Evaluation. Workbook*. Weinheim: Beltz.

- Hager, W., Patry, J.-L. & Brezing, H. (Hrsg.). (2000). *Evaluation psychologischer Interventionsmaßnahmen*. Bern: Huber.
- Moosbrugger, H. & Schweizer, K. (2002). Evaluationsforschung in der Psychologie. *Zeitschrift für Evaluation*, 1, 19–37.
- Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. San Francisco, CA: Wiley.

Literatur

- Balzer, L. (2005). *Wie werden Evaluationsprojekte erfolgreich?* Landau: Verlag Empirische Pädagogik.
- Birckmayer, J. D. & Weiss, C. H. (2000). Theory-based evaluation in practice. *Evaluation Review*, 24 (4), 407–431.
- Deutsche Gesellschaft für Evaluation (2002a). *Angefragte Gutachten/Meta-Evaluationen auf Basis der DeGEval-Standards*. Verfügbar unter: <http://www.degeval.de/> [18. 6. 2008].
- Deutsche Gesellschaft für Evaluation (2002b). *Standards für Evaluation*. Verfügbar unter: <http://www.degeval.de/> [18. 6. 2008].
- 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.
- Döring, N. (2009). Phasen der Evaluationsforschung. In H. Holling (Hrsg.), *Grundlagen und statistische Methoden der Evaluationsforschung* (Enzyklopädie der Psychologie, Serie Evaluation, Band 1, S. 99–134). Göttingen: Hogrefe.
- Döring, N. & Ingerl, A. (2008). Medienkonzeption. In B. Batinic & M. Appel (Hrsg.), *Medienpsychologie* (S. 403–424). Heidelberg: Springer.
- Gollwitzer, M. & Jäger, R. S. (2007). *Evaluation. Workbook*. Weinheim: Beltz.
- 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.
- Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.
- Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne & M. Scriven (Eds.), *Perspectives on curriculum evaluation* (pp. 39–83). Skokie, IL: Rand McNally.
- Scriven, M. (1991). Prose and cons about goal-free evaluation. *American Journal of Evaluation*, 12 (1), 55–62.
- Scriven, M. (2000). *The logic and methodology of checklists*. Verfügbar unter: http://www.wmich.edu/evalctr/checklists/papers/logic&methodology_oct05.pdf [18. 6. 2008].
- Stahl, E., Bromme, R., Pieschl, S., Hölzenbein, S. & Kiffe, K. (2005). Qualitätssicherung bei der Lernsoftware-Entwicklung durch partizipative, formative Evaluation. Das Projektbeispiel „Cyperaceae-Online“. *Zeitschrift für Psychologie*, 213 (1), 23–33.
- Stufflebeam, D. L. (2007). *CIPP evaluation model checklist*. Verfügbar unter: <http://www.wmich.edu/evalctr/checklists/cippchecklist.pdf> [18. 6. 2008].
- Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: Wiley.

- Uhl, A. (1997). Probleme bei der Evaluation von Präventionsmaßnahmen im Suchtbereich. *Wiener Zeitschrift für Suchtforschung*, 20 (3/4), 93–109.
- Ulrich, S. & Wenzel, F. (2003). *Partizipative Evaluation. Ein Konzept für die politische Bildung*. Gütersloh: Verlag Bertelsmann Stiftung.
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation* (2. Aufl.). Bern: Huber.