

STUDIUM

Ekkehard Nuissl

Evaluation in der Erwachsenenbildung

» STUDIENTEXTE FÜR ERWACHSENENBILDUNG «



Ekkehard Nuisl

Evaluation in der Erwachsenenbildung

Studientexte für Erwachsenenbildung

Eine Buchreihe des Deutschen Instituts für Erwachsenenbildung (DIE)

Bei der gelben Reihe des DIE handelt es sich um didaktisch strukturierte Bestandsaufnahmen zu Kernthemen der Erwachsenenbildung. Die Studientexte vermitteln fachliches Begründungswissen vor dem Hintergrund des wissenschaftlichen Forschungsstands und einer reflektierten Praxis. Sie eignen sich als Begleitmaterial in Fortbildungen und als Ausbildungsliteratur im Studium. Die Studientexte sind als Selbstlernmaterialien konzipiert und ermöglichen Neueinsteiger/innen im Handlungsfeld, erfahrenen Fachkräften und Studierenden die selbständige Erschließung des Themas.

Wissenschaftliche Betreuung der Reihe am DIE: Dr. Thomas Vollmer

Bisher in der Reihe Studientexte für Erwachsenenbildung erschienene Titel (Auswahl):

Karin Dollhausen, Regine Mickler

Kooperationsmanagement in der Weiterbildung

Bielefeld 2012, ISBN 978-3-7639-5060-7

Svenja Möller

Marketing in der Erwachsenenbildung

Bielefeld 2011, ISBN 978-3-7639-4902-1

Horst Siebert

Theorien für die Praxis

3. akt. und überarbeitete Auflage,

Bielefeld 2011, ISBN 978-3-7639-4266-4

Stefanie Hartz, Klaus Meisel

Qualitätsmanagement

3. akt. und überarbeitete Auflage,

Bielefeld 2011, ISBN 978-3-7639-4248-0

Susanne Lattke, Ekkehard Nuissl, Henning Pätzold

Europäische Perspektiven der Erwachsenenbildung

Bielefeld 2010, ISBN 978-3-7639-4248-0

Ekkehard Nuissl

Empirisch forschen in der Weiterbildung

Bielefeld 2010, ISBN 978-3-7639-4246-6

Dieter Gnahs

Kompetenzen – Erwerb, Erfassung, Instrumente

2. akt. und überarbeitete Auflage,

Bielefeld 2010, ISBN 978-3-7639-4244-2

Stefan Hummelsheim

Finanzierung der Weiterbildung in Deutschland

Bielefeld 2010, ISBN 978-3-7639-1976-5

Steffen Kleint

Funktionaler Analphabetismus –

Forschungsperspektiven und Diskurslinien

Bielefeld 2009, ISBN 978-3-7639-1975-8

Mona Pielorz

Personalentwicklung und Mitarbeiterführung

in Weiterbildungseinrichtungen

Bielefeld 2009, ISBN 978-3-7639-1965-9

Wiltrud Gieseke

Bedarfsorientierte Angebotsplanung in der

Erwachsenenbildung

Bielefeld 2008, ISBN 978-3-7639-1955-0

Peter Faulstich, Erik Haberzeth

Recht und Politik

Bielefeld 2007, ISBN 978-3-7639-1949-9

Claudia de Witt, Thomas Czerwionka

Mediendidaktik

Bielefeld 2007, ISBN 978-3-7639-1914-7

Horst Siebert

Lernmotivation und Bildungsbeteiligung

Bielefeld 2006, ISBN 978-3-7639-1931-4

Weitere Informationen zur Reihe unter

www.die-bonn.de/st

Bestellungen unter

wbv.de

Studientexte für Erwachsenenbildung

Ekkehard Nuisl

**Evaluation in der
Erwachsenenbildung**

Herausgebende Institution

Deutsches Institut für Erwachsenenbildung – Leibniz-Zentrum für Lebenslanges Lernen e.V.

Das Deutsche Institut für Erwachsenenbildung (DIE) ist eine Einrichtung der Leibniz-Gemeinschaft und wird von Bund und Ländern gemeinsam gefördert. Das DIE vermittelt zwischen Wissenschaft und Praxis der Erwachsenenbildung und unterstützt sie durch Serviceleistungen.

Lektorat: Dr. Thomas Vollmer

Korrekturat: Manuela Hentschel

Wie gefällt Ihnen diese Veröffentlichung? Wenn Sie möchten, können Sie dem DIE unter www.die-bonn.de ein Feedback zukommen lassen. Geben Sie einfach den Webkey **42/0033** ein. Von Ihrer Einschätzung profitieren künftige Interessent/inn/en.

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.

Verlag:

W. Bertelsmann Verlag GmbH & Co. KG

Postfach 10 06 33

33506 Bielefeld

Telefon: (0521) 9 11 01-11

Telefax: (0521) 9 11 01-19

E-Mail: service@wbv.de

Internet: wbv.de

Bestell-Nr.: 42/0033

© 2013 W. Bertelsmann Verlag GmbH & Co. KG, Bielefeld

Umschlaggestaltung und Satz: Christiane Zay, Potsdam

Herstellung: W. Bertelsmann Verlag, Bielefeld

ISBN 978-3-7639-5266-3 (Print)

ISBN 978-3-7639-5267-0 (E-Book)



Inhalt

| | |
|--|-----------|
| Vorbemerkungen | 7 |
| 1. Evaluation – Was bedeutet sie in der Erwachsenenbildung? | 9 |
| 2. Evaluation – Worum handelt es sich eigentlich dabei? | 14 |
| 2.1 Der Begriff „Evaluation“ und seine Nachbarn | 15 |
| 2.2 Geschichte der Evaluation | 18 |
| 2.3 Ziele, Aufgaben und Prinzipien der Evaluation | 24 |
| 2.4 Theoretische Ansätze der Evaluation | 27 |
| 3. Evaluation – Was wird in der Erwachsenenbildung evaluiert? | 33 |
| 3.1 Lernprozesse und Lernergebnisse | 37 |
| 3.2 Projekte und Programme | 45 |
| 3.3 Bildungseinrichtungen | 49 |
| 3.4 Systeme der Erwachsenenbildung | 51 |
| 4. Evaluation – Wie läuft sie im Bildungsbereich idealtypisch ab? | 56 |
| 4.1 Der Auftrag zur Evaluation | 56 |
| 4.2 Konzept und Planung | 60 |
| 4.3 Realisierung | 65 |
| 4.4 Ergebnisverwertung | 71 |
| 5. Evaluation – Welche Daten werden wie ermittelt? | 75 |
| 5.1 Evaluationsdesign | 76 |
| 5.2 Vorhandene Daten | 79 |
| 5.3 Erhebung neuer/zusätzlicher Daten | 84 |
| 5.4 Datenaufbereitung | 92 |

| | |
|---|------------|
| 6. Evaluation – Wie werden die Daten bewertet? | 95 |
| 6.1 Kriterien der Interpretation und Bewertung | 95 |
| 6.2 Auswertungsfelder und -verfahren | 99 |
| 6.3 Auswertung und Interpretation von Daten | 102 |
| 6.4 Empfehlungen | 106 |
| 7. Evaluation – Verbessern ihre Ergebnisse das Lernen Erwachsener? | 111 |
| 7.1 Zugänglichkeit der Ergebnisse | 112 |
| 7.2 Interessen und Verfahren | 116 |
| 7.3 Nutzung und Wirkung | 119 |
| Glossar | 123 |
| Annotierte Literatur | 128 |
| Literatur | 129 |
| Abbildungen und Tabellen | 132 |
| Autorenportrait | 133 |
| Zusammenfassung/Abstract | 134 |

Vorbemerkungen

Evaluation ist zugleich ein traditionelles und aktuelles Thema der Weiterbildung. In der Traditionslinie gehört die Evaluation schon lange zum Grundbestand professionellen pädagogischen Handelns und ist spätestens mit dem Aufkommen der Curriculum-Debatte in den 1970er Jahren Allgemeingut: Eine Lernzielformulierung und eine daran ausgerichtete Lernerfolgskontrolle sind nichts anderes als Evaluation. Es ging dabei vorrangig um didaktische Reflexion, um Selbstvergewisserung der Lehrenden und Planenden. Evaluation ist in diesem Verständnis eher ein Instrument der Selbst- als der Fremdkontrolle, bar jeden Bedrohungspotentials.

Aktuell hat das Thema neuen Schub durch die Qualitätsdebatte erhalten, bei der die Evaluation ein wichtiges Element im Qualitätskreislauf darstellt. Überdies hat auch das Vordringen einer neuen Steuerungslogik in der Bildungspolitik den Evaluationsgedanken neu akzentuiert: Es geht um die Wirkungskontrolle politischer Maßnahmen, um Effizienz- und Kostendenken, um die Bereitstellung von Steuerungswissen. Die Verwendung staatlicher Mittel wird einer strengen Kontrolle unterzogen, um ihre Allokation möglichst zielgenau vornehmen zu können. In diesen Kontexten sind Evaluationen in gewissem Sinne „Prüfungen“, die mit positiven oder negativen Konsequenzen verbunden sind.

Vor dem Hintergrund der beschriebenen bildungspolitischen Entwicklungen und in Anbetracht des entfalteten Instrumentariums sieht sich die Weiterbildung mit der Frage konfrontiert, ob die Wirkungen ihrer Arbeit nachweisbar sind bzw. sogar gemessen werden können. Von weiten Teilen der Weiterbildungsszene, insbesondere aus den Bereichen der politischen, kulturellen und allgemeinen Bildung, wird eine zunehmend kritische Haltung gegenüber Evaluationsmaßnahmen zum Ausdruck gebracht. Eine bloße Orientierung am Output wird abgelehnt. Es wird betont, dass die Evaluationsforderung oft im Kontext bildungspolitischer Sparanstrengungen steht und somit instrumentalisiert werden kann. Auf der anderen Seite ist der Wunsch vieler Bildungsträger nach einer weiteren empirischen Absicherung der Bildungsarbeit erkennbar, um im politischen Diskurs um knappe Fördermittel gewappnet zu sein, aber auch, um den Stellenwert der eigenen Arbeit anschaulich zu dokumentieren.

Die oben aufgezeigte Skepsis, die Wirkungen von Bildungsprozessen messen zu können, vor allem auch solche Wirkungen, die sich im Praxisfeld erst mit zeitlicher Verzögerung einstellen, ist ein Problem, das den gesamten Weiterbildungsbereich betrifft. Bei allen Bildungsprozessen, selbst bei der beruflichen Bildung, muss damit gerechnet werden, dass ihre Effekte durch intervenierende Variablen überlagert, konterkariert oder verstärkt werden. Kausalitätsunterstellungen sind deshalb immer nur mehr oder weniger plausibel, nie zwingend.

Dennoch braucht wegen dieser methodischen Schwierigkeit nicht auf Wirkungsforschung im weiteren Sinne verzichtet werden. Die Befragung von Lehrenden und

Teilnehmenden liefert erste Aufschlüsse, gibt Informationen über Verunsicherungen, Reflexionen, Verhaltensabsichten. Nachgehende Befragungen der Teilnehmenden können dann Einschätzungen und Annahmen liefern über mögliche Verkettungen von Ereignissen und Wirkungszuschreibungen. Über diesen direkten Zugriff auf ehemalige Teilnehmende hinaus, kann die Analyse von Bildungs- bzw. Lernbiografien Hinweise und Erklärungsmuster liefern. Und in weiten Teilen der Weiterbildung lassen sich sogar Kompetenzmessungen über psychometrische Verfahren verwirklichen, weil dort die Lehr-/Lernziele vergleichsweise gut operationalisiert werden können. Zu denken ist hier etwa an die Bereiche „Fremdsprachen“ oder „EDV“. Alle Methoden haben ihren Stellenwert und können sich gerade in der Zusammenschau gut ergänzen und wichtige Anhaltspunkte für Verbesserungen der makro-, meso- und mikro-didaktischen Gestaltung und auch der Rahmenbedingungen von Lehr-/Lernprozessen liefern.

Der Studientext gibt eine erste Einführung in das Thema. Sein Schwerpunkt liegt indes nicht im Methodisch-Instrumentellen, sondern in der Problemsensibilisierung. Sowohl in der Planungsphase als auch in der Berichtsphase sind viele Feinheiten zu beachten, wie der Autor immer wieder durch Rückgriff auf seine vielfältigen Erfahrungen mit Evaluationsstudien deutlich macht. Das eigentliche „Evaluationsgeschäft“, die datengestützte Analyse von Strukturen und Prozessen in der Weiterbildung mit den daran hängenden methodologischen und methodischen Fragen, wird daher nur angerissen. Zur detaillierten Betrachtung verweist der Autor auf einschlägige Lehrbücher zur empirischen Bildungsforschung.

Der anschaulich, leicht und verständlich geschriebene Text verliert sich nicht in Detailbetrachtungen, sondern lädt die Leserinnen und Leser zur Reflexion ein und natürlich auch zum kritischen Umgang mit Evaluationsansätzen und -ergebnissen. Er liefert somit sowohl für Studierende als auch in der Praxis Tätige wichtige Orientierungshilfen beim Umgang mit Evaluationen und den daraus abgeleiteten Konsequenzen.

Das DIE setzt mit diesem Studientext eine Reihe von Texten fort, die besondere Akzente im Methodischen setzen. Hinzuweisen ist in diesem Zusammenhang auf den Band von Stefanie Hartz und Klaus Meisel zum Qualitätsmanagement, der ein eigenes Kapitel zu „Evaluation als traditionell pädagogischer Zugang zu Qualitätsentwicklung“ beinhaltet, die Bände von Dieter Gnahs („Kompetenzen – Erwerb, Erfassung, Instrumente“) und Strauch/Jütten/Mania („Kompetenzfassung in der Weiterbildung. Instrumente und Methoden situativ anwenden“), die sich mit Fragen der Kompetenzerfassung auseinandersetzen, sowie schließlich der Band „Empirisch forschen in der Weiterbildung“ des Autors des hier eingeleiteten Studientextes. Diese Häufung ist sicher eine Reaktion auf aktuelle wissenschaftliche Trends und Dispute, aber auch auf die entsprechende Nachfrage von Studierenden und Praktikern.

Dieter Gnahs

*Deutsches Institut für Erwachsenenbildung –
Leibniz-Zentrum für Lebenslanges Lernen*

1. Evaluation – Was bedeutet sie in der Erwachsenenbildung?

Das gesamte Bildungssystem ist voll von Evaluationen, immer mehr und immer häufiger wird alles Mögliche evaluiert. Für Schulen und Hochschulen kennt man das schon länger, aber in den letzten beiden Dekaden ist das auch in der Erwachsenenbildung immer selbstverständlicher geworden. Evaluiert werden Angebote, Lehrkräfte, Einrichtungen, Verbände, Gesetze und Systeme der Weiterbildung. Es ist notwendig, sich mit Evaluationen und ihren Ergebnissen systematisch auseinanderzusetzen, um sie einordnen und verstehen zu können – in ihren Reichweiten und ihren Grenzen.

Evaluation ist ein Fremdwort, auch wenn das, was damit gemeint ist, alltäglich und überall stattfindet: Die Vergewisserung, ob und was erreicht worden ist, die Vergewisserung, ob die Dinge so verlaufen, wie man sich das wünscht. Es ist der bewertende Blick auf einen Prozess oder ein Ergebnis. Und dieser Blick ist keineswegs neu. So heißt es in der Schöpfungsgeschichte am Ende des fünften Tages: „Und Gott sah an alles, was er gemacht hatte, und siehe, es war sehr gut“ (Genesis 1, 31).

Eine Evaluation erbringt ein bewertendes Urteil. Wir finden solche bewertenden Urteile praktisch in allen uns betreffenden Angelegenheiten, von den Noten in der Schule über die Arbeitszeugnisse bis hin zu Rückblicken und Erinnerungen, wie sie heute schon in jungen Jahren und in endloser Zahl verfasst werden. Das bewertende Urteil ist uns so sehr eingepflegt, dass es beizeiten schwierig ist, ihm eine eher neutrale und sachliche Darstellung, eine „Beschreibung“, voranzustellen. Viele Gespräche, gerade auch Streitgespräche, im Alltag sind nur Auseinandersetzungen über unterschiedliche bewertende Urteile, weniger Auseinandersetzungen mit der Sache oder solche, die sich auf eine gemeinsam akzeptierte Beschreibung stützen.

Mit dem Begriff „Evaluation“ erhält das bewertende Urteil gewissermaßen seine wissenschaftliche Weihe. Wenn man evaluiert, folgt man einem Regelsystem, dessen Verletzung die Qualität des Werturteils mindert, drastisch mindert, wenn der Verstoß gravierend ist. Das Regelsystem dient dazu, dass auch andere als diejenigen, die evaluieren, nachvollziehen können, wie es zu dem bewertenden Urteil kam. Das Regelsystem ist, wenn man so will, objektiv und nachprüfbar, es folgt den Regeln empirisch-analytischer Sozialwissenschaft.

Zu einem solchen Regelsystem gehören Kriterien darüber, 1. was und 2. wie es beurteilt wird, 3. offengelegte Ziele und 4. Verfahren, die nachweislich eingehalten sind. In der Regel gehören dazu auch Grundprinzipien der Intersubjektivität, vor allem durch eine Balance zwischen unterschiedlichen individuellen Sichtweisen (wie sie etwa im Peer-Review-Verfahren angewandt werden). Dadurch unterscheidet sich Evaluation auch vom alltäglichen Bewerten.

Diese vier Anforderungen an Kriterien einer wissenschaftlich vertretbaren Evaluation sind in der Erwachsenen- und Weiterbildung alles andere als leicht zu erfüllen. Dies liegt zum einen an deren Pluralität, ihrer Vielgestaltigkeit und ihrer Deutungsoffenheit. Zum anderen daran, dass in der Erwachsenenbildung das Prinzip der Freiwilligkeit herrscht, eine Evaluation also, wenn sie akzeptiert und wirksam sein soll, auf Einigung und Partizipation angewiesen ist. Die Einigung auf und die Festlegung von Kriterien ist auch deshalb schwierig, weil in der Erwachsenenbildung von lernerorientierten Konzepten über normative Setzungen bis hin zu bildungspolitischen Prioritäten alles zu finden ist und meist auch offensiv vertreten wird.

Die Schwierigkeiten in der Erwachsenenbildung, sich auf gemeinsame bewertende Kriterien zu einigen, finden sich ebenso dort, wo es um die Ziele der Evaluationen geht. Ziele in der Erwachsenenbildung sind sehr heterogen und folgen gänzlich unterschiedlichen gesellschafts- und lerntheoretischen Prämissen – von der kognitiven Lehre fachlicher Details bis hin zur Aneignung subjektiver und emotionaler Kompetenzen lässt sich alles finden. Hier spielen natürlich die Interessen eine große Rolle, die gegenüber den Zielen von Evaluation formuliert werden und sich in ihnen wiederfinden sollen; gelegentlich werden Ziele auch bereits als vorweggenommene Ergebnisse formuliert (was herauskommen soll ...).

Auch die Gegenstände der Evaluation sind kaum eingrenzbare. „Personen, Umwelt/ Umgebungsfaktoren, Produkte, Techniken/Methoden, Zielvorgaben, Programme, Projekte, Systeme/Strukturen, Forschungsergebnisse“ und die Evaluationen selbst können evaluiert werden – und werden es auch. Auch in der Erwachsenenbildung wird so ziemlich alles evaluiert; fast jede Einrichtung führt Kurs- und Seminarevaluationen durch, vielfach mit „handgestrickten“ Verfahren, um die eigene Arbeit verbessern und den Lernenden die Möglichkeit zum Feed-Back geben zu können. Dies dient auch der Evaluation der Einrichtungen selbst, die zu Förderzwecken weit verbreitet stattfindet: Dort wiederum gilt die Implementation eines Qualitätsmanagement-Systems mit Anwendung von evaluativen Methoden als wichtiges Kriterium. Ebenfalls werden Weiterbildungsgesetze und ihre Förderregelungen (vgl. Nuissl/Schlutz 2001), Bildungsprogramme wie das der „Lernenden Regionen“ sowie die Einführung von Programmen und Maßnahmen – etwa das Modell des Bildungsschecks in Nordrhein-Westfalen oder der Auf- und Ausbau von Bildungsberatungsstellen (vgl. unter Evaluationsgesichtspunkten auch die Hinweise zur „Qualität guter Beratung“ → www.forum-beratung.de) – evaluiert. Und schließlich finden sich in jüngster Zeit Evaluationen zu Bildungssystemen (auch der Erwachsenenbildung) in vergleichender Sicht in den Large-Scale-Studien insbesondere der OECD, zuletzt in der vergleichenden Analyse zu Kompetenzen Erwachsener in verschiedenen Ländern „Programme for the International Assessment of Adult Competencies“ (PIAAC).

Natürlich finden alle diese Evaluationen nicht im luftleeren Raum statt; sie sind eng eingebunden in den wissenschaftlichen und den bildungspolitischen Diskurs zur Weiterbildung. Und sie basieren sehr stark auf den jeweiligen Interessen von Auftraggebern und Evaluatoren. In der Diskussion der Weiterbildung finden sich dabei vor allem folgende Aspekte als Antworten auf die entsprechenden Fragen:

- *Zieldebatte* (Warum soll evaluiert werden?)
Dabei geht es darum, was mit den Evaluationen erreicht werden soll und was unter Evaluationsaspekten überhaupt sinnvoll ist. So wird etwa in der politischen Bildung diskutiert, ob dort Lernergebnisse messbar sind oder überhaupt sein sollen (und damit Gegenstand von Evaluationen sein können).
- *Gegenstandsdebatte* (Was soll evaluiert werden?)
Dabei geht es um die Definition des Objektes der Evaluation, seine Wichtigkeit im Kontext und die Relevanz der Evaluation für diesen Gegenstand, etwa eine Bildungseinrichtung oder einen Kurs.
- *Verlässlichkeitsdebatte* (Welche Daten werden erhoben?)
Dabei geht es nicht nur um die Frage, welche Daten zu erheben sind, um die Werturteile abzugeben, sondern ob sie überhaupt erhoben werden können und das zu evaluierende Objekt angemessen darstellen.
- *Methodendebatte* (Wie müssen die Daten erhoben werden?)
Dabei geht es um die geeigneten methodischen Zugänge in den jeweiligen Feldern (Lernen, Einrichtungen, Systeme etc.), also um die entsprechenden wissenschaftlichen Verfahren und deren Überprüfbarkeit.
- *Nutzendebatte* (Wem und wofür nützt die Evaluation?)
Dabei geht es um die Frage, welche Instanz welchen Nutzen aus den Evaluationsergebnissen ziehen kann – und zieht –, also um das, was letztlich Konsequenz und Perspektive sind. Gerade in der Erwachsenenbildung haben Evaluationen in der Vergangenheit nicht selten Budget-Kürzungen zur Folge gehabt.

Eine weitere Linie der Debatte findet sich in der Wissenschaft der Weiterbildung. Dort geht es besonders um die Frage der *Unabhängigkeit der Forschung* im Evaluationskontext. Diese Frage ist in der Weiterbildungswissenschaft hoch relevant, weil hier viele Forschungen – gerade auch die größeren Evaluationen von Programmen und Projekten – im Auftrag und als Evaluationen stattfinden (→ Kap. 3.2). Das Selbstverständnis der Disziplin ist eng damit verbunden, dass autonome und eher grundlagenorientierte Anteile in solchen Evaluationsstudien enthalten sind. Es ist schwer, Mittel für umfangreichere Forschungsarbeiten zur Weiterbildung in anderen Kontexten zu erhalten.

Bereits vor dreißig Jahren wurde eine Arbeit speziell zur „Evaluation in der Erwachsenenbildung“ vorgelegt (vgl. Gerl/Pehl 1983); sie betonte zum damaligen Zeitpunkt die Evaluation von Lehr-/Lernprozessen mit Blick auf Verbesserung der Lehre

und fokussierte auf die Reflexion pädagogischer Arbeit (so auch Hans Tietgens damals in den Vorbemerkungen zu dem Buch). Heute hat sich die Diskussion, wie auch die Evaluationstätigkeit im Allgemeinen, deutlich erweitert, es geraten deutlich mehr „Gegenstände“ in den Blick.

Der hier vorliegende Studententext will in die weiterbildungspolitische und -wissenschaftliche Debatte zum Thema „Evaluation“ einführen, die wichtigsten Grundlagen und Argumentationsstränge ausbreiten und Sachkenntnisse zum Umgang mit dem Thema „Evaluation“ liefern. Er bindet dabei die spezifische Diskussion der Weiterbildung in die allgemeine Evaluationsdiskussion ein, deren wesentliche Elemente sie spiegelt. Der Aufbau des Studententextes führt dabei über allgemeine (historische und theoretische) Grundlagen (→Kap. 2) und die Auflistung der Evaluationsfelder in der Weiterbildung (→Kap. 3) hin zu den eher methodischen Detailfragen der Datenerhebung, Dateninterpretation und Datenbewertung (→Kap. 4 bis 6). Das abschließende Hauptkapitel befasst sich mit der gerade im Evaluationskontext höchst sensiblen Frage des Umgangs mit den Ergebnissen (→Kap. 7).

Lehr-/Lernziele des Studententextes

Wer diesen Studententext bearbeitet, sollte am Ende folgende Kompetenzen haben:

- Die Ursachen für die Zunahme und weite Verbreitung von Evaluation in weiten Feldern der Gesellschaft können beschrieben werden.
- Die wichtigsten Ansätze für Evaluationen können benannt und gegeneinander abgegrenzt werden.
- Der Bezug von Evaluation zur Qualität und Entwicklung von Lehr-/Lernprozessen Erwachsener kann erläutert werden.
- Die wichtigsten Begriffe und Methoden der Evaluation sind bekannt.
- Die Probleme und Widersprüche im Falle von Auftragsevaluationen können dargestellt werden.
- Die Sensibilitäten im Umgang mit Evaluationsergebnissen sind bekannt und können an Beispielen erläutert werden.

Einführend zu den einzelnen Kapiteln finden sich detailliertere Lehr-/Lernziele, die sich auf das jeweilige konkrete Thema beziehen. Zur besseren Übersicht befinden sich im Anhang ein Glossar mit den wichtigsten Begriffen sowie ein Verzeichnis der verwendeten Literatur mit Lektüreempfehlungen.

Es wird zudem empfohlen, diesen Studententext zusammen mit zwei weiteren Studententexten der gleichen Reihe zu bearbeiten: dem Studententext von Hartz/Meisel zum „Qualitätsmanagement“ (2011, 3. Aufl.) und dem Studententext von E. Nuissl zum „Empirisch Forschen“ (2010a). Dort finden sich Kapitel zur Einbindung von Evaluation in

das institutionelle Qualitätsmanagement bzw. zur Anwendung empirischer Methoden bei der Datenbeschaffung und -interpretation. Insbesondere auf letzteren Punkt, die Erhebung und Auswertung empirischer Daten im Rahmen von Evaluationen, wird in diesem Studientext nicht ausführlich eingegangen, sondern entsprechend auf die genannten Texte verwiesen.

2. Evaluation – Worum handelt es sich eigentlich dabei?

Evaluation ist ein Begriff in Politik und Wissenschaft, der noch keine lange Geschichte hat; das – und vieles mehr – verbindet ihn mit dem Begriff „Innovation“, der noch vor dreißig Jahren in keinem Lexikon enthalten war. Evaluation ist eine Ausformung des Wortstammes „val-“, der in vielen Fremdwörtern enthalten ist und in etwa der Grundbedeutung des Wortes „Wert“ entspricht. Evaluation hat also mit dem Wert zu tun, mit dem Wert eines Objekts, eines Prozesses, einer Innovation.

In der gesellschaftlichen Praxis der vergangenen Dekaden wurde Evaluation immer häufiger realisiert, vielfach verbunden mit innovativen Programmen und Projekten, mit Genehmigungs- und Akkreditierungsverfahren, mit der Systematisierung bereits praktizierter Feed-Back- und Rückkoppelungsverfahren, wie sie in der Erwachsenenbildung schon lange üblich sind. Die Zunahme von Evaluationen wurde und wird immer wieder kritisch gesehen, gerade wissenschaftliche Expertinnen und Experten klagen über das große Ausmaß von Evaluationen, Gutachten, Kommissionen und Gremien, die der Kontrolle, Überprüfung und Bewertung von Gegenständen dienen. Andererseits finden – gerade in der Bundesrepublik Deutschland – viele grundlegende Innovationen und Programme ohne jede Evaluation statt, wie etwa die bundesweite Einführung der konsekutiven BA-/MA-Studiengänge an den Hochschulen, welche das gesamte Studienwesen veränderte.

Im Zuge der Ausbreitung von Evaluationen haben sich naturgemäß auch deren Ansätze, Verfahren und Bewertungen verändert, ausdifferenziert und entwickelt. Unterstützt wird dies durch die Tatsache, dass im Evaluationsbereich auch beträchtliche Fördergelder fließen, für Wissenschaftler/innen und Forscher/innen gerade auch in weniger üppig ausgestatteten Bereichen wie dem der Erwachsenenbildung eine attraktive Ressource.

Dies alles muss bedacht sein, wenn man über Evaluation spricht. Vor allem aber sollte in solchen Gesprächen auch klar sein, was Evaluation überhaupt ist. In diesem Kapitel geht es daher zunächst um den Begriff und seine begrifflichen „Nachbarn“ (→ Kap 2.1), ferner um einen kurzen Überblick zur Geschichte der Evaluation (→ Kap 2.2), überdies um einen ersten Einblick in Ziele und Aufgaben der Evaluation (→ Kap 2.3) sowie um Hinweise zur theoretischen Einbettung der Evaluation (→ Kap 2.4).

Nach der Bearbeitung dieses Kapitels sollten Sie

- Evaluation begrifflich, wissenschaftlich und historisch einordnen,
- Aussagen zu den Zielen und Aufgaben von Evaluation machen und
- zentrale Aspekte der theoretischen Reflexion von Evaluation benennen können.

2.1 Der Begriff „Evaluation“ und seine Nachbarn

Die Gestalt des Wortes „E-valuation“ betont, dass etwas „heraus“ gefunden, gewichtet und bewertet wird. Der Begriff basiert auf dem Lateinischen und kam über das amerikanische Englisch nach Deutschland. Evaluation benötigt Daten aus dem zu evaluierenden Gegenstand heraus bzw. über den zu evaluierenden Gegenstand, die einer Bewertung zugrunde gelegt werden können. Es handelt sich dadurch auch um einen Begriff, der eng an eine empirische und wissenschaftliche Herangehensweise angelehnt ist. In den meisten Handbüchern zur empirischen Forschung der jüngeren Sozialwissenschaft ist der „Evaluation“ ein eigenes Kapitel, zumindest ein Abschnitt gewidmet.

Die in Deutschland gängigste Definition von Evaluation lautet: „Evaluation [ist] ein Instrument zur empirischen Generierung von Wissen (...), das mit einer Bewertung verknüpft wird, um zielgerichtete Entscheidungen zu treffen“ (Stockmann/Meyer 2010, S. 64). Evaluation umfasst also drei Aspekte: Informationen, Bewertungen, Entscheidungen. Speziell für die Erwachsenenbildung lohnt sich ein Blick auf die Definition von Reischmann:

DEFINITION

Evaluation

„[D]er Fachbegriff Evaluation enthält im Wesentlichen drei Elemente, die sich in den meisten Evaluationsdefinitionen finden; diese werden in folgender Definition zusammengefasst: Evaluation meint

- das methodische Erfassen und
- das begründete Bewerten von Prozessen und Ergebnissen zum
- besseren Verstehen und Gestalten einer Praxismaßnahme im Bildungsbereich durch Wirkungskontrolle, Steuerung und Reflexion“ (Reischmann 2006, S. 18).

Evaluation wäre kein weit verbreiteter wissenschaftlicher Bereich, wenn es nicht Definitionen in großer Anzahl gäbe, die jeweils bestimmte Aspekte betonen. Hier sind nur einige von ihnen:

- „Educational Evaluation is the process of delineating, obtaining, and providing useful information for judging decision alternatives“ (Stufflebeam u.a. 1971, S. 43).
- „Evaluation sollen alle jene Handlungen heißen, die dazu dienen, den Grad der Reflexivität von oder in Lernsituationen zu erhöhen“ (Gerl/Pehl 1983, S. 19).
- „Evaluation zielt auf die Verbesserung gesellschaftlicher Praxis, indem sie Programme, Maßnahmenbündel oder Materialien systematisch, d.h. methodisch angeleitet und an Gütekriterien überprüfbar, beschreibt und bewertet“ (Beywl 1988, S. 1).

- „Evaluation is the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision making“ (Mertens 1998, S. 219).
- „Wir plädieren dafür, ‚Evaluation‘ und ‚evaluieren‘ für das wissenschaftlich abgestützte, systematische Beschreiben und Bewerten zu reservieren“ (Beywl/Balzer o.J. S. 5).

Das Sammeln, Sichten, Ordnen und Bewerten von Daten dient naheliegend einem Zweck, etwa dem, den evaluierten Gegenstand zu verbessern oder in einem weitergehenden Kontext zu beurteilen hinsichtlich Nutzen, Wirkung und Folgen. Insofern Evaluation wissenschaftlich anspruchsvoll genannt werden soll, ist es eine angewandte Wissenschaft, die sich – nicht zweckfrei – auf Fragen und Bedarfe der Praxis bezieht. Damit sind auch bereits die beiden Haupttypen der Evaluation benannt: die bewertende Analyse, die im Verlaufe eines Prozesses diesen verbessern soll (*formative* Evaluation) und die abschließende Analyse, die das Ergebnis eines Prozesses festhält und bewertet (*summative* Evaluation). Im Prinzip kann auch die summative Evaluation zum Teil einer formativen Evaluation werden, wenn sich – auf ihr aufbauend und sie verwertend – der evaluierte Gegenstand weiterentwickelt und erneut evaluiert wird.

In der einen wie in der anderen Richtung gibt es jedoch auch andere Ansätze, mit denen eine bewertende Bestandsaufnahme vorgenommen wird. Und Ansätze, die bereits vor Beginn eines Programms oder Projekts oder einer Maßnahme bewertende Aspekte betonen. Dabei handelt es sich um das Monitoring, das Assessment, die Validierung, das Controlling, die Validation und die Akkreditierung.

Monitoring

Der (formativen) Evaluation sehr ähnlich ist das Monitoring, mit dem ein Prozess begleitet und bewertet werden kann. Das Monitoring zielt darauf ab, den Fortgang eines Prozesses zu überprüfen, mögliche Fehlentwicklungen und Verzögerungen zu erkennen und frühzeitig zu beheben. Im Monitoring werden in der Regel „*milestones*“ definiert, an denen überprüft wird, ob der Weg zum erwünschten Ergebnis zeitlich und inhaltlich noch eingehalten ist. Monitoring wird überwiegend in betrieblichen Arbeitsabläufen, aber auch bei der Realisierung von Projekten angewandt.

Assessment

Relativ ähnlich der summativen Evaluation ist das Assessment, am ehesten bekannt aus Verfahren der Personalrekrutierung. Im Assessment werden anhand ausgearbeiteter Verfahren Stärken und Schwächen erfasst und analysierbar gemacht. Es bietet daher Grundlagen für Entscheidungen, etwa für die Einstellung bestimmter Personen in einen gegebenen Arbeitszusammenhang.

Validierung

Einen ähnlichen Zugang wie das Assessment hat die Validierung; dieser Begriff wird meist für die Erfassung und Bewertung von Kompetenzen gebraucht, die nicht bereits durch Zertifikate nachgewiesen werden können. Dabei werden nach systematisierten Verfahren vorhandene Kompetenzen festgestellt und in ihrer Stärke und Bedeutung gewichtet und bewertet. Höchst ausdifferenzierte Verfahren (vgl. Strauch/Jütten/Mania 2009) bestehen hier vor allem im Zusammenhang mit „Bildungspässen“ wie dem „ProfilPASS“ (→ www.profilpass-online.de).

Controlling

Im betrieblichen Kontext findet sich – teilweise hoch differenziert – auch das Controlling, das nicht mit „Kontrolle“ verwechselt oder gleichgesetzt werden darf. Controlling-Konzepte dienen in der Regel dazu, anhand von konkreten Zielvorgaben (etwa Produktziffern, Kosten etc.) die Zielerreichung zu überprüfen. Controlling-Konzepte können prozessorientiert angelegt sein (und ähneln dann dem Monitoring), aber auch zu jeweils kleineren Einheiten das Ergebnis betrachten, sodass sie dann eher Aspekte der summativen Evaluation oder des Assessments anwenden.

Validation

Ein neuer Begriff hat sich, insbesondere in der Förderpolitik der Europäischen Union, mittlerweile durchgesetzt, derjenige der Validation. Auch hier findet sich der lateinische Stamm „val-“, für „Wert“, der Begriff selbst kommt aus dem Französischen. Es geht aber weniger um die Analyse des Wertes, sondern um die Vorsorge, dass ein bestimmter Wert erreicht wird; meist – mit Blick auf Projekte – der Wert der Nachhaltigkeit des entstandenen Produkts. Die Validation ist daher eine zentrale Aufgabe in der Planung und Konzeption von Projekten und Programmen, um die Wertschöpfung der Ergebnisse sicherzustellen.

Akkreditierung

Schließlich ist der Begriff der Akkreditierung zu nennen. Akkreditierung ist zwar nicht die Evaluation selbst, setzt eine solche aber voraus. Akkreditierung umfasst auch die Konsequenz der Evaluation, die Entscheidung. Es ist die (rechtliche und formale) Anerkennung einer Einrichtung oder eines Curriculums, verbunden mit der Berechtigung, bestimmte Dinge auszuführen (z.B. als Anbieter von Seminaren) oder zu implementieren (z.B. Curricula an Hochschulen).

Die unterschiedlichen Begriffe und Ansätze zeigen, dass die Notwendigkeit der Beschreibung und Bewertung, der Wunsch nach Kenntnissen über Abläufe und erreichte Ergebnisse in unterschiedlichsten Kontexten besteht und dafür unterschiedliche, wenn auch sich überschneidende, Begriffe verwendet werden. Nimmt man die große Varianz

hinzu, mit welcher der Begriff der Evaluation selbst gebraucht wird, wird die große gesellschaftliche Relevanz und die enorme Differenziertheit des Gemeintem deutlich, aber auch dessen Unschärfe und Interpretierbarkeit.

2.2 Geschichte der Evaluation

Evaluation ist hervorgegangen aus politischen Kontexten, nicht aus ökonomischen. In der in Deutschland und den europäischen Staaten praktizierten und diskutierten Form geht sie zurück auf US-amerikanische Vorbilder, bei denen die Regierung die Wirksamkeit von Reformprogrammen überprüfen wollte. Besonders herausragend waren dabei Programme im Rahmen des „New Deal“ zur Verringerung der Arbeitslosigkeit und Verbesserung der sozialen Sicherheit.

Der eigentliche Boom der Evaluationsforschung setzte in den USA zu Beginn der 60er Jahre im Gefolge umfangreicher Sozial-, Bildungs-, Gesundheits-, Ernährungs- und Infrastrukturprogramme ein. Von Anfang an waren diese mit dem Auftrag verknüpft, die Wirkungen dieser Programme zu überprüfen (Stockmann/Meyer 2010, S. 24).

Es ist bezeichnend, dass die Anfänge der Evaluation im heutigen Sinne, also einer systematischen Überprüfung der Wirkung und des „Erfolgs“, im Wesentlichen in „weichen“ gesellschaftlichen Bereichen stattfanden wie etwa im Bildungs- und Sozialbereich. Hier sind Ergebnisse am unklarsten, und ihre Feststellung ist am unsichersten. Für die Politik ist dies eine große Herausforderung: Einerseits werden erhebliche Mittel investiert, um Verbesserungen zu erreichen, andererseits sind die Effekte der Investitionen nur schwer messbar.

Im „mikrodidaktischen“ Bereich der Bildung, dem des Lernens und der Lernleistungen, entwickelte sich eine eigene Geschichte. Einen Meilenstein signalisierten dabei die empirischen und experimentellen Forschungen von Kurt Lewin zu Lernleistungen in den 1930er Jahren. Auch in ihnen zeigte sich, dass der „Nutzen“, also die positive Wirkung, von Investitionen in Bildung nur schwer messbar ist. Es ist darüber hinaus schwierig, kausale Zusammenhänge zwischen konkreten Bildungsmaßnahmen und ihren Effekten festzustellen. Von den ursprünglichen Kosten-Nutzen-Analysen wurde daher im Bildungsbereich bald abgegangen.

Nach Guba und Lincoln (1989) lassen sich vier „Generationen“ von Evaluationen (in den Vereinigten Staaten) festhalten:

- Das „*Measurement*“ als erste Generation, bei welcher bestimmte Aspekte des Evaluationsgegenstands gemessen werden, etwa Übergangsquoten zwischen Schultypen oder das Verhältnis von Investition und Bildungsabschlüssen.

- Die „*Description*“ als zweite Generation, in der zusammenfassend die Daten der einzelnen Messungen zu einem beschreibenden Gesamtbild zusammengefügt werden.
- Das „*Judgement*“, also das Urteil als dritte Generation, in der auf der Basis der Beschreibung bewertende Aussagen zum Evaluationsgegenstand gemacht werden.
- Die „*responsive constructivist evaluation*“, in der über einen kommunikativen Zusammenhang die Evaluation einer gemeinsamen Entwicklung dient.

Auch wenn diese Darstellung – wohl zu Recht – als nicht sehr realitätsadäquat und empirisch kaum haltbar kritisiert wird (vgl. Stockmann/Meyer 2010, S. 102ff.), weisen die „Generationen“ doch jeweils auf Elemente von Evaluationen hin, die zu unterschiedlichen Zeiten betont werden. Evaluation ist kontextgebunden, in ihrer Vorgehensweise, ihrem Selbstverständnis und ihren Kriterien jeweils ein Kind ihrer Zeit und ihres Umfeldes.

Der vierten Generation der Evaluation in der Darstellung von Guba und Lincoln folgte eine Dekade der Konzentration der Evaluationen auf die Wirkungen von Investitionen im Bildungsbereich im Sinne von Outputs und direkten Effekten. In der UN-Dekade „Bildung für eine nachhaltige Entwicklung“ (seit 2005)

konzentrieren sich auch Evaluationen vielfach auf die Ergebnisse, Resultate und nachhaltigen Wirkungen eines Programms oder eines bestimmten Lehr-Verhaltens und nicht mehr so sehr auf die Erfassung von Input- und Prozess-Variablen, wie z.B. die Zahl von Lehr-/Lerneinheiten, die Qualität von Lehrmaterialien oder den Grad momentaner Zufriedenheit der Lernenden (Wessler 2010, S. 1033).

Das Verständnis der Wirkung pädagogischer Maßnahmen und Prozesse ist weiter geworden. Man sieht heute mehr die Zusammenhänge und wechselseitigen Beziehungen, also neben „Output“ und „Outcome“ auch den „Impact“.

DEFINITION

Output, Outcome, Impact

Output

ist das direkte Ergebnis einer Maßnahme, Investition oder Handlung: So ist etwa der Zuwachs an Vokabular und grammatikalischen Kenntnissen ein möglicher Output eines Seminars zu einer Fremdsprache.

Outcome

ist das Ergebnis bzw. die Wirkung in einem erweiterten Sinne: So kann sich möglicherweise ein Lernender nach Besuch eines Fremdsprachen-Seminars besser im Zielland der Sprache bewegen.

Impact

ist die Wirkung im Kontext; etwa eine erhöhte Reisetätigkeit in das Zielland des Sprachlern-Seminars oder ein anderes Verhalten im interkulturellen Kontext.

Es ist naheliegend, dass gerade in dieser erweiterten Betrachtungsweise die Anforderungen an die Wissenschaftlichkeit der Konzepte und Methoden der Evaluation steigen. Und dass die Sicht der Evaluation als eine bestimmte Ausprägung empirischer sozialwissenschaftlicher Forschung zunehmend real ist, auch wenn sie sich meist selbst als „wissenschaftliche Dienstleistung“ versteht.

BEISPIEL**Wirkungsevaluation**

In einem der Länder Deutschlands wird ein neues, nachfragegebundenes System der staatlichen Weiterbildungsförderung eingeführt, wonach alle Erwachsenen einen Gutschein für Weiterbildung erhalten, den sie bei Angeboten ihrer Wahl (sofern diese akkreditiert sind) einlösen können. Komplexere Wirkungsevaluationen stellen nicht nur die Frage danach, ob und wie dadurch die Teilnahmequote an Weiterbildung steigt, sondern auch danach, welchen Einfluss das auf die Anbieter-Einrichtungen, die Lehrkräfte und die Lehre hat, wie sich die Qualität der Angebotsplanung entwickelt, ob „Mitnahmeeffekte“ erzielt werden, ob sich die Teilnahmestruktur ändert, ob das Vergabe- und Abrechnungsverfahren funktional ist und ob sich Qualität und Intensität des Diskurses um Weiterbildung erhöhen.

In Deutschland wie in anderen europäischen Staaten haben Tätigkeiten der Evaluation in den vergangenen zwanzig Jahren stark zugenommen, auch wenn noch nicht der Professionalisierungsgrad der USA erreicht ist. In einem Vergleich der „Evaluationenkultur“ spielen nicht nur die Quantitäten der Evaluationen eine Rolle, sondern auch deren Verbreitung in verschiedenen gesellschaftlichen Feldern, deren politischer Stellenwert, deren wissenschaftliche Qualität (nachgewiesen in einem entsprechenden Diskurs) und deren institutionelle Verankerung. In einer Ranking-Liste der Staaten mit der am weitesten entwickelten Evaluierungskultur befand sich Deutschland 2002 auf dem siebenten Platz:

| | I | II | III | IV | V | VI | VII | VIII | IX | Total |
|----------------|----|----|-----|----|----|----|-----|------|----|-------|
| United States | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 18 |
| Canada | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 17 |
| Australia | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 16 |
| Sweden | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 16 |
| Netherlands | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 15 |
| United Kingdom | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 15 |
| Germany | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 13 |
| Denmark | 2 | 2 | 2 | 1 | 1 | 0 | 2 | 1 | 1 | 12 |
| Korea | 1 | 1 | 2 | 2 | 2 | 0 | 2 | 1 | 1 | 12 |
| Norway | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 12 |
| France | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 0 | 11 |
| Finland | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 |
| Israel | 1 | 1 | 1 | 2 | 1 | 0 | 1 | 1 | 1 | 9 |
| Switzerland | 1 | 1 | 2 | 2 | 0 | 0 | 2 | 0 | 0 | 8 |
| New Zealand | 1 | 0 | 1 | 2 | 0 | 0 | 1 | 1 | 1 | 7 |
| Ireland | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 7 |
| Italy | 1 | 1 | 1 | 2 | 0 | 0 | 1 | 1 | 0 | 7 |
| China | 1 | 1 | 0 | 0 | 2 | 0 | 1 | 0 | 1 | 6 |
| Spain | 1 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 5 |
| Zimbabwe | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 4 |
| Japan | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 3 |
| Total | 32 | 26 | 27 | 30 | 26 | 11 | 28 | 23 | 20 | |

I) Evaluation takes place in many policy domains
 II) Supply of domestic evaluators in different disciplines
 III) National discourse concerning evaluation
 IV) Professional organizations
 V) Degree of institutionalization – Government
 VI) Degree of institutionalization – Parliament
 VII) Pluralism of institutions or evaluators performing evaluations within each policy domain
 VIII) Evaluation within the Supreme Audit Institution
 IX) Proportion of outcome evaluations in relation to output and process evaluations

Tabelle 1: Ranking of Countries on the Indicators of an Evaluative Culture (Quelle: Furubo/Rist/Sandahl 2002, S. 10)

Die große Dynamik der Evaluationsaktivitäten in den letzten beiden Jahrzehnten in Deutschland auch im Bildungsbereich lässt sich an mehreren Tatsachen beobachten: So sind etwa Lehrevaluationen an allen Hochschulen implementiert, vermehrt peer-reviewte Zeitschriften auf dem Markt, Programme (europäische wie nationale) werden durchweg evaluiert, Curricula akkreditiert und Forschungseinrichtungen systematisch begutachtet. Vor fünfzehn Jahren wurde die „Deutsche Gesellschaft für Evaluation“ (DeGEval) gegründet, welche 2004 „Standards für Evaluation“ verabschiedete.

Auszug aus den „Standards für Evaluation“ der Deutschen Gesellschaft für Evaluation (DeGEval)

Evaluationen sollen vier grundlegende Eigenschaften aufweisen:

Nützlichkeit – Durchführbarkeit – Fairness – Genauigkeit

Nützlichkeit

Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an den geklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet.

Durchführbarkeit

Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird.

Fairness

Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird.

Genauigkeit

Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt.

Quelle: www.degeval.de/degeval-standards/standards

Die wesentliche Triebkraft der zunehmenden Bedeutung von Evaluation ist im politischen Bereich zu verorten. Der Staat und öffentliche Förderinstitutionen sind vor allem aus drei Gründen an verlässlichen Evaluationen interessiert:

- Zum ersten sind die öffentlichen Mittel knapper geworden, sie müssen, wenn sie weiter wirksam sein sollen, effektiv alloziert und effizient verwendet werden. Ein gutes Beispiel ist hier die Weiterbildung: Die Investitionen in Weiterbildung sind seit vielen Jahren rückläufig (vgl. DIE Trendanalyse 2008), absolut und relativ, es bedarf gezielter und überprüfbarer Wege, die beabsichtigten Wirkungen zu erzielen; entsprechend sind auch Evaluationen von Weiterbildungsgesetzen als Förderinstrumente in Auftrag gegeben worden (vgl. DIE 2011).
- Zum zweiten ist der Legitimationsbedarf für die Verwendung öffentlicher Mittel gestiegen; öffentliche Ausgaben werden zunehmend daraufhin kritisch betrachtet, ob sie auch öffentlichen Zielen dienen. Politik muss sich angesichts der gewachsenen Transparenz aller Vorgänge im politischen Bereich immer häufiger fragen (lassen), ob die Mittel für die „richtigen“ Aufgaben angemessen aufgewendet wurden.
- Zum dritten hat sich die politische Steuerungsstruktur geändert; immer mehr greifen öffentliche und private Sektoren ineinander, übernehmen Netzwerke und nicht-

staatliche Organisationen gesellschaftliche Aufgaben. Moderne „Governance“-Konzepte werden gerade auch im Bildungsbereich wirksam (vgl. Hartz/Schrader 2008). Es ist komplexer geworden, die Wirkung öffentlicher Mittel im gesellschaftlichen Raum zu ermitteln.

Diese relativ enge Bindung von Evaluation an die Vergabe öffentlicher Mittel hat natürlich auch ihre Schattenseiten. Evaluation kann dadurch auch dazu tendieren, dass sie überwiegend nur dort stattfindet, wo öffentliches Geld fließt, weniger dort, wo ein gesellschaftlicher Bedarf an Evaluation bestünde. Es gibt keine Garantie dafür, dass gesellschaftlich wichtige Bereiche auch regelmäßig evaluiert werden.

Die Evaluation in der deutschen Erwachsenenbildung hat eine eigene Tradition, die erst seit etwa drei Jahrzehnten im Kontext der nationalen und internationalen, bereichsübergreifenden Evaluationsdebatte einzuordnen ist. Genau genommen wird die entsprechende Tätigkeit auch erst seit dieser Zeit „Evaluation“ genannt – die Arbeit von Gerl/Pehl (1983) markiert hier den Beginn der Verwendung des Begriffs.

Traditionell hat die Evaluation in der Erwachsenenbildung bis hinein in die 1990er Jahre drei Gegenstandsbereiche im Blick:

1. Die Evaluation von *Lehr-/Lernprozessen* vor allem über zunehmend systematisierte Rückmeldeverfahren der Lernenden (Fragebögen, Abschlüsse etc.); frühe Ansätze dazu finden sich bereits in den Volkshochschulen der Weimarer Republik.
2. Die Evaluation von *Angebots- und Teilnehmerstrukturen* über Teilnahmezahlen und soziodemografische Daten der Teilnehmenden, auch hier liegen schon aus den 1920er Jahren Erhebungen vor.
3. Die Evaluation von *Einrichtungen* hinsichtlich Angebotsumfang, Qualifikation der Leitung und Dauer der Tätigkeit, begonnen auf der Basis der Landesgesetze zur Weiterbildung seit Anfang der 1970er Jahre.

Schon die staatlich geregelten Strukturen des Schul- und Hochschulwesens verfügten über nur wenige empirische Daten etwa zu Abschlüssen, Übergängen und Teilnahmequoten, was Bestandteil dessen war, was Mitte der 1960er Jahre schmerzlich als „Bildungskatastrophe“ empfunden wurde¹ (Dahrendorf 1965; Picht 1964); auch heute ist die Datenlage nicht wirklich zufriedenstellend. Umso schwieriger ist die Situation in der

1 Die apostrophierte „Bildungskatastrophe“ führte in Deutschland nicht nur zur Gründung des „Deutschen Bildungsrates“, der über acht Jahre lang Perspektiven für das Bildungswesen beriet und ausarbeitete, und zur Verfassungsänderung (§91a), wonach Bund und Länder kooperativ im „Bund-Länder-Ausschuss für Bildungsfragen und Forschungsförderung“ (BLK) fast 50 Jahre zusammenarbeiteten, sondern auch zur Gründung von Instituten, einige von ihnen über eine Anschubfinanzierung der Stiftung Volkswagenwerk. Sie hatten nicht nur die Verbesserung einzelner Teile des Bildungssystems zum Ziel, sondern allesamt auch die Verbesserung der Datenlage. Zu diesen gehörten das „Hochschulinformationssystem“ (HIS), die „Gesellschaft zur Förderung quantitativer Modelle des Bildungswesens“ (AfeB) und – für den Schulbereich – das „Institut für die Pädagogik der Naturwissenschaften“ (IPN) an der Universität Kiel (vgl. dazu Edding 1987).

Erwachsenenbildung, die in einer nur vordergründig systematisierten Form zwischen Staat und Markt außerordentlich heterogene Strukturen aufweist. Allerdings: Anders als der staatlich verantwortete Bereich war die Erwachsenenbildung schon immer darauf angewiesen, in ihren fluiden Strukturen Erhebungen durchzuführen und zur Steuerung der Attraktivität ihrer Angebote die Zufriedenheit der Lernenden zu erfassen.

2.3 Ziele, Aufgaben und Prinzipien der Evaluation

Ziele und Aufgaben der Evaluation sind im Grunde schon in ihrer o.g. Definition enthalten. Mit der Evaluation soll ein Gegenstand, ein Objekt, erfasst, analysiert und bewertet werden, um Grundlagen für anstehende Entscheidungen zu erhalten, z.B. ob ein Programm fortgeführt, eine weitere Investition getätigt oder eine Tätigkeit verändert oder anders ausgerichtet werden soll. Evaluation ist also Bestandteil eines konkreten Handlungs- und Entscheidungskontextes und dient, auch wenn sie sich abgesicherter Forschungsmethoden bedient (und bedienen muss), nicht einer „freien“ Erkenntnisgewinnung. Sie ist also, nimmt man die Wissenschaft als Bezugsrahmen, eine Form von Handlungs- oder angewandter Forschung. In einer Gegenüberstellung zwischen Grundlagenforschung und Evaluationsforschung, die Stockmann/Meyer (2010, S. 59) vornehmen, sind es besonders Ziel, Zweck, Ressourcen und Nutzen, welche die beiden unterscheiden:

| Kriterium | Grundlagenforschung | Evaluationsforschung |
|--------------------------|------------------------------|---|
| Erkenntnisziel | theorieorientiert | verwertungsorientiert |
| Zweck | zweckfrei | zweckbestimmt |
| Untersuchungsgegenstand | freie Auswahl | extern definiert |
| Ressourcenbereitstellung | von der Gesellschaft | von Auftraggebern |
| Zeitrahmen | zeitungebunden | zeitgebunden |
| Erkenntnisgewinn | für die Gesellschaft | um Entscheidungen treffen zu können |
| Schlussfolgerungen | Basis für neue Erkenntnisse | positive oder negative Folgen für Stakeholder |
| Nutzen | alle (gesamte Welt) | Auftraggeber, Zielgruppen, generelle: „Stakeholder“ |
| Ergebnisse | Interpretation | Interpretation und Bewertung |
| Kontext | in der Regel unproblematisch | politisch sensibel |

Tabelle 2: Unterschiede zwischen Grundlagen- und Evaluationsforschung (Quelle: Stockmann/Meyer 2010, S. 59)

Neben ihrem Forschungsbezug hat Evaluation eine starke Nähe zur Politik. Sie entstand aus dem politischen Interesse an den Wirkungen von Politik heraus, und sie hat diesen Bezug stets beibehalten. Die Entscheidung, einen Gegenstand für die Evaluation

auszuwählen, die Entscheidung, wer mit welcher Methode die Evaluation durchführen soll, und die Entscheidung, was mit den Ergebnissen der Evaluation geschieht – das alles liegt in den Händen der Auftraggeber, die nach wie vor überwiegend staatliche Institutionen sind. Dabei ist es unerheblich, ob es sich um kommunale Behörden, Landesregierungen, nationale Instanzen oder die transnationale Europäische Union handelt – sie alle verfügen über das Instrument der Evaluation als ein Instrument praktischer Politik.

Allerdings ist der Einfluss des (öffentlichen) Auftraggebers eingeschränkt, sobald die Evaluation einmal begonnen hat. Dies liegt nicht nur an Widerständigkeiten der Evaluatoren, die in Grenzen wissenschaftliche Freiheit in Anspruch nehmen (auch um ihren Ruf zu wahren), sondern gleichermaßen am Interesse der Auftraggeber selbst, eine Akzeptanz des Evaluationsergebnisses nach innen und nach außen zu erhalten. „Gefälligkeitsgutachten“ sind dabei eher schädlich. Man kann daher die Beziehung der Evaluation zur Politik als eine Balance zwischen Abhängigkeit im Auftrag und relativer Unabhängigkeit in der Durchführung verstehen.

Diese Balance ist jedoch vor dem Hintergrund langjähriger Praxis und wechselseitiger Kenntnis zu sehen und zu interpretieren. Dabei spielen nicht nur die Kosten des Angebots (öffentliche Evaluationsaufträge werden fast immer im Wettbewerb vergeben) eine Rolle, sondern auch das Renommee des Anbieters, die Qualität des Angebots und die gemeinsamen Vorerfahrungen, etwa zu den Positionen des Anbieters in politischen Fragen oder seinen früheren Evaluationen hinsichtlich Linientreue und Verwendbarkeit. Gerade im Bildungsbereich, wo Forscher einwerbbarer Drittmittel überwiegend für Evaluationsaufträge erhalten, spielen diese Zusammenhänge eine große Rolle.

In der Verwertung der Ergebnisse allerdings besteht – ohne sensible Balance – eine völlige Abhängigkeit der Evaluatoren vom Auftraggeber, er ist gewissermaßen der „Besitzer“ des Evaluationsoutputs. Er hat ihn auch „bezahlt“.

Im funktionenbezogenen Ansatz von Stockmann und Meyer hat die Evaluation hauptsächlich vier Funktionen (Stockmann/Meyer 2010, S. 145ff.):

1. *Erkenntnisgewinn* über den evaluierten Gegenstand durch zielgerichtete Datensammlung und Analyse,
2. *Entwicklung* des evaluierten Gegenstands anhand der ermittelten Daten und Erkenntnisse,
3. *Kontrolle* über die effiziente und effektive Vorgehensweise des evaluierten Gegenstands, sowie
4. *Legitimation* der Arbeit und der Investition hinsichtlich des behandelten Gegenstands, wie z.B. einer Einrichtung oder eines Lehrgangs.

In der Realität finden Evaluationen meist mit dem Schwerpunkt in der einen oder anderen Funktion statt. So betonen bspw. Akkreditierungen (von Curricula oder Einrich-

tungen) eher die Kontrollfunktion, während Lernprozess-Evaluierungen eher der Entwicklung dienen.

Letztlich dient die Evaluation – ob in Kontrolle, Entwicklung, Erkenntnis oder Legitimation – dazu, die *Qualität* des evaluierten Gegenstands sicherzustellen oder zu verbessern. Qualität ist ursprünglich ein neutraler Begriff, der eigentlich nur die genaue Beschaffenheit eines Gegenstands bezeichnet. Im Zuge der Qualitätsdebatten hat sich der Begriff allerdings immer mehr mit seiner positiven Konnotation verbunden, steht also implizit für „gute“ Qualität. Wie bei der Debatte um die Kriterien der Evaluation (→ Kap. 6) gibt es auch unterschiedliche Sichtweisen auf das, was „gute Qualität“ ist; in der Regel richtet sich das Verständnis danach, was erreicht werden soll, ob der beschriebene Gegenstand effektiv („Tue ich das Richtige?“) und effizient („Tue ich das Richtige richtig?“) ist. Gerade in der Weiterbildung hat die Qualitätsdebatte Anfang der 1990er Jahre einige Wellen geschlagen, mit weitreichenden Folgen für Einrichtungen, Angebote, Lehrpersonal und Lernende (vgl. Hartz/Meisel 2011, S. 24ff.).

Ein zentraler Aspekt von Evaluation und eng verbunden mit Zielen und Aufgaben sind die Prinzipien, unter denen sie stattfinden soll und darf. Hier sind es vor allem die Assoziationen der Evaluation, die – Berufsverbänden vergleichbar – auf ethische und professionelle Bedingungen hinweisen, die einzuhalten sind. Besonders betont hat dies die „American Evaluation Association“ (www.eval.org), aus gutem Grund: Evaluation ist nicht nur ein Zugang, der nach US-amerikanischem Vorbild überall in der Welt Eingang gefunden hat, dort fließt auch das meiste Geld in Evaluationen mit – folgerichtig – großem Risiko für die Solidität und Professionalität des Bereichs. Die Association hat fünf Prinzipien formuliert, denen man auch entnehmen kann, dass mit ihnen konkrete Probleme eingedämmt werden sollen. Die fünf Prinzipien sind die folgenden:

1. *Systematic inquiry*

Hier geht es nicht nur um die Methodenwahl, sondern auch um die Abstimmung mit dem Feld und die Kommunikation über Vorgehensweise und Ergebnisse.

2. *Competence*

Hier geht es nicht nur darum, auf die notwendige Kompetenz der Evaluatoren und ihres Teams hinzuweisen, sondern auch einzufordern, dass ohne ausreichend vorliegende Kompetenz Evaluationen nicht durchgeführt werden sollen. Auch die Notwendigkeit der Fortbildung wird betont.

3. *Integrity and honesty*

Hier geht es darum, explizit Interessen zu bekunden, ehrenhaft zu verhandeln und Interessenkonflikte zu vermeiden.

4. *Respect for the people*

Dieses Prinzip stellt die Würde der Menschen und den Personenschutz in den Vordergrund.

5. *Responsibility for general and public welfare*

Dieses Prinzip verweist auf die immer gegebene Verantwortlichkeit der Evaluatoren gegenüber der Gesellschaft, wer immer auch ihnen den Evaluationsauftrag erteilt hat.

Der ethische Duktus der Prinzipien unterstreicht die Schwierigkeiten, in einem Feld – in dem faktisch Sonden in Interna gesenkt und innere Vorgänge und Abläufe transparent und bewertet werden, andererseits aber auch einiges an finanziellen Mitteln eingesetzt und aufgebracht wird –, die professionelle und qualitätsvolle Linie einzuhalten, die für Seriosität bürgt.

2.4 Theoretische Ansätze der Evaluation

Über Evaluation und Evaluierung wird – insbesondere im US-amerikanischen Raum – enorm viel geschrieben, diskutiert und systematisiert. Es mangelt nicht an Entwürfen, die für sich in Anspruch nehmen, Theorien der Evaluation zu entwickeln. Hier ähnelt Evaluation dem „Management“, das in immer wieder anderen Varianten, die zur Vermarktung geeignet sind, theoretisch und begrifflich neu modelliert wird. Es ist nicht zu übersehen, dass es einen großen „Evaluations-Markt“ gibt, in dem auch das Vorzeigen eigener konzeptioneller Entwürfe bei der Akquisition von Aufträgen bedeutsam ist.

Es zeigt sich aber auch, dass Evaluation, sofern sie wissenschaftlich betrieben wird, im Wesentlichen von ihrer Anwendung lebt. Theoretische Ansätze zur Evaluation ordnen bislang eher Zweckzusammenhänge und Verfahren, als dass sie grundlagenorientierte Fragen beantworten, wie etwa die nach dem Zustandekommen von Werturteilen.² Auch beziehen sie sich sehr aufeinander und verstehen die historische Abfolge unterschiedlicher Ansätze als Teil der eigenen Identität, wie dies die Guba-Lincolnschen Generationen tun (→ Kap. 2.3).

Genau dies ist auch ein wichtiger Punkt der Kritik von Stufflebeam und Shinkfield (2007) an der theoretischen Diskussion zur Evaluation; sie fordern wissenschaftliche Theorien, die sich von der bis dato dominierenden Modelldebatte absetzen. Für theoretisch gehaltvollere Evaluationsansätze fordern sie eine übergreifende Kohärenz, ein Kernkonzept, überprüfte Hypothesen dazu, wie Evaluationen erwünschte Ergebnisse haben, bearbeitbare Verfahren, ethische Prinzipien und ein übergreifendes Verständnis der Forschung im Rahmen von Evaluationen.

2 Eine Ausnahme in der theoretischen Diskussion zur Evaluation bietet hier Michael Scriven; ihm zufolge ist Evaluation „the process of determining the merit, worth and value of things, and evaluations are the product of that process“ (Scriven 2002, S. 1). Oder, mit anderen Worten: „Evaluation is the judgement of value, worth and impact“ (Scriven 1974, S. 4), entsprechend macht er grundsätzliche Ausführungen zum Entstehen von Werturteilen.

Aber auch in der an Modellen orientierten Diskussion existieren einige Eckpunkte, die allgemein benannt werden können. Der wichtigste von ihnen, der Evaluationen in zwei unterschiedliche „Stränge“ teilt, ist das Paradigma des Wissenschaftsverständnisses und der darin liegenden Normativität. Es lassen sich hier zwei unterschiedliche Hauptlinien finden: Eine Richtung betrachtet Evaluation als ein Verfahren, das der kritisch-rationalen Forschungslogik folgt und insofern einen gewissermaßen „objektiven“ Gegenstand antizipiert (vgl. Stockmann/Meyer 2010, S. 83). Die andere Richtung ist eher der konstruktivistischen Sichtweise verpflichtet und interpretiert Realität als soziale Konstruktion; Evaluationen dieser Richtung setzen eher auf die „Viabilität“ (Passung) des Evaluierten im Kontext und konzentrieren sich auf diskursive und kommunikative Evaluationsverfahren. In Bezug auf Lehr-/Lernprozesse der Erwachsenenbildung hat die letztere Richtung sicher einige Vorzüge, da sie eher der Mehrdeutigkeit des Evaluierten entgegenkommt. Beide Ansätze unterscheiden sich jedoch nicht durch die Anwendung empirisch-analytischer Forschungsverfahren quantitativen oder qualitativen Zuschnitts.

DEFINITION

Viabilität

Der Begriff „Viabilität“ spielt in der Konstruktivismus-Debatte eine bedeutende Rolle. Er benennt letztlich die treibende Kraft, die hinter der jeweiligen Wirklichkeitskonstruktion der Individuen steht: Das Handeln muss ein Überleben in einer bestimmten Umwelt ermöglichen, einen „Weg“ (via) aufzeigen; Handlungen und Deutungen dürfen keine Hindernisse im Durchleben der Welt erzeugen.

Die theorieorientierte Modellierung von Ansätzen hat lange Zeit eine Theoriebildung zur Evaluation substituiert, insofern steht das „Generationenmodell“ von Guba und Lincoln, eine der ersten Systematisierungen der noch kurzen Theoriegeschichte der Evaluationswissenschaft (s.o.), stellvertretend für solche Ansätze. Ihr theoretischer Fokus ist der jeweilige Arbeits- und Aufmerksamkeitsschwerpunkt im Evaluationsprozess; am deutlichsten wird dies in ihrer vierten, „eigenen“, Generation, die sie als Abfolge von Evaluationselementen darstellen (→ Abbildung 1).

Das Modell enthält eine starke Betonung der Positionierung der Evaluation in das interessierte Umfeld, also Vertragsabschluss, „Stakeholder“, Bericht, Verhandlung und Umsetzung. Die „eigentlichen“ wissenschaftlichen Evaluationsarbeiten wie Ziel- und Fragestellung, Erhebungsplan („Organizing“), Kriterien, Prioritäten usw. sind hier in wenigen Schritten zusammengefasst. In das Modell gingen die Erfahrungen darüber ein, dass zwischen dem Ermitteln von Sachverhalten und ihrer Bewertung sowie auch und den Konsequenzen, die sich daraus ergeben und gezogen werden, oft eine große Lücke klafft.

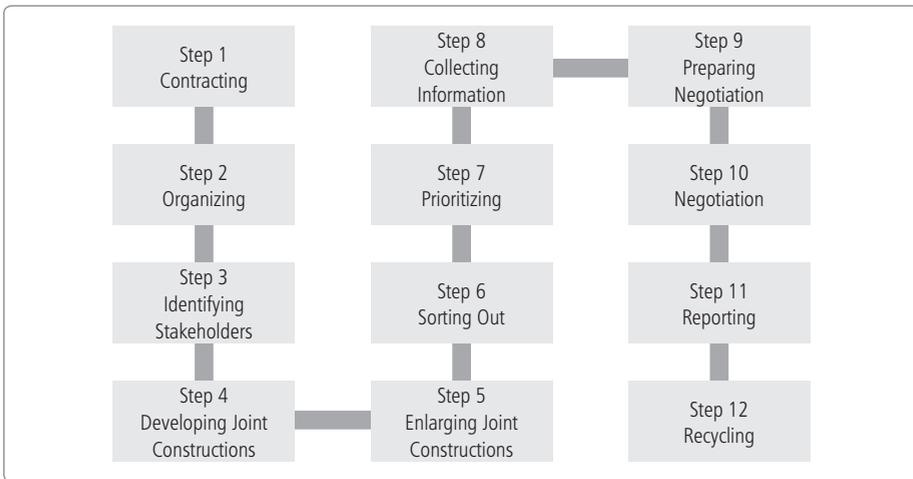


Abbildung 1: Ablaufplan Evaluation der 4. Generation (Quelle: verkürzte Abb. von Stockmann/Meyer 2010, S. 111 nach Guba/Lincoln 1989, S. 186f.)

Eine andere, enger auf die deutsche Evaluationsforschung bezogene Abfolge haben Steinert/Klieme skizziert, indem sie acht Schritte der Evaluation in ihrer Abfolge benennen (→ Abbildung 2):

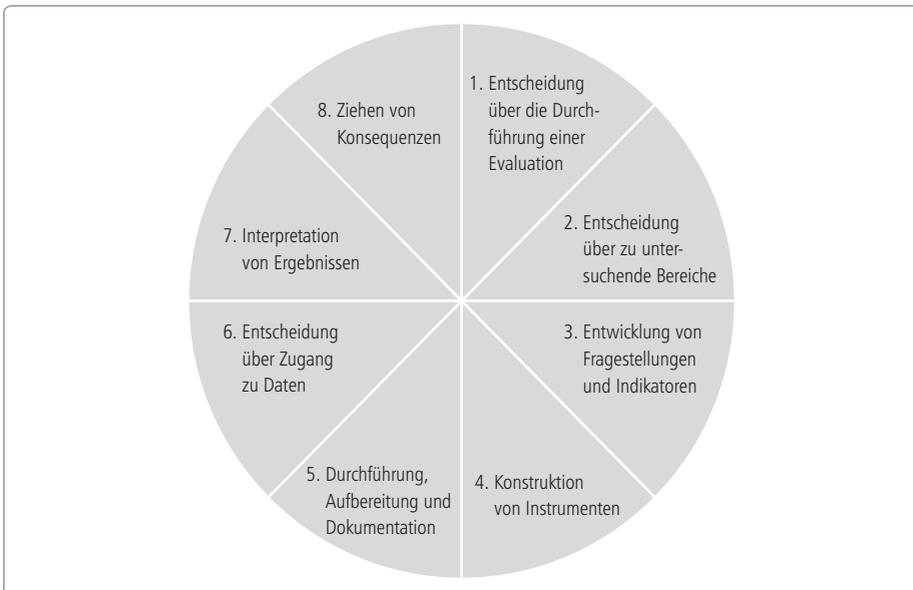


Abbildung 2: Acht Schritte der Evaluation (nach Steinert/Klieme 2008, S. 646)

Während Guba und Lincoln als wesentliches Kriterium die jeweilige Arbeit im Evaluationsverfahren für ihre Klassifikation zugrundelegen, betonen Fitzpatrick, Sanders und Worthen (2004) das Kriterium der „driving force“ hinter der Evaluation; in diesem Sinne strukturieren auch sie eher das Umfeld der Evaluation als deren wissenschaftliche Qualität selbst. Sie unterscheiden nach Stockmann/Meyer 2010, S. 126:

1. *„Zielorientierte Ansätze*, bei denen der Fokus auf der Identifizierung von Programmzielen und einer Beurteilung liegt.
2. *Managementorientierte Ansätze*, deren Schwerpunkt die Identifizierung und Erfüllung der Informationsbedürfnisse von Entscheidungsträgern ausmacht.
3. *Konsumentenorientierte Ansätze*, deren Hauptaufgabe darin besteht, produktbezogene Informationen und Bewertungen z.B. anhand von Produkt-Checklisten zur Verfügung zu stellen.
4. *Expertenorientierte Ansätze*, bei denen eine Evaluation durch ausgesuchte Experten eines Praxisfeldes erfolgt.
5. *Partizipative Ansätze*, die durch eine besondere Schwerpunktsetzung auf die Einbeziehung der verschiedenen an einer Evaluation beteiligten oder davon betroffenen Interessengruppen (Stakeholder) bei der Evaluationsplanung und -durchführung charakterisiert sind“.

Ein anderer Klassifizierungsansatz findet sich bei Beywl (1988, S. 45), der sich insbesondere auf die Arbeiten von Stufflebeam bezieht; er unterscheidet zwischen Evaluationsmodellen, die

- *effizienzgesteuert* vorgehen, also einen (überwiegend quantitativen) Input-Output-Vergleich vornehmen;
- *zielgesteuert* vorgehen, in dem die Effekte des evaluierten Gegenstands mit den ursprünglich formulierten Zielen abgeglichen werden;
- *richtliniengesteuert* vorgehen, in dem bestehende Vorgaben (etwa Akkreditierungs-Richtlinien) überprüft werden, um Genehmigungen zu erteilen;
- *gegnerschaftgesteuert* vorgehen, in dem zwei „Anwälte“ gegensätzliche bewertende Positionen gegenüber eine Jury zur Entscheidung vortragen;
- *kennerschaftgesteuert* vorgehen, in dem die Expertise des Evaluators oder der Evaluatoren die Kriterien für das Werturteil bilden;
- *entscheidungsgesteuert* vorgehen, indem die entscheidungsrelevanten Daten ermittelt und bewertet werden;
- *konsumentengesteuert* vorgehen, indem die Urteilkriterien aus den (bekannten oder vermuteten) Bedürfnissen der Klienten abgeleitet werden;
- *interessengesteuert* vorgehen, indem die Interessen der Beteiligten abgewogen und in ein Bewertungsmodell transformiert werden.

Für die Erwachsenenbildung, in der Evaluation traditionell eher ein Ansatz zur Reflexion, Interpretation und Sensibilisierung (vgl. Tietgens u.a. 1986) war und vielfach noch ist, bietet sich eher die „Funktionstheorie“ von Stockmann und Meyer an, wie sie auch den Standards der Deutschen Gesellschaft für Evaluation (vgl. DeGEval 2002) zugrunde liegt. Dabei stehen „Erkenntnis“ und „Entwicklung“ als gleichberechtigte Evaluationsfunktionen neben „Kontrolle“ und „Legitimation“ (Stockmann/Meyer 2010, S. 144ff.). Diese vier Funktionen lassen sich – je nach konkreter Evaluation in der Weiterbildung – in die unterschiedlichsten Schwerpunkte integrieren.

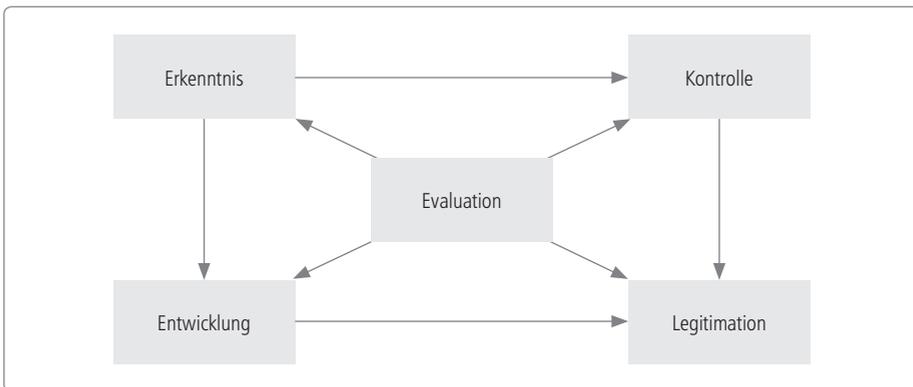


Abbildung 3: Leitfunktionen von Evaluation (Quelle: Stockmann/Meyer 2010, S. 73)

So betonen Evaluationen von Einrichtungen zur Durchführung von Bildungsurlaub oder Migrant/inn/enkursen etwa die Kontrollfunktion, während Seminarevaluationen durch die Lernenden den Aspekt der Erkenntnis und Entwicklung in den Vordergrund stellen. Das bestätigt die Auffassung aller, die jemals mit Evaluation zu tun hatten: Jede Evaluation ist ohnehin einzeln zu konzipieren und auszuhandeln.

ZUR REFLEXION

- Was sind aus Ihrer Sicht die wichtigsten Ansatzpunkte für Evaluationen in der Erwachsenenbildung?
- Welche Vorteile sehen Sie in der Evaluation, verglichen etwa mit Monitoring, Assessment und Validierung?
- Wie realistisch sind in der Erwachsenenbildung Ansätze der Kontrolle und der Zielerreichung, besonders bei Lernprozessen?
- Welche Interessen und Interessenvertreter sehen Sie in der Erwachsenenbildung, die Evaluationen vorantreiben und initiieren?

**Lektüreempfehlungen**

- Gerl, H./Pehl, K. (1983): Evaluation in der Erwachsenenbildung. Bad Heilbrunn
- Stockmann, R./Meyer, W. (2010): Evaluation. Eine Einführung. Opladen
- Wesseler, M. (2010): Evaluation und Evaluationsforschung. In: Tippelt, R./Hippel, A. v. (Hg.): Handbuch Erwachsenenbildung/Weiterbildung. 4. Aufl. Wiesbaden, S. 1031–1048

3. Evaluation – Was wird in der Erwachsenenbildung evaluiert?

Im Grunde kann alles und zu jeder Zeit evaluiert werden, das gilt auch für die Erwachsenenbildung. Und in den vergangenen Dekaden haben vielerorts auch immer mehr Evaluationen stattgefunden. Nicht selten wurden sie als lästig, als überflüssige und zusätzliche Arbeit empfunden; Begriffe wie „Evaluitis“, „Inquisition“ und „Ritualisierung“ machten die Runde (vgl. Hornbostel 2008).

Heute hat sich jedoch eher eine positive Sicht durchgesetzt, der Nutzen von Evaluationen für die Entwicklung der evaluierten Gegenstände wird überwiegend anerkannt. Das gilt sowohl für Evaluationen im Kontext von Akkreditierungen (z.B. bei Hochschulen oder Erwachsenenbildungseinrichtungen als Bildungsurlaubs-Anbieter), im Kontext von Institutionen (z.B. bei Forschungsinstituten oder Erwachsenenbildungseinrichtungen) sowie im Kontext von Produkten (z.B. bei Artikeln in peer-reviewten Zeitschriften oder bei Seminaren und Kursen einer Volkshochschule). Auch die inzwischen verpflichtenden Evaluationsauflagen bei Projekten und Programmen werden als nützlich akzeptiert, auch wenn immer wieder der erforderliche hohe Aufwand beklagt wird.

Die Nützlichkeit von Evaluationen wird seitens der „Evaluationsgegenstände“, also der evaluierten Personen, Einrichtungen und Organisationen, vor allem in fünf Punkten gesehen:

- Die Evaluation zwingt dazu, sich zur Effektivität, Effizienz und Qualität des eigenen Tuns Gedanken zu machen, Fehler und Probleme zu identifizieren und Lösungswege zu erörtern, noch bevor die Evaluation selbst eigentlich stattfindet. Hilfreich dabei ist vor allem eine Liste der Evaluationskriterien, die abgearbeitet werden kann.
- Die Evaluation ermöglicht Veränderungs- und Verbesserungsschritte, die ohne Evaluation möglicherweise gar nicht oder erst sehr viel später stattgefunden hätten.
- Die Evaluation erhöht – gerade bei Einrichtungen und Programmen – das Bewusstsein der Beteiligten von der Aufgabe, den Zielen und intensiviert den Dialog.
- Die Evaluation verbessert die Transparenz des Gegenstands, zumindest in denjenigen Bereichen, die durch die Evaluationskriterien erfasst sind.
- Positive, aber nicht selten auch negative, Evaluationsergebnisse lassen sich werbewirksam einsetzen und auf dem Markt und im politischen Raum verwerten.

Ein sechster Nutzen von Evaluationen kann – gerade bei Einrichtungen und Systemen – aus der Sicht des Managements darin liegen, dass eine beabsichtigte Änderung oder Entscheidung durch die Evaluationsergebnisse unterstützt und so leichter durch- und umgesetzt werden kann.

Allerdings sind auch heute noch die Vorbehalte beträchtlich. Sie liegen vor allem darin, dass

- der Aufwand zur Vorbereitung und Durchführung der Evaluation für die vorhandenen Kapazitäten meist eine große Herausforderung darstellt,
- die Vorbereitung nicht punktgenau die Kriterien trifft und die Beschreibung der Realität misslingt,
- das Verfahren fehlerhaft abläuft und im Endeffekt irreversibel ist („Nachevaluatio-nen“ sind die absolute Ausnahme!),
- Empfehlungen gegeben werden, die dem eigenen Interesse und den eigenen Mög-lichkeiten nicht gerecht werden, und
- die Verwertung der Evaluationsergebnisse und -empfehlungen immer Risiken ent-hält – sie können sich leicht verselbständigen.

Es gibt eigentlich nur drei Aspekte, welche das Durchführen von Evaluationen ein-schränken:

Anlass

Zum ersten ist ein Anlass notwendig, verbunden mit einer Person oder Instanz, die aus einem Grund heraus eine Evaluation fordert oder in Auftrag gibt. So sind etwa alle Eva-luationen von Landesgesetzen zur Weiterbildung auf der Basis von Landtagsbeschlüssen entstanden, denen bestimmte Ziele zugrunde lagen. Ein Beispiel ist die geplante Absicht der Landesregierung, die Mittel für einen Bereich – etwa die Erwachsenenbildung – zu kürzen. Die Evaluation wird dann hauptsächlich aus dem Grund in Auftrag gegeben, Streichungsmöglichkeiten zu ermitteln und die Konsequenzen von Streichungen zu erwägen (vgl. Nuissl/Schlutz 2001). Der Anlass zur Evaluation ist immer auch verbunden mit einem Interesse, dessen Richtung und Qualität es zu analysieren gilt. Am genann-ten Beispiel bedeutet dies: Gerade in der Erwachsenenbildung ist aufgrund historischer Erfahrungen bei einer Evaluation die Befürchtung groß, dass sie letztlich Vorbereitung und Legitimation einer Mittelkürzung ist.

Mittel

Zum zweiten sind dies die erforderlichen Mittel für die Evaluation. Evaluationen binden, wenn sie gut gemacht werden sollen, erhebliche Ressourcen. Und das nicht nur bei den Auftraggebern, welche die Personen und Prozesse der Evaluation finan-zieren müssen, sondern auch bei den evaluierten Einrichtungen und Organisationen, welche die entsprechenden Daten sammeln und bereitstellen müssen oder sollen. Die Entscheidung für die Durchführung einer Evaluation ist daher auch immer eine Entscheidung über Investitionen. Die Kosten für Evaluationen stehen nicht selten im (scheinbaren) Widerspruch zum Anlass, also der Reduktion von Mitteln; es handelt

sich dann um eine Investition, die sich durch zu erwartende Einsparungen „amortisieren“ muss.

Akzeptanz

Zum dritten müssen die zu evaluierenden „Gegenstände“, also Einrichtungen, Lehrpersonen etc., ihrerseits die Evaluation akzeptieren. Dies kann einerseits dann erfolgen, wenn das Evaluationsobjekt in gewisser Weise vom Auftraggeber der Evaluation abhängig ist – etwa finanziell (der Hauptgeldgeber) oder instrumentell (Vergeber des bearbeiteten Projekts mit Evaluationsaufgabe), organisatorisch (bei einer untergeordneten oder nachgeordneten Organisation) oder rechtlich (es besteht per Gesetz – z.B. in Akkreditierungsfällen – die Evaluationsverpflichtung). Es kann andererseits aber auch dadurch erfolgen, dass der Evaluationsgegenstand mit der Evaluation einverstanden ist, sie vielleicht selbst wünscht und konstruktiv mitwirkt. Insbesondere letzteres ist generell wünschenswert, vor allem dann, wenn die Ermittlung vieler Daten erforderlich ist. Es ist daher auch dann Werbung und Überzeugung für die Evaluation angebracht, wenn durch die rechtliche oder ökonomische Dominanz des Auftraggebers ein freies Spiel der Kräfte gar nicht nötig wäre.

Im Bereich des Lehrens und Lernens von Erwachsenen wird heute auf die eine oder andere Weise so ziemlich alles evaluiert, was sich als „Evaluationsgegenstand“ definieren lässt. Das ist vielfach schwierig, weil sich die Gegenstände (Kurse, Lehrende, Einrichtungen etc.) in den unterschiedlichen gesellschaftlichen Bereichen anders zueinander ordnen – einmal sind Einrichtungen auch Träger, Lehrende auch Manager, Finanzierer auch Anbieter.

Wichtig ist, dass jeder Evaluationsgegenstand, wenn er definiert wird, ein Ziel oder eine klare Aufgabe erkennen lassen muss. Ohne die Ziele des Gegenstands zu kennen, ist Evaluation praktisch nicht möglich.

DEFINITION

Ziele des Evaluationsgegenstands

Ein Seminar in der Erwachsenenbildung verfolgt das Ziel, das Lernen der Teilnehmenden zu befördern und zu unterstützen, auch Lehrende leisten hier Dienstleistungen. Eine Weiterbildungseinrichtung will gute Angebote erstellen und möglichst vielen Menschen die Teilnahme ermöglichen. Ein Programm oder Projekt will eine Innovation erproben oder implementieren, in einem System werden soziale, politische, kulturelle und ökonomische Ziele in Bezug auf das Lernen von Erwachsenen verfolgt. Letztlich bezieht sich der Evaluationsauftrag immer auf die Ziele des Evaluationsgegenstands und deren Erreichung.

Nun ist der Bildungsbereich immer voller Ziele, und sie sind nicht immer miteinander vereinbar oder überhaupt zu vereinbaren. Nicht selten liegen auch in unterschiedlichen Teilen des Untersuchungsgegenstands unterschiedliche Ziele vor, so etwa in Programm-

bereichen einer Volkshochschule andere als in der Gesamteinrichtung oder in Lehr-/Lernprozessen bei den Lehrenden andere als bei den Lernenden (was übrigens die Normalität ist). Der Zuschnitt des Evaluationsgegenstands ist daher auch für das Gelingen der Evaluation von großer Bedeutung. Hier sind also – schon im Auftrag – evaluationsrelevante Entscheidungen zu treffen.

Nicht selten werden Evaluationen analog zu den Gegenstandsbereichen in der „Mehrebenenstruktur des Bildungssystems“ (Kuper 2005, S. 135) gruppiert: nach den Leistungen von Personen (Schülern und Lehrern), nach den Settings pädagogischer Intervention (Unterricht), nach den Institutionen (Schulen, Erwachsenenbildungseinrichtungen), nach den Institutions- und Organisationsformen (Schulform), nach den Territorien und Regionen sowie nach nationalen und internationalen Systemen. Evaluationen finden hier überall statt, nach unterschiedlichen Ansätzen und mit unterschiedlichsten Zielen.

In Bezug auf die Weiterbildung als Teilbereich des Bildungswesens liegt es nahe, sich auf bestehende Gliederungsverfahren zu besinnen. Ein solches liegt insbesondere vor im „Forschungsmemorandum Weiterbildung“ (Arnold u.a. 2000), in dem fünf Bereiche definiert werden: die Lehr-/Lernprozesse, das Wissen, die Lehrenden, die Einrichtungen und das System. Nimmt man das Kriterium des explizierten Zieles, dann konzentrieren sich die Evaluationsverfahren auf Lernprozesse und Lehrende (→ Kap. 3.1), auf die Einrichtungen der Weiterbildung (→ Kap. 3.3) und das System der Weiterbildung national und international (→ Kap. 3.4). Hinzu kommen die Projekte und Programme, die jeweils befristete Interventionen mit eigenen Zielen sind (→ Kap. 3.2). Diese vier „Gegenstände“ betreffen den innersten Kern der Erwachsenenbildung ebenso wie ihren Kontext. Sie lassen sich in Form konzentrischer Kreise darstellen:

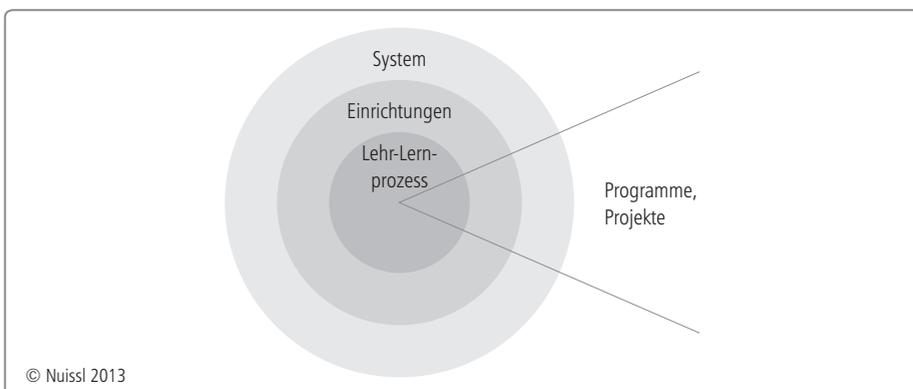


Abbildung 4: Evaluationsgegenstände in der Weiterbildung (Quelle: eigene Darstellung)

Das vorliegende Kapitel handelt von den Gegenständen der Evaluation in der Erwachsenenbildung und deren jeweiligen Erfordernissen bzw. konzeptionellen Voraussetzun-

gen. Es zeigt die Schnittstellen, an denen Evaluation in der Erwachsenenbildung ansetzen kann – und auch ansetzt.

Nach der Bearbeitung dieses Kapitels sollten Sie in der Lage sein, darzustellen

- welche Aspekte der Erwachsenenbildung Gegenstand von Evaluation sind bzw. sein können,
- welche jeweiligen Interessen gegenüber einer Evaluation bestehen und
- mit welchen Problemen und Einschränkungen jeweils zu rechnen ist.

3.1 Lernprozesse und Lernergebnisse

Lehr-/Lernprozesse sind genuine Gegenstände erziehungswissenschaftlicher Forschung und erziehungswissenschaftlich orientierter Evaluation. Anders als bei anderen Gegenständen der Evaluation, in denen sozialwissenschaftliche, ökonomische und politikwissenschaftliche Ziele, Zugänge und Verfahren eine Rolle spielen, sind die Erziehungswissenschaften (zusammen mit der pädagogischen Psychologie) in der Betrachtung dieses Gegenstands weitgehend allein.

Dabei haben sich – generell gesehen – historisch vor allem zwei Schwerpunkte herausgebildet: die pädagogische Diagnostik, also vor allem die Evaluation des Outputs von Lehr-/Lernprozessen, und die Handlungsforschung, also vor allem die begleitende Evaluation der Prozesse (vgl. Kuper 2005, S. 51ff.). Letztere kommt der erziehungswissenschaftlichen Disziplin besonders entgegen – sie ist ohnehin eine anwendungsbezogene Disziplin, die mehr der Verbesserung der Lehr-/Lernprozesse dient als der vermehrten Erkenntnis über sie. Und Erstere kommt der pädagogischen Psychologie entgegen, die sich verstärkt auf das individuelle Lernen und dessen Ergebnisse konzentriert. Beide Traditionen haben – unter erziehungswissenschaftlicher Perspektive – große Bedeutung auch in der Erwachsenenbildung. Deren Lehr-/Lernprozesse werden jedoch breiter in den Blick genommen.

Erst in jüngerer Zeit ist deutlich geworden, dass – vor allem als Teil der Prozessevaluation – die Lehrenden eine wesentliche Rolle spielen. Paradoxerweise geschah dies zu einem Zeitpunkt, als „Lehre“ als antiquiert galt gegenüber „Lernarrangements“ und (virtuellen) Selbstlernprozessen. Die Evaluation des Lehr-/Lernprozesses ist immer auch eine Evaluation der Lehrenden, mit ganz praktischer Absicht: Sie sollen selbst weiterlernen und ihre Lehrpraxis verbessern.

Das Lernen Erwachsener wird traditionell in unterschiedlichster Weise evaluiert. So ist die traditionellste Form der Lernevaluation diejenige, dass Teilnehmende Feedback-Bögen zu Kursen ausfüllen – dies ist seit Jahrzehnten üblich, schon lange bevor entsprechende Verfahren in Hochschulen implementiert wurden (vgl. Gerl/Pehl 1983). Grund dafür war und ist vor allem das Interesse der Einrichtungen, Informa-

tionen darüber zu erhalten, wie die Lehre der (meist nebenberuflichen) Lehrkräfte bei den Lernenden „ankam“ und – darauf aufbauend – Konsequenzen daraus zu ziehen (etwa positive oder negative Sanktionen gegenüber dem Lehrenden, Verwendung in Fortbildungen, Impulse für die kommenden Programme und deren didaktische Umsetzung). In der Erwachsenenbildung sind solche Rückmeldungen deshalb besonders wichtig, weil die Teilnahme an Kursen, Seminaren und dergleichen freiwillig ist, das Angebot für die Adressaten also attraktiv sein muss. Aber auch die Lehrenden haben ein Interesse an diesen Evaluationen. Sie erhalten ein unmittelbares Feed-Back zu ihrer Tätigkeit, das sie beim nächsten Kurs berücksichtigen können, sie erhalten Bestätigung und vielleicht auch einen Motivationsschub. Für die Lernenden ist es eher eine Vergewisserung darüber, was sie erlebt und gelernt haben, und sie bekommen eine präzisere Vorstellung davon, wie sie einen Lernprozess gestaltet haben möchten.

Das Lernen der Erwachsenen wird aber mittlerweile in vielen unterschiedlichen Formen und zu unterschiedlichen Aspekten analysiert und evaluiert. Zu unterscheiden ist die Lehr-/Lernforschung einerseits und die Prozessevaluation andererseits.

Lehr-/Lernforschung

Die Lehr-/Lernforschung betrachtet Lernprozesse vorrangig mit Erkenntnisinteresse – wie laufen die Prozesse ab, welche lernrelevanten Beobachtungen lassen sich machen, wo liegen Problemzonen und „gelungene“ Einheiten. Die Evaluation betrachtet die Lernprozesse eher unter der Frage, wie diese beurteilt werden (von Lehrenden und Lernenden), wo Verbesserungsmöglichkeiten liegen und welche Erfolge erzielt werden. In der praxisorientierten Erziehungswissenschaft lassen sich diese beiden Ansätze nicht immer scharf trennen; die erziehungswissenschaftliche Tradition ist, wie oben bereits bemerkt, genuin darauf gerichtet, Lernprozesse zu verbessern. Von daher haben auch die Forschungen zu den Lernprozessen Erwachsener Anteile einer Prozessevaluation. Das gilt vor allem für diejenigen Arbeiten, in denen Lernprozesse mittels Videoaufnahmen erfasst und analysiert werden – sie gehen entsprechend aufbereitet als Material ein in Fortbildungsmaßnahmen für Lehrkräfte und die Ausbildung an der Hochschule (vgl. die Arbeiten im Umkreis von J. Kade in Frankfurt, J. Schrader in Tübingen/Bonn und S. Nolda in Dortmund). Ein wesentlicher Unterschied zwischen den wissenschaftlichen Analysen von Lernprozessen und ihren Evaluationen besteht heutzutage vor allem auch in der Priorität der benutzten Verfahren: Die Analyse bedient sich dabei vor allem der Beobachtung, die Evaluation vor allem der Befragung.

Prozessevaluation

In der Prozessevaluation dominiert nach wie vor die Evaluation durch die Teilnehmenden, die Lernenden. Sie hat sich aber gegenüber früheren Formen und Verfahren der Evaluation ausdifferenziert und in den Instrumenten „professionalisiert“, folgt stärker

wissenschaftlichen Standards. Die Evaluationen durch die Teilnehmenden können prozessbegleitend („formative Evaluation“) oder abschließend („summative Evaluation“) erfolgen. Im professionellen pädagogischen Kontext kommen stets beide Formen aufeinander bezogen vor.

Die formative Evaluation konzentriert sich auf den Fortgang des Lernprozesses und ist jeweils an Schnittstellen platziert (etwa – bei mehrtägigen Kursen – jeweils abends oder nach Abschluss einer inhaltlichen Einheit oder bei Wechsel des Pädagogen etc.). Sie richtet sich in der Regel retrospektiv auf die zurückliegende Kurseinheit und bewertet Inhalt, Methode, Lehrverhalten, Praxisnähe u.a. mit dem Ziel, im weiteren Verlauf des Kurses negativ kritisch angemerkte Punkte zu verbessern und positiv kritisch angemerkte Punkte zu verstärken. Die Verfahren dieser formativen „Zwischenevaluation“ von Lernprozessen können höchst unterschiedlich sein; so finden sich in der Praxis insbesondere diskursive und visualisierte Verfahren wie:

- Blitzlicht, eine kurze mündliche Rückmeldung, die nach bestimmten Regeln erfolgt, zu einer präzise formulierten Frage;
- Stimmungsbarometer, meist in visualisierter Form (mit „Smilies“) zu einzelnen Aspekten des Lerngeschehens (z.B. Klima, Thema, Methode etc.);
- Partnerinterview, in dem Einstellungen eines anderen erfragt und dann der Lerngruppe berichtet werden;
- Gruppenphase, in der sich Kleingruppen zu Aspekten des Erlebten beraten und das Ergebnis anschließend der Gesamtgruppe berichten;
- Moderation, in der das „Plenum“ (also die gesamte Lerngruppe) über den Ablauf des Lernprozesses diskutiert;
- Gruppensituation in der Gruppe malen oder darstellen (Gerl/Pehl 1983, S. 38ff.), das Gemalte wird im Plenum vorgestellt oder eine Statue wird von den Gruppenmitgliedern gebildet, beides wird unter dem Aspekt der Lernsituation und des Lernstands interpretiert;
- Sätze schreiben mit Bezug zum Lernprozess, die anschließend vorgetragen, erklärt und diskutiert werden (es geht auch mit erzählten oder geschriebenen Kurzgeschichten, teilnehmergruppenabhängig);
- Fragebogen, der den Verlauf, nicht das Ergebnis im Blick hat (ein Beispiel findet sich bei Gerl/Pehl 1983, S. 43) und dessen Ergebnisse bei direkter Auswertung im Kurs gemeinsam erörtert werden.

Die meisten dieser Verfahren der Zwischenevaluation verfolgen das Ziel, nicht nur ein Zwischen-Feed-Back zu erhalten, sondern kritische Kommentare *frei* zu ermöglichen (anonym), sie *konkret* auf den Lernprozess zu richten und sie als *persönliche* Eindrücke zu formulieren. Auch wird Wert darauf gelegt, die jeweilige Perspektive für den weiteren gemeinsamen Lernprozess zu benennen.

Auch die Schlussevaluation durch die Teilnehmenden hat eine retrospektive und eine perspektivische Dimension (vgl. Geißler 2005). Die retrospektive Komponente bewertet den Lernprozess aus unterschiedlicher Perspektive, bezogen auf die Inhalte, die Gruppe, die Lehrpersonen, das Gelernte etc.; die perspektivische Komponente richtet sich auf den geplanten und erwartbaren Umgang mit dem Gelernten, die Verwendbarkeit und die Bedeutung für die einzelnen Lernenden in der Zukunft. Anders als bei der Zwischenevaluation werden hier häufig standardisierte Verfahren eingesetzt (Fragebogen), deren Ergebnisse den Vorteil haben, auch in Kontexten außerhalb des Lernprozesses verwendbar zu sein. Einige Verfahren der Schlussevaluation sind (vgl. ebd., S. 83ff.):

- der Fragebogen: Er ist das am häufigsten verwendete Instrument, wobei die Anforderungen an eine „richtige“ Gestaltung des Fragebogens in der Praxis oft unterschätzt werden;
- das Kofferpacken: Hierbei legen die Lernenden Zettel in einen Koffer, auf denen steht, was sie aus dem Seminar mitnehmen (Betonung der perspektivischen Dimension; ebd., S. 87);
- die „Abschiedsgeographie“: Die Lernenden erstellen eine eigene Landschaft, die ihren Lernerfahrungen (z.B. in Gestalt einer Küstenlandschaft: „im Meer versunken?“ oder „festen Boden unter den Füßen?“) entspricht (vgl. ebd., S. 90).

BEISPIEL

| | stimmt völlig | | stimmt gar nicht | |
|---|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| 1. Ich glaube, dem Kursleiter ist es wichtig, was die Teilnehmer wollen. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |
| 2. Was hier gemacht wird, ist mehr theoretisch. Die Praxis sieht ganz anders aus. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |
| 3. In der Lerngruppe fühle ich mich frei und entspannt. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |
| 4. Ich habe das Gefühl, der Kursleiter will den Teilnehmern seine Meinung aufdrängen. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |
| 5. In der Pause wird über die Fragen offener geredet als in der Lerngruppe. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |
| 6. Wenn ich hier in der Gruppe einen Fehler mache, ist mir das sehr unangenehm. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |
| 7. Mit meiner Meinung halte ich hier lieber hinter dem Berg. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |
| 8. Hier werden die wirklichen Probleme beim Namen genannt. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |
| 9. Jeder kann hier Einfluss nehmen, was im Rahmen des Themas gemacht wird. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |
| 10. Der Kursleiter hört anderen genau zu. | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 |

(Quelle: Gerl/Pehl 1983, S. 43)

In der Schlussevaluation geht es meist weniger um eine kritische Würdigung des Abgelaufenen (die natürlich auch erfolgt, was viele ängstliche Lehrende veranlasst, ihre Lehre nicht abschließend evaluieren zu lassen, um keine „schlechte“ Kritik zu erhalten) als vielmehr um eine Bewertung des Gelernten. Dabei spielen naturgemäß die Erwartungen der Lernenden vor dem Hintergrund ihrer individuellen Arbeits- und Lebenssituationen eine Rolle. Das wichtigste Bewertungskriterium zum Gelernten hängt eng damit zusammen; hier handelt es sich auch um die Frage, ob es nützlich und anwendbar ist. Die Anwendbarkeit setzt die Brauchbarkeit des Gelernten voraus, z.B. zur Lösung von Alltagsproblemen, zum besseren Verständnis von Abläufen und Konflikten oder konkret für den Beruf und die Arbeit im Betrieb. Die Nützlichkeit kann direkt im „Output“ und „Outcome“ des Kurses liegen, aber auch auf „Umwegen“ entstehen – z.B. in der Verbesserung des Images des Lernenden oder des Fortbestehens der Lerngruppe als wichtiger sozialer Kontakt (sogenannte „Impact“-Faktoren).

Für die Teilnehmenden ist es weniger bedeutsam, ob sich die Lernprozesse für andere Lernende und kommende Generationen von Lernenden verbessern, sondern wie sich für sie der Lehr-/Lernprozess „angefühlt“ hat und ob sie mit dem Gelernten etwas anfangen können. Hier unterscheiden sich auch deutlich die Interessen der Lehrenden von denen der Lernenden hinsichtlich der Ergebnisse der Schlussevaluation.

Wichtig zur Beurteilung der Schlussevaluation vor allem bei Lernprozessen, in denen gruppendynamische Prozesse möglich waren und stattfanden, ist die positive Verklärung des Geschehenen, die dabei immer festzustellen ist. Der gemeinsame Abschied wirft ein mildes Licht auf das Vergangene. Dieser Effekt kann teilweise ausgeglichen werden, wenn die Zwischenevaluationen und die Schlussevaluation in einem systematischen Zusammenhang stehen, vergleichbare Kategorien verwenden und Kriterien vorgeben. So kann etwa das durchgängig visualisierte Stimmungsbarometer, wenn es ausreichend differenziert ist, das Ergebnis der Schlussevaluation einordenbar machen.

Anders als mit der Evaluation des *Lernprozesses* verhält es sich mit der des *Lernprodukts*. Produkte von Lehr-/Lernprozessen sind, einfach gesagt, zunächst einmal das Lernergebnis, also das Gelernte. Es gibt zwar auch andere Produkte – wie etwa geänderte Ansichten, die nicht kognitiv ermittelbar sind, die möglicherweise speziell angefertigten Materialien, das entwickelte Curriculum oder Reflexionen, Aufzeichnungen und eben die Prozessevaluation –, sie gelten jedoch eher als „Beiprodukte“, die dem „eigentlichen“ Ziel, dem Lernergebnis, dienen.

Das „eigentliche“ Lernergebnis, das Lernprodukt, der „Lernoutput“ kann direkt zum Abschluss des Lehr-/Lernprozesses erhoben werden durch Prüfungen, Tests, qualitative Verfahren oder auch – nach einer Überprüfung auf Tauglichkeit in der Praxis – in einem angemessenen zeitlichen Abstand. Im ersten Fall spielt die Prospektion eine Rolle, d.h. die Annahme, das Gelernte (abgefragt!) sei im Leben verwendbar. Im zweiten Falle ist es gewissermaßen eine doppelte Evaluation: des Lernoutputs und -outcomes selbst

sowie deren Praxisprüfung. Dabei zeigen sich meist ernüchternde, manchmal aber auch überraschende Ergebnisse: So etwa, dass die Lernenden im Seminar miteinander diskutieren, ohne Notiz von den intensiv intervenierenden Pädagogen zu nehmen, oder dass die Prüfung des Lehrangebots an der eigenen Praxis eine kollektive theorieorientierte Lernleistung ist (vgl. Kejcz/Nuissl u.a. 1979–1981).

Der zweite wichtige Unterschied liegt in der Frage, ob das Lernergebnis, also „Output“ und „Outcome“, vom Lehrenden intendiert war oder nicht. Die Frage der Lehrintention spielt dabei in doppelter Weise eine Rolle (vergleichbar den Projekten und Programmen, → Kap. 3.2): zum einen als Raster für das, was in der Evaluation erfasst wird, und zum anderen als Grundlage von Bewertungen, als Raster der Bewertungskriterien. Gerl stellte diese „Korridore“ der Ergebnisevaluation in einer Grafik dar:

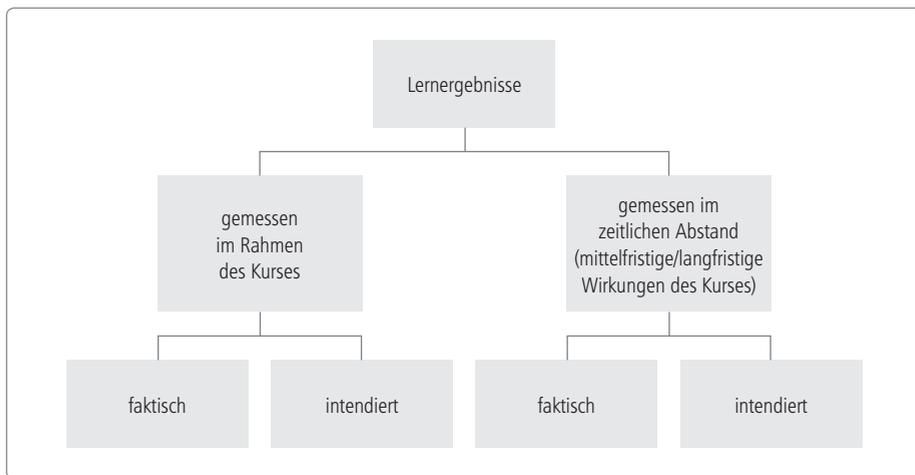


Abbildung 5: „Korridore“ der Ergebnisevaluation (Quelle: Gerl/Pehl 1983, S. 63)

Eine Evaluation, die darauf abzielt zu überprüfen, ob das intendierte Lehrziel erreicht ist, setzt natürlich ein solches, explizit formuliert, voraus. Eine solche explizite Lernzielformulierung ist in der Erwachsenenbildung insbesondere dort zu finden, wo es um Qualifikationen geht, also um zu vermittelnde Elemente von Wissen und Fähigkeit, etwa in der beruflichen Weiterbildung, beim Sprachenlernen, beim Erwerb von Lizenzen und Zertifikaten usw. Auch sind sie, allerdings weniger prüfbar, zu finden in gesinnungsorientierten Bildungsangeboten wie der Gewerkschaften oder der Kirchen. Dort geht es vor allem um das praktische Umsetzen des Gelernten im Alltag, im Beruf und in der Familie. Dies ist so bedeutsam, dass oft der „Praxistest“ an das Ende der Veranstaltung selbst angefügt wird, weil eine wirkliche Nachbefragung (etwa nach einem halben Jahr) erhebungstechnisch zu kompliziert ist.

Die Überprüfung des Lernergebnisses anhand der Lehrziele ist – aus Sicht der Lehrenden – sinnvoll, hat aber vor allem im Lichte der konstruktivistischen Sichtweise nur eine begrenzte Aussagekraft. Das Lernergebnis, soweit ist die erwachsenenpädagogische Diskussion mittlerweile übereinstimmend der Meinung, ist ein Produkt des Lernenden, bestenfalls ein gemeinsames Produkt von Lehrenden und Lernenden. Lernende sind nach diesem Verständnis „Prosumenten“ des Lernergebnisses, sie konsumieren (das Lehrangebot) und produzieren (das Lernergebnis) gleichzeitig (vgl. Hartz/Meisel 2011, S. 14). Diese eindeutige (was das Lernen angeht) und zugleich verschränkte (was den Lehr-/Lernprozess angeht) Blickweise macht auch die Evaluation des Ergebnisses sehr komplex.

Dies betrifft nicht nur den „Output“, also das direkt Gelernte, sondern auch den „Outcome“ und vor allem auch den „Impact“. Am einfachsten hierbei ist noch die Evaluation des „Outputs“. Hier kann man auf unterschiedlichste Test- und Prüfverfahren zurückgreifen, in denen feststellbar ist, was gelernt wurde. Weniger valide sind dabei allerdings Rückschlüsse darüber, wann es gelernt wurde und ob zwischen dem Gelernten und dem Lehrangebot ein kausaler Zusammenhang besteht, wie er viele Jahrzehnte lang als selbstverständlich angenommen wurde. Schwieriger ist es schon beim Outcome, der zu einem hohen Anteil Dimensionen des Nutzens und der Verwertung des Gelernten hat. Und Impact-Analysen sind hier – wie generell vor allem bei gesellschaftswissenschaftlichen Aspekten – überhaupt nur schwer möglich.

Ein wesentlicher Aspekt bei der Beurteilung von Lehr-Lernprozessen und -ergebnissen ist der „Erfolg“. Dies ist eine Kategorie, die vor allem interessenabhängig ist. Zunächst stellt sich die Frage des Interesses der Beteiligten, der Lehrenden und Lernenden. Sodann die Frage nach dem Interesse des Anbieters und schließlich diejenige nach dem anderer Beteiligter, etwa der Betriebe, in denen Lernende beschäftigt sind und die den Lernprozess aus betrieblichen Interessen fördern. Da alle näher und ferner Beteiligten an einem „Erfolg“ interessiert sind, stellt sich die Frage, wie genauer zu qualifizieren und zu messen ist.

DEFINITION

Erfolg

Erfolg wird verstanden als Erfüllung der gestellten Anforderungen, auch im Wettbewerb. Gemeinhin gilt als Erfolg, was außengesetzte Anforderungen erfüllt. Gerade aber im Kontext des Lernens kann ein Erfolg auch aus der individuellen Leistung im biografischen Kontext bestehen.

Der Erfolg für den Lehrenden liegt in der Regel darin, dass die Kursteilnehmenden etwas Gelerntes aus seinem Kurs mitnehmen, von dem angenommen werden kann, dass

es in seiner Lehrintention lag. Er ist also vor allem daran interessiert zu überprüfen, ob Lehrziel und Lernergebnis in einem nachweislichen und kausalen Zusammenhang stehen – das wäre eine Bestätigung seiner eigenen Lehrleistung. Der Lehrende stellt daher am ehesten Evaluationsfragen, die sich auf diesen Zusammenhang beziehen, und er bezieht seine Bewertungskriterien auch vor allem darauf.

Der Erfolg für die Lernenden liegt in der Regel darin, dass sie das gelernt haben, was sie lernen wollten, also ihre Lernziele erreicht haben. Da die Teilnahme an Erwachsenenbildungskursen bekanntlich multi-motivational ist (vgl. Nuissl/Siebert 2013), kann sich das auch nur auf eine gute kommunikative und soziale Lernatmosphäre beziehen, die im Ergebnis ein verbessertes Selbstwertgefühl mit sich bringt. Auch ist die Bewertung sehr individuell – Lernende bewerten eher bezogen auf ihre eigene Anspruchshaltung als auf generelle Lernanforderungen.

DEFINITION

„Gelungenes Lernen“

In ihrem Studientext „Gelungenes Lernen – Qualität und Qualitätsmanagement in der Weiterbildung“ definieren Zech und Tödt so: „Gelungenes Lernen ist ein Lernen, das der Lernende selbst wertschätzt, weil er dadurch seine eigenen selbstbestimmten Ziele erreicht und seine Handlungsfähigkeit erhöht hat. Durch gelungenes Lernen erhöht sich die Lebensqualität des Subjekts“ (2012, S. V).

Der Erfolg für die anbietende Bildungseinrichtung liegt vor allem darin, über positive Lernergebnisse der Lernenden die Effektivität und Qualität der eigenen Arbeit nachzuweisen – je höher der Anteil der erfolgreichen Lernenden ist, desto attraktiver muss das Angebot für potentielle neue Lernende sein und desto höher wird die Einrichtung von Auftrag- und Geldgebern eingeschätzt.

Der Erfolg für die Betriebe, deren Beschäftigte an einem Erwachsenenbildungskurs teilnehmen, liegt vor allem darin, dass sie von der Bildungsmaßnahme mit einem nützlichen und betrieblich verwertbaren Ergebnis „zurück“ kommen. Das kann auch erhöhte individuelle Zufriedenheit sein, meist aber handelt es sich um erworbene Kompetenzen für die Bewältigung der Arbeitsanforderungen.

Sofern es um die einfachste Form der Ergebnisevaluation von Lehr-/Lernprozessen geht, das Zertifikat, treffen sich diese unterschiedlichen Interessen dann, wenn es sich um ein (auf dem Markt) anerkanntes und die Lernergebnisse angemessen dokumentierendes Papier handelt. Geht es um nicht derart zertifizierte Ergebnisse, bedarf es der Aushandlung darüber, welches bewertende Paradigma im Vordergrund stehen soll.

Bei den gängigen Lernerfolgsfeststellungen handelt es sich um eine summative Evaluation, um die Bewertung eines Ergebnisses am Ende eines Prozesses. Aufgrund der

schwierigen Messbarkeit anderer Aspekte, aber auch aufgrund traditioneller Verständnisse von Lernprozessen und -erfolgen beziehen sie sich vor allem auf die kognitive Ebene. In der Testforschung sind die Vor- und Nachteile entsprechender Verfahren vielfach empirisch überprüft und erörtert.

Gerade sie aber zeigen auch, dass Evaluationen, die einen komplexeren Ansatz haben, mehr Aussagekraft besitzen als scheinbar objektive Tests und Prüfungen, die den Stand von Kenntnissen festhalten. Die Prognosekraft von Kognitionstests ist eher gering, gerade deshalb entstand in den vergangenen zwei Jahrzehnten die Dominanz des Kompetenzbegriffs, bei dem es nicht nur um das Wissen selbst als vielmehr um die Fähigkeit der Anwendung von Wissen geht.

Die heutigen Prüfungs- und Testverfahren legen daher einen großen Wert auf die Outcome-, gelegentlich auch auf die Impact-Dimension des Gelernten.

3.2 Projekte und Programme

Projekte und Programme sind – wie die Lernprozessevaluationen – die „Klassiker“ der Evaluationsgegenstände im Bildungsbereich.

Dies hat verschiedene Gründe, und sie sind vor allem solche, die aus dem Kontext (politisch, gesellschaftlich) und weniger aus dem Bildungsbereich selbst kommen. Die Gründe sind (ohne dass in der Reihenfolge eine Hierarchie liegt):

- Der Bildungsbereich ist (wenn man einmal die Erwachsenenbildung als „Zwitzer“ zwischen Staat und Markt ausklammert) überwiegend staatlich strukturiert und finanziert. Der Staat ist (zunehmend) an einer Kontrolle und Legitimation seiner Investitionen interessiert und zu derselben verpflichtet, man lese nur die jährlichen Berichte des Bundesrechnungshofs. Der Bundesrechnungshof ist die umfassendste und gefürchtetste Instanz der Evaluation staatlichen Handelns, weil präzise auf die Finanzen konzentriert, weniger auf Inhalte, Kontexte, Prozesse etc.
- Der Bildungsbereich unterliegt (leider) einem fortwährenden Innovationsdruck; dieser erklärt sich aus Modernisierungsprozessen der Gesellschaft, neuen wissenschaftlichen Erkenntnissen über das Lernen, wachsende Konkurrenz im globalen Rahmen, Profilierungen politischer Stakeholder u.a.; Innovationen (d.h. Veränderungen) bedürfen einer besonderen Förderung (zu Beginn) und der Überprüfung ihrer Wirksamkeit.
- Wirkungen von Innovationen im Bildungsbereich sind außerordentlich schwer abzuschätzen, kaum messbar und ergebnisunsicher. Vielfach zeigen sich Wirkungen erst nach einer langen Zeit – lange nachdem die Innovation als Programm oder Projekt in Gang gesetzt worden ist. Daher überschneiden sich im Bildungsbereich die

- Aspekte von Nachhaltigkeit und Wirkung besonders deutlich. Kurzfristige Wirkungen sind feststellbar, aber die nachhaltigen Wirkungen sind viel wichtiger.
- Der Legitimationsdruck bei Maßnahmen, die im Bildungsbereich getätigt werden (meist staatlicherseits) ist besonders hoch; die Mehrheit der Bevölkerung ist in der einen oder anderen Weise mit dem Bildungsbereich verbunden (durch Kinder, als Lehrende, Arbeitgeber etc). Es gibt auch keinen anderen gesellschaftlichen Bereich, in dem alle Menschen individuell und kollektiv so viele eigene Erfahrungen gesammelt haben. Und es ist nur folgerichtig, dass bei gegebenem Anlass (wie etwa der Debatte um PISA) diese Erfahrungen, verbunden mit eigenen Interessen, breit und bewertend in die Diskussion eingebracht werden.
 - Staatliches Handeln richtet sich mit vielen Zielen auf den Bildungsbereich, etwa in Fragen der Migration, der sozialen Integration, der Regulierung sozialer Konflikte, der Entwicklung der „work force“ für den globalen Wettbewerb sowie der Anerkennung von Leistungen und Qualifikationen (wie aktuell intensiv in den nationalen Qualifikationsrahmen); der Staat hat ein genuines Interesse daran, diesen Bereich weiter zu entwickeln.³

Dies alles sind Gründe für eine immense Aktivität von Projekten und Programmen im Bildungsbereich, auf Länder- und auf Bundesebene, in den letzten Jahren aber auch insbesondere auf europäischer Ebene (vgl. Lattke/Nuissl/Pätzold 2010). Genau genommen begann die Geschichte der Evaluation auch mit Projekten und Programmen im Bildungsbereich in den Vereinigten Staaten von Amerika (vgl. Beywl 1988), es geht dabei also nicht um ein *deutsches* Phänomen.

Projekte und Programme – das ist im bereits Gesagten ersichtlich – sind vor allem solche des Staates, sofern es um den Bildungsbereich geht. In wirtschaftlichen und technologischen Kontexten dominiert die private Wirtschaft. Seit auch die Europäische Union Bildung und vor allem Weiterbildung als ihr Politikfeld definiert hat, also seit den Maastrichter Verträgen von 1992 (vgl. Lattke/Nuissl/Pätzold 2010), haben Projekte im Bildungsbereich (und damit deren Evaluation) in den Mitgliedsstaaten deutlich zugenommen. In diesen europäischen Projekten sind seit vielen Jahren sowohl die „Validation“ (also die Sicherung der Nachhaltigkeit zu Beginn) als auch die „Evaluation“ konstitutiver Bestandteil der Förderung. Dabei hat sich in den letzten Jahren neben die „interne“ Evaluation verpflichtend die „externe“ Evaluation gesellt.

3 Wobei dieses staatliche Interesse auch gelegentlich eigenartige Volten schlägt. Bei der Einführung des konsekutiven Bachelor/Master-Programms an allen deutschen Universitäten und in allen Disziplinen erfolgte keine Erprobung, kein Projekt, kein Programm, keine Evaluation – ein Zeichen dafür, dass auch Interessen an Viabilität und Legitimation von politischen Kalkülen dominiert werden können.

DEFINITION

Interne und externe Evaluation

Unter „interner“ Evaluation wird im Kontext europäischer Projekte die Evaluation der Prozesse und der Erstellung der Produkte verstanden (es ist immer ein „workpackage“ im Rahmen des Projektplans), während die „externe“ Evaluation (durchgeführt von einer Person, die nicht zum Projektteam gehört) sich auf die Bewertung der Produkte in Bezug auf Qualität, Relevanz und Wirkung konzentriert.

Die Verschränkung beider Blickweisen hat den Vorteil, die „blinden Flecken“ der internen Evaluation aufhellen zu können, gewissermaßen die interne Evaluation auch zu evaluieren, und zugleich den Vorteil, eine Innen- und Außensicht zu kombinieren.

Die Projekte und Programme im Bildungsbereich sind nicht nur generell staatlich initiiert, sondern auch – von Thematik und Ziel – „innovativ“. Mit ihnen sollen neue Wege gesucht und erprobt werden, Bildungsaufgaben zu realisieren und zu verbessern. Es handelt sich demnach fast immer um „Modellversuche“ (vgl. Kolwe-Jung u.a. 1980). Solche Modellversuche verfolgen ein bestimmtes Ziel, basieren auf einem entwickelten Konzept und erproben dessen Umsetzung. Große Modellversuche in der Bundesrepublik Deutschland waren in den Jahren 1974 bis 1981 das „Bildungsurlaubs-Versuchs- und –Entwicklungsprogramm“ (BUVEP) (Kejcz/Nuissl u.a. 1979–1981) und in den Jahren 2000 bis 2009 das Programm „Bildungsnetzwerke in der Region“ (Nuissl u.a. 2006). Beide wurden ausführlich evaluiert. Projekte und Programme lassen sich in drei Phasen einteilen:

- Die erste Phase ist diejenige der Konzeption und des Aushandelns; an Projekten und Programmen des Bildungsbereichs sind unterschiedlichste Akteure beteiligt, nicht nur die Bildungseinrichtungen, sondern vor allem auch staatliche Instanzen (Ministerien), regionale Körperschaften (Kommunen), Betriebe und Gewerkschaften sowie vielerlei Organisationen. In dieser Phase werden nicht nur die Konzeptionen erstellt und die Ablauf- und Förderungsbedingungen ausgehandelt, sondern auch wesentliche Elemente des Ablaufs sowie die zu beteiligenden Personen festgelegt. In dieser Phase wird auch in aller Regel die Aufgabe der Evaluation (oder „wissenschaftlichen Begleitung“) präzisiert und einer Instanz (Person, Institut, Organisation) übertragen.
- Die zweite Phase ist diejenige der Realisierung des Projekts oder Programms, etwa der Implementation der (neuen) Konzepte oder der Entwicklung neuer Produkte. Curriculumentwicklung oder die Erstellung von Handbüchern und Fortbildungsmodulen sind solche projektförmig realisierten Aktivitäten. In dieser zweiten Phase findet die Erhebung für Evaluationsdaten statt, aber auch – im Falle einer formativen, begleitenden Evaluation – die fortlaufende Beratung und Unterstützung des Prozesses durch die Evaluatoren oder „wissenschaftlichen Begleiter“.

- Die dritte Phase ist diejenige der Umsetzung der Ergebnisse, so etwa die nachhaltige Implementation etwa eines neuen Curriculums, der Transfer und die Dissemination erstellter Produkte. In dieser Phase richtet sich das Interesse der Evaluation (sofern sie auch für diese Phase noch eingesetzt ist) auf die Wirkung, den „Impact“ des Projekts. In vielen Fällen aber endet die Evaluation mit dem Ende des Programms, sodass die Wirkung antizipativ zu erfassen ist.

Da solche Projekte und Programme dezidiert ein Ziel verfolgen, das von großem Interesse für den Auftraggeber Staat ist, richten sich Ziel und Auftrag der Evaluation auch sehr eng nach der Absicht, die mit dem Programm verbunden war und ist. Die Evaluation hat daher im Wesentlichen die Aufgabe festzustellen, ob das intendierte Programmziel auch tatsächlich erreicht wurde oder in welchem Ausmaß es erreicht wurde. In diesem Sinne ist die Evaluation auch eine „Controlling“-Maßnahme, vor allem dann, wenn sie mit dem Ende des Programms *endet*.

Die wesentlichen Kriterien, welche die Evaluation von Programmen und Projekten zu überprüfen hat, ergeben sich aus deren Qualität; sie lassen sich nahezu immer finden:

- *Ziel und Ertrag/Ergebnis/Produkt*
Wird das angestrebte Ziel erreicht?
- *Investition und Ertrag/Ergebnis/Produkt*
Ist die Relation vertretbar und sinnvoll?
- *Prozess und Ertrag/Ergebnis/Produkt*
Entspricht das Ergebnis dem Prozessablauf?
- *Produkt und Qualität*
Hat das Produkt die angestrebte Qualität?
- *Projekt und Wirkung*
Hat das Projekt die intendierte Wirkung?
- *Projekt und Kontext*
Gibt es Kollateralschäden oder Nutzen?

Da der Auftraggeber Staat in besonderem Maße daran interessiert ist, aus sachlichen und legitimatorischen Gründen die nötigen Informationen zu diesen Fragen zu erhalten, ist er auch selbst evaluierend tätig: in der begleitenden Kontrolle und Unterstützung der Programme und ihrer Evaluation. Nicht selten wechselt der Staat auch die Evaluatoren im Verlaufe des Programms, wenn er – aus welchen Gründen auch immer – mit der Evaluation nicht zufrieden ist.

3.3 Bildungseinrichtungen

Bildungseinrichtungen können höchst unterschiedlich sein – in der Erwachsenenbildung sind sie es auch. Sowohl nach Größe als auch nach Angebotsvielfalt, nach Einzugsbereich und Selbstverständnis, nach Rechtslage und Finanzierung, nach Geschichte und organisationaler Einbettung, nach Zielsetzung und Struktur sowie nach Methodik und Adressatenbezug. Es ist daher kaum möglich, die Evaluation „der“ Einrichtung der Erwachsenenbildung darzustellen. Allein die Feststellung, wie viele Weiterbildungseinrichtungen in Deutschland überhaupt existieren, erforderte ein eigenes Forschungsprojekt mit differenziertem Ergebnis: im Jahre 2008 waren es ca. 17 Tausend (Dietrich/Schade 2008).

Es ist eher möglich zu charakterisieren, welche Gründe und Anlässe zu einer Evaluation von Weiterbildungseinrichtungen führen können und führen. Der Grund kann zum einen darin liegen, dass die übergeordnete Organisation der Bildungseinrichtung (z.B. ein Verband, ein Betrieb) überprüfen möchte, ob sich die Investitionen in die Einrichtung auch weiterhin lohnen; viele Erwachsenenbildungseinrichtungen gehören Organisationen wie Kirchen, Gewerkschaften und Wohlfahrtsverbänden an. Der Grund kann zum anderen auch darin liegen, dass ein Hauptgeldgeber eine solche Überprüfung wünscht, z.B. die Kommune bei Volkshochschulen oder im Kammer-Bereich. Zum dritten kann der Grund darin liegen, dass das Einrichtungsmanagement eine Überprüfung wünscht, sei es, weil es gewechselt hat, sei es, weil tiefgreifende Änderungen anstehen oder geplant sind, sei es, weil Optimierungsprobleme in den Arbeitsabläufen oder in der Finanzierung bestehen. Auch sind regelmäßige Evaluationen in manchen Konzepten des Qualitätsmanagements vorgesehen.

Einrichtungsevaluationen finden fast immer, nicht nur in der Erwachsenenbildung, nach dem Peer-Review-Verfahren statt. Das liegt vor allem daran, dass der Gegenstand „Einrichtung“ sehr komplex ist, betriebliche, ökonomische, inhaltliche, soziale und personale Elemente hat, die jeweils unterschiedlichen Sachverstand erfordern. Einrichtungsevaluation ist auch meist eine „Expertenevaluation“, bei der die Bewertungskriterien (→ Kap. 6.1) aus fachwissenschaftlichen Zusammenhängen stammen. Solche Evaluationen finden nach festgelegten Verfahren statt, wie etwa anhand der Richtlinien des Deutschen Instituts für Erwachsenenbildung (DIE):

Verfahrensschritte bei der Peer-Review-Evaluation anhand der Richtlinien des Deutschen Instituts für Erwachsenenbildung (DIE)

Das DIE, das auf zahlreiche Erfahrungen mit Peer-Review-Evaluationen in Weiterbildungseinrichtungen verweisen kann, strukturiert diese grob in folgende Verfahrensschritte:

- Zunächst wird gemeinsam vereinbart, was das Ziel und was der Gegenstand der Evaluation ist und wie das Verfahren geregelt wird.

- Daraufhin wird ein Evaluationskontrakt geschlossen, der Kosten, Zeiten, Pflichten und Arbeitsformen beschreibt.
- Seitens des DIE wird eine fachkundige Evaluationskommission berufen.
- In einem weiteren Schritt stellt die Organisation der Evaluationskommission eine Reihe von Daten und Dokumenten zur Verfügung: Satzung, formulierte Ziele, Programme, Statistiken, Haushaltspläne, Organigramm etc.
- Auf der Grundlage der Dokumentenanalyse erstellt die Kommission einen Statusbericht und kristallisiert Fragen an die Organisation heraus. Diese werden von der Einrichtung schriftlich beantwortet. Sie erhält damit einen Anstoß zur Selbstevaluation zu Themen, die von außen als relevant angesehen werden.
- Bei einer Begehung der Einrichtung werden offene Fragen mit den Mitarbeitenden und der Leitung besprochen.
- Die Einrichtung erhält dann einen Bewertungsbericht mit Entwicklungsempfehlungen. Diese werden mit der Einrichtung abschließend kommuniziert.

Quelle: Hartz/Meisel 2011, S. 51

Bei Evaluationen von Weiterbildungseinrichtungen stellen sich dabei (mit dem Bedarf an entsprechenden Daten und Informationen) in der Regel folgende Fragen:

Evaluationsfragen zu Einrichtungen

- Wodurch finanziert sich die Einrichtung? Wie hat sich die Finanzierungsstruktur entwickelt?
- Wo liegen Angebotsschwerpunkte? Wie verhalten sie sich zu den Bedarfen des „Marktes“ der Einrichtung (z.B. Region)?
- Welche Adressaten sollen erreicht werden, welche werden erreicht?
- Welche Kostendeckungsbeiträge werden in den einzelnen Programmbereichen erzielt?
- Wie haben sich die Teilnahmen in den letzten Jahren entwickelt (differenziert nach Teilnehmergruppen und Programmbereichen)?
- Welche Ausfallquoten und welche Drop-Out-Quoten sind zu verzeichnen? Wie wird damit umgegangen?
- Gibt es eine kontinuierliche Evaluation durch die Teilnehmenden, wie sind die Ergebnisse hinsichtlich Zufriedenheit und Lernerfolgen?
- Welche konkurrierenden Einrichtungen erbringen welche Angebote, welche Formen der Kooperation und Vernetzung gibt es?
- Wo liegen aus Sicht der Beschäftigten und des Managements die größten Herausforderungen für die Zukunft?

Bei Evaluationen von Einrichtungen hängt viel davon ab, welche Daten und Informationen schon vorhanden und welche möglicherweise erst zu erheben sind (→ Kap. 5.2

und 5.3). Auch ist von Bedeutung, inwieweit nicht nur das Management der Einrichtung hinter der Evaluation steht, sondern auch die Beschäftigten einbezogen sind. Dies ist relevant für den Erhalt der Daten, für die Interviews und Gespräche im Zuge des Besuchs der Einrichtung und für die anschließende Umsetzung der Evaluationsergebnisse.

In diese Umsetzung gehen die Evaluationsdaten in der Regel ein als Grundlage der Organisations- und Qualitätsentwicklung (vgl. Zech/Tödt 2012, S. 93ff.) und müssen dafür nicht nur belastbar, sondern auch akzeptiert sein. Dabei haben sie unterschiedliche Funktionen, sie geben Impulse, unterstützen die Analyse und erleichtern die Festlegung eines Zielkorridors.

3.4 Systeme der Erwachsenenbildung

Systemevaluationen sind eine Spezialität der Erwachsenen- und Weiterbildung. Sie bieten sich vor allem deshalb an, weil es streng genommen gar kein System der Erwachsenenbildung gibt. Historisch betrachtet ist Erwachsenenbildung in Deutschland, aber auch in anderen europäischen Ländern, aus vier Ursprüngen entstanden:

- der Arbeiterbildung, die Mitte des 19. Jahrhunderts große Bedeutung für den Kampf der Arbeiterklasse gegen Unterdrückung und Ausbeutung gewann („Wissen ist Macht“); insbesondere Einrichtungen der gewerkschaftlichen Erwachsenenbildung stehen in dieser Tradition;
- dem öffentlichen (und popularisierten) Hochschulwesen, das – nach englischem Vorbild – in Vorlesungen und Seminaren seit Ende des 19. Jahrhunderts vor allem bürgerliche Kreise ansprach und nach dem ersten Weltkrieg in der Gründung der Volkshochschulen einen wichtigen Beitrag zur Demokratisierung der Gesellschaft versprach;
- der konfessionellen Weiterbildung, getragen von kirchlichen Organisationen, welche Seelsorge und Verkündung immer häufiger seit der Aufklärung auch in Form von Weiterbildung der Gläubigen gestaltet hat; sowie
- den berufs- und betriebsbezogenen Weiterbildungen von Beschäftigten, die seit den 1980er Jahren immer wichtiger wurden und heute den Bereich der Weiterbildung dominieren.⁴

Zu diesen Wurzeln der heutigen Weiterbildung gesellten und gesellen sich andere Elemente wie etwa die Fortbildung von Trainer/inne/n und Funktionären in Verbänden

4 Bezeichnend dafür ist die Kategorisierung der statistischen Erhebung der Weiterbildungsteilnahme im AES (Adult Education Survey), in der zwei von drei Kategorien (beruflich-betriebliche Weiterbildung, individuelle berufsbezogene Weiterbildung) einer dritten Kategorie (nicht-berufsbezogenen Weiterbildung) gegenüberstehen – und von den Teilnehmezahlen her auch dominieren.

und Organisationen (Sport- und Wohlfahrtsverbände sind hier führend), die Fortbildung von Lehrer/inne/n an den allgemeinbildenden Schulen (mittlerweile in allen Bundesländern verpflichtend) oder die Weiterbildung in Heimvolkshochschulen nach dänischem Vorbild (Grundtvig) vor allem im norddeutschen Raum (vgl. Nuissl/Brandt 2008).

Erst in den sechziger und siebziger Jahren des letzten Jahrhunderts kam Weiterbildung in den Blick der Bildungspolitik, getrieben vor allem von der damaligen Sorge um eine „Bildungskatastrophe“ und eine ungenügende Ausschöpfung der Bildungsreserven der Bevölkerung. Auf der Grundlage eines Strukturplanes für das Bildungswesen (BLK 1973) entwickelte sich Weiterbildung seitdem als „vierter Bildungsbereich“, in dem der vorherige „Wildwuchs“ staatlich geordnet wurde. Weiterbildungsgesetze wurden erlassen und Regelungsgrundsätze verordnet (vgl. Nuissl 2010b). All dies bedeutete jedoch nicht eine Verstaatlichung des Weiterbildungsbereichs, sondern eine „mittlere Systematisierung“, in der die Prinzipien der Pluralität unterschiedlichster Anbieter und der Subsidiarität (Eingreifen des Staates nur bei vorhandenen strukturellen Defiziten) galten (vgl. Faulstich 2003). Und öffentliche Gelder spielen im Bereich der Weiterbildung eine wichtige, aber in weiten Teilen nicht die entscheidende Rolle.

In diesem „System“ können staatliche Ziele nicht einfach verordnet werden, sondern bedürfen eines komplizierten Impuls-, Motivations- und Regelungsverfahrens. Dabei ist es nicht immer sicher oder auch nur transparent, ob spezifische Fördermittel das angestrebte Ergebnis erreicht haben, etwa eine verstärkte politische Bildung oder vermehrte Angebote für einzelne Zielgruppen. Naturgemäß ist vor allem der Staat daher daran interessiert, Genaueres über Wirkung und Wirksamkeit seiner förder- oder ordnungspolitischen Initiativen im Weiterbildungsbereich zu erfahren, um z.B. Korrekturen vorzunehmen oder weniger Mittel gezielter einzusetzen.

Seit Beginn der 1980er Jahre ist es daher in vielen Bundesländern üblich, das „System“ der Weiterbildung zu evaluieren bzw. evaluieren zu lassen und daraus Folgerungen für die weitere Gestaltung und Förderung dieses Politikfeldes zu ziehen. Den Anfang der Gesamtevaluationen machte Baden-Württemberg 1984 mit dem Weiterbildungsgutachten. Damals war der Begriff der Evaluation noch nicht geläufig: Dem Land ging es dabei darum, die Verantwortung und Förderung der Weiterbildung nicht einseitig als staatliche, sondern auch als private und betriebliche Aufgabe zu konstatieren. Viele weitere Systemevaluationen folgten (vgl. Nuissl/Schlutz 2001) und folgen weiterhin. Aber auch einzelne Teile des Systems werden evaluiert – so etwa insbesondere in den 1970er und 1980er Jahren die Wirkung und Wirksamkeit der Regelungen zum Bildungsurlaub (vgl. Kejcz/Nuissl u.a. 1979–1981).

Fragen der Systemevaluation:

- Welche Einrichtungen/Träger erbringen das Angebot?
- In welchem Verhältnis stehen die Angebote quantitativ und qualitativ zueinander?
- Sind Qualitätsmanagementverfahren implementiert?
- Gibt es Defizite räumlicher Art (unterversorgte Regionen)?
- Gibt es Defizite hinsichtlich der Zielgruppen?
- Werden die öffentlichen Mittel zielgerichtet alloziert?
- Wie effizient ist der Mitteleinsatz in den einzelnen Angebotsbereichen?

Ein Beispiel für eine solche Systemevaluation ist diejenige in Nordrhein-Westfalen, die in den Jahren 2009 und 2010 stattfand. Es war die dritte Systemevaluation im Land nach 1998 und 2004, mit veränderten Fragen und Evaluierungsverfahren. Alle drei Evaluationen basierten auf unterschiedlichen Interessen bzw. Anlässen: Die erste war verbunden mit der Absicht, die öffentlichen Mittel zu kürzen; die zweite sollte die erfolgte Entwicklung überprüfen; die dritte wurde vom Landtag beschlossen, um die Zielgenauigkeit bei der Verwendung der öffentlichen Fördergelder zu erhöhen. De facto handelte es sich – trotz unterschiedlicher Fragen, unterschiedlicher Akteure und unterschiedlicher Evaluationsmethoden – insgesamt in der raschen Abfolge der drei Einzelevaluationen um eine „formative“ Evaluation. Im Bericht zur dritten Evaluation heißt es dazu:

Formative Systemevaluation

„So gab die Landesregierung bereits im Jahre 1997 ein Evaluationsgutachten in Auftrag, welches das Ziel hatte, angesichts knapper werdender öffentlicher Mittel Effektivität und Effizienz der Förderung nach dem WbG zu überprüfen und Vorschläge für eine Novellierung zu unterbreiten. Nach der erfolgten Novellierung wurde (2004) in einer weiteren Evaluation überprüft, wie sich die Regelungen des neuen Gesetzes auswirken. Wenig überraschend werden in beiden Evaluationsgutachten ähnliche Punkte thematisiert wie in dem hier vorliegenden, auch die Empfehlungen weisen eine ähnliche Kontinuität auf. (...) In den Einzelempfehlungen werden viele Aspekte angesprochen, die auch im hier vorliegenden Evaluierungsgutachten eine Rolle spielen. So geht es um die Qualitätssicherung über Anerkennungskriterien von Einrichtungen und Implementation eines Qualitätsmanagements, die Stärkung der Professionalität und Hauptberuflichkeit, die Stärkung regionaler Weiterbildungsstrukturen die Entwicklung eines Supportsystems – allerdings noch mit deutlich geringerer Betonung von Beratung und Information. (...) Viele dieser Empfehlungen wurden umgesetzt (...) Es gibt jedoch auch Empfehlungen, die – mangels erkennbaren Fortschritts – hier faktisch wiederholt werden. Zu ihnen gehören die Empfehlungen hinsichtlich einer konzeptionellen Zusammenführung der politischen Aktivitäten der Landesregierung insgesamt im Bereich der Weiterbildung sowie die Verbesserung der Daten- und Berichtslage über die Entwicklung im Bereich der Weiterbildung des Landes“ (DIE 2011, S. 40f.).

Die Evaluationen der Jahre 1997, 2004 und 2009/10 unterschieden sich methodisch deutlich voneinander. Die erste Systemevaluation erfolgte nach dem Peer-Review-Prinzip: eine Gruppe von Experten (aus anderen Ländern, um Interessenkollisionen vorzubeugen) analysierte und begutachtete die Situation der Weiterbildung im Land NRW. Die zweite war gewissermaßen eine interne Systemevaluation, vorgenommen vom Landesinstitut für Schule und Weiterbildung, einer nachgeordneten Einrichtung des Landes. Die dritte erfolgte als Drittmittelauftrag an das bundesweit agierende „Deutsche Institut für Erwachsenenbildung“ in Bonn und enthielt in vermehrtem Maße Anteile der empirischen Forschung (eigene Datenerhebungen). Dabei wurden „Leitfragen“ definiert und ausdifferenziert dem Evaluator DIE an die Hand gegeben:

Evaluationsfragen zum System

Aus diesen Leitfragen ergeben sich folgende Evaluationsbereiche, die – vom Auftragnehmer ausgehend – mit einer Analyse der Wirksamkeit der Weiterbildungsmittel des Weiterbildungsgesetzes zu bewerten und mit Handlungsvorschlägen zusammenzuführen sind.

Hierzu sind folgende Schritte nötig:

1. Es ist zu prüfen, ob und wie das in § 11 Abs. 2 WbG definierte Pflichtangebot und das entsprechende Angebot nach § 16 Abs. 2 WbG seit 2000 ausgestaltet und entwickelt wurde. Gleichzeitig werden Empfehlungen erwartet, ob noch eine stärkere Konzentration der Förderung auf staatliche Kernaufgaben, an denen ein besonderes gesellschaftliches Interesse besteht (Gemeinwohlorientierung) geboten ist. Es sind Wege aufzuzeigen, wie mit diesen Angeboten gesellschaftliche Gruppen erreicht werden, die besonderer Unterstützung bedürfen.
2. Es sind Vorschläge zu entwickeln, wie insbesondere mit den Weiterbildungsmitteln des Landes die Stärken der pluralen Bildungslandschaft ausgebaut und ihre Innovationskraft entfaltet werden kann. Hierzu werden auch Aussagen erwartet, welche Auswirkungen eine Berücksichtigung bisher nicht nach dem Weiterbildungsgesetz geförderter freier und privater Träger, die sich aber bereit erklären, die Landesförderung vollumfänglich im gemeinwohlorientierten Förderbereich (§§ 11 Abs. 2 und 16 Abs. 2 WbG) einzusetzen, haben würden.
3. Es ist zu prüfen, ob die Gestaltungsfreiheit der Kommunen bei der Wahrnehmung der Pflichtaufgabe noch weiter erhöht werden kann. Dazu gehört auch die Frage, ob und wenn ja, welche vom Land für die Kommunen gesetzte Standards und Normen entfallen bzw. abgesenkt werden können, z.B. indem die Zuweisungen des Landes an die Kommunen als Pauschalen gezahlt werden. Dabei sind auch die Konsequenzen aufzuzeigen, die sich aus der Systematik des Weiterbildungsgesetzes für die Förderung der Weiterbildungseinrichtungen in anderer Trägerschaft ergeben.
4. Ferner ist zu prüfen, ob und gegebenenfalls wie die besondere Förderung der Angebote zum Nachholen von Schulabschlüssen im Rahmen des Weiterbildungsgesetzes (Zweiter Bildungsweg) durch Einbeziehung der Prüfmitteilung des Landesrechnungshofs und im Kontext der vergleichbaren Bildungsangebote des Berufskollegs, des Weiterbildungskollegs und andere Anbieter durch Nutzung fachlicher, organisatorischer und finanzieller Synergien optimiert werden kann.

5. Vorschläge sind zu unterbreiten, ob die ressortspezifischen weiterbildungspolitischen Ansätze noch stärker zu einer landeseinheitlichen Weiterbildungspolitik fortentwickelt werden können.

Quelle: Auszug aus der Leistungsmittelbeschreibung der „Evaluation der Wirksamkeit der WbG-Weiterbildungsmittel“, S. 1ff.

Systemevalutationen in der Weiterbildung unterliegen einer eigenen Dynamik. Da es keinen staatlich verordneten Zugriff zu den Einrichtungen und Trägern der Weiterbildung gibt (ein wichtiger Unterschied zum Schul- und Hochschulbereich), basiert eine Datenerhebung auf deren freiwilliger Teilnahme. Diese setzt voraus, dass Sinn und Ziel der Evaluation akzeptiert sind und dass im Verfahren der Evaluation selbst ein Höchstmaß an Transparenz herrscht. Insbesondere in der dritten Evaluation in NRW wurde dies in Form einer „partizipativen“ Evaluation umgesetzt, einem Ansatz, der weitgehend der „responsiven“ Evaluation der allgemeinen Typologie entspricht (→ Kap. 2.4).

Auch die Verwendung der Evaluationsergebnisse hat eine eigene Dynamik, ebenso wie Anlass und Auftragserteilung: Beides findet im politischen Raum statt, auf den sich die Evaluation bezog. Gerade in ihm aber haben auch die Träger und Einrichtungen der Erwachsenenbildung (meist über ihre jeweiligen Verbände) eine starke Vertretung und einen starken Einfluss.

ZUR REFLEXION

- Gegenüber welchen Gegenständen und Objektbereichen der Erwachsenenbildung spielt das Evaluationsverfahren eine wichtige Rolle?
- In welchen Fällen trägt die Evaluation mehr zur Entwicklung und Verbesserung bei als zur Kontrolle und Legitimation?
- Welche Interessenstrukturen und Interessenten gibt es in der Erwachsenenbildung gegenüber Evaluationen?
- Kennen Sie Beispiele für die Evaluation von Lehr-/Lernprozessen, Lernergebnissen, Einrichtungen, Projekten und Programmen sowie für das Weiterbildungssystem als Ganzes? Für wie wichtig halten Sie diese?



Lektürecmpfehlungen

- Hartz, S./Meisel, K. (2011): Qualitätsmanagement. 3. akt. und überarb. Aufl. Bielefeld
- Nuissl, E./Schlutz, E. (Hg.) (2001): Systemevaluation und Politikberatung. Gutachten und Analysen zum Weiterbildungssystem. Bielefeld
- Royse, D. u.a. (2001): Program Evaluation. An introduction. Belmont

4. Evaluation – Wie läuft sie im Bildungsbereich idealtypisch ab?

Trotz aller Unterschiede in der konzeptionellen Anlage und im Ablauf folgen Evaluationen doch einem gewissermaßen idealtypischen Modell. Dieses ergibt sich durch die sachlichen Notwendigkeiten: Evaluationen sind zeitlich befristete Aktivitäten, deren Ergebnis ein begründetes Werturteil enthalten soll. Sie haben also eine analoge Gestalt wie Lehr-/Lernprozesse: Beginn (Input), Durchführung (Throughput) und Ende (Output). Im Fall der Evaluation ist jedoch der Beginn genauer zu betrachten, da hier der „Auftrag“ stärker dominiert als das „Lehrziel“ im pädagogischen Prozess. Auch Outcome und Impact als Folgen der Evaluation bedürfen ihrer politischen und kontextuellen Qualität wegen einer ausführlicheren Betrachtung.

In diesem Kapitel werden die Stufen des Evaluationsprozesses nachgezeichnet. Erläutert werden zunächst die beiden wesentlichen Aspekte des Beginns, der Auftrag (→ Kap. 4.1) sowie Konzept und Planung (→ Kap. 4.2). Der darauf folgende Abschnitt erläutert wesentliche Aspekte der Realisierung (→ Kap. 4.3) und der abschließende Abschnitt schließlich enthält Hinweise auf das Evaluationsergebnis und den Umgang damit (→ Kap. 4.4).

Nach der Bearbeitung des Kapitels sollten Sie in der Lage sein,

- die Grundstruktur einer Evaluation nachzuzeichnen,
- die wesentlichsten Schnittstellen einer Evaluation zu benennen,
- die Anforderungen an die Qualität einer Evaluation zu erläutern und
- Evaluationsergebnisse hinsichtlich ihres Zustandekommens zu überprüfen.

4.1 Der Auftrag zur Evaluation

Evaluation beginnt mit dem Evaluationsauftrag. Dieser kann sich an interne Personen oder Instanzen richten oder nach außen vergeben werden. Im Falle der internen Vergabe handelt es sich in der Folge um eine „interne Evaluation“ (gelegentlich auch „Selbstevaluation“ genannt), im Falle der Vergabe nach außen um eine „externe Evaluation“ (gelegentlich auch „Fremdevaluation“ genannt). Eine Mischung von beiden entsteht, wenn die über- oder vorgeordnete Instanz (z.B. ein Betrieb gegenüber Abteilungen, ein Ministerium gegenüber nachgeordneten Behörden) den Auftrag zu einer Evaluation erteilt, die selbst durchzuführen ist – also letztlich eine von außen initiierte interne Evaluation. Nicht selten werden interne und externe Evaluation auch kombiniert; so

vergibt die Europäische Union etwa seit einigen Jahren Projekt- und Programmaufträge mit der Auflage, beide Evaluationen durchzuführen, wobei sie jeweils unterschiedliche Aufgaben beinhalten.

BEISPIEL

Kombinierte interne und externe Evaluation. Das Projekt CReBus

Im Projekt „CReBus“ (Creating a Business in the Digital Age), gefördert in den Jahren 2011 bis 2012 von der Europäischen Union im Rahmen des „Lifelong Learning Programme“, hieß das Workpackage 8 „Evaluation and Quality Assurance“.

Die Ziele des Arbeitspakets waren die interne und die externe Evaluation und werden folgendermaßen beschrieben:

„The aim of this workpackage is to provide the partnership with an evaluation plan and the tools for internal and external evaluation, both formative and summative, of the work undertaken and of the project outcomes. It will also develop a Quality Assurance Plan (criteria and benchmarks). The evaluation will be ongoing throughout project lifetime. A quality management plan of the didactic, organizational, support and technical concepts well as for the guide content and discussion boards from the website and social network will be set up in the first three months of the project and will include formative evaluation measures which integrate feedback from final users, intermediate users and stakeholders in a rather qualitative approach (questionnaire which will be completed at the end of each meeting, using online tools like Google Forms to facilitate the application and analyze of the data). The external evaluator will also organize focus groups with the partners after every project using SWOT analysis and other specific methods.

The Evaluation will focus on the:

- transnationality – the success of transnational working and the effectiveness of partners' contribution,
- functionality of the partnership – the overall management and administration of the partnership and the prospect of future collaboration,
- dissemination and volarization – whether the project outputs have reached a wide audience,
- quality and validity of outputs – whether the needs addressed have been met, and whether the products fulfill the quality criteria agreed at the beginning of the partnership's work,
- innovation – how genuinely new the products are and what their European added value is.“

Zur externen Evaluation heißt es bei CReBus:

„The external evaluator will provide a report after each project meeting which will include conclusions, recommendations, suggestions, and directions for improving the project and a SWOT analysis of the meeting. The external evaluator will have access to the internal part of the website and will upload the documents there. His evaluation will not focus on the partnership/process aspect, but will provide an objective feedback on the relevance of the products for the end users.“

Interne und externe Evaluation haben jeweils Vor- und Nachteile. Die Entscheidung darüber, welche Evaluation stattfinden soll, hat diese Vor- und Nachteile zu reflektieren mit Blick auf das Evaluationsziel.

Vor- und Nachteile interner und externer Evaluation in der Erwachsenenbildung

interne Evaluation

- mögliche Vorteile:
Nähe zum Gegenstand, Identifikation der Beschäftigten/Betroffenen, unmittelbarer Nutzen, besserer Zugang zu Daten und Personen, kürzere Dauer, Kostenersparnis
- mögliche Nachteile:
Verbleib „blinder Flecken“, „Schmoren im eigenen Saft“, unprofessionelle Methoden, fehlende Distanz, geringere Glaubwürdigkeit und Legitimation

externe Evaluation

- mögliche Vorteile:
professionelle Arbeit, systematisches Vorgehen, wissenschaftliche Akzeptanz, „Objektivität“, Legitimation, mehr „Impact“
- mögliche Nachteile:
Zugangs- und Akzeptanzprobleme, längere Dauer, höhere Kosten, Zielkonflikte, Kommunikationsbarrieren

Der Auftrag zur Evaluation setzt gewöhnlich einen Anlass voraus. Dies kann, aus der Einrichtung heraus, etwa ein Budgetproblem sein (zu hohe Kosten in einzelnen Arbeitsfeldern) oder ein Problem des (strategischen) Managements: Zweifel an der Richtung der Einrichtung. Es kann der Abschluss eines Kurses sein, über dessen Qualität man sich informieren möchte (Teilnehmer-Schlussbefragung), oder das Fehlen von Entscheidungswissen für Programmplanung und Angebotssteuerung.

Der Anlass für eine externe Evaluation kann in der (vorgeschriebenen) Notwendigkeit der Kontrolle liegen (z.B. bei Akkreditierungen), in der regelmäßigen Überprüfung von Zieleinhaltung und Arbeitsschwerpunkten, in den Steuerungsverfahren bei Programmen und Projekten, in dem Bedarf an Daten und Urteilen für bevorstehende Entscheidungen (z.B. beim Erlass von Gesetzen oder beim Entwickeln von Förderprogrammen).

Im Anlass für die Vergabe eines Evaluationsauftrages ist immer auch der Grund für die Evaluation enthalten, manchmal nicht explizit, aber doch erkennbar. Die Kenntnis dieses Grundes ist wichtig zur Einordnung der Zielsetzung der Evaluation – er bestimmt in der Regel auch die Definition des Zieles. Das Ziel wiederum bestimmt letztlich alle weiteren Aspekte der Evaluation: erforderliche Daten, Vorgehensweise, Bewertungskriterien und Behandlung des Ergebnisses. Es erfüllt damit bei Evaluationen die Funktion der Forschungsfrage bei Forschungsprojekten.

Zielvorgaben

Das *Evaluationsziel* kann exakt formuliert sein mit Blick auf die Interessen, die mit dem Auftrag verbunden sind. Diese Exaktheit besteht in folgenden Dimensionen:

- Angabe des Inhalts der Evaluationsergebnisse (Zu welchen Aspekten werden Aussagen gemacht?)
- Angabe des Abstraktionsniveaus der Ergebnisse (Wie allgemein werden die Ergebnisse formuliert?)
- Angabe der zu verwendenden Daten und Informationen (Welche Daten liegen den Ergebnissen zugrunde?)
- Angabe der Reichweite des Ergebnisses (Was ist in welchem Zeitrahmen und welchem Umfang von den Ergebnissen betroffen?)
- Angabe der Verfahrensschritte der Evaluation (In welchen Abschnitten und mit welchen Methoden erfolgt die Evaluation?)
- Angaben möglicher Modifikationen im Verlaufe der Evaluation (Inwieweit ist das Ziel aufgrund der Feldbedingungen und Zwischenerkenntnissen modifizierbar?)
- Angabe der erwünschten Empfehlungen/Aussagen (In welche Richtung sollen die Ergebnisse und die Empfehlungen gehen?)

Viele Evaluationsaufträge enthalten Aussagen nur zu einem Teil dieser Dimensionen. Das ist nicht selten Absicht, um den Spielraum der Evaluatoren nicht zu sehr einzuzengen. Im Extremfall können Evaluationen ohne Zielangabe in Auftrag gegeben werden; dies hat Michael Scriven als „Goal-free Evaluation“ schon 1974 (vgl. S. 137f.) beschrieben. Danach sollen *alle* Elemente eines Programms und seiner Wirkungen in den Blick der Evaluation kommen. Der Evaluator hat während der Konzeptionsphase der Evaluation sogar den Kontakt zu den jeweiligen Verantwortlichen (Programme, Einrichtungen etc.) zu meiden, um seine Analyse- und Bewertungsfragen nicht zu beeinflussen. Da letztlich eine Evaluation ganz ohne Ziel und Kriterien aber nicht möglich ist, tritt hier an die Stelle der (fehlenden) Vorgabe der Auftraggeber der vermutete (und analysierte) Zielgruppenbedarf, also der Bedarf der Nutznießer des evaluierten Gegenstands. Solche „Goal-Free-Analysen“ fanden und finden in der Erwachsenenbildung weniger als Evaluationen sondern eher als Forschungen – meist zu Lehr-/Lernprozessen – statt.

Zur Einschätzung der Auftragsziele ist zu berücksichtigen, in welchem Verhältnis sie zum Evaluationsgegenstand stehen. Zielvorgaben können, ob sie nun intern oder extern erfolgen, den Gegebenheiten des Evaluationsobjekts näher oder ferner sein. So erfolgen Zielvorgaben (verbunden mit ihren Kriterien, → Kap. 6.1) häufig auf der Basis normativer Setzungen, nicht jedoch auf der Basis der konkreten Situation des Gegenstands. Das kann zu Diskrepanzen und letztlich auch wenig konsistenten Ergebnissen führen, wenn die Normvorgaben und der konkrete Gegenstand zu weit auseinanderliegen, praktisch Äpfel anhand von Birnenkriterien beurteilt werden.

Ein weiterer wichtiger Aspekt zur Beurteilung der Auftragsziele ist der Spielraum, den sie für die Evaluation eröffnen. Auch wenn die Zielformulierungen alle genannten

Dimensionen abdecken, können sie doch Varianten und Modifikationen zulassen, die sich bei der Konzeption der Evaluation oder ihrem Verlaufe ergeben. Solche Spielräume sind meist erforderlich, um bei der „Passung“ der Evaluation an die Bedingungen des Gegenstands den praktischen Erfordernissen Rechnung zu tragen. So sind etwa Datenerhebungen zu bestimmten Zeiträumen sinnvoll, in anderen nahezu unmöglich⁵ – Verfahren und Zeitabläufe haben sich darauf einzustellen.

Schließlich ist bei der Auftragsvergabe zu berücksichtigen, welche Kommunikationsvorgaben die Evaluatoren zur Rückkoppelung mit dem Auftraggeber bekommen, welche Berichts- und möglicherweise Publikationspflichten vorgesehen sind. Dies gilt auch – etwa bei Programm- und Systemevaluations – für die Rückkoppelung mit den Evaluationsobjekten, mit dem „Feld“. Berichte und andere Kommunikationsformen erhöhen die Transparenz, Zielgenauigkeit und Akzeptanz, sind aber auch arbeitsaufwendig und erhöhen die Möglichkeiten von Interessenten (Auftraggeber, Feld), starken Einfluss auf die Evaluationsergebnisse zu nehmen.

Ein überaus wichtiger Punkt beim Auftrag, insbesondere natürlich beim externen Auftrag, ist die Finanzierung. Sie ist meist an Zeiträume und Meilensteine (also Berichtszeitpunkte) gekoppelt. Die Finanzierung ist daraufhin zu überprüfen, ob sie ausreichend ist für die vorgesehenen Arbeiten, ob in ihr Auflagen enthalten sind, welche die Evaluation in der einen oder anderen Weise beeinflussen und ob sie flexibel ist für Veränderungen oder Verlängerungen im Verlaufe der Evaluation. Insbesondere die Finanzierung der Nacharbeit, also des Publizierens und Vertretens der Evaluationsergebnisse, wird meist nicht ausreichend im Evaluationsbudget berücksichtigt.

4.2 Konzept und Planung

Ist der Auftrag einmal erteilt, geht es an die Konzeption und die Planung der Evaluation. Meist erfolgt dies bereits vor der Auftragsvergabe in einem gewissen Umfang: Evaluationen werden, wenn es sich um externe Evaluationen handelt, im Prinzip immer „ausgeschrieben“. Das bedeutet, dass das geplante Evaluationsvorhaben vom Auftraggeber veröffentlicht und zur Bewerbung freigegeben wird (manchmal auch nur in einer „begrenzten“ Ausschreibung, d.h. nur eine kleine Anzahl bestimmter, als qualifiziert angesehener potentieller Bewerber erhalten die Ausschreibung), und dass der „Zuschlag“, also die Auftragsvergabe, auf der Basis eines bereits eingereichten Konzepts erfolgt. Allerdings sind in der Ausschreibung bereits die wesentlichen Elemente des geplanten

5 Ein einfaches Beispiel dafür ist im Bildungsbereich die Ferienzeit, insbesondere die Sommermonate: In dieser Periode liegt der Rücklauf von Fragebögen im nicht auswertbaren Anteilsbereich. Daher sind Erhebungen, die möglicherweise einen „Nachlauf“ benötigen, mit ausreichendem Abstand vor den Ferien zu beginnen.

Auftrags enthalten, so dass sich auch in diesem Fall das Konzept auf den Auftrag bezieht – zumal dann im Zuge der Auftragsvergabe noch „nachgesteuert“, das Angebot also noch enger mit dem Auftrag abgestimmt wird. Der erste Schritt der Konzeption besteht darin, sich ein „Bild“ von den zu leistenden Aufgaben und ihrem Verlauf zu machen. Ein solches Bild könnte aussehen wie der Phasenaufbau des partizipativen Evaluationsansatzes des „Centrum für Evaluation“ (CEval) (wobei „partizipativ den Einbezug der „Akteure“ zu einzelnen Phasen bedeutet) (→ Abbildung Seite 62).

Die Konzeption kann detailorientiert, aber auch allgemeiner sein. Sie gibt jedoch immer Antwort auf folgende Fragen:

- Was genau wird evaluiert? (Beschreibung des Gegenstands → a)
- Welcher theoretische Ansatz wird zugrunde gelegt? (Bezug zur Wissenschaft → b)
- Mit welchem Ziel wird evaluiert? (Detaillierung des „Zielkorridors“ → c)
- Welchen Zweck verfolgt die Evaluation? (Angaben zu Umsetzung und Nutzen → d)
- Anhand welcher Kriterien wird evaluiert? (Erläuterung des Bewertungsrasters → e)
- Welche Informationen werden herangezogen? (Angaben zur geplanten Datenbasis → f)
- Wie wird der Zugang zum Evaluationsgegenstand geregelt? (Akzeptanzfragen → g)
- Wie wird evaluiert? (Angaben zum methodischen Vorgehen → h)
- Welche „Meilensteine“ sind vorgesehen? (Angaben zu Berichten, Kommunikationen, Zwischenergebnissen, Abstimmungen → i)
- Wie sind die Zuständigkeiten geregelt? (zwischen Auftraggebern, Evaluatoren und Evaluationsobjekten → j)
- Wie wird kommuniziert? (während der Evaluation mit dem Auftraggeber, dem „Feld“ und der Wissenschaft, im Anschluss mit den Nutzern → k)

Betrachten wir nun im Folgenden diese elf Grundfragen des Konzepts einzeln und genauer (unten werden sie in einer Grafik zusammengefasst):

- a) Der Gegenstand ergibt sich durch das Evaluationsinteresse. Um ihn für eine Evaluation bearbeitbar zu machen, ist er zu differenzieren. Es geht es nicht nur um Gegenstandsbereiche (Lernprozesse, Einrichtungen, Lernergebnisse, Programme und das System), sondern um eine detailliertere Festlegung der Elemente des Gegenstands, die in der Evaluation betrachtet werden sollen. Beim Lehr-/Lernprozess können das etwa die Lehrenden sein oder die eingesetzten Methoden, die Lernzufriedenheit der Teilnehmenden und die Dienstleistungsqualität des Kurses. Beim System kann das die Kooperation der Einrichtungen ebenso sein wie die Finanzierungsstruktur oder die Angebotsschwerpunkte. Die präzise Festlegung des Gegenstands ist Grundlage für alle weiteren konzeptionellen Schritte, wie etwa die Selektion von Daten oder die methodische Vorgehensweise.

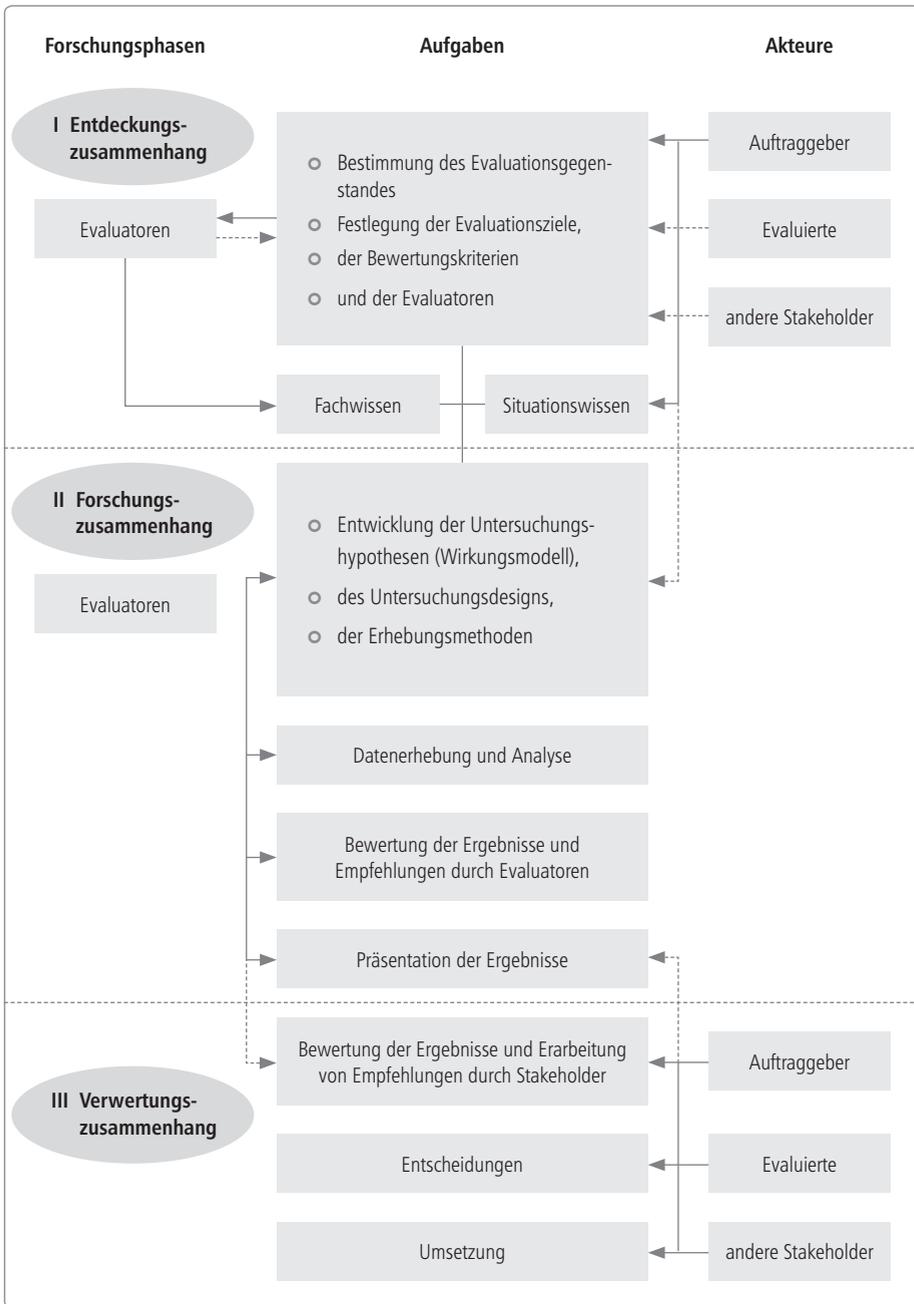


Abbildung 6: Partizipativer Evaluationsansatz (Quelle: Stockmann/Meyer 2010, S. 189)

- b) Jede Konzeption lässt sich in einen theoretischen und wissenschaftshistorischen Kontext einordnen, baut auf früheren Arbeiten und Ergebnissen auf. Diese Zusammenhänge sollten dargelegt werden (leider fehlen sie bisweilen), um die Bonität des vorliegenden Konzepts beurteilen zu können. So lassen sich Konzepte in vorliegende Modelle und Theorien einordnen. Die theoretische und systematische Einordnung erlaubt zudem auch die Möglichkeit eines wissenschaftlichen und/oder methodischen Fortschritts durch die Evaluation.
- c) Die präzise Angabe des Evaluationszieles ist die Arbeitsgrundlage; ohne das Ziel genau zu kennen, kann man nicht abschätzen, ob man auf dem richtigen Weg ist. Das Ziel bestimmt auch die zentralen Fragen der Evaluation. Gerade in größeren Evaluationsvorhaben (wie etwa bei solch großen Programmen wie den „Lernenden Regionen“) sind Zielvorgaben als Entscheidungshilfen unerlässlich. Sie ermöglichen zudem, die Evaluation selbst an ihrem eigenen Anspruch zu messen (Ziel erreicht?), sie also selbst zu evaluieren (quasi eine „Meta-Evaluation“, → Kap. 7.3).
- d) Der Zweck der Evaluation ist nicht identisch mit deren Ziel. So wurde etwa in den 1970er Jahren ein großangelegtes Programm zur Evaluation des Bildungsurlaubs mit dem Ziel gefördert, Erkenntnisse über Vorgehensweise und Wirkungen dieses Bildungsansatzes zu erhalten; der Zweck bestand jedoch hauptsächlich darin, mit dem Verweis auf das Programm ein gefordertes Bundesgesetz hinauszuschieben – ein Auseinanderfallen von Ziel und Zweck, das sich allmählich im Verlauf des Programms selbst entwickelte. Für den Zweck sind die Interessen der Beteiligten, insbesondere der Auftraggeber, zu hinterfragen.
- e) Die Festlegung und Begründung der Bewertungskriterien ist eine zentrale Aufgabe der konzeptionellen Arbeit an der Evaluation. Anders als in manchen empirischen Kontexten können sie nicht „induktiv“ aus dem Datenmaterial abgeleitet werden, sondern sind dem Evaluationsverfahren von Anfang an immanent. Ableitung und Begründung der Kriterien sind eine eigene, hoch anspruchsvolle Tätigkeit (→ Kap. 6.1).
- f) Ohne Informationen kann es keine Evaluation geben. Im Prozess der Evaluation werden diese Informationen zu „Daten“ (analog zu anderen Verfahren empirischer Sozialwissenschaft). Welche Informationen benötigt werden, hängt ab von der detaillierten Beschreibung des Zieles und den darauf aufbauenden Fragen. Danach richtet sich auch die Entscheidung darüber, ob eine eigene Datenerhebung oder eine „Sekundäranalyse“ der vorhandenen Daten erfolgen soll.
- g) Der Zugang zum Evaluationsgegenstand richtet sich nach der Beziehung des Auftraggebers zum Gegenstand (hat er das Direktionsrecht? Verfügt er über Sanktionen?), aber auch nach der Akzeptanz der Evaluatoren. Das Image derselben (im Feld oder als Wissenschaftler) spielt hier eine große Rolle. Schließlich hängt der Zugang zum Feld auch davon ab, welche Arbeiten mit welchem Ressourcenaufwand

vom Untersuchungsgegenstand abverlangt werden. So werden sich Teilnehmende an Kursen weigern, einen zwanzigseitigen Fragebogen auszufüllen, Einrichtungen werden sich schwer tun, Budgetzahlen aus den vergangenen zwanzig Jahren zusammenzutragen, wenn sie diese nicht bereits fortlaufend erhoben hatten.

- h) Das methodische Vorgehen bildet sich vor allem in den Kommunikationen und den eingesetzten Instrumenten ab – gegenüber dem Evaluationsgegenstand. Datenerhebungen sind auf dessen Gegebenheiten abzustellen, Auswertungen und Aufbereitungen von Daten haben den Prinzipien empirischer Sozialforschung zu entsprechen. Die Konzeption muss entsprechende Angaben enthalten (vgl. Nuissl 2010a, S. 52ff.).
- i) Als „Meilensteine“ gelten gewöhnlich die Zeitpunkte, an denen Arbeitsschritte abgeschlossen, begonnen oder berichtet werden. Auch solche, an denen – auf der Grundlage nachgewiesener Leistungen – eine weitere Tranche der Finanzierung erfolgt (so etwa bei einer gestaffelten Drittmittelfinanzierung). Meilensteine sind – je nach Konzeption – auch immer die Schnittstellen, an denen überprüft werden kann, ob sich die Evaluation noch auf dem richtigen Weg zum Ziel befindet und die geplanten Leistungen vorliegen.
- j) Bei der Frage der Zuständigkeiten ergibt sich ein differenziertes und teilweise auch sensibles Bild. Im Prinzip sind die Evaluatoren für die Evaluation zuständig. Es kann aber sein, dass auch die beiden wesentlichen anderen Beteiligten, die Auftraggeber und der evaluierte Gegenstand, für Einzelnes zuständig sind. So liegt die Zuständigkeit für die Vertretung der Evaluation im politischen Raum bei Systemevaluationen etwa beim politischen Auftraggeber (z.B. Ministerium), die Zuständigkeit für das zeitlich und inhaltlich korrekte Zusammenstellen vorhandener Daten bei den Untersuchungsobjekten. Die Konzeption hat sicherzustellen, dass es keine „Überlappungen“ und keine Konflikte bei den Zuständigkeiten gibt. Auch ist klarzustellen, in wieweit Zuständigkeiten auch Verantwortlichkeiten enthalten.
- k) Die Kommunikation im Verlauf der Evaluation, aber auch danach, ist nicht selbstverständliches Beiwerk, sondern muss hinsichtlich Funktion, Verfahren und Verbindlichkeit festgeschrieben werden. Bezüglich des Auftraggebers erfolgt dies durch Berichte und Abstimmungen, aber auch durch Problemlösungen und Außenvertretungen, bezüglich der Evaluierten durch Information, Transparenz und Beteiligung, etwa auch durch die Möglichkeit der kommentierenden Stellungnahme zum Evaluationsergebnis. Auf jeden Fall bedarf eine gelingende Evaluation immer eines ausgewiesenen Anteils von Kommunikation. Dies gilt besonders dann, wenn – aus unterschiedlichen Gründen – das evaluierende Vorgehen verändert werden muss.

Die Konzeption der Evaluation ist die Grundlage für deren Realisierung, aber – das zeigt sich immer wieder – nicht mit ihr gleichbedeutend. Evaluationen sind Prozesse, in denen immer wieder Probleme und neue Sachverhalte auftreten können und zu Veränderungen führen. Dies bedeutet auch, dass das Konzept im Verlauf einer Evaluation, besonders wenn sie prozessbegleitend von längerer Dauer ist (wie im Falle des Programms „Lernende Regionen“ acht Jahre), modifiziert werden kann. Dies kann aber auch nur eine Erhebungsmethode betreffen, wenn etwa bei einer Systemevaluation der Fragebogen-Rücklauf nicht ausreicht und durch andere Erhebungsverfahren substituiert werden muss.

Die Qualität der Konzeption und ihrer Umsetzung ruft bei Evaluationen immer wieder eine kritische Diskussion hervor. Anders als viele Forschungen steht Evaluation stets auf dem Prüfstand der Praxis und der Tauglichkeit, der Zuverlässigkeit und Akzeptanz. Dies gerade auch deshalb, weil Evaluation ein „Geschäft“ ist, mit dem ordentlich Geld verdient werden kann. Es ist daher naheliegend, dass die Deutsche Gesellschaft für Evaluation (DeGEval) schon kurz nach ihrer Gründung Standards formulierte, die nach „innen“ (gegenüber den Evaluatoren) Qualitätsansprüche ebenso wie moralische Prinzipien enthält und nach außen die Seriosität von Evaluation betont (Kurzform der DeGEval-Standards; vgl. ebenfalls Stockmann/Meyer 2010, S. 183).

4.3 Realisierung

Neben der Flexibilität, die eine Konzeption im Bedarfsfall aufweisen muss, hat die Umsetzung des Geplanten ihre eigenen Anforderungen. Zunächst ist relevant, wer an der Evaluation beteiligt ist, konkret: Welche Personen evaluieren, leisten die mit der Evaluation verbundene Arbeit? Da es bei Evaluationen in der Regel um befristete Aktivitäten geht mit kürzeren (einige Monate) über mittlere (etwa drei Jahre) und langen (bis zu zehn Jahren) Laufzeiten, handelt es sich dabei auch immer um befristete Arbeitsverträge. Dies sowie andere Eckwerte (Finanzierung, Einsatzbedingungen etc.) führen meist dazu, dass jüngere Akademiker und Akademikerinnen hiermit ihre ersten wissenschaftlichen Erfahrungen sammeln. Sie verfügen über eine hohe Arbeitsmotivation und großes Engagement, aber zumeist geringe Feldkenntnisse und Erfahrungen in empirischen Arbeiten. Auch ist es für sie dadurch nicht immer leicht, Zugang zum evaluierten Feld zu finden.

Gerade in Evaluationen mit einem hohen Anteil empirischer Forschung sind hier Anleitungstrukturen gefragt, die personale Kombinationen mit erfahreneren Kollegen implizieren. Nicht immer leicht, diese Kombination, und nicht immer spannungsfrei, aber auch voller Chancen und Entwicklungen. Es gibt kaum Berichte aus und über Evaluationen, welche diese personale Seite thematisieren, obwohl sie doch für die Realisierung der Evaluierung sehr bedeutsam ist.

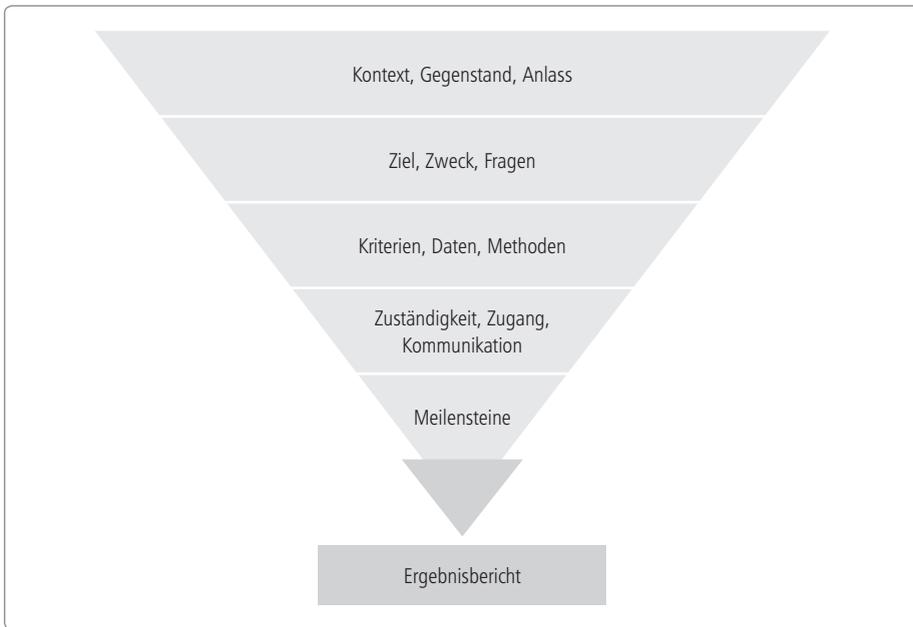


Abbildung 7: Realisierung der Evaluation (Quelle: eigene Darstellung)

Ist das Evaluationsteam zusammengestellt, sind die Aufgaben definiert und verteilt, beginnt der eigentliche Evaluationsprozess. Er beginnt – wie üblicherweise bei zielgerichteten Vorhaben – mit einem Plan. Es ist ein Plan, der das Konzept in einen zeitlichen, personalen, organisatorischen sowie einen Ablaufplan übersetzt. Zeitlich bedeutet, festzulegen, welche Arbeitsschritte in welcher Reihenfolge zu erledigen und wann sie jeweils abzuschließen sind (das sind dann die „milestones“). Personal bedeutet, festzulegen, wer die Arbeit leistet und wer dafür verantwortlich ist. Organisatorisch bedeutet, Orte und Mittel (z.B. für Reise oder Kommunikation) zu fixieren. Der Ablaufplan ist schließlich die Zusammenstellung des Ganzen in einer für alle Beteiligten verbindlichen Struktur.

Der erste rein inhaltliche Schritt im Evaluationsablauf ist die Präzisierung der Fragen, die anhand von Daten und Informationen zu beantworten sind. Ohne präzise Fragen kann man nicht wissen, welche Daten benötigt werden. Die Fragen sollten einen mittleren Abstraktionsgrad aufweisen, nicht zu allgemein, aber auch nicht zu detailliert sein (vgl. Nuissl 2010a, S. 18ff.). Sie sollten auch zu einzelnen Blöcken zusammengefasst sein, welche die Aspekte des Evaluationsgegenstands beschreiben. Bei einer Evaluation von Erwachsenenbildungseinrichtungen könnte das wie folgt aussehen:

BEISPIEL

Evaluationsfragen Bildungseinrichtung**Budget:**

- Wie ist die Einnahmenstruktur? (Quellen, Höhe, Auflagen, Sicherheit)
- Wie verteilen sich die Ausgaben auf Kostenstellen, Kostenarten und Kostenträger?
- Wo liegen unterfinanzierte, wo überfinanzierte Bereiche? (Organisationsstruktur)
- Welcher Anteil der Kostendeckung wird jeweils erzielt? (drei Deckungsstufen)⁶

Personal:

- Wie viele Beschäftigte mit welchem Einkommen/Arbeitsvertrag hat die Einrichtung?
- Wie ist das Verhältnis von planendem, lehrendem, verwaltendem Personal?
- Wie und wie klar sind die Aufgaben definiert und verteilt?

Angebot:

- Wo liegen Schwerpunkte des Angebots inhaltlicher und organisatorischer Art?
- Welche Bedeutung haben sie jeweils für die Ziele der Einrichtung?
- In welchem organisatorischen und inhaltlichen Verhältnis stehen sie zueinander?

Dieser Fragenkatalog ist nur beispielhaft (somit unvollständig) zu verstehen; er zeigt die Ebene der Fragen und ihre immanente Struktur. Die einleitenden Fragen sind meist deskriptiv, richten sich auf eine Zustandsbeschreibung. Darauf folgen eher analytische Fragen zu Zusammenhängen und Entwicklungen. In der Regel enthalten die Fragen auch bereits Hinweise auf die anvisierten Merkmale bzw. Indikatoren sowie auf geplante Korrelationen, also Prüfungen von Zusammenhängen.

Üblicherweise werden die Fragen in der Evaluation – ob explizit oder implizit – bereits auf der Basis von Hypothesen, d.h. von Annahmen über zu betrachtende Aspekte, gestellt. Etwa die Frage „Wo liegen unterfinanzierte, wo überfinanzierte Bereiche?“ basiert auf der Annahme, dass es solche Bereiche gibt, und zugleich auch auf der Annahme, dass das ein für die Entwicklung der Einrichtung wichtiger und zu überprüfender Sachverhalt ist.

Auf der Basis dieser Fragen wird überprüft, welche benötigten Daten vorhanden und gegebenenfalls zu bearbeiten sind (→ Kap. 5.2) und welche Daten für die Evalua-

6 „Deckungsstufe“ bezeichnet die betriebswirtschaftliche Seite einer Maßnahme bezogen auf die Deckung der entstehenden Kosten – in der Regel Deckungsstufe 1 = Finanzierung des Dozenten honorars, Deckungsstufe 2 = Finanzierung aller maßnahmebezogenen Kosten (also auch Raum, Heizung etc.), Deckungsstufe 3 = alle maßnahmebezogenen Kosten und zusätzlich die umgelegten Kosten des „Overheads“ (also Öffentlichkeitsarbeit, Direktor etc.).

tion gezielt erhoben werden müssen (→ Kap. 5.3). Im ersteren Fall werden die Daten bei der Einrichtung angefordert, im letzteren ein Erhebungsplan erstellt mit Angaben dazu: welche Art von Erhebung, wann und wie durchzuführen, wie aufzubereiten.

Sind Einrichtungen der Evaluationsgegenstand, hat sich inzwischen ein Verfahren bewährt, nach dem die Einrichtungen einen Fragenkatalog erhalten, den sie zu beantworten und mit entsprechenden Daten in der Anlage zu versehen haben. Dies erbringt nicht nur die meisten erforderlichen Daten, sondern erlaubt zugleich einen Blick darauf, welche Aspekte der Einrichtung in welcher Weise wichtig sind. Ein solches Verfahren ist aber nur möglich, wenn der Untersuchungsgegenstand ein artikulationsfähiges Subjekt ist – das „System“ als Untersuchungsgegenstand wird sich schwerlich gezielt äußern können.

Der Erhebungsplan erfordert in der Regel die Entwicklung eigener Erhebungsinstrumente (vor allem: Fragebogen) nach Prinzipien empirisch-analytischer Forschung (→ Kap. 5.3). Auch der Einsatz der Instrumente unterliegt diesen Prinzipien.

In der Sammlung und Aufbereitung der erhobenen Daten besteht meist die Aufgabe, die erhobenen Daten mit den zusammengetragenen, schon existierenden Daten zusammenzubringen. Das impliziert genaue Betrachtungen, Arbeit an identischen Kategorien und Datenprüfung. Dabei ist oft genug auf die „Rohdaten“ zurückzugehen, also die Daten ohne kategoriale Aggregationen (z.B. alle einzelnen Altersangaben der Beschäftigten, nicht gruppiert nach Alterskohorten). Schon vorhandene Daten bedeuten von daher nicht notwendig einen geringeren Zeitaufwand, allenfalls einen geringeren Erhebungsaufwand (→ Kap. 5.4).

Nachdem die Daten gesammelt und aufbereitet sind, kann mit der Analyse begonnen werden. Auch sie folgt im Prinzip den Regeln der empirischen Forschung (→ Kap. 6.2). Die Analyse enthält zunächst nur geringe Anteile von Bewertung; der bewertende Aspekt steckt hauptsächlich in der Selektionsentscheidung, was genau analysiert wird. Aber auch hier unterscheidet sich die Evaluationsforschung nicht von der empirischen Sozialforschung.

Anhand der vorliegenden Analyseergebnisse (z.B. Relation von Beschäftigtenkapazität und Unterrichtsstunden) beginnt die eigentliche Bewertung; am genannten Beispiel heißt das: zu viele Unterrichtsstunden pro Beschäftigte oder zu wenig? Hier werden die Kriterien wirksam, die herangezogen werden oder werden sollen, und deren Ableitungszusammenhang. „Zu viel“ oder „zu wenig“ kann man in diesem Beispiel danach beurteilen, wie dies in vergleichbaren Einrichtungen geregelt ist; danach beurteilen, wie die Qualität des Angebots ist; danach, wie sich die Relation über die Jahre hinweg verändert hat; danach, wie hoch die Investition in einen bestimmten Bereich sein soll; danach, wie zufrieden die Teilnehmenden sind; danach, was als Norm vorgeschrieben ist; usw.

Je nach gewähltem Evaluationsverfahren unterliegt die Bewertung noch einer „Schleife“, einer Rückkoppelung mit Vertretern des Untersuchungsgegenstands, die

ihre Sicht der Bewertung und der angewandten Bewertungskriterien vortragen. In der Regel dient dies nicht nur zu einer Schärfung der Bewertung, sondern auch zu ihrer höheren Akzeptanz, was vor allem bei dann folgenden Empfehlungen von Bedeutung ist. Spätestens in dieser „Rückkoppelungsschleife“ bestehen die Möglichkeit und die Notwendigkeit, Widerstände gegen die Evaluierung aufzulösen, zumindest aber transparent und bearbeitbar zu machen. Denn

es gibt zahlreiche Widerstände gegen Evaluationen von Bildungsmaßnahmen, die häufig gar nicht bemerkt oder aber in ihren Auswirkungen unterschätzt werden. Überschätzt wird dagegen wohl eher die Kooperationsbereitschaft der Beteiligten (B. Zuschlag in Will/Winteler/Krapp 1987, S. 75).

Das Erkennen, Bearbeiten und Auflösen von Widerständen im Prozess der Datenerhebung ist meist derjenige Teil der praktischen Realisierung einer Evaluation, der bei der Konzeption am wenigsten bedacht wurde.

Vielfach sind Widerstände gegen die Evaluation während des Verfahrens selbst motiviert durch die Sorge um das Ergebnis, nicht selten auch zu Recht. Das Ergebnis enthält ja meist nicht nur die bewertenden Urteile zum evaluierten Gegenstand, sondern auch Empfehlungen, die mit dem Ziel der Umsetzung formuliert werden. Insbesondere bei Evaluationen, die von Geldgebern eines Evaluationsobjekts in Auftrag gegeben wurden, sind solche Empfehlungen höchst verbindlich und folgenreich, selbst dann, wenn sie in sich widersprüchlich oder unangemessen gegenüber dem Evaluationsobjekt sind. Wenn etwa einer Einrichtung, die sehr gute Arbeit in der Weiterbildung von Migrant/inn/en gemacht hat, empfohlen wird, diese zwar fortzusetzen, aber zukünftig Angebote für Manager/innen von Betrieben in den Mittelpunkt zu stellen, dann fehlt die Kohärenz der Empfehlungen als auch deren Passung zur Institution. Der Grund für solche Widersprüche liegt meist außerhalb des Evaluationsprozesses – so ist, in diesem Beispiel, aktuell die Weiterbildung von Führungspersonal in Betrieben politisch sehr bedeutsam.

Die Formulierung von Empfehlungen auf der Basis der bewertenden Urteile erfolgt selten als kausale und logische Ableitung, sondern ist noch einmal ein eigener Schritt der bewertenden Betrachtung. In die Empfehlungen gehen weitere Betrachtungen (mit jeweils eigenen und möglicherweise neuen Kriterien) ein, die auch außerhalb des Evaluierungsverfahrens liegen können. Es sind vor allem:

- *Opportunität*

Empfehlungen zeigen die gewünschte Richtung an, in welche die Entwicklung gehen soll; dabei ist äußerst bedeutsam, um wessen Wunsch es sich handelt. Die häufigste Umsetzung von Wünschen in Empfehlungen folgt – erwartungsgemäß – den Auftraggebern, selbst dann, wenn sich die Evaluatoren während des Verfahrens um

Unabhängigkeit und Objektivität bemüht haben. Es spielt die – richtige – Überlegung eine Rolle, dass eine Empfehlung die höchsten Chancen auf Umsetzung hat, wenn sie vom Auftraggeber (der ja meist auch Geldgeber des Evaluationsgegenstands ist) mitgetragen wird.

- *Praktikabilität*
Nach Ziel, Zeitfenster und Anspruchsniveau nehmen die Empfehlungen aus gutem Grund die Möglichkeiten und Grenzen des Evaluationsgegenstands in den Blick, um die Chancen der Umsetzung zu erhöhen und nicht zu überfordern. Es ist eine schwierige Balance, zwischen weiterreichenden Empfehlungszielen und dem Status quo eine praktikable Zielmarke zu formulieren.
- *Kontextualität*
Empfehlungen betrachten nicht nur den (soeben noch) evaluierten Gegenstand, sondern in weiterem Sinne dessen Einbettung in einen Kontext. Das kann in der Erwachsenenbildung die Kommune sein (bei Volkshochschulen) oder die Kirche (bei konfessionellen Einrichtungen) oder die Situation der Zielgruppen (bei zielgruppenorientierten Projekten). Hier gehen Einschätzungen über die Relevanz und die Problematik der entsprechenden Bildungsaktivitäten ein, die auch im Gegensatz zu den Einschätzungen der Evaluierten stehen können
- *Aktualität*
Dies ist ein freundlicher Begriff für das Phänomen, dass Evaluatoren in ihren Empfehlungen dazu neigen, das gerade „Moderne“ als Maßstab und Ziel zu nehmen. Insbesondere bei Peer-Review-Evaluationen zeigt sich diese durchaus negativ einzuschätzende Orientierung an der aktuell dominierenden, an der „herrschenden“ Meinung. Evaluationen unterliegen dabei dem gleichen Problem wie das gesamte wissenschaftliche System, in dem dominierende Strömungen Altbewährtes verwerfen und zugleich Neues bekämpfen (vgl. Max Weber 1922).

Mit der Abfassung des Evaluationsberichts und der darin enthaltenen Empfehlungen ist das eigentliche Evaluationsverfahren abgeschlossen. Die Erstellung dieses Textes kann dann eine Wissenschaft für sich sein, wenn es sich um eine Gruppenevaluation (meist Peer-Review) handelt. Der Text ist dann in der Regel so zu formulieren, dass ihn alle Gruppenmitglieder mittragen – was gerade angesichts des zumeist bewusst und gezielt plural (fachlich und positionell) zusammengesetzten Gremiums ein zähes und langwieriges Unterfangen sein kann mit der großen Gefahr einer fehlenden Kohärenz.

Der Evaluationsbericht hat in der Regel vier Teile: Sachstandsbericht, Analyse, Bewertung und Empfehlungen. Gehen diese Ebenen im Bericht durcheinander, ist kritischer Zweifel angebracht. Die Wissenschaftlichkeit von Evaluationen zeichnet sich gerade in dieser Trennung aus, die der Überprüfbarkeit dient (→ Kap. 7.1).

In den meisten Fällen aber ist die Evaluation mit dem Bericht, seiner Abgabe und Erläuterung (oft auch im öffentlichen Raum) noch nicht beendet. Es kommt die „Nacharbeit“, die Vertretung und Erläuterung der Ergebnisse und Empfehlungen, die Diskussion mit den Stakeholdern und Interessenvertretern, der Transfer in den wissenschaftlichen Raum (der Evaluationsbericht erfüllt nicht immer die Kriterien, die an wissenschaftliche Veröffentlichungen angelegt werden). Letztlich gehört auch diese „Nacharbeit“ zur Evaluation, auch wenn sie (besonders bei Drittmittelprojekten) nicht mehr in den „Förderungszeitraum“ fällt. Ohne sie zeitigt die Evaluation weniger Wirkung, sie kann auch leichter zu anderen als den beabsichtigten Wirkungen führen.

4.4 Ergebnisverwertung

Evaluationen dienen grundsätzlich dazu, zu verwertbaren Ergebnissen zu führen. Der Evaluationsauftrag, die Evaluationsfragen und die Vorgehensweise sind eng an Praxis und Realität gebunden, ebenso auch das, was mit dem erzielten Ergebnis geschieht. Vielfach ist die Verwertungsdimension bereits impliziter Teil des Auftrags oder des Evaluationskonzepts. In der Zielangabe der Evaluation ist implizit oder explizit der Verwertungszusammenhang mitformuliert. Unterscheiden lassen sich insbesondere folgende Verwertungen, die auch in Kombination auftreten können:

- *prozessessteuernde Verwertung*
Die Ergebnisse werden in einen laufenden Prozess eingespeist, um dessen Richtung und Prozedere zu beeinflussen (meist als formative Evaluation), z.B. in der Programmevaluation.
- *entscheidungsbasierende Verwertung*
Die Evaluationsergebnisse dienen dazu, bestimmte Entscheidungen zu begründen oder auch nur zu legitimieren, z.B. in der Akkreditierung.
- *entwicklungsorientierte Verwertung*
Die Evaluationsergebnisse dienen dazu, Problemlagen aufzuzeigen und Entwicklungen anzustoßen, z.B. in der Systemevaluation.
- *qualitätssteigernde Verwertung*
Die Ergebnisse werden verwendet, um die Qualität von Prozessen und Produkten zu verbessern, z.B. in der Evaluation von Lehr-/Lernprozessen.
- *konfliktlösende Verwertung*
Die Ergebnisse werden als Argumente in Streitsituationen verwendet, bspw. in ökologischen oder politischen Auseinandersetzungen.

Es kann aber auch sein, dass die Evaluationsergebnisse überhaupt nicht verwertet werden. Das sind dann Evaluationen, die keine direkte Wirkung zeitigen. Seit Ende

der 1970er Jahre werden verstärkt auch die Wirkungen von Evaluationen untersucht. Dabei wurde festgestellt, dass das Ausbleiben direkter Wirkungen hoch korreliert mit Defiziten der Evaluationsergebnisse bzw. ihrer Präsentation (vgl. Beywl 1988, S. 34ff.):

- Ein Problem der Evaluationsergebnisse kann darin liegen, dass kein messbares Ergebnis erzielt worden ist, etwa dadurch, dass kein Unterschied zwischen dem vorigen und dem folgenden Zustand existiert oder ein Programm oder ein Lernprozess keine erkennbaren Ergebnisse haben.
- Ein Problem der Evaluationsergebnisse kann darin liegen, dass keine oder keine umsetzbaren Empfehlungen gegeben werden; die Umsetzbarkeit von Empfehlungen hängt von ihrer Praxistauglichkeit, ihrer Präzision, ihrer Widerspruchsfreiheit (Kohärenz) und ihrer Reflexion von Interessenstrukturen ab.
- Ein Problem der Evaluationsergebnisse kann darin liegen, dass sie nicht zu Entscheidungen führen und führen können, weil sie nicht eindeutig sind oder in unterschiedliche Richtungen hin ausgelegt werden können.

Diese Probleme der Verwertung liegen überwiegend direkt in der Qualität der Evaluationsergebnisse und ihrer Formulierung. Es gibt jedoch auch andere Gründe, warum Evaluationen keine direkten Wirkungen zeitigen; gerade in Kontexten, in denen die Evaluation mit ihren Ergebnissen Impulse geben, Dinge verändern und Innovationen einleiten soll, gelten alle Widerstände, die auch aus traditionellen Organisationsentwicklungs-Prozessen bekannt sind. Mit ihnen muss man beim Vorlegen von Evaluationsergebnissen immer in der einen oder anderen Weise oder bei dem einen oder anderen Adressaten rechnen. Sie sind besonders stark, wenn bereits während der Realisierung der Evaluation Widerstände bestanden, die nicht geklärt wurden oder werden konnten (→ Kap. 4.3). Stockmann/Meyer (2010, S. 176ff.) haben die häufigsten Widerstände gegen Evaluationsergebnisse zusammengestellt, sie sind gleichermaßen gültig für den Bildungsbereich und die Erwachsenenbildung:

- *„Alles bekannt“!*
Hier wird das Ergebnis als letztlich überflüssig und redundant abgewehrt.
- *„Methodisch falsch“!*
Hier wird das Ergebnis als nicht sachgemäß und begründet entstanden abgewehrt.
- *„Es kann nicht sein“!*
Hier wird das Ergebnis nach dem Prinzip: „Es kann nicht sein, was nicht sein darf!“ abgewehrt.
- *„Fehlerhaft durchgeführt“!*
Hier wird das Ergebnis wegen wirklicher oder behaupteter Verfahrensmängel abgewehrt.

Auch dann, wenn die Ergebnisse der Evaluation nicht direkt Wirkung zeigen, hat die Evaluation dennoch eine Wirkung: Das Zusammentragen der Daten, die Debatte um Sinn und Ziel der Evaluation, die Thematisierung von Problemen und Interessen im Verlauf der Evaluation haben oft einen nachhaltigen Einfluss auf den Untersuchungsgegenstand – gelegentlich mehr als die formulierten Ergebnisse. Ein solcher Einfluss, ein solcher Nutzen wird auch als „Prozessnutzen“ benannt, der oft zu Unrecht deutlich unterschätzt wird (vgl. Stockmann/Meyer 2010, S. 181).

Dennoch: Das wesentliche Ziel der Evaluation muss sein, eine direkte Wirkung der formulierten Ergebnisse zu bewirken und gemachte Empfehlungen Realität werden zu lassen. Am Ende zeigt sich der Wert einer Evaluation darin, dass sie und ihre Ergebnisse „nützlich“ sind. Entsprechend wird diese Nützlichkeit auch in den „Nützlichkeitsstandards“ der Gesellschaft für Evaluation (DeGEval 2002) betont. Diese Nützlichkeit kann auf unterschiedlichen Ebenen erfolgen – auf operativer Ebene als direkte Handlungsanleitung, auf konzeptioneller Ebene als Anleitung zur Schwerpunktbildung und auf kommunikativ/sozialer Ebene als Mittel zur Argumentation und Überzeugung.

Nützlichkeit in den DeGEval-Standards (Übersicht)

- N1 Identifizierung der Beteiligten und Betroffenen
- N2 Klärung der Evaluationszwecke
- N3 Glaubwürdigkeit und Kompetenz der Evaluierenden
- N4 Auswahl und Umfang der Informationen
- N5 Transparenz von Werten
- N6 Vollständigkeit und Klarheit der Berichterstattung
- N7 Rechtzeitigkeit der Evaluation
- N8 Nutzung und Nutzen der Evaluation

Quelle: www.degeval.de/degeval-standards/standards

In vorliegenden Studien zur Nutzung von Evaluationsergebnissen (vgl. Übersicht bei Stockmann/Meyer 2010, S. 180ff.) zeigen sich folgende wichtige Faktoren für die Nutzung von Evaluationsergebnissen, mit anderen Worten: für deren direkte Wirkungen:

- Die Evaluation muss insgesamt für die entsprechenden Entscheidungsträger (etwa Vertreter von Bildungsministerien oder Leiter von Bildungseinrichtungen) und an-

dere „Stakeholder“ (etwa Bildungswissenschaftler) relevant sein, was im Vorfeld der Evaluation schon zu klären ist (*Relevanzfaktor*).

- Die Betroffenen, insbesondere die Entscheidungsträger und die „Stakeholder“, sind in den Prozess der Evaluation (z.B. über die Diskussion von Berichten) permanent einzubeziehen (*Partizipationsfaktor*).
- Die Kommunikation über die Evaluation, ihren Ablauf, ihre Fortschritte, aber auch ihre Schwierigkeiten muss fortlaufend erfolgen gegenüber allen Beteiligten und Betroffenen (*Transparenzfaktor*).
- Die Ergebnisse müssen zum passenden Zeitpunkt vorgelegt werden und nach Umfang, Sprache und Klarheit adressatengerecht, verständlich und eindeutig sein (*Präsentationsfaktor*).
- Die Diskussion über die Ergebnisse muss begleitet und unterstützt, Handlungsempfehlungen müssen gegeben und vertreten werden (*Disseminationsfaktor*).

Die genannten Faktoren machen deutlich, dass eine Evaluation von Beginn an darauf ausgelegt sein muss, die Wirkung der Ergebnisse zu reflektieren und zu steuern. Die Wirkung entsteht nicht erst bei der und durch die Formulierung des Ergebnisberichts. Entsprechend gelten Regeln der „Validation“, der Sicherung der nachhaltigen Wirkung schon zu Beginn, nicht nur bei Projekten und Programmen, sondern auch bei Evaluationen, die ja oft mit ihnen verbunden sind.

ZUR REFLEXION

- Von welchen Faktoren hängt die Entscheidung für eine externe oder interne oder gemischte Evaluation ab?
- Welche Rolle spielen Moral und Legitimation bei der Vergabe eines Evaluationsauftrags?
- Welche Spannungsfelder sind zwischen Geld- und Auftraggebern der Evaluation einerseits und den Evaluationsgegenständen andererseits denkbar?
- Welche Kompetenzen müssen Evaluatoren haben, wie sind sie zu überprüfen, wo kann man sie lernen?



Lektüreempfehlungen

- Stockmann, R./Meyer, W. (2010): Evaluation. Eine Einführung. Opladen
- Steinert, B./Klieme, E. (2008): Evaluation im Bereich der empirischen Bildungsforschung. In: Faulstich-Wieland, H./Faulstich, P. (Hg.): Erziehungswissenschaft. Ein Grundkurs. Reinbek, S. 641–654
- Kuper, H. (2005): Evaluation im Bildungssystem. Eine Einführung. Stuttgart

5. Evaluation – Welche Daten werden wie ermittelt?

Evaluation benötigt Daten. Diese Daten sind entweder schon vorhanden oder müssen für die Evaluation gesondert erhoben werden. In der Regel ist der größere Teil der Daten für die Evaluation bereits vorhanden. Dies liegt vor allem daran, dass in der Evaluation ja wesentliche Elemente des „Gegenstands“, des Evaluationsobjekts, in den Blick genommen werden. Zu diesen wesentlichen Elementen liegen im jeweiligen Eigeninteresse Daten vor.

Geht es bspw. um die Evaluation einer Bildungseinrichtung, so sind Daten zu den Finanzen, den Angebots- und Teilnehmendenzahlen erforderlich, wie auch solche zu den Beschäftigten. Geordnet arbeitende Einrichtungen schreiben diese Datenbestände kontinuierlich fort, schon um für das eigene Management die notwendigen Planungsdaten zur Verfügung zu haben. Zu ergänzen sind diese möglicherweise – jeweils abhängig von der Evaluationsfrage – um Daten zur Region und zum Umfeld, Entwicklungsreihen, Teilnehmendenevaluationen, zu Organisationszielen und -konzepten sowie anderes. Geht es um eine Evaluation des „Systems“, ist die Datenlage meist begrenzter, hier sind zusätzliche Erhebungen die Regel. Im Falle von pädagogischen Prozessen bedarf es, sofern es um den Verlauf geht, ebenfalls weiterer Erhebungen. Im Falle von Projekten und Programmen sind die erforderlichen Daten meist bereits im Programm- bzw. Projektplan als Lieferdesiderate eingeschlossen.

Es ist also letztlich mit zwei Gruppen von Daten zu rechnen – den bereits erhobenen und den noch zu erhebenden –, welche je nach Evaluationsgegenstand in einem unterschiedlichen Verhältnis zueinander stehen. Welche Daten verwendet werden sollen und können, betrifft die Klärung im Rahmen des Evaluationskonzepts (→ Kap. 5.1). Auch bereits vorhandene Daten bedürfen einer Bearbeitung; Ausführungen hierzu finden sich im Abschnitt 5.2. Der Datenerhebung, die auch im Evaluationskontext den Regeln empirischer Sozialforschung folgt – mit einigen Einschränkungen – widmet sich Abschnitt 5.3. Und schließlich sind die vorliegenden Daten für die Analyse und Interpretation noch aufzubereiten; hierzu wird auf Abschnitt 5.4 verwiesen.

Nach der Lektüre sollten die Leser in der Lage sein,

- die Qualität vorliegender Daten für eine Evaluation einzuschätzen,
- angemessene Erhebungsmethoden auszuwählen und anzuwenden sowie
- das Evaluationsdesign kritisch hinsichtlich der erforderlichen und benutzten Daten zu prüfen.

5.1 Evaluationsdesign

Vor der Untersuchung, welche die bewertenden Urteile der Evaluation begründen soll, bedarf es eines Designs, in dem die wesentlichen Fragen, deren Beantwortung eine solche Bewertung erlaubt, formuliert werden. Damit ist nicht die zentrale Evaluationsfrage gemeint, welche das Ziel der Evaluation in Frageform repliziert, sondern gemeint sind die Fragen, die detailliert an das Material und die Daten zu richten sind.

DEFINITION

Evaluationsdesign

Das Untersuchungsdesign bezeichnet eine Skizze, welche Untersuchungsziel, Untersuchungsfragen, Untersuchungsmethode, Untersuchungsprozess und erwartete Untersuchungsergebnisse skizziert. Sie ist vor Beginn aller Untersuchungstätigkeiten zu erstellen und die Grundlage für die Arbeit im gesamten Untersuchungszeitraum.

Im Design geht es darum, die einzelnen Teile und Schritte der Evaluation zusammen zu sehen, zu bezeichnen und in einen vernünftigen und bearbeitbaren Zusammenhang zu bringen. Das Design beantwortet etwa folgende Fragen:

- Welche Evaluationsfrage (einschließlich Unterfragen) habe ich?
- Wie lauten erste und vorläufige, aber leitende Annahmen?
- Wie könnte am Ende die Antwort auf die Ausgangsfragen aussehen?
- Welche Evaluationsgegenstände liegen vor, wie sind sie beschaffen?
- Ist der Zugang zum Untersuchungs-(Evaluations-)objekt gesichert?
- Welche Daten sind für die Evaluation zu erheben?
- Welche Daten sind bereits vorhanden und brauchbar?
- Welche Datenmengen benötige ich, mit welchen muss ich rechnen?
- Wie gehe ich mit unterschiedlichen Datenquellen um?
- Welche Schritte setzen den Abschluss welcher anderen Schritte voraus?
- Welche Arbeiten lassen sich parallelisieren?

Das Design dient dazu, über einen realistischen Arbeits- und Zeitplan zu verfügen sowie unnötige Datenberge und irreparable Datenlücken zu vermeiden. Neben einer Beschreibung der erforderlichen Daten und der Festlegung, woher man sie erhält, ist daher auch wichtig, den verfügbaren Zeitrahmen realistisch zu planen. Evaluationen sind immer befristet, die Ergebnisse haben zu einem angegebenen Zeitpunkt vorzuliegen, und der Zeitraum bis dahin ist nicht beliebig dehnbar. Meist wird daher auch ein „Zeitstrahl“ erstellt, auf dem die wichtigsten Schritte eingetragen sind (vgl. Nuissl 2010a, S. 49).

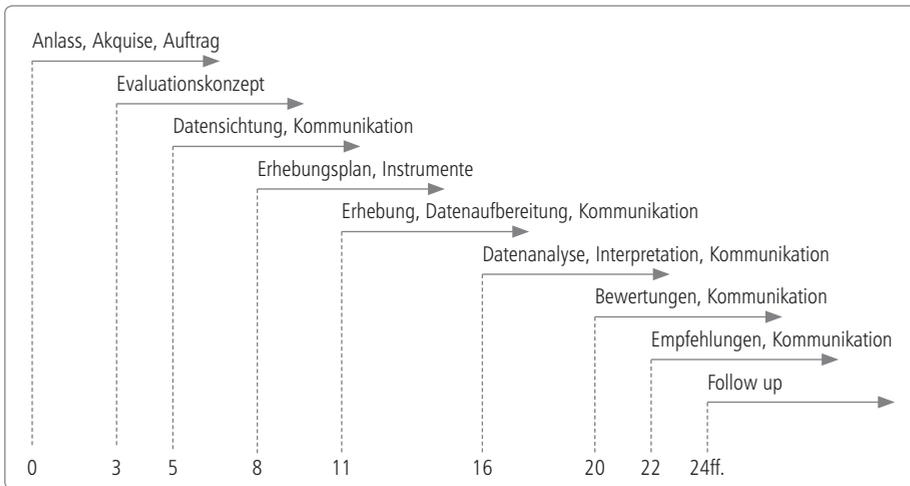


Abbildung 8: Zeitstrahl Evaluation (Modell 24 Monate, angelehnt an Nuissl 2010a, S. 49)

Anders als bei Forschungsarbeiten in anderen Kontexten ist der Aufwand für Kommunikation im Feld und für die Formulierung der Ergebnisse sowie vor allem der Empfehlungen bei Evaluationen größer. Beides hat sensibel den Kontext und die Wirkungen abzuschätzen. Zeiten für Rückkoppelungs- und Abstimmungsverfahren spielen daher eine große Rolle.

Der wichtigste Teil des Designs sind zunächst die Fragen, die Hauptfrage der Evaluation und die Fragen, deren Beantwortung die Daten zur Beantwortung der Hauptfrage liefern. Es ist daher zunächst zwischen der primären „Evaluationsfrage“ und den ihr zuarbeitenden „Untersuchungsfragen“ zu unterscheiden. Die Evaluationsfrage zielt meist auf die Wirkung des Evaluationsgegenstands ab. Wenn die Evaluationsfrage zu einem Programm etwa hieß: Hat sich die Einführung des Bildungsgutscheines auf das Teilnahmeverhalten der Menschen ausgewirkt?, dann könnten die Untersuchungsfragen wie folgt lauten: Nehmen mehr Menschen an Weiterbildung teil?, Welche Personengruppen nehmen den Bildungsgutschein in Anspruch?, Werden andere Angebote wahrgenommen? usw. Im gleichen Programm könnte die Evaluationsfrage auch lauten: Hat sich die Einführung des Bildungsgutscheines auf die Angebote der Weiterbildungseinrichtungen ausgewirkt?, dann richteten sich die Untersuchungsfragen eher auf die Institutionen und Strukturen, etwa: Hat sich die Klientel von Einrichtungen geändert?, Wird stärker für die Angebote geworben?, Ist die Qualität der Angebote gestiegen?, Hat sich die Planungssicherheit der Einrichtungen geändert?

Die Funktion der Fragen, dies ist leicht ersichtlich, ist es vor allem, den Blick auf die erforderlichen Daten, die zu erhalten und zu analysieren sind, zu richten. Damit steuern sie auch die Wahl der geeigneten Erhebungsmethode, sofern eine Erhebung nötig ist und

die Daten nicht bereits vorliegen. Dieser Zusammenhang wird gelegentlich auch anders herum gesucht, die Fragen werden danach formuliert, was an Daten vorliegt, um nicht zeit- und ressourcenaufwändige Primärerhebungen durchführen zu müssen. Die Fragen haben aber darüber hinaus noch andere Funktionen: Sie führen zielgerichtet durch das Literaturstudium, sie strukturieren die Analyse der Daten, sie rahmen den Ergebnistext und sie helfen, Umwege und überflüssige Arbeiten zu vermeiden (vgl. auch Bryman 2004, S. 31).

Es ist nicht einfach, Fragen für die Evaluation zu formulieren, wenn sie nicht bereits durch den Auftrag (zumindest in der Hauptlinie) gegeben sind. Präzise Fragen setzen entweder ein sehr entschiedenes Interesse oder relative gute Kenntnisse des Gegenstands voraus – im ersteren Fall liefert das einen punktgenauen Blick, im zweiten bereits vorab formulierte Annahmen („Hypothesen“) über mögliche Ergebnisse. Da sich die Fragen von Evaluationen oft über den Gegenstand hinaus auf dessen Wirkungen richten, die in einem weiteren Kontext festzustellen sind, ergeben sich auch komplexere „Fragenbatterien“, die folgerichtig auch komplexere Datenbestände erfordern.

Diese Daten sind zudem die *conditio sine qua non* zu stellender Fragen – in Evaluationen wie in der empirischen Sozialforschung generell. Formuliert Fragen müssen mit Hilfe von Daten (die es zu finden oder zu erheben gilt) beantwortbar sein; sind solche Daten nicht erwartbar, ist es unsinnig, die Frage zu stellen. So ist z.B. die Frage: Beeinflusst Gott das Lernverhalten Erwachsener? unsinnig, weil empirisch nicht überprüfbar. Überprüfbar hingegen ist die Frage: Beeinflusst der Glaube an Gott des Lernen Erwachsener? – der Glaube von Menschen an etwas ist empirisch erfassbar und (wie in diesem Fall) mit großer Wahrscheinlichkeit auch einflussreich. Der Weg von den Evaluationsfragen zu den erforderlichen Daten ist daher keine Einbahnstraße, deren fehlende Verfügbarkeit kann auch eine Umformulierung der Fragen erfordern.

Bei der Formulierung der Untersuchungsfragen in der Evaluation (welche die Beschaffung der Daten für die Evaluationsfrage steuern) ist es möglich, unterschiedliche Aspekte des Evaluationsgegenstands in den Blick zu nehmen. Der wichtigste Unterschied ist dabei derjenige zwischen den gegenstandsimmanenten Aspekten (sie betreffen Input, „Throughput“, Output und teilweise Outcome) und den gegenstandsexternen Aspekten (sie betreffen Kontext, teilweise Outcome und „impact“, also Wirkung im weiteren Sinne). Beide Richtungen haben weitreichende Konsequenzen für die erforderlichen Daten und deren Qualität.

Im Falle einer gegenstandsimmanenten Evaluation (meist bei Lehr-/Lernprozessen, Unterrichtsgestaltung und Einrichtungen) kann sich die Definition der erforderlichen Daten auf Bestände richten, die möglicherweise schon vorhanden oder aber mit einfacheren Erhebungen zu erhalten sind. Richtet sich die Evaluation auf Wirkungen (etwa bei Lernergebnissen, Programmen, Systemen) sind schwierigere Reduktionen und Selektionen von Datenerhebungen vorzunehmen, die sich in der Regel auf Stichproben,

Fallanalysen oder repräsentative Befragungen konzentrieren. Das Design hat entsprechend die erforderlichen Zeiträume und die Menge von Daten abzuschätzen.

Generell tendieren Evaluationen dazu, auf vorhandene Daten zurückzugreifen, sofern diese der Beantwortung der Fragen dienlich sind, eine entsprechende Qualität (Validität, Reliabilität) aufweisen und (auch) als Rohdaten verfügbar sind. Dies dient der Kosten- und Zeitersparnis und belastet den Evaluationsgegenstand weniger als gesonderte Erhebungen, auch hat es kaum „verzerrende“ Effekte auf die Alltagsrealität. Für die Prüfung, welche Daten vorhanden und verwendbar sind, werden bei Evaluationen in der Regel bereits in der Design-Phase entsprechende Rückkoppelungen mit den Vertretern des Untersuchungsgegenstands vorgenommen (zu weiteren Aspekten → Kap. 5.2).

Für alle weiteren erforderlichen Daten werden eigene Erhebungen vorgenommen, die den Regeln empirischer Sozialforschung folgen, sofern es sich nicht um den (formativen) Ansatz der Handlungsforschung handelt. Die wichtigsten Aspekte dabei sind in Abschnitt 5.3 unten erläutert.

5.2 Vorhandene Daten

Es gibt kaum Evaluationen, bei denen nicht schon bestimmte Daten, die benötigt werden, vorhanden sind. Dies vor allem deshalb, weil sich Evaluationen auf wichtige Aspekte richten, und wichtige Aspekte auch ohne Evaluation erfasst sind. Dies gilt in allen vier Gruppen von Evaluationsgegenständen (Lehr-/Lernprozesse, Einrichtungen, Projekte und das System), wenn auch vor allem für die beiden ersten, und für nahezu alle Objekte in diesen Gruppen.

Lehr-/Lernprozesse

Evaluationen von Kursen, Lehr-/Lernprozessen und Lernergebnissen werden in großem Maße durch die Evaluierungen erfasst, welche die Einrichtungen (teilweise aufgrund übergreifender Vorschriften wie bei den Hochschulen) aus eigenem Interesse zu Kontroll- und Entwicklungszwecken unternehmen. In der Erwachsenenbildung sind dies vor allem die Ergebnisse der weit verbreiteten Evaluationsbögen von Kursen und Seminaren; sie sind allerdings nicht selten dann wenig brauchbar, wenn es um empirisch qualitätsvolle Daten geht – viele von ihnen sind „selbstgebastelt“ und genügen nur unzureichend den Standards empirischer Sozialforschung. Der Nachteil oder besser: die Probleme vorhandener Daten liegen vor allem auf folgenden Ebenen:

- Die Daten entsprechen möglicherweise nicht den erforderlichen qualitativen Standards, sind nicht valide und reliabel (→ Kap. 5.3). So werden in der Praxis oft Fragebögen zur Kursevaluation eingesetzt, die fehlerhaft sind, unklare Fragen und

unpassende Antwortvorgaben enthalten. Oder Planungsdaten zu Kursangeboten sind ungenau. Oder Erhebungsbögen (etwa zu Teilnehmerzahlen) werden unzureichend ausgefüllt.

- Die Daten sind unvollständig. Bei Personaldaten fehlen Familienstands-, Alters- oder Ausbildungsangaben, bei Kursdaten Belegungszahlen, bei Programmdateien die Finanzierungsdaten und bei Systemdateien Zahlen zur regionalen Versorgung.
- Die Daten sind veraltet. Die Informationen zu einzelnen Aspekten des Evaluationsgegenstands wurden vor längerer Zeit erhoben und bilden die Realität nicht mehr angemessen ab, etwa Daten zur Teilnehmerstruktur oder zu konkurrierenden Angeboten.
- Die Daten liegen nicht (mehr) als „Rohdaten“ vor (also einfachen Auszählungen), sondern nur in Korrelationen, also in Prozent- und Verhältniszahlen; diese Korrelationen sind nur dann verwendbar, wenn sich auch die Evaluation für sie interessiert.
- Die Daten entsprechen nicht den Evaluationskategorien – dies ist eigentlich das häufigste Problem. So untersucht etwa eine Evaluation die Bildungsbeteiligung von Menschen zwischen 60 und siebenzig Jahren, die verwendeten Kategorien bilden aber Altersgruppen ab von 55 bis 65 sowie älter als 65 (übrigens häufig verwendet).

Der Umgang mit den Problemen der vorhandenen Daten ist meist recht aufwendig. So lassen sich etwa Aggregationen auf bestimmte Kategorien hin (wie bei den Altersgruppen) nur noch dann auf die Kategorien der Evaluation hin korrigieren, wenn man auf die Basisdaten zurückgehen kann (also die einzelnen Teilnehmerfragebögen), die aber meist nicht mehr vorliegen. Dies gilt auch oft für vorgenommene Korrelationen, zu denen die Ausgangsdaten nicht mehr vorhanden sind.

Ein besonderes Problem stellen Zeitreihen dar; man kann in einer Evaluation nur in Ausnahmefällen historische Datenbestände rekonstruieren. Viele Evaluationen sind aus diesem einfachen Grund auf den „Status quo“, den Evaluationszeitpunkt, ausgerichtet; nur größere Evaluationen, etwa Programmverlaufs- Evaluationen, bilden eigene Zeitreihen.

Auch besteht das Problem, dass sie möglicherweise in allen Angeboten einer Einrichtung eingesetzt sind, in einer anderen Einrichtung jedoch wieder andere Instrumente verwendet werden; einrichtungübergreifende Evaluationen müssen dann eigene Evaluationen einsetzen, die nicht selten mit den bestehenden in Konflikt stehen und – vor allem aus Sicht der Lernenden – ein doppeltes Erhebungsverfahren sind.

Weitere Daten zu Lehr-/Lernprozessen werden über Erhebungsverfahren gewonnen, welche sich vor allem auf die Struktur der Teilnehmenden und ihre Verteilung auf die Angebote konzentrieren. Für den Volkshochschulbereich ist dies etwa die seit fünfzig Jahren realisierte jährliche Erhebung im Rahmen der Volkshochschulstatistik (Huntemann/Reichart 2012), die verlässliche Daten zu Teilnehmenden und Angeboten

liefert. Auch andere Träger sind mittlerweile in der Verbundstatistik des DIE erfasst, mit der auch versucht wird, einheitliche Kategorien für alle angeschlossenen Einrichtungen und Bereiche anzuwenden. Einzelne Weiterbildungsorganisationen und -organisationsbereiche wie der Fernunterricht oder die wissenschaftliche Weiterbildung erstellen (meist jährlich) eigene Statistiken, die auch Informationen meist formaler Art über Lehr-/Lernprozesse und deren Ergebnisse liefern.

Andere Datenquellen zu Lehr-/Lernprozessen liegen in den Ergebnissen der Erhebungen zur Weiterbildung wie dem „Berichtssystem“, das bis vor kurzem in Deutschland (seit 1979) alle drei Jahre das Weiterbildungsverhalten der Bevölkerung repräsentativ erfasste (zuletzt Berichtssystem Weiterbildung 2008). Seit einigen Jahren ist dieses rein deutsche Berichtssystem in einer europaweit einheitlichen Erfassung aufgegangen, dem „Adult Education Survey“ (AES), mit dem auch internationale Vergleiche möglich sind.

Allerdings: Diese statistischen Daten enthalten nur wenige Angaben zu didaktischen und methodischen Aspekten der Weiterbildung und der Angebote in der Weiterbildung. Auch ist ihre Verwendbarkeit durch die längere Periode der Erfassung (ca. alle drei Jahre) und die Repräsentativität der Erfassung eingeschränkt. Es gibt noch andere Datenquellen zu Teilnehmenden, Teilnahmen und Lernprozessen, für die diese Einschränkungen noch vermehrt gelten. Dennoch ist bei allen auf die Lerner bezogenen Evaluationen damit zu rechnen, dass eigene Erhebungen im Rahmen von Evaluationen bereits realisierte Dauererhebungen aus Sicht der Befragten „verdoppeln“.

Man kann sagen, dass Evaluationen von Lehr-/Lernprozessen dann, wenn sie genauere Daten zu didaktischen und methodischen Fragen evaluieren wollen und sollen, an eigenen Erhebungen schwerlich vorbeikommen, die vorhandenen Daten jedoch mit erfassen und berücksichtigen müssen.

Einrichtungen

Evaluationen von Einrichtungen stützen sich in großen Teilen auf den gleichen Datenbestand, wie er für das strategische ebenso wie das operative Management erforderlich ist: Daten zu den Finanzen, zum Personal, zu den Produkten, zum Nutzungs- und Marktbereich. Solche Daten sind vor allem auch dann vorhanden, wenn eine Rechenschaftspflicht für öffentliche Gelder besteht, wie dies etwa bei Volkshochschulen der Fall ist. Folgende Datengruppen sind fast immer vorhanden:

- *Personaldaten*

Hier liegen fast immer die Beschäftigtenzahlen, Eingruppierungen, Alter, Beschäftigungsdauer, Qualifikation, Arbeitszeiten vor.

- *Finanzdaten*

Hier liegen immer die (gesetzlich erforderlichen) buchhalterischen oder kaufmännischen Grunddaten vor, die eine Übersicht über Einnahmen und Ausgaben, Her-

kunft (z.B. Projekte, institutioneller Zuschuss) und Verwendungszweck (z.B. Personal, Material) der Mittel, Zeiträume und Zeitpunkte sowie Belege liefern.

- *Angebotsdaten*
Hier liegen fast immer Programme, Stunden- und Belegungsnachweise, Planungen und Konzepte vor.
- *Strukturdaten*
Hier liegen fast immer das Organigramm, die Satzung, Unterschriftsberechtigungen, Ablaufdiagramme und Strukturpläne vor.
- *Kommunikationsdaten*
Hier liegen fast immer Protokolle von Gremiensitzungen, Berichte, Newsletter und Informationsbroschüren etc. vor.

Das Problem bei vorliegenden einrichtungsbezogenen Daten ist weniger deren Existenz selbst als vielmehr deren Zugänglichkeit, Vollständigkeit und Qualität. Die Zugänglichkeit ist in der Regel eingeschränkt durch Datenschutzbestimmungen (beim Personal), durch Vertraulichkeitsregeln (bei Betriebsinterna) oder durch Wettbewerbskautelen (bei Planungen, Konzepten). Nur selten sind die Unterlagen vollständig, sei es, dass sie nicht vollständig erstellt wurden (z.B. Protokolle von Sitzungen), sei es, dass sie nicht systematisch gesammelt und archiviert wurden. Und die Qualität der Datenbestände ist unterschiedlich und unterschiedlich verwertbar; bspw. wurden die Daten nicht zu Evaluationszwecken erhoben, sondern zum Zweck betrieblicher Abläufe, Verfahren und Entscheidungen. Sie unterlagen also von Beginn an nicht den Anforderungen an Daten im Kontext empirischer Forschungen.

Üblicherweise werden bei Evaluationen von Einrichtungen zunächst die vorhandenen Daten geprüft, und sodann wird entschieden, an welchen Stellen sie aufbereitet und welche weiteren Daten erhoben werden müssen. Die Klärung der Datenlage setzt bei Einrichtungsevaluationen eine enge und vertrauensvolle Kooperation der Evaluatoren mit den Vertretern der Einrichtung bereits zu Beginn der Designphase voraus.

Projekte und Programme

Evaluationen von Projekten und Programmen erhalten ihre Daten gewissermaßen naturwüchsig über den Fortgang des Programms als verordnete Zulieferung, die bereits in der Vergabe des Programmauftrags geregelt war. Der Vorteil vorhandener Daten liegt auf der Hand: Sie müssen nicht (zeit- und kostenraubend) erhoben werden, es entstehen keine entsprechenden Engpässe, und sie enthalten meist eine brauchbare historische Dimension, d.h. sie wurden schon seit Beginn des Projekts oder Programms, erhoben und gesammelt. Die meisten Projekte und Programme beginnen gleichzeitig mit ihrer Evaluation, so dass sich beide Prozesse aufeinander beziehen können.

Wichtig ist bei Projekten und Programmen, dass die Festlegung der erforderlichen Daten bereits zu Beginn erfolgt – zum einen, um mit dem Fortgang des Projekts zugleich die Daten zu erheben, zum anderen, weil nach Abschluss des Projekts die Daten nicht mehr erhoben werden können. Wenn das Projekt einmal beendet ist, sind auch die Akteure meist in anderen Kontexten eingebunden.

In der Regel richtet sich das Evaluationsinteresse bei Projekten und Programmen auf die Wirkungen, setzt also bereits eine gewisse „Laufzeit“ voraus. Aus den Zielen dieser meist innovativen Vorhaben ergeben sich auch die Anforderungen an notwendige Daten. Ist bspw. das Ziel eines Programms, regionale Bildungsnetzwerke aufzubauen, so sind in seinem Verlauf die beteiligten Personen und Organisationen auf Interessen, Barrieren und Fortschritte hin zu befragen, nach Ablauf aber die nachhaltige Existenz zu überprüfen. Letzteres erfolgt – wenn überhaupt – in Form einer zeitlich versetzten Nachbefragung, die auf den Daten der programmbegleitenden Evaluation aufbaut.

In Projekten und Programmen sind demnach möglicherweise diejenigen Daten vorhanden, welche die Ausgangslage dokumentieren (also etwa die Kooperationsstruktur zu Beginn eines Netzwerkprogramms), die weiteren Daten werden fortlaufend ermittelt und dabei entsprechend als neue Daten erhoben (→ Kap. 5.3).

System

„System“ ist ein weiter Begriff und kann Daten aus allen anderen Gegenstandsbereichen erfordern, etwa Teilnehmerstrukturen, Einrichtungsdaten und Angebotsdichte. Gewöhnlich richtet sich das Interesse der Systemevaluation auf die Funktionsfähigkeit des Zusammenhangs all dieser Elemente und die Wirkungen, die einzelne Aktivitäten und Faktoren dabei für die Entwicklung des Systems haben. So zeigen Systemevaluationen zur Erwachsenenbildung immer auch – je nach Schwerpunkt – Einblicke in andere Elemente des gesamten Bereichs (vgl. Nuissl/Schlutz 2001).

Systemdaten, also Daten über das „Gesamte“ der Weiterbildung, liegen in unterschiedlichster Weise bei den Ministerien, den statistischen Ämtern und den einzelnen Sektoren der Verbands- und Trägerbereiche vor. Auch können hier oft frühere Erhebungen und Fallstudien herangezogen werden. Beispiele dafür sind die genannte Volkshochschul-Statistik oder Statistiken zum Bildungsurlaub, wie sie in manchen Ländern mit entsprechendem Gesetz alle zwei Jahre vorgelegt wurden. Man kann in der Erwachsenenbildung in Deutschland allerdings davon ausgehen, dass sie als „System“ nur in geringem Umfang datenmäßig erfasst ist (vgl. DIE 2010) und die vorliegenden Daten zudem oft wenig verlässlich und wenig vergleichbar sind.

Evaluationen zum System sind daher – unabhängig von der konkret gestellten Frage – meist mit eigenen Erhebungen verbunden. Dabei können zwar vorliegende Daten herangezogen werden, sie dienen jedoch meist eher zur Schärfung einzusetzender

Erhebungsinstrumente und zum Abgleich neu erhaltener Daten. In der Trendanalyse des DIE von 2010 heißt es einleitend:

Es gilt also, in einer Trendanalyse alle nur denkbaren Daten heranzuziehen, um Trends in der Realität aufspüren und nachweisen zu können. Dabei ist oft festzustellen, dass es an empirischen Daten mangelt; solange es keine spezifischen Primärerhebungen in diesen Fällen gibt, wird es vielfach für einen vermuteten oder beobachtbaren Trend keinen wirklichen Beleg geben (DIE 2010, S. 9).

Da Evaluationen zum System der Weiterbildung (bzw. zum nur halb systematisierten Bereich der Weiterbildung) ausschließlich von staatlichem Interesse sind, handelt es sich bei solchen Erhebungen auch immer um staatliche Aufträge. Das Evaluationsinteresse richtet sich demnach auch vorrangig auf die Funktionsfähigkeit des Systems und die Wirkung der Instrumente, die dem Staat zur Regulierung des Bereichs zur Verfügung stehen. Meist steht letzteres – aus naheliegenden Handlungsinteressen heraus – im Vordergrund und ist dann oft mit Programmen verbunden, mit denen das System in einer bestimmten Richtung hin verändert werden soll, ist also eine Evaluation innovativer Programme. Es lassen sich aber auch Systemevaluationen finden, mit denen durch eine Evaluation des Status quo Problemfelder entschlüsselt werden sollen, die einen staatlichen Einfluss erfordern.

5.3 Erhebung neuer/zusätzlicher Daten

Evaluationen benötigen Daten, mit denen die Fragen, die sie stellen, beantwortet werden können. Sind sie nicht bereits vorhanden, müssen sie „erhoben“, also beschafft werden. Wenn es sich um ernstzunehmende Evaluationen handelt (und dieser Anspruch wird immer erhoben), dann müssen die Daten in einem systematisch geplanten und begründeten sowie nachprüfbar Verfahren erhoben werden. Dieses folgt dem differenziert entwickelten System empirischer Forschungsmethoden, die korrekt angewendet werden müssen. Anders zu bewerten ist nur der Ansatz der handlungsorientierten Evaluation, die meist formativ als (wissenschaftliche) Begleitung eines Prozesses erfolgt und eine Balance von empirischer Genauigkeit, zeitlicher Nähe und direkter Umsetzbarkeit herstellen muss.

Bei aller Betonung der Tatsache, dass die empirischen Erhebungsmethoden bestehenden Standards entsprechen und entsprechend auch angewendet werden müssen, ist doch darauf hinzuweisen, dass es sich „nur“ um die Instrumente handelt, mit denen eine Frage begründet beantwortet werden kann; auch die perfekte Umsetzung empirischer Erhebungsmethoden kann fehlgehen, wenn die zugrundliegende Frage nicht (wissenschaftlich) abgesichert ist.

Ausgehend vom Primat der Fragestellung(en) an den Evaluationsgegenstand stellt sich die Frage danach, welche Daten erforderlich sind, welche Daten der Evaluationsfrage *adäquat* sind. Dabei steht natürlich die inhaltliche Seite der Daten im Vordergrund, also: Zu welchen Aspekten der Evaluation sind welche Kenntnisse über Fakten erforderlich? Es stellen sich aber auch weitere Fragen: die Frage nach der Qualität der Daten – sind sie in der notwendigen Qualität überhaupt erhältlich und welche Erhebungsmethoden sind dazu erforderlich? Die Frage nach dem zu erwartenden und bearbeitbaren Umfang: Welche Datenmenge wird vorhanden und auszuwerten sein? Und schließlich (→ Kap. 5.4) die Frage danach, ob die Daten angemessen aufbereitet werden können.

Mit der Formulierung der Frage(n), welche die Evaluation steuert, ist nicht nur die Richtung der Evaluation angegeben (und damit auch das bestehende Interesse), sondern sind auch Annahmen verbunden, die begründet sind. Sie sind nicht im strengen Sinne Hypothesen, also Vermutungen, die aufgrund der ausgewerteten Daten bestätigt oder verworfen werden können. Sie stellen vielmehr den Zusammenhang zwischen einzelnen Aspekten des Untersuchungsgegenstands her, der dann ebenfalls in den Blick der Evaluation kommt.

BEISPIEL

Evaluationsannahmen

Wenn in der Evaluation von Lehr-/Lernprozessen Daten zu Lehrverhalten, Lernumgebung und Lernergebnis erhoben werden, dann besteht die Annahme darin, dass sie in einem Zusammenhang stehen und dass die beiden ersteren das letztere beeinflussen. Oder wenn bei der Evaluation einer Einrichtung nach Kooperationen mit Betrieben und anderen Bildungseinrichtungen gefragt wird, dann besteht die Annahme darin, dass das vorteilhaft für die Perspektiven der Bildungseinrichtung wäre. Oder wenn bei der Systemevaluation überprüft wird, ob einzelne Regionen und Zielgruppen weniger an Weiterbildung teilnehmen, dann besteht die (normative) Annahme darin, dass ein System weder Personen noch Regionen benachteiligen soll.

Evaluationen betrachten niemals den gesamten Evaluationsgegenstand, sondern immer nur dasjenige, was im Interesse der Evaluation heraus in den Blick genommen und für wichtig gehalten wird. Insofern unterscheidet sie sich nicht von der Forschung im Allgemeinen, bei der ebenfalls immer nur die – mehr oder weniger abgeleitete und begründete – Frage beantwortet wird. Das Besondere bei der Evaluation liegt aber darin, dass ihre Ergebnisse (und die daraus resultierenden Empfehlungen, vgl. → Kap. 6.4) die Weiterentwicklung des Evaluationsgegenstands stark beeinflussen.

Es ist daher bei Evaluationen zu prüfen, welche – oft nicht explizierten – Vermutungen oder normativen Vorstellungen hinter den abgefragten Daten stehen. Daraus ergeben sich oft überraschende Einsichten in das Problem der Objektivität. Auch Evaluationen unterliegen in ihrer Blickrichtung den Prinzipien des Konstruktivismus, mit gelegent-

lich fatalen Folgen für die Weiterentwicklung des Evaluationsgegenstands, wenn die entsprechenden Konsequenzen verbindlich gemacht werden. In „guten“ Evaluationen werden die Bereiche, zu denen Daten erhoben werden sollen, nicht nur aus dem Auftrag heraus begründet, sondern auch aus der übergeordneten Vorstellung eines „optimalen“ Zustands des Evaluationsgegenstands, letztlich anhand einer normativen Idee, die mit der Identität des Evaluationsgegenstands korrespondiert.

Wenn feststeht, zu welchen Bereichen (etwa Finanzen, Teilnehmerzahlen, Angebote, Beteiligungsquoten, Teilnehmerzufriedenheit etc.) Daten benötigt werden, stellt sich die Frage, wie sie ermittelt oder erhoben werden, um die notwendige Qualität zu besitzen. Dabei ist meistens die Entscheidung von Bedeutung, ob es sich um „quantitative“ oder „qualitative“ Daten handelt. Gerade in Deutschland spielte diese Frage in den vergangenen Dekaden eine große Rolle (vgl. Nuissl 2010a, S. 57ff.). Vielfach besteht das Vorurteil, quantitative Daten seien genauer als qualitative Daten. Dies entstand vermutlich durch die Überzeugungskraft von Tabellen mit Zahlen und Grafiken mit Entwicklungslinien. Eine gewisse Faszination ist diesen quantitativen Konstruktionen nicht abzusprechen, und gerade im Bereich der Politik spielen Mengen immer eine größere Rolle als Differenzierungen.

Entscheidend in der Evaluation ist, ob ein quantitatives Abbild des Evaluationsgegenstands die Fragen der Evaluation besser beantwortet als ein qualitatives Abbild. Dies hängt im gleichen Maße vom Ziel der Evaluation wie vom Charakter des Gegenstands ab. Bei Lehr-/Lernprozessen etwa stoßen quantitative Analysen rasch dort an ihre Grenzen, wo der Inhalt des Prozesses mit einbezogen wird – ein bei Lernvorgängen eigentlich selbstverständliches Anliegen. Bei Systemevaluationen hingegen sind eher quantitative Daten vonnöten, etwa die Zahl der Institutionen, die Höhe der investierten Mittel oder die Quote der Teilnahme.

Qualitative Verfahren ermöglichen es, mit den erhobenen Daten (z.B. aus Interviews) Zusammenhänge, Verläufe, Kausalitäten, Interpretationen, Selbstkonzepte und Einschätzungen zu erfassen. Bei Evaluationen zeigt sich ihre Stärke darin, auch Erklärungen für verfehlt Ziele oder Realisierungsprobleme zu liefern.

Stärken qualitativer Verfahren

Die besondere Stärke qualitativer Verfahren liegt darin, dass es mit ihnen möglich ist,

- „geheime“ oder „versteckte“ Erkenntnisse über Personen und Sachverhalte zu gewinnen, die diesen entweder
- selbst gar nicht bekannt sind oder die sie nicht offen äußern würden;
- Zusammenhänge und Einflüsse identifizieren zu können, die „von außen“ nicht erkennbar wären;
- plausible Hypothesen über Zusammenhänge und Faktoren formulieren zu können, zu denen bis dato noch

- keine weitergehenden Erkenntnisse vorlagen;
- Fakten und Erkenntnisse in soziale und biografische Zusammenhänge einordnen zu können, die
- kompliziert sind (z.B. bei Personen mit Migrationshintergrund);
- mit den „Untersuchungsobjekten“ ein partizipatives und an gemeinsamen Erkenntnisgewinnen interessiertes Verhältnis aufbauen zu können – ein gerade im Bereich der Erwachsenenbildung wichtiger Aspekt (Nuisl 2010a, S. 59).

Quantitative Verfahren haben immer etwas mit Zahlen zu tun und unterliegen daher, wenn es um komplexe und inhaltliche Sachverhalte und Prozesse geht, immer der Gefahr der eingeengten Sichtweise. Im Bildungsbereich ist es besonders wichtig, geeignete Variablen zu definieren und Indikatoren für diese Variablen zu finden, um auch mit quantitativen Verfahren die komplexe Realität angemessen abzubilden. Das Definieren und Finden dieser Variablen und Indikatoren setzt eine genaue Kenntnis des Gegenstands voraus, die in der Regel auf der Basis qualitativer Forschung entstand. Ein genaues Abbild gelingt vor allem auch dort, wo entsprechende Variablen und Indikatoren naheliegen, wie etwa bei Übergängen im Schulsystem, Teilnahmefällen an Weiterbildungsveranstaltungen, Finanzdaten und kognitiven Lernleistungsmessungen.

Mittlerweile ist in der Diskussion der (scheinbare) Gegensatz zwischen qualitativen und quantitativen Ansätzen, der besonders in Deutschland heftig thematisiert wurde, dem Verständnis der unterschiedlichen Leistungsfähigkeiten gewichen. In vielen Forschungen und Evaluationen werden sowohl quantitative als auch qualitative Erhebungs- und Auswertungsverfahren angewandt und miteinander kombiniert. Dieses Mischverfahren nennt man „Triangulation“. Man kann unterscheiden zwischen vier Arten von Triangulation:

- *Datentriangulation*,
bei der unterschiedlich erhobene Daten miteinander verbunden werden.
- *Forschertriangulation*,
bei der verschiedene Personen unterschiedlicher Provenienz einen Gegenstand analysieren und beurteilen (besonders beim Peer-Review und der Peer Group Evaluation von Bedeutung).
- *Theorientriangulation*,
in der Evaluation besonders als Zusammenfügen von Zielen und normativen Setzungen aus unterschiedlichen Kontexten angewandt.
- *Methodische Triangulation*,
bei der unterschiedliche Methoden (auch innerhalb eines Instruments) miteinander verknüpft werden.

Allerdings setzen solche Triangulationen, insbesondere die Daten- und die Methoden-triangulation, bereits im Evaluationsdesign eine gemeinsame konzeptionelle Basis voraus. Wenn etwa in der Evaluation eines Lehr-/Lernprozesses im Beobachtungsleitfaden das Merkmal „Intervention des Lehrenden“ auftaucht, aber in den Fragebögen für die Lehrenden und die Lernenden keine dieses Merkmal betreffenden Fragen enthalten sind, ist eine Triangulation kaum möglich.

Damit sind bereits die Methoden der Datenerhebung angesprochen. Im Kern gibt es nur vier unterschiedliche Methoden der Datenerhebung:

- *Befragung*,
bei der Menschen direkt befragt werden,
- *Beobachtung*,
bei der Menschen, Sachverhalte und Prozesse indirekt, also beobachtend, „befragt“ werden,
- *Tests und Experimente*,
bei denen über eine konstruierte Situation Sachverhalte und Änderungen ermittelt werden, und
- *Inhaltsanalyse*,
die sich in unterschiedlichen Verfahren hauptsächlich mit Texten und Dokumenten beschäftigt.

In Form einer Grafik haben Stockmann/Meyer die Methoden der Datenerhebung visualisiert, wobei dort unter der Rubrik „nicht reaktiv“ Tests, Experimente und Inhaltsanalyse mit enthalten sind (dem Evaluationsparadigma: „Beteiligung und Interaktion“ entsprechend) (→ Abbildung 9).

Allen vier Methoden sind Instrumente zuzuordnen, deren Konstruktion und Einsatz nach wissenschaftlichen Regeln erfolgen muss. Solche Instrumente sind bei der Befragung etwa der Fragebogen (der meist die Grundlage einer quantitativen Auswertung ist) oder das leitfadengestützte Interview (meist im Kontext qualitativer Ansätze verwendet). Ein häufig angewandtes Instrument der Beobachtung ist der Beobachtungsleitfaden, bei Tests und Experimenten wird vielfach mit Vergleichsgruppen gearbeitet, und bei der Inhaltsanalyse haben hermeneutische Verfahren eine große Bedeutung (vgl. Nuissl 2010a).

Alle vier Ansätze der Datenerhebung sind bei Evaluationen – jeweils abhängig vom Evaluationsgegenstand und der Evaluationsfragestellung – sinnvoll, auch in unterschiedlichen Kombinationen. Dabei spielen, neben der erwähnten Menge und Qualität der Daten, auch die verfügbare Zeit und die bereitgestellten Ressourcen eine Rolle.

Bei der Evaluation von Lehr-/Lernprozessen werden insbesondere Kombinationen von Beobachtung und Befragung eingesetzt. Beobachtungsdaten werden oft zur Kontrolle der Befragungsdaten oder als deren Hintergrund herangezogen, können aber

auch nur für sich eine angemessene Datengrundlage sein. Für die Evaluationsfrage: Arbeitet der Lehrende mit vielfältigen kohärenten Methoden? sind Beobachtungsdaten ausreichend. Bei der Evaluationsfrage: War/ist die Methodenwahl des Lehrenden den Interessen und Voraussetzungen der Lernenden angemessen? sind Befragungsdaten praktisch unabdingbar.

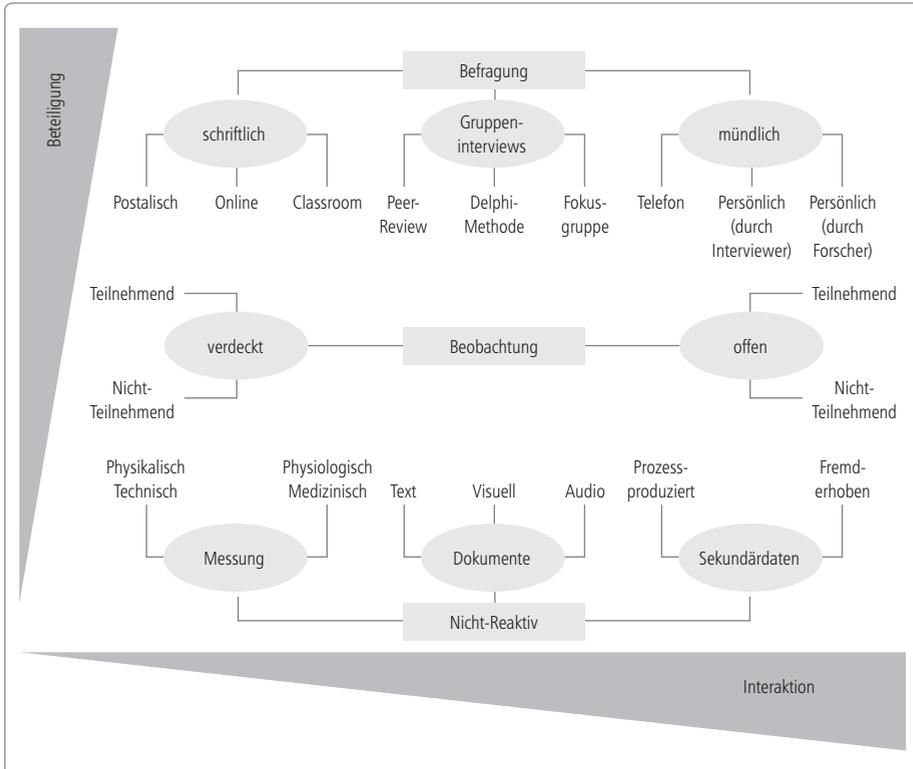


Abbildung 9: Datenerhebungsmethoden (Quelle: Stockmann/Meyer 2010, S. 207)

Bei der Evaluation von Einrichtungen und Systemen werden am häufigsten die Ansätze der Inhaltsanalyse und der Befragung gewählt. Die Inhaltsanalyse ist meist eine der Befragung vorgeschaltete Dokumentenanalyse, welche hilft, die Präzision der Fragen und die Interpretation der Antworten zu unterstützen. Die Befragungen richten sich an die agierenden Personen, wobei deren Auswahl eine eigenständige wissenschaftliche Aufgabe ist: die *Samplebildung*. Anders als bei Analysen von Lehr-/Lernprozessen, in denen es in der Regel nur zwei Akteursgruppen gibt (Lehrende und Lernende), sind diese bei Einrichtungen und Systemen vielschichtiger und differenzierter. Eine begründete und abgesicherte Auswahl für die Befragung ist unerlässlich.

Bei der Evaluation von Projekten und Programmen sind, je nach Ziel und Gegenstand, alle Ansätze gleichermaßen bedeutsam. Besonders wichtig ist hier auch der Ansatz von Test und Experiment. Bei einem Projekt etwa, in dem es um neue Lehrmethoden geht, kann die Konstruktion einer Vergleichs- oder Kontrollgruppe von tradierten Seminarmodellen sinnvoll sein. Und bei der Bewertung des Ergebnisses von Programmen werden meist Vergleichssituationen und -gegenstände herangezogen.

Bei den Instrumenten, die in der Erhebung eingesetzt werden, sind die gesicherten Regeln wissenschaftlicher Forschung zu berücksichtigen. Bei der Fragebogenkonstruktion etwa sind dies (vgl. Nuissl 2010a, S. 53ff.):

- die Grundsätze der Standardisierung und Einführung,
- die Verfahren des Aufbaus und der Führung,
- die inhaltlichen Regeln für „geschlossene“ Fragen (z.B. Eindeutigkeit, Trennschärfe und Residualkategorie),
- die formalen Regeln für geschlossene Fragen (Vorgaben, Intervall, Odd/Even) sowie
- die Codierungsregeln.

Bei den Interviewleitfäden ist auf die Konzeption der Interviews bezogen eine engere oder weitere Führung (von Einzelfragen über Leitfragen mit Stichworten) möglich. Bei sogenannten „narrativen“ Interviews (die allerdings nur selten bei Evaluationen verwendet werden) kann auf einen Leitfaden ganz verzichtet werden. Wichtig sind bei Interviews die Auswertungsregeln, die in mancher Hinsicht eng mit den inhaltsanalytischen Verfahren verwandt sind, aber auf das engste verbunden sind mit den „Transkriptionsregeln“, also der Art und Weise, wie das gesprochene Wort in die Schriftform übertragen wird. Interviews werden bei Evaluationen recht häufig geführt, da es sich aber meist um „Experteninterviews“ handelt, geht es eher um Inhalte und Informationen als um feine sprachliche Abstufungen.

Die Vergleichstests und Experimente, meist im Kontrollgruppen-Verfahren, die bei Programmevaluationen gerne verwendet werden, setzen genaue Definitionen der zu vergleichenden Variablen voraus (etwa Selbständigkeit, Beteiligung, Aufmerksamkeit) sowie deren Messbarkeit. Zugleich müssen die Variablen auch die wesentlichen Innovationen des Programms oder Projekts abbilden.

In Programmen und Projekten werden auch Beobachtungen immer wieder angewandt. Dies besonders bei Curriculum-Entwicklungen, aber auch bei Projektteams in europäischen Projekten (durch die „externe“ Evaluation). In Lehr-/Lernprozessen sind diese meist durch Beobachtungsleitfäden gestützt, in denen die zu beobachtenden Variablen explizit benannt und die Beobachter entsprechend geschult sind. In Projektgruppen geht es eher um eine inhaltliche Expertise, deren Schwerpunkte sich nach dem Fortgang und den Problemen der Arbeit richten.

Welche Art der Datenerhebung auch gewählt wird, welche Instrumente auch eingesetzt werden: Fehler und Verzerrungen sind immer möglich und können im Extremfall auch das bewertende Urteil der Evaluation beeinflussen. Viele solcher Fehler können daraus entstehen, dass der „Evaluationsgegenstand“ nicht ausreichend in Planung, Konzeption und Durchführung der Evaluation einbezogen war. Die wesentlichen Fehlerquellen können in den folgenden fünf Bereichen liegen:

- im *Design*,
d.h. in der geplanten Kombination von Daten und ihrem Bezug zueinander. Solche Probleme entstehen weniger bei Evaluationen von Projekten und Programmen oder Systemen, wo die Fragestellung sich in der Regel aus dem Untersuchungsgegenstand oder der Formulierung des Auftrags ergibt, sondern eher bei Evaluationen von Lehr-/Lernprozessen und Einrichtungen, wo die Gefahr einer Datenerhebung, die gar nicht zum Gegenstand passt, größer ist.
- in den *Instrumenten*,
d.h. in einer fehlerhaften Konstruktion eines oder mehrerer Erhebungsinstrumente oder ihrer „Triangulation“ – etwa wenn Validität und Reliabilität nicht gesichert oder die Merkmale nicht aufeinander abgestimmt sind. Auch das Sample kann fehlerhaft ermittelt sein.
- im *Erhebungs- und Auswertungsprozess*,
wenn etwa der Prozess der Datenerhebung den Evaluationsgegenstand deutlich verzerrt (leicht möglich in der Evaluation von Lehr-/Lernprozessen), ungleiche Zeitpunkte oder Zustände abbildet oder durch Streuungseffekte beeinträchtigt ist.
- im *Informationsverhalten* des Evaluationsgegenstands,
d.h. wenn – etwa aus Widerstand, Unwillen oder Überforderung – erforderliche Informationen nicht, nur lückenhaft oder falsch gegeben werden; gerade bei Evaluationen, die ja in der Regel Folgen für den Evaluationsgegenstand haben, sind die Interessen der Informationsgeber von großer Bedeutung und ein eigener Analysegegenstand.
- im *Interpretationsinteresse* der Evaluatoren,
das sich ja nicht nur nachprüfbar in den Instrumenten und den Auswertungsverfahren spiegelt, sondern auch, wenig nachprüfbar, die Datenerhebung, Datenauswahl und Datenzuordnung beeinflusst. Dies erfolgt besonders dann, wenn die Evaluation aus einem kontrollierenden Interesse heraus erfolgt oder ihr ein normativer Anspruch zugrunde liegt.

In der Regel werden Fehler und Probleme bei den Evaluationsergebnissen mitgenannt, jedenfalls sollte das so sein nach den Selbstverpflichtungen der Evaluierungsgesellschaften wie DeGEval. Dies ist jedoch nicht immer der Fall. Dann lässt sich einiges durch die Widersprüchlichkeit der Evaluationsergebnisse, der Datenqualität und Datenauswertung erschließen.

5.4 Datenaufbereitung

Ein oft unterschätzter Zwischenschritt zwischen der Datenerhebung und der Datenauswertung ist die Datenaufbereitung. Sie kann – je nach Menge und Qualität der vorhandenen und erhobenen Daten – von außerordentlicher Zeitintensität sein. Bei Evaluationen handelt es sich hauptsächlich um die folgenden Aufbereitungsaufgaben:

- Sichtung, Prüfung und Strukturierung der vorhandenen Daten (→ Kap. 5.2),
- Codieren und Auszählen von standardisierten Fragebögen,
- Fehlerbereinigung in den erhobenen Datensätzen,
- Transkription von Interviews und Protokollen sowie
- Ordnung und Prüfung aller vorhandenen Materialien.

Einige der Aufbereitungsschritte sind aus der Logik der Daten heraus notwendig und unabänderlich. Dazu gehören etwa die Fehleranalyse und die Auszählung codierter Fragebögen. Andere sind abhängig von Forschungsfrage und Auswertungsziel. So hängt etwa die Intensität und Reichweite der Transkription von Interviewdaten davon ab, um welche Art von Interview es sich handelt und bis zu welchem Grad die enthaltenen Informationen ausgewertet werden sollen (erfahrungsgemäß werden Interviews oft viel intensiver transkribiert als von der Fragestellung her notwendig wäre) (vgl. Nuissl 2010a, S. 87).

Letztlich gilt das Ablaufmodell, das Mayring 1993 für die Inhaltsanalyse strukturiert hat, mit geringfügiger Modifikation für alle aufzubereitenden Daten (siehe Abbildung 10).

Das Modell zeigt, dass auch die Datenaufbereitung – ebenso wie das Erhebungsdesign und die Erhebung selbst – der Fragestellung folgt, die der Untersuchung oder Evaluation zugrunde liegt. So wie die Tiefenschärfe der Transkription von Interviews ist auch die Zusammenfassung von Fragebogen-Daten oder, natürlich, die kategoriale Zuordnung von Interview- und Protokollauschnitten davon abhängig, wo das größte Interesse liegt. Das Schließen von offenen Fragen aus Fragebögen etwa, das in der Regel dem Prinzip der induktiven Kategorienbildung folgt, erfolgt stets mit Blick auf die zugrundeliegende Fragestellung.

Die Datenaufbereitung endet dort, wo die eigentliche Auswertung beginnt. Man verwendet hier meist den Begriff der „Rohdaten“. Dies sind bereits aufbereitete Daten, die auf Fehler, Lücken und Ungenauigkeiten überprüft wurden und – gelegentlich – bereits zu Gruppen zusammengefasst sind.

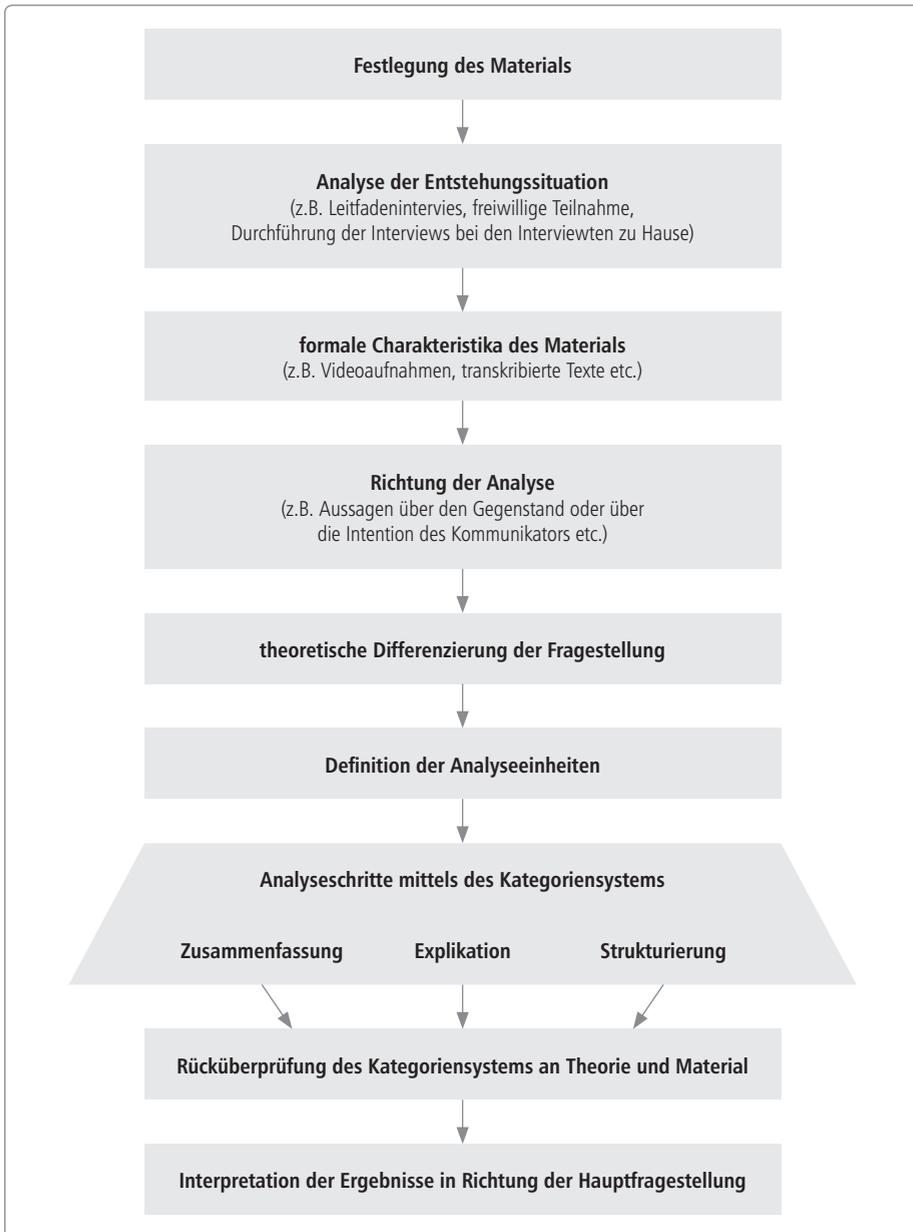


Abbildung 10: Datenaufbereitung. Allgemeines inhaltsanalytisches Ablaufmodell
(Quelle: Nussli 2010, S. 96 in Anlehnung an Mayring)

BEISPIEL**Rohdaten aus Interviews**

Die eigentlichen Rohdaten des Interviews bestehen im Interview selbst. Meist ist es mitgeschnitten, d.h. als Tonkassette verfügbar. Für die Auswertung sind jedoch „aufbereitete“ (und transkribierte) Textpassagen die Rohdaten, noch im Wortlaut, aber segmentiert und noch nicht interpretiert. Dies gilt sowohl für eine quantitative wie qualitative Auswertung. Die quantitative Weiterarbeit enthält Codierungsverfahren und statistische Operationen, die qualitative enthält Paraphrasen und Interpretationen.

Letztlich gehen Aufbereitung und Auswertung ineinander über, wenn es um die Bearbeitung der vorliegenden Daten geht. Der wesentliche Unterschied ist – normativ –, dass in der Aufbereitung noch keine Auswertung vorgenommen wird. Es sind nicht wenige Fälle, in denen auch dieser Unterschied nicht entschieden eingehalten werden konnte. Dies muss der Qualität des Ergebnisses jedoch keinen Abbruch tun, wenn es entsprechend dargelegt und begründet ist.

ZUR REFLEXION

- Welches sind spezifische Daten, die für Evaluationen in der Erwachsenenbildung wichtig sind?
- In welchen Feldern ist es schwierig, überhaupt an evaluationsrelevante Daten heranzukommen?
- Inwieweit verzerren bei welchen Evaluationsgegenständen die Datenerhebungen die Realität, den Alltag?
- Welche Anforderungen sind an ein Evaluationsdesign in der Erwachsenenbildung zu stellen (insbesondere in Bezug auf die Lerner!)?
- Inwieweit spielen Anforderungen an die gesellschaftliche Aufgabe der Erwachsenenbildung eine Rolle beim Untersuchungsdesign?

**Lektüreempfehlungen**

- Bryman, A. (2004): Social Research Methods. Oxford
- Nuissl, E. (2010a): Empirisch forschen in der Weiterbildung. Bielefeld
- Bortz, J./Döring, N. (2002): Forschungsmethoden und Evaluation für Sozialwissenschaftler. 3. Aufl. Berlin/Heidelberg/New York
- Aeppli, J./Gasser, L./Gutzwiller, E./Tettenborn, A. (2011): Empirisches wissenschaftliches Arbeiten. Ein Studienbuch für die Bildungswissenschaften. Bad Heilbrunn

6. Evaluation – Wie werden die Daten bewertet?

Schon bei der Frage, welche Daten zur Evaluation herangezogen werden sollen, haben „Bewertungskriterien“ eine wichtige Rolle gespielt. Je nach dem, worauf sich das Werturteil richten soll, benötigt man auch entsprechende Daten. Geht es bei der Evaluation einer Einrichtung etwa um Leistungsgrößen wie Personalkapazität und Angebotsquantitäten, so benötigt man andere Daten als bei einer Evaluation der regionalen Versorgung oder Zufriedenheit der Lernenden mit dem Angebot der Einrichtung. Ähnlich ist dies bei der Evaluation von Lehr-/Lernprozessen: Wenn es vor allem um das Lehrverhalten des Pädagogen geht, sind andere Daten erforderlich als bei den Lernstrategien oder Lernproblemen der Teilnehmenden.

Es liegt also bereits in der Entscheidung darüber, was genau evaluiert werden und anhand welcher Daten das geschehen soll, ein „vorgängiges“ Werturteil vor, das in Evaluationsziel und Datenselektion mündet. Es gibt jedoch nur die allgemeine Richtung vor, gewissermaßen den Bewertungskorridor, die Bewertung selbst erfolgt anhand konkreterer Kriterien und nach einem geregelten Analyse- und Bewertungsverfahren.

Darum geht es in diesem Kapitel. Es werden Interpretationsverfahren vorgestellt. Insbesondere wird die Differenz von Interpretieren und Bewerten verständlich gemacht. Im Zentrum befinden sich hier die Kriterien; sie sind notwendig, denn ohne sie kann keine Bewertung stattfinden. Sie ermöglichen es, plakative Aussagen wie „gut“ zu differenzieren und zu begründen – und damit auch verständlich und nachvollziehbar zu machen. Aber auch das Verfahren ist wichtig, denn es ermöglicht die Überprüfbarkeit des Werturteils.

Nach der Lektüre sollten Sie in der Lage sein,

- vorliegende Evaluationsergebnisse auf die zugrundeliegenden Kriterien zu überprüfen,
- die Schlüssigkeit des Interpretationsverfahrens zu beurteilen,
- Evaluationen mit Augenmaß in den jeweiligen Kontext einzuordnen und
- selbst Kriterien für eine Evaluation zu entwickeln.

6.1 Kriterien der Interpretation und Bewertung

Die Bewertungskriterien stehen im Mittelpunkt des Evaluationsvorhabens. Genau genommen steht und fällt mit ihnen die gesamte Evaluation. In Bezug auf Fragen und

Methoden sind die Unterschiede zwischen Forschungen und Evaluationen nicht wesentlich, jedoch in Bezug auf die Bewertungen und deren Kriterien sind sie es. Da Evaluationen immer einen „Gegenstand“ haben, der über eine eigene Stimme verfügt, und immer im „Feld“ angesiedelt sind, gibt es keine „objektiven“ Kriterien und letztlich keine Kriterien, die nicht auch bei dem einen oder anderen Adressaten oder in der einen oder anderen Weise „anstößig“ sind.

Die Frage, wie die Kriterien entstehen bzw. entstehen sollen, wird unterschiedlich beantwortet. So fällt etwa nach Scriven (2007) die Formulierung und Festlegung der Kriterien in die Kompetenz der Evaluatoren. Nach Stake (1995) obliegt die Festlegung der Kriterien den Betroffenen, also letztlich dem Evaluationsgegenstand selbst, der am ehesten abschätzen kann, was an Erkenntnissen bei der Verfolgung der eigenen Ziele erforderlich ist. Andere Ansätze weisen die Kompetenz der Kriterienfestlegung dem Auftraggeber zu, in dessen Interesse ja die Evaluation stattfindet. Kromrey (2007) schließlich lehnt – von Ausnahmen abgesehen – Bewertungen in Evaluierungsverfahren aus forschungsethischen Gründen gänzlich ab – „damit droht allerdings Evaluation ihren spezifischen Unterschied zur wertedistanzierten Forschung zu verlieren“ (Beywl/Balzer o.J. S. 67).

In der Diskussion um die Frage, wer mit welchem Umfang zuständig ist für die Festlegung der Bewertungskriterien spielen vor allem vier Aspekte eine Rolle: die Legitimation der Kriterien, ihr Ableitungskontext, ihre Kohärenz und ihr Bezug zum Gegenstand.

Legitimation

Kriterien benötigen eine Legitimation, um das Werturteil, das auf ihrer Grundlage entsteht, zu rechtfertigen und akzeptabel zu machen. Diese Legitimation kann aus unterschiedlichen Quellen kommen. Die nächstliegende Legitimation ist die des Nutzens für die „Zielgruppe“, hier: für die lernenden Erwachsenen. So kann eine Evaluation etwa aufzeigen, dass ein angemessener Methodenwechsel den Lernprozess unterstützt – und sie hat das auch schon getan. Eine Legitimation kann auch aus normativen Vorgaben kommen, etwa aus den Richtlinien eines Akkreditierungsverfahrens. Eine häufige Legitimation bei Drittmittel-Aufträgen ist die der sinnvollen Investition (etwa bei Projektvergaben), das Kriterium also die Zielerreichung des Projekts. Schließlich kann die Legitimation auch aus übergeordneten Expertise-Kontexten kommen, etwa dem wissenschaftlichen System, demzufolge die wissenschaftliche Begründung des Evaluierten im Mittelpunkt steht.

Ganz gleich wie die Kriterien im Einzelnen legitimiert sind oder werden: Die Legitimation ist die Voraussetzung für die Akzeptanz der Evaluation im Verlauf und im Ergebnis. Werturteile, die auf der Grundlage fremder, nicht legitimierter und nicht akzeptierter Kriterien gefällt werden, haben letztlich keine Wirkung in der Realität.

Ableitungskontext

Wo auch immer die Kriterien herkommen: Sie kommen von irgendwo her und sie haben einen Ableitungskontext. Und dieser ist gewöhnlich mit einem theoretischen Bezug versehen. Ein Kriterium, das Lehrtätigkeit am angemessenen Methodenwechsel misst und bewertet, ist aus einer Lerntheorie abgeleitet, der zufolge (erwachsene) Menschen eben dies für ein besseres Lernen benötigen, weil die aneignenden Zugänge zu unterschiedlichen Gegenständen unterschiedlich sein müssen, weil die Aufmerksamkeit erhalten bleibt und die Nähe des Lernprozesses bei den Lernern belegt ist. Das Gleiche gilt für Kriterien, die etwa Transparenz, Partizipation und Kommunikation im Lernprozess beurteilen.

Aber auch andere Kriterien haben Ableitungskontexte. Vor allem dann, wenn sie nach dem Prinzip des Vergleichs strukturiert sind (besser/schlechter als ...). Vedung (2000) hat zehn solcher Ableitungskontexte benannt:

1. historischer Vergleich („besser als früher“)
2. nationaler Vergleich („besser als andere im Land“)
3. internationaler Vergleich („besser als andernorts“)
4. Benchmarkvergleich („besser als Richtwerte“)
5. Zielvergleich („besser als die Zielvorgabe“)
6. Adressatenabgleich („bedarfsgerecht“)
7. Interessenabgleich („den Interessen der Stakeholder gerecht“)
8. Professionsvergleich („professionellen Standards entsprechend“)
9. Minimalabgleich („Mindestansprüche erfüllt“)
10. Optimalvergleich („besser als andere entsprechende Modelle etc.“)

Bedeutsam wegen ihrer Reichweite in Förderprogrammen sind auch die Normen der Europäischen Union oder der OECD für ihre Projekte und Programme, die sich sehr ähneln. Sie formulieren als Bezugsrahmen der Kriterien die Relevanz, die Effektivität, die Effizienz, den Impact und die Nachhaltigkeit.

Wieder andere Kriterien entstammen einem wissenschaftlichen Ableitungskontext, der dann zu einem normativen Paradigma wird. Sie finden sich etwa in Evaluationen von Einrichtungen hinsichtlich von Strukturen, Arbeitsabläufen und Ressourcenverteilungen und stammen aus betriebswirtschaftlichen oder aus Managementwissenschaften. Und eher aus politik- und sozialwissenschaftlichen Kontexten abgeleitet sind Kriterien bei der Systemevaluation; dies sind etwa Kriterien wie „Erreichen benachteiligter Zielgruppen“ oder „Flächendeckendes Weiterbildungsangebot“.

Kohärenz

Kohärenz bedeutet in diesem Zusammenhang vor allem, dass die Kriterien widerspruchlos zueinander sein müssen, wenn das Ergebnis der Evaluation nicht unklar

erscheinen und die Konsequenzen nicht chaotisch sein sollen. Ein „Klassiker“ in Weiterbildungseinrichtungen ist dabei die schwer zu vereinbarende Ambivalenz zwischen pädagogischen und ökonomischen Zielen. Sind beide gleichzeitig und gleichwertig Evaluationskriterien, wird es zu keiner sinnvollen Weiterentwicklung der Einrichtung kommen. Das Evaluationskriterium muss sich daher auf die Art der Bewältigung des potentiellen Widerspruchs in der Einrichtung richten.

Es gibt bei allen Evaluationsgegenständen in der Erwachsenenbildung solche Kohärenzprobleme. Dies liegt nicht zuletzt am Charakter der Erwachsenenbildung – sie ist ein deutungsöffener und differenzierter gesellschaftlicher Bereich. Kohärenz setzt in gewisser Weise die Eindeutigkeit eines Zieles voraus. In Programmevaluationen ist sie daher gewissermaßen *per se* gegeben, das Programm verfolgt ein definiertes Ziel, auf das sich die Kriterien richten. Aber selbst dort kann es Probleme mit der Kohärenz der Bewertungskriterien geben, wenn es entweder eine Zielpluralität (ohne Prioritäten) oder gar eine Zieloffenheit gibt. Programme, die etwa nur das Ziel „Verbesserung der Lernsituation von Migrant/inn/en“ verfolgen, sind insofern ergebnisoffen, als das Ziel nicht operationalisiert ist. Und Programme, die das Ziel „Verbesserung der Lernsituation von Migrant/inn/en und Arbeitslosen“ verfolgen, erlauben eine Zielvarianz auch in den Kriterien.

Ähnlich ist es bei den Kriterien für die Bewertung von Lernprozessen und Lernerfolgen. Hier kann es auf der einen Seite um die Erhöhung von Wissensbeständen gehen, auf der anderen Seite um verstärktes Selbstbewusstsein und erhöhte Motivation der Lernenden. Oder auch im Systemkontext: Regionale Versorgungsgerechtigkeit einerseits und Qualitätssicherung der Einrichtungen andererseits sind unterschiedliche Ziele, die auch miteinander konfliktieren können.

Evaluationsberichte, die auf der Grundlage weniger kohärenter Kriterien entstanden, lesen sich oft sehr überraschend. Trotz guter Bewertung zu Aspekten des einen Ziels/Kriteriums enthalten sie dann Passagen, die aus einem anderen Ziel- und Kriteriensystem stammen. Solche ambivalenten Evaluationsergebnisse können naturgemäß zur Konfusion führen, wenn es um die Umsetzung von Empfehlungen geht.

Bezug zum Evaluationsgegenstand

Eine weitere Anforderung an die Kriterien, aufgrund derer beurteilt wird, liegt in der Beziehung zum Untersuchungsgegenstand, etwa der Einrichtung, dem Lernprozess oder dem System. Bewertet man den Gegenstand mittels Kriterien, die gar nicht zum Zielsystem des Gegenstands gehören, tut man ihm mit Sicherheit Unrecht. Es ist daher immer auch zu prüfen, was denn überhaupt Ziele und Selbstverständnis des Evaluationsobjekts sind.

BEISPIEL**Passende Evaluationskriterien**

Eine Einrichtung, die sich der politischen Bildung verschrieben hat, sollte nicht daraufhin evaluiert werden, ob und welchen Beitrag sie zur „employability“ der Teilnehmenden leistet. Ein Weiterbildungssystem, das auf Daseinsvorsorge und regionale Weiterbildungsdichte orientiert ist, wird sich schwerlich bei Evaluationskriterien der Exzellenz wiederfinden (das gilt bspw. auch für Hochschulen). Und Lernprozesse, die sich der beruflichen Leistungsfähigkeit der Lernenden verpflichtet fühlen, sind zu Unrecht einem Kriterium unterworfen, das die Selbstentfaltung der Lernenden in den Mittelpunkt stellt.

Bereits in der Anlage der Evaluation sind daher die Ziele und Gegebenheiten des Evaluationsgegenstands so ausführlich in den Blick zu nehmen, dass eine angemessene Passung hergestellt wird. Dies erhöht die Treffsicherheit des Urteils ebenso wie die Akzeptanz von Empfehlungen.

Die Festlegung der Kriterien, die dem Urteil zugrunde liegen, hat noch weitere Implikationen. Die wichtigste von ihnen ist die Abfrage der erforderlichen Daten. Die genaue Angabe, was zur Beurteilung anhand welcher Kriterien ansteht, ist die Voraussetzung, die dafür notwendigen Daten zu erhalten. Dabei zeigt sich nicht selten, dass auch hier Passungsnotwendigkeiten bestehen. Es kann sein, dass vorgesehene Beurteilungskriterien deshalb nicht greifen, weil die benötigten Daten gar nicht verfügbar sind – und auch nicht erhoben werden können (das gilt etwa für längere Zeitreihen).

Schließlich ist auch von großer Bedeutung, ob die Evaluatoren „Herren“ der Kriterien sind. Ob sie einerseits festlegen können, welche Kriterien (operationalisiert!) angelegt werden sollen (nach Verhandlungen mit den Auftraggebern und dem Evaluationsgegenstand) und ob sie andererseits die Kriterien auch anwenden können. Letzteres ist nicht nur eine Frage der Bedingungen während der Evaluation (etwa bei Begehungen von Einrichtungen), sondern auch der Kompetenz der Evaluatoren. Genau genommen müssen Evaluatoren (vor allem im Peer-Review) auf die festgelegten Kriterien hin vereinheitlicht, jüngere Mitarbeitende in Evaluationsteams entsprechend geschult werden.

6.2 Auswertungsfelder und -verfahren

Wenn alle Daten erhoben bzw. ermittelt und aufbereitet worden sind, die Fragen operationalisiert und die Kriterien festgelegt sind, kann es – dem Evaluationsdesign entsprechend – an die Auswertung gehen. Die Auswertung erfolgt grundsätzlich in zwei voneinander getrennten Schritten: der Beschreibung als erstem Schritt und der Interpretation und Bewertung als zweitem Schritt (→ Kap. 6.3).

Erkenntnislogisch ist dies, was die Trennbarkeit angeht, durchaus umstritten. Natürlich fließen in die deskriptiven (beschreibenden) Arbeiten immer schon bewertende und interpretierende Aspekte mit ein, schon allein durch die Auswahl der Fragen, die Formulierung von Hypothesen und die Definition von Merkmalen und Indikatoren. Spätestens seit dem Positivismusstreit in der deutschen Sozialforschung ist dieser Sachverhalt bekannt und immer wieder zu reflektieren. Dies findet allerdings eher bei Grundlagenforschung als bei Evaluations- und angewandten Forschungen statt.

Beginnen wir mit der Beschreibung. Dazu ist es zunächst notwendig, die vorhandenen Informationen, welche nun – über die Methode ihrer Erfassung – zu überprüf- baren „Daten“ wurden, zu sichten und zu überprüfen. Es handelt sich um eine letzte Kontrolle der Vollständigkeit, der Qualität und der Bearbeitbarkeit des vorliegenden Datenmaterials. Im Falle einer „Triangulation“ umfasst dieser Schritt auch das gegenstandsbezogene Zusammenführen der Daten aus den unterschiedlichen Erhebungen.

Sodann vollzieht sich die Beschreibung des Evaluationsgegenstands anhand der vorliegenden Daten in mehreren Stufen, die zwischen einer quantitativen und einer qualitativen Herangehensweise kaum Unterschiede aufweisen. Diese Stufen sind (in Anlehnung an Weiss (1998):

- das *Zählen*

In diesem Schritt wird der Gesamtumfang der vorliegenden Daten erfasst sowie die Größenordnung einzelner für die Evaluation wichtiger Aspekte und Merkmale. Äußerungen in Interviews, Teilnahmezahlen, Finanz- und Angebotszahlen, Beobachtungsdaten etc. werden – je nach Evaluationsgegenstand und ausgewählten Merkmalen – quantifiziert. In qualitativen Ansätzen endet dieser Schritt mit Aussagen wie „mehr“ oder „seltener“, in quantitativen Ansätzen mit respektablen Skalen auf Nominalniveau.

- das *Zerlegen*

In diesem Schritt wird die Totalität des Evaluationsgegenstands in seine Einzelteile, soweit sie datenmäßig erfasst sind, buchstäblich „zerlegt“. Ein Lehr-/Lernprozess besteht danach aus Sequenzen, eine Einrichtung aus Programmen und Budgets, ein System aus Gesetzen und Einrichtungen sowie regionalen Strukturen. Die Kunst des Zerlegens liegt darin, das Gesamte immer im Blick und rekonstruierbar zu halten.

- das *Gruppieren*

In diesem Schritt werden Daten, die ein ähnliches Merkmal betreffen oder ausweisen, zusammengefasst. Im Prozess der qualitativen Auswertung ist dies etwa die Kategorienbildung, die über Paraphrasen gesteuert induktiv aus dem Material entsteht, im Falle einer quantitativen Auswertung werden hier Faktoren- und Clusteranalysen sowie Skalierungen vorgenommen.

- *das Vergleichen*
In diesem Schritt wird der schon beim „Gruppieren“ angelegte Prozess der Gruppierung insofern fortgesetzt, als nun zwischen den Gruppen nach Unterschieden, vor allem auch nach Gemeinsamkeiten gesucht wird. Bei Untersuchungen des Lehr-/Lernprozesses etwa geht es um ein unterschiedliches Verhalten von Lernenden in unterschiedlichen Lernsettings, bei Einrichtungen um Ähnlichkeiten und Unterschiede in Programmplanungen und Marketingstrategien. Vor allem quantitative Ansätze ermöglichen hier meist Ergebnisse, so beim Vergleich von Lernergebnissen oder Aktivitätsprofilen.
- *das Verknüpfen*
In diesem Schritt geht es darum, einzelne Befunde zu Zusammenhängen zu verknüpfen. In der qualitativen Auswertung sind dies plausible Zusammenhänge, in der quantitativen bi- oder multivariate Korrelationen. So kann man etwa mit beiden Verfahren die Frage überprüfen, ob eine höhere Investition in die Kompetenz der Lehrenden zu einem verbesserten Ergebnis bei den Lernenden führt. Dieser Schritt setzt Annahmen („Hypothesen“) voraus, um nach sinnvollen Verknüpfungen suchen zu können; diese stellen sich vielfach auch erst bei Betrachtung der Basisdaten ein.
- *die Kontextprüfung*
In diesem Schritt wird nach dem gesucht, was möglicherweise fehlt. Welche Zusammenhänge und Auffälligkeiten auch feststellbar sind, man muss immer überprüfen, ob nicht ein anderer, nicht in den Daten abgebildeter Faktor dafür verantwortlich ist oder zumindest sein könnte. So kann etwa ein Lehr-/Lernprozess völlig unterschiedlich verlaufen abhängig davon, wie die Außentemperatur ist – die aber nicht erfasst wurde.

Das Ergebnis dieser ersten Stufe des Auswertungsverfahrens ist eine Beschreibung des Untersuchungsgegenstands. Sie ist nie vollständig, dies ist weder möglich noch intendiert. Aber sie muss diejenigen Merkmale des Untersuchungsgegenstands abbilden, die für die Analyse und Bewertung notwendig sind.

Nicht selten gibt es bereits bei dem Ergebnis dieser ersten Stufe, der Beschreibung, Unklarheiten und Widerspruch seitens des Betroffenen. Das kann sich auf die Makroebene so wie die Mikroebene der Beschreibung beziehen. Auf der Mikroebene handelt es sich um einzelne Begriffe, die als nicht passend angesehen, oder um logische Verknüpfungen, die nicht akzeptiert werden. Auf der Makroebene geht es meist um die Gewichtung der einzelnen Elemente, ihrer Wertigkeit für den Evaluationsgegenstand und untereinander.

Vielfach verfahren Evaluationen so, dass sie hier einen Zwischenschritt vorsehen, in dem sie eine Rückkoppelungsschleife zu der Beschreibung mit dem Evaluationsgegenstand einbauen und dessen Kommentare dann gegebenenfalls noch zur Korrektur verwenden. Das erhöht die Präzision und Akzeptanz, kann aber auch zu Verfälschungen führen.

6.3 Auswertung und Interpretation von Daten

Wurden die vorliegenden Daten geprüft, aufbereitet, sortiert und geordnet, geht es an die Auswertung und Interpretation. Natürlich ist diese Reihenfolge idealtypisch; nur zu oft ist es notwendig, während der Auswertung noch einmal zurückzugehen zu den „Rohdaten“ und sie zu überprüfen oder neu zu ordnen.

Die Auswertung und die Interpretation sind zwei Seiten der gleichen Medaille. Vorab-Interpretationen („Hypothesen“) steuern die Auswertung, Auswertungsergebnisse kreieren neue Interpretationen. Wichtig ist, immer die zugrunde liegenden Kriterien (→ Kap. 6.1) zu reflektieren.

In der Vorgehensweise bei Auswertung und Interpretation unterscheiden sich der quantitative und der qualitative Ansatz deutlich. Im quantitativen Ansatz dominiert die statistische Operation mit Zahlen, im qualitativen Ansatz die Interpretation von Begriffen und strukturellen Zusammenhängen. Beide haben jedoch Schnittmengen: Der quantitative Ansatz kommt ohne hypothesenorientiertes „Verstehen“ der ermittelten Daten nicht aus, der qualitative Ansatz verwendet häufig „weiche“ (Worte wie „oft“, „meist“, „selten“ etc.) oder „härtere“ (Quantitäten etwa von genannten Kausalitäten) Quantifizierungen.

In beiden Fällen muss der Plan der Datenauswertung und -interpretation den gestellten Fragen folgen, es müssen wissenschaftlich vertretbare Antworten auf diese Fragen erbracht werden.

BEISPIEL

Teilnehmer-Zufriedenheit

Die Frage, ob sich die Teilnehmer-Zufriedenheit an Seminaren eines Anbieters in den vergangenen zehn Jahren verbessert hat, lässt sich am ehesten quantitativ beantworten; entsprechende Items eines standardisierten Fragebogens sind in einer Zeitreihe darstellbar. Die Frage, ob das Lehrverhalten eines Pädagogen dem eigenen Lernstil liegt, lässt sich besser in einer qualitativen Befragung der Teilnehmenden ermitteln und entsprechend differenziert auswerten.

Quantitative Verfahren der Auswertung und der Interpretation sind – gerade auch mit den modernen Mitteln der Datenverarbeitung – mittlerweile hoch ausdifferenziert und komplex. Sie hier darzustellen sprengt den Rahmen des Buches; daher wird hier auf die praxisbezogene Darstellung multivariater Verfahren bei Backhaus u.a. (2008, 12. Aufl.) sowie das Kapitel „Evaluationsstudien zur Prüfung von Effekten“ in Bortz/Döring (2002, 3. Aufl.) verwiesen. In Bezug auf qualitative Auswertungen empfiehlt es sich, den Text zur Lehr- und Trainingsevaluation von Kuckartz u.a. (2008, 2. Aufl.)

sowie den „Klassiker“ von Mayring zur Inhaltsanalyse (2007, 9. Aufl.) zu studieren. Das Buch von Aeppli u.a. (2011) ist eine gute Einführung.

Der quantitative Ansatz

Im quantitativen Ansatz geht es darum, Erkenntnisse über – wie der Name sagt – die Messung von Mengen zu gewinnen. Natürlich setzt das entsprechende Erhebungsverfahren voraus, Mengen müssen vorhanden sein. Man erzeugt solche Mengen am einfachsten dadurch, dass man Items oder Variablen des Evaluationsgegenstands festlegt, die erfassbar und messbar sind. Sind sie nicht *direkt* erfassbar (wie „Zufriedenheit“), sind *Indikatoren* („Anzeiger“) zu finden, die auf die zu messende Variable verweisen.

Bei der Erhebung wird gewöhnlich mit einem standardisierten Verfahren (z.B. einem standardisierten Fragebogen) gearbeitet, das unmittelbar zu einer ausreichenden Menge von Daten führt. In diesem standardisierten Verfahren wird die statistische Auswertung in der Regel bereits durch eine entsprechende *Codierung* vorbereitet, also eine binär abbildbare Verschlüsselung der Daten. Mit ihr kann die statistische Auswertungsoperation erfolgen.

Wichtig ist, dass die Menge der erhobenen Daten ausreicht, um statistische Operationen überhaupt durchführen zu können. Dabei ist die erforderliche Datenmenge abhängig von der Komplexität der Rechenoperationen: Bei bivariaten Korrelationen (also der Beziehung zwischen zwei Items) sind weniger Daten nötig als bei multivariaten Korrelationen.

Die Datenmenge ist aber auch noch aus einem anderen Grund relevant, aus dem der Repräsentativität. Will man sicher sein, dass die vorhandenen Daten den Gegenstand abbilden, so muss die bearbeitete Datenmenge die Gesamtheit (die „Grundgesamtheit“) angemessen repräsentieren. Dazu gibt es gültige Erfahrungs- und Belegwerte (vgl. Nuissl 2010a, S. 100ff.).

BEISPIEL

Image einer Weiterbildungseinrichtung

Will man das Image einer Weiterbildungseinrichtung bei den Teilnehmenden erfassen (als Evaluationskriterium: positives Image), so ist die Zahl der Teilnehmenden in einem definierten Zeitraum (z.B. Arbeitsjahr) festzustellen und von ihnen ist dann ein akzeptabler Prozentsatz zu befragen (z.B. 20% bei 1000 Teilnehmenden oder 10% bei 5000 Teilnehmenden). Um die Repräsentativität der erhaltenen Daten noch zu verbessern, kann man überprüfbare Verfahren der „Samplebildung“ anwenden, also einer geschichteten Auswahl der Teilnehmenden z.B. nach Alter und Geschlecht. Dieses „Sample“ wird auch oft „Stichprobe“ genannt.

Neben der Repräsentativität ist bei statistischen Auswertungen vor allem die Frage der *Signifikanz* von Bedeutung. Sie besagt, dass bei einem Vergleich zwischen zwei Mengen der gemessene Unterschied so groß ist, dass er aussagefähig ist. Bei einem Vergleich etwa zwischen zwei Seminargruppen, die mit unterschiedlichen Methoden arbeiten, müssen die Unterschiede bei den festgelegten Merkmalen (etwa Lernerfolg, Zufriedenheit, Dauer) groß genug sein, um eine bewertende Aussage zur Güte der gewählten Methode machen zu können.

Signifikante Ergebnisse müssen auch praktisch bedeutsam sein (...) [es gibt] einen wichtigen Zusammenhang zwischen der Wahl eines angemessenen Stichprobenumfangs und der Wahrscheinlichkeit, ein praktisch bedeutsames Ergebnis auch statistisch absichern zu können (Bortz/Döring 2002, S. 599).

Der quantitative Ansatz wird in der Evaluation vor allem bei den Gegenständen angewandt, die quantifizierbar sind, ohne dass ihre inhaltliche Bedeutung darunter leidet. Dies etwa bei Zeitreihen der Finanzierung (bei Systemen und Einrichtungen) oder der Teilnahme (bei Seminaren und Zielgruppen), Leistungsunterschiede zwischen Einrichtungen (Mitarbeiterzahl/Angebotsumfang) oder Systemen (Abbruchquoten, Teilnehmerquoten, Übergangsquoten). Auch hinsichtlich Lernergebnissen und Systemleistungen liefern quantitative Analysen relevante Erkenntnisse, wie die großen Kompetenzstudien der Vergangenheit zeigen (z.B. PISA, in Kürze PIAAC).

Ergebnisse des quantitativen Zugangs sind jedoch nicht immer unumstritten. So sind etwa Leistungsparameter bei Wissenschaftlern, die sich auf den *citation index* beziehen (also die Häufigkeit des Zitierens in anderen Arbeiten) je nach Disziplin und Sprache sehr unterschiedlich einzuschätzen. Auch Vergleiche von Einrichtungen, die Leistungsparameter (wie Personalstärke und Angebotsumfang) zugrunde legen, ohne die Spezifika des Einzugsgebiets zu berücksichtigen, gehen an den Realitäten vorbei.

Der qualitative Ansatz

Bei diesem Ansatz geht es, wie der Begriff „Qualität“ betont, darum, die Beschaffenheit des Gegenstands zu erkennen, zu verstehen und zu bewerten – nach Möglichkeit in seiner ganzen Komplexität. Er arbeitet insbesondere an der Interpretation von Aussagen in Texten, Interviews und Gesprächen, versucht, dem jeweiligen Sinn auf die Spur zu kommen und die Zusammenhänge wieder herzustellen.

Eine wichtige Rolle spielt beim qualitativen Ansatz das sinnerfassende Verstehen, sei es über strukturiertes *Zuhören* und Nachfragen oder über Auswertungsverfahren, die mit *Paraphrasen* arbeiten. Die größte Schwierigkeit beim qualitativen Ansatz liegt darin, in einem Universum von verwendeten Wörtern und logischen und kausalen Ver-

knüpfungen erklärende Begriffe und Kategorien zu finden, welche die erfasste Wirklichkeit möglichst strukturiert und zugleich komplex wiedergeben.

Der zentrale Begriff des qualitativen Ansatzes ist die Paraphrase. Sie wiederholt mit anderen Worten, was gesagt oder geschrieben worden ist. In dieser Wiederholung verwendet der Evaluator „seinen“ Begriff und prüft, ob er dem – mit anderen Worten – Gesagten entspricht.

BEISPIEL

Sachlichkeit

In der Bewertung einer Lehrmethode sagt ein Teilnehmer: „Ich fand diese Arbeitsweise zu verspielt“, ein anderer: „Ich finde, man sollte das Thema ernst nehmen“, ein dritter: „Ich habe mich gut unterhalten“. In der Paraphrase könnte möglicherweise das – unterschiedlich bewertete – Prinzip der „Sachlichkeit“ der Begriff sein, auf den sich die drei Wahrnehmungen und Urteile beziehen lassen.

In der Regel geht der qualitative Ansatz in geordneten Schritten vor: Er beginnt mit einer Zusammenfassung und Gliederung des Gesagten oder des Textes (vor allem des transkribierten Interviews) und teilt ihn in Abschnitte ein, die entweder nach Themenwechseln beginnen oder nach Wechseln der Perspektive etc. Diese Abschnitte werden dann einzeln interpretiert, mit Paraphrasen-Technik analysiert und kategorisiert.

Der dritte Schritt ist die eigentliche Interpretation, in der auch andere Informationsquellen als Text oder Protokoll herangezogen werden.

In der „eigentlichen“ Interpretation sind – trotz aller strukturierenden und systematisierenden Analysen – immer mehrere „Lesarten“ möglich, also unterschiedliche Deutungen. Sie werden in einer ganzheitlichen Re-Analyse durch eine „*Plausibilitätsprüfung*“ reduziert; in dieser Prüfung zeigt sich, dass einige der möglichen Interpretationen des Textes oder der Interviewpassage zwar möglich, aufgrund des Kontextes aber weitgehend ausschließbar sind.

Im Verfahren der Evaluation ist die Systematik von Mayring dabei weitgehend anwendbar. Sie verwendet ein klar vorgegebenes Regelsystem, das sich hauptsächlich auf den Inhalt bezieht, die weniger beachteten subjektiven, biografischen oder deutenden Aspekte sind bei Evaluationen nur selten im Mittelpunkt.

Qualitative Verfahren werden bei Evaluationen vor allem dann angewandt, wenn der zu bewertende Gegenstand kaum standardisiert und messbar zu erfassen ist – vor allem bei Lehr-/Lernprozess-Analysen, aber auch häufig bei Systemevaluationen.

„Triangulation“

In den meisten Evaluationen findet sich eine Mischung beider Ansätze, vielfach auch „Triangulation“ genannt. Dabei werden die Vorteile nach Möglichkeit kombiniert. Ein Beispiel dafür sind etwa Peer-Review-Evaluationen bei Einrichtungen; hier werden die Daten zusammengetragen, die als wichtig für die Beurteilung der Einrichtung angesehen werden, von den „Peers“ studiert und analysiert und anschließend in einer „Begehung“ hinterfragt – diese dient dann der Überprüfung der Daten und einer Erklärung ihres Entstehens und Zusammenhangs. Auch dies sind „Plausibilitätsprüfungen“, wie sie in vielen Varianten kommunikativer Rückkoppelung sinnvoll und nötig sind.

6.4 Empfehlungen

Die „Königsdisziplin“ bei Evaluationen ist die Formulierung von Empfehlungen. Sie umfasst die schwierigsten und oft auch strittigsten Elemente der Evaluation. Nicht selten wird – der besonderen Probleme wegen – auch auf eine explizite Formulierung von Empfehlungen seitens der Evaluatoren verzichtet. Large-Scale-Evaluationen wie das „Programme for International Student Assessment“ (PISA) sind dafür gute Beispiele: Was die Konsequenz aus den Daten der PISA-Studie sein soll, ist – wie man anhand der deutschen PISA-Diskussion sehen kann – hoch umstritten.

Zunächst sind Evaluationsergebnisse daraufhin zu prüfen, ob sie Empfehlungen enthalten. Sind Empfehlungen explizit ausgewiesen, ist dieser Schritt einfach. Aber auch dann kann es sein, dass darüber hinaus implizite Empfehlungen in den formulierten Ergebnissen enthalten sind. So enthalten die meisten bewertenden Aussagen bei Ergebnissen eine Handlungsempfehlung, auch wenn die Richtung und das Ziel unbestimmt sind. Textstellen wie „Ein Problem ist ...“, „... zu wenig ...“ oder „... zu viel ...“ oder „... nicht angemessen ...“ verweisen stets auf einen Handlungsbedarf, während Textstellen wie „... unter einer vergleichbaren Einrichtung ...“ oder „... stark zugeschnitten auf ...“ zwar eine Bewertung enthalten, diese aber kaum eine Handlungsoption aufzeigt.

Bewertende Aussagen des ersten Typs, die aus einer normativen Sicht stammen (→ Kap. 6.2), legen auch ohne explizite Empfehlung eine zu verfolgende Handlungsoption nahe. Sie sind insofern kritisch zu sehen, als sie das normativ hinterlegte Modell nicht offenlegen und für den Fall einer Handlung auch keine Zielmarke definieren. Ihr Vorteil ist, dass sie die Operationalisierung der Konsequenz den Lesern der Evaluationsergebnisse überlassen, insofern also noch handlungsoffen sind.

Anders verhält es sich mit expliziten Empfehlungen, welche den Evaluationsergebnissen angefügt oder in sie integriert werden. Sofern die Evaluationsergebnisse öffentlich zugänglich sind, sind es auch die Empfehlungen. Bei Evaluationen im staatlichen Auftrag sind die Ergebnisse und damit die Empfehlungen fast immer öffentlich zugänglich und

somit transparent, während bei privat initiierten Evaluationen keine Veröffentlichung erfolgt. Veröffentlichte Empfehlungen haben in der Regel eine größere Wirkkraft.

Explizite und ausformulierte Empfehlungen sind zwar weniger flexibel, was die Operationalisierung angeht, haben jedoch den Vorteil, dass sie kritisch betrachtet und analysiert werden können. Eine solche Analyse von Empfehlungen, vor allem vorgenommen vom „Untersuchungsgegenstand“, prüft zunächst immanent, ob

- die Empfehlungen im Rahmen der Evaluation legitimiert sind,
- die Empfehlungen durch die Ergebnisse belegt sind,
- die Empfehlungen in sich widerspruchsfrei sind,
- die Empfehlungen eine realistische Reichweite haben,
- die Empfehlungen einen Grad von Verbindlichkeit haben,
- die Empfehlungen Umsetzungsverfahren enthalten und
- die Empfehlungen ein normativ gesetztes Ziel anvisieren.

Legitimation

Die Legitimation von Empfehlungen richtet sich danach, wie gehalt- und qualitätsvoll die Evaluation ausgeführt und ob sie mit der notwendigen Objektivität und Distanz versehen wurde. Schon im Vorfeld gehen diese Kriterien in das Design der Evaluation und die Auswahl der Evaluatoren ein; bei Peer-Review-Evaluationen etwa ist streng darauf zu achten, dass Befangenheiten der Peers (gleichgültig ob positiver oder negativer Art) ausgeschlossen werden können. Empirisch-analytische Evaluationen basieren auf professionell erstellten und im Feld abgestimmten Erhebungsinstrumenten. Die Verlässlichkeit der Evaluation hinsichtlich dieser Normen ist Voraussetzung dafür, dass Empfehlungen als legitim betrachtet werden.

Beleg

Die Frage, ob die Empfehlungen durch die Ergebnisse der Evaluation gedeckt sind, ist von großer Bedeutung. Vielfach lassen sich aus Ergebnissen ganz unterschiedliche Empfehlungen ableiten, die nicht selten auch mit unvollständigen Indikatoren der Evaluation zusammenhängen. Niedrige Kompetenzen von Schüler/inne/n bspw. können an der Schule liegen, am Einzugsbereich, am Unterricht, am fehlenden Support; sofern nicht zu allen Aspekten Ergebnisse vorliegen, sind Empfehlungen kritisch zu sehen. Liegen zu allen Aspekten Ergebnisse vor, kann es immer noch zweifelhaft sein, ob sich die Empfehlungen auf eine Veränderung der Schule oder des Schultyps richten sollen oder auf einen stärkeren Support im Umfeld der Schule oder auf alles gleichzeitig. Auch wenn die Ableitung von Empfehlungen mehrdeutig sein kann, so ist doch immer erforderlich, dass sich das, was letztlich empfohlen wird, aus den Untersuchungsergebnissen ableiten lässt. Empfehlungen, die sich nicht auf Untersuchungsergebnisse stützen können, werden mindestens kritisch hinterfragt werden.

Widerspruchsfreiheit

Die Widerspruchsfreiheit der Empfehlungen ist eine wesentliche Voraussetzung für ihre Akzeptanz. Empfehlungen können widersprüchlich sein hinsichtlich der Richtungen, die sie anzeigen, aber auch hinsichtlich der Abstraktionsebenen und Zielperspektiven, die sie ansteuern. So kann etwa die Empfehlung, den Lehr-/Lernprozess partizipativer zu gestalten, im Widerspruch zu der (gleichzeitigen) Empfehlung stehen, die dort verwendeten Materialien zu aktualisieren und verbindlicher einzusetzen. Oder es kann die Empfehlung, ein breiteres Angebot für Migrant/inn/en zu erstellen, im Widerspruch stehen zu der (gleichzeitigen) Empfehlung, einen höheren Anteil der Kosten seitens der Teilnehmerentgelte zu finanzieren. Gerade bei größeren Evaluationen, die einen komplexen Sachverhalt in den Blick nehmen, können auf diese Art Empfehlungen, die in unterschiedliche Richtungen gehen, nebeneinander stehen. Kritisch zu sehen ist dann, ob überhaupt eine und welche der Empfehlungen im weiteren Verlauf berücksichtigt wird.

Realistische Reichweite

Die Reichweite der Empfehlungen sollte realistisch sein, wenn sie wirken sollen. Realistisch heißt, dass sie nicht einen Zustand des Evaluationsgegenstands ansteuern, der in zu weiter zeitlicher Ferne liegt oder eine zu große Änderung gegenüber dem ermittelten Status quo bedeutet. Empfehlungen solcher Art können demotivieren oder zu unvorhersehbaren Entwicklungen führen. Empfehlungen haben daher oft Ähnlichkeiten mit der Struktur von Fünf-Jahrs-Plänen: Sie richten sich auf einen ähnlichen Zeitraum, visieren ein konkretes Ziel und knüpfen an den Stand der Evaluation an.

Verbindlichkeit

Empfehlungen werden mit dem Ziel ausgesprochen, dass sie eine Leitlinie für das weitere Handeln des Evaluationsgegenstands sind, sie implizieren also auch eine gewisse Verbindlichkeit. Diese Verbindlichkeit drückt sich meist auch in der unterschiedlichen Verwendung der Begriffe „soll“, „sollte“, „könnte“ oder „würde“ aus („muss“ wird nicht verwandt). Die verbindlichste Art der Formulierung ist die kausale Ableitung aus einem Ergebnis, wie z.B. „Die Einrichtung hat eine strukturelle Unterdeckung der Finanzen im Bereich der apparativen Ausstattung. Es wird daher empfohlen, diese zu verschlanken und kostengünstiger zu betreiben“. Letztlich jedoch hängt die Verbindlichkeit von der Beziehung ab, die der Auftraggeber der Evaluation zum Evaluationsgegenstand hat, genauer: Welche Einfluss- und Sanktionsmöglichkeiten bei ihm liegen.

Umsetzungsverfahren

Empfehlungen können, zur Erhöhung ihrer Wirksamkeit, auch Angaben dazu enthalten, wie sie umzusetzen sind. Das *Umsetzungsverfahren*, also die Art und Weise, wie mit den Empfehlungen umgegangen wird, ist dann selbst Gegenstand der Empfehlung. In diesen Fällen kann darin ein Ratschlag liegen, aber auch ein Kontrollraster, anhand dessen die Umsetzung überprüft werden kann. Das Umsetzungsverfahren benennt meist die Schritte in ihrer notwendigen Abfolge und – als „milestones“ – Ergebnisse, die zu bestimmten Zeitpunkten erreicht werden sollen.

Normatives Ziel

Empfehlungen liegen darüber hinaus in aller Regel Vorstellungen darüber zu Grunde, welches die erforderliche „Qualität“ des Evaluationsobjekts sein sollte – wie etwa ein Lehr-/Lernprozess beschaffen sein, welche Aufgabe und Identität eine Einrichtung haben, was ein System leisten muss. Es ist selten, dass diese normative Vorstellung in den Empfehlungen explizit formuliert oder ihnen erklärend vorangestellt wird. Sie muss sich auch nicht notwendig mit denjenigen Vorstellungen decken, die das Evaluationsobjekt vertritt. Sind solche unterschiedlichen normativen Setzungen nicht während des Evaluationsprozesses formuliert und diskutiert worden, kann hier ein wesentliches Problem im Nachgang zur Evaluation entstehen.

In komplexeren Evaluationsprozessen sind die genannten Aspekte der immanenten Prüfung der Empfehlungen eingebettet in ein übergeordnetes Verfahren, in dem Ergebnisse und Empfehlungen noch von Dritten (Gremien oder Personen) geprüft, beraten und korrigiert werden (können). Bei Lehr-/Lernprozess-Evaluationen gehen die Ergebnisse oft in die Beratungen der Lehrkräfte, bei Einrichtungsevaluationen befasst sich etwa ein Beirat oder das Aufsichtsgremium damit, bei Systemevaluationen beschäftigen sich Ministerien und Parlamente mit den Konsequenzen der Evaluation. Die Frage des Sinnes der Empfehlungen und ihrer Umsetzbarkeit spielt dabei immer eine zentrale Rolle.

Deshalb, aber auch ganz prinzipiell, ist die Frage von entscheidender Bedeutung, an wen sich die Empfehlungen letztlich richten. Im öffentlichen Bereich sind es meist übergeordnete Instanzen, welche die Evaluation in Auftrag gegeben haben. Und an diese richten sich auch die Empfehlungen, selbst wenn sie ein Handeln des Evaluationsgegenstands erfordern. Auftraggeber der Evaluation und Subjekt der Ausführung der Empfehlungen fallen hier selten in eins zusammen. Folgerichtig ergibt sich auch daraus ein Spannungsverhältnis und die Aufgabe, im Nachgang zur Evaluation eine eigene Struktur zu schaffen, in der die Ergebnisse bearbeitet werden können (→ Kap. 7.2).

ZUR REFLEXION

- Welche Schwierigkeiten gibt es beim Finden von Bewertungskriterien, wenn es sich um eine internationale und vergleichende Evaluation handelt?
- Wie lassen sich Evaluationsziele und Bewertungskriterien kohärent entwickeln?
- Sind Bewertungskriterien gegenstandsabhängig zu differenzieren und wenn ja, wie?
- Wie wichtig sind (und welche) Theorien bei der Interpretation der Evaluationsdaten?
- Wie genau und verbindlich sollten Empfehlungen sein und wie sollten sie vermittelt werden?

**Lektüreempfehlungen**

- Stockmann, R./Meyer, W. (2010): Evaluation. Eine Einführung. Opladen
- Kuper, H. (2005): Evaluation im Bildungssystem. Eine Einführung. Stuttgart
- Deutsche Gesellschaft für Evaluation (DeGEval) (Hg.) (2001): Evaluation – Reformmotor oder Reformbremse? Köln

7. Evaluation – Verbessern ihre Ergebnisse das Lernen Erwachsener?

Nach all den Betrachtungen zur Evaluation stellt sich noch einmal die Frage, wozu sie gut ist. In der Erwachsenenbildung heißt diese Frage: Verbessert sie die Erwachsenenbildung oder, noch genauer, verbessert sie das Lernen der Erwachsenen (auch das ist, nebenbei, ein Evaluationsziel, gewissermaßen für eine „Metaevaluation“, eine Evaluation der Evaluation). Ziel aller Bemühungen um Erkenntnis, Entwicklung, Kontrolle und Legitimation sollte vor allem sein, dem Gegenstand und seiner Qualität zu dienen.

Nun gibt es unmittelbare Wirkungen auf die Qualität, die sich leichter feststellen lassen. Wenn etwa nach einer Kursevaluation festgestellt werden kann, dass ein Folgekurs genauer auf die Lernbedürfnisse und -strategien der erwachsenen Teilnehmenden abgestimmt ist, dann ist dies sicher ein positiver Effekt der Evaluation; eine kritische Frage würde hier allenfalls die Relation von Aufwand und Ertrag betreffen – steht die positive Wirkung der Evaluation in einem vertretbaren Verhältnis zum betriebenen Aufwand?

Schwieriger gestaltet sich die Klärung der Frage bei mittelbaren Wirkungen. Systemevaluationen etwa, welche die Passgenauigkeit der Allokation öffentlicher Mittel überprüfen, werden zunächst nicht als positiver Effekt bei den Lernenden und der Gestaltung ihrer Lernsituation „ankommen“. Wichtig ist es hier deshalb, die Folgen der Evaluation zu betrachten und zu analysieren, zu welchen Ergebnissen die Verhandlungen und Nacharbeiten im Anschluss an die Evaluation kommen.

In diesem Kapitel geht es darum, wie die Wirkungen der Evaluationen einzuschätzen sind und welche Verfahren erforderlich und geeignet sind, einen nachweislichen Nutzen für die Erwachsenenbildung und die Lernenden herzustellen. Dabei ist zunächst wichtig, zu betrachten, welche Qualität die Ergebnispräsentation hat und wie transparent die Ergebnisse für alle Betroffenen sind (→ Kap. 7.1). Sodann ist zu klären, wie und von wem über die Evaluationsergebnisse verhandelt wird, wie sie letztlich verwertet werden. Auch geht es darum, in welcher Weise Interessen der Auftraggeber und wissenschaftliche Interessen möglicherweise konfliktieren (→ Kap. 7.2). Und schließlich wird dargestellt, wo die Gefahren für eine sinnvolle und nützliche Verwertung der Ergebnisse lauern, worin aber auch der Erfolg und die Wirksamkeit von Evaluationen besteht (→ Kap. 7.3).

Nach der Bearbeitung dieses Kapitels sollten Sie:

- im Umgang mit Evaluationsergebnissen sensibel sein und ausreichend Vorsicht walten lassen können,
- die Notwendigkeit einer strukturierten Nacharbeit zur Evaluation kennen und erläutern können,
- bestehende Interessenstrukturen kennen und beschreiben können sowie
- den Umgang mit bisherigen Evaluationsergebnissen verstehen und einordnen können.

7.1 Zugänglichkeit der Ergebnisse

Ergebnisse von Evaluationen werden gewöhnlich mit einem Bericht zugänglich gemacht. Dieser Evaluationsbericht weist in der Regel den Auftraggeber, das Ziel, das Verfahren und die Ergebnisse aus, häufig ergänzt um Empfehlungen. Auch finden sich zunehmend Zusammenfassungen der Berichte, welche die Ergebnisse und Empfehlungen auf den Punkt bringen (sogenannte „Management Summary“) und Diskussion und Umsetzung erleichtern sollen, vor allem dann, wenn es sich um eine aufwendige Evaluation gehandelt hat.

Aufbau Evaluationsbericht

- A) Titelblatt
Titel der Evaluation (incl. klare Benennung des Evaluationsgegenstands), Verfasser, Institution, Datum der Fertigstellung, Auftraggebende, Finanziers, Erscheinungsort und -datum
- B) Management Summary
Benennung des Evaluationsgegenstands, Zweck, Ziel, Fragen, Methoden (Überblick), Hauptergebnisse, Schlussfolgerungen, Empfehlungen
- C) Inhaltsverzeichnis
mindestens zwei, maximal vier Gliederungsebenen, Überschriften, Seitenzahlen, Verweis auf den Anhang
- D) Einführung/Ausgangslage
Zweck, Fragen, Kontext (Anlass!), Forschungsstand, Evaluationsgegenstand detailliert beschrieben, Übersicht über Aufbau und Roten Faden des Berichts
- E) Methodik
Zweck und Fragen expliziert, differenziert und operationalisiert, Ansatz genannt und begründet, Ablauf und Prozess beschrieben, Erhebungsdesign, Methoden und Instrumente, Quellen, Probleme und Einschränkungen
- F) Ergebnisse
Ergebnisse, Diagramme, Tabellen und Graphen, Bezug zu den Fragen, kohärente Struktur, kausaler Aufbau, deskriptiver Duktus, Quellennachweise
- G) Bewertung
Ergebnisse werden interpretiert, Zusammenfassungen, Bewertungen, Empfehlungen für eine Verbesserung in Bezug auf die Ziele
- H) Anhang
Instrumente, Tabellenverzeichnis, Glossar, Akronymenverzeichnis, Dank an Sponsoren, Informanten, Mitwirkende (falls nicht im Vorwort), Literatur

Evaluationen haben immer einen Gegenstand, die Evaluationsergebnisse und -empfehlungen, also der Bericht, auch immer einen *Adressaten*. Meist ist der Adressat der Evaluation auch deren Auftraggeber, das muss aber nicht so sein. Manchmal ist der Adressat auch der Gegenstand selbst, manchmal auch ein „Dritter“, ein Gremium oder eine Entscheidungsinstanz. Neben der Frage des Adressaten ist diejenige nach der *Transparenz*, also der Zugänglichkeit des Berichts für einen breiteren Interessentenkreis, von Bedeutung. Schließlich ist die *Präsentation* der Evaluationsergebnisse von einer eigenen Qualität.

Adressaten

Bei Evaluationen von Lehr-/Lernprozessen sind Auftraggeber und Gegenstand selten identisch, oft sind der Auftraggeber und die Einrichtung, an welcher sie stattfinden, identisch. Die Einrichtung ist auch der hauptsächliche Adressat der Ergebnisse. Weitere Adressaten sind in der Regel die Lehrpersonen der evaluierten Prozesse und nicht selten auch die Lernenden; auch ein Geldgeber oder ein Aufsichtsgremium kann hier Adressat sein.

Bei Evaluationen von Einrichtungen kann die Einrichtung selbst den Auftrag erteilt haben, häufiger jedoch sind es Geldgeber oder übergeordnete Instanzen. Adressat der Ergebnisse sind hier die Auftraggeber, Adressat der Empfehlungen naheliegender die Einrichtungen.

Bei Evaluationen von Projekten und Programmen unterscheiden sich die Adressaten meist danach, ob es um Ergebnisse der formativen oder der summativen Evaluation geht. Die formativen Evaluationsergebnisse richten sich direkt an die Projektakteure und dienen der Verbesserung des Verlaufs. Die summativen Ergebnisse richten sich an die Auftraggeber der Projekte und dienen hauptsächlich der Legitimation und Kontrolle. Im letzteren Falle gibt es nur selten Empfehlungen.

Bei Systemevaluationen sind die Ergebnisse vor allem für den öffentlichen Auftraggeber interessant, an den sie sich auch richten. In der Regel sind sie jedoch Grundlage einer erweiterten Diskussion aller Akteure im System, so dass hier von einem breiten Adressatenkreis sowohl der Ergebnisse als auch der Empfehlungen gesprochen werden kann.

Transparenz

Der wichtigste Unterschied in Bezug auf die Transparenz liegt nicht in der Diversität der Evaluationsgegenstände, sondern in der öffentlichen Funktion des Auftrags bzw. Auftraggebers. Man kann sagen, dass Evaluationen, die von öffentlichen Institutionen in Auftrag gegeben wurden, durchweg zu öffentlichen oder besser: veröffentlichten Ergebnissen führen, während das bei privat initiierten Evaluationen die Ausnahme ist. Das entspricht im Regelfall auch dem „öffentlichen Interesse“, das sich einerseits auf Dinge des Allgemeinwesens, andererseits auf gemeinsam finanzierte Aktivitäten richtet.

Die Transparenz hat jedoch nicht nur eine mediale Seite, die Verfügbarkeit des Berichts als Printprodukt oder im Internet, sondern auch eine inhaltliche Seite. Nicht immer sind alle Ergebnisse der Evaluation in solchen veröffentlichten Evaluationsberichten enthalten, auch nicht immer alle wichtigen Ergebnisse. Auch wurden nicht immer alle möglichen Zugangsbarrieren wie Angabe des Ortes (auch virtuell) und des Zeitpunkts der Veröffentlichung vermieden.

Der Regelfall der Transparenz ist allerdings eine offensive Publizierung, die das Augenmerk auf den Gegenstand lenkt und die öffentliche Förderung legitimiert. Evaluationsberichte haben immer auch eine Funktion der Public Relation, der öffentlichen Vermittlung und Werbung. „Tue Gutes und sprich darüber“ – in der Würdigung öffentlicher, hier vor allem bildungspolitischer Aktivitäten, ist dieses Sprichwort richtig. Dies gilt natürlich vor allem dann, wenn eine Evaluation hauptsächlich die Stärken des Evaluationsgegenstands positiv betont.

Präsentation

Die Veröffentlichung eines Textes ist nur eine mögliche Form, Inhalte zu verbreiten. Sie ist jedoch nicht immer die wirkungsvollste. In der Regel sieht eine Evaluation daher nicht nur die Abfassung und Veröffentlichung der Ergebnisse in Textform vor, sondern auch eine Präsentation in anderen Formen und Kontexten.

Die gängigste weitere Form dabei ist die „Übergabe“ des Evaluationsberichts an den Auftraggeber; ein Auftrag, der erteilt wurde, wird formal und symbolisch abgeschlossen. Dabei wird meist eine Zuspitzung der Ergebnisse erwartet, das Wichtigste, das Überraschendste, das Dringlichste. Da solche „Übergaben“ bei öffentlichen Auftraggebern meist medial und öffentlich sind, entsteht so eine zusätzliche Dimension der Evaluation, die Priorisierung. Sie kann im weiteren Verlauf der Diskussion durchaus eine eigene Dynamik entfalten.

Immer dann, wenn die Evaluation ein politisch brisantes oder für eine breite Öffentlichkeit interessantes Objekt zum Gegenstand hatte (das ist vor allem bei System- und Programmevaluationen der Fall) oder wenn die Ergebnisse in unterschiedlichen Kontexten zu thematisieren sind (wie bei Einrichtungsevaluationen), ergibt sich eine Reihe von „Follow-Up-Präsentationen“, die auch zeitlich eigene Dimensionen (bis zu einigen Jahren) umfassen können.

Solche Follow-Up-Präsentationen sind entweder geplant oder ungeplant. Als geplante Präsentationen sind sie Teil des Evaluationsauftrags, wobei die Finanzierung dieser Folgeaktivitäten im Bildungsbereich eher schmal ist und viel auf das Eigeninteresse der Evaluatoren gesetzt wird, ihre Ergebnisse zu verbreiten und zu „vermarkten“.

Geplante Präsentationen stehen meist im Kontext des Evaluationsauftrags. Form, Inhalt und Umfang hängen davon und der Beschaffenheit des Evaluationsgegenstands

ab, weniger vom Inhalt und Gehalt der Evaluationsergebnisse oder der Qualität der Empfehlungen. Meist werden solche Präsentationen (und Diskussionen) schon vor oder während der Evaluation angekündigt, um die Motivation zur Teilnahme zu gewinnen oder zu erhalten.

Präsentationen bei Evaluationen von Lehr-/Lernprozessen finden üblicherweise in den entsprechenden Einrichtungen und vor den betroffenen Lehrkräften statt, die Lernenden werden nicht selten hinzugebeten. Zumeist handelt es sich um eine einmalige Veranstaltung, die diskursiv angelegt ist. In ihr wird auch die Konsequenz thematisiert, die sich aus dem Evaluationsbericht ergibt.

Präsentationen bei Evaluationen von Einrichtungen finden meist in differenzierten Kontexten statt, oft auch getrennt. So erfolgen sie bei den Auftraggebern praktisch immer auch gegenüber Vertretern der Einrichtung. Häufig auch ergänzend vor Gremien (wie Vorstand, Aufsichts-/Verwaltungsrat und Beirat) und vor der Belegschaft. Die Diskussion über das Follow-Up ist hier weniger an die Präsentation gekoppelt als vielmehr in der Folge zwischen den unterschiedlichen Adressaten.

Präsentationen von Systemen erfolgen fast immer in größerem Rahmen und vor größeren Gremien, Landtagen, Beiräten, Kuratorien und Ausschüssen. Vielfach fragen auch die betroffenen Organisationen der Weiterbildung – manchmal gemeinsam, oft getrennt – nach einer Präsentation vor ihren Vertretern.

Präsentationen von Projekt- und Programmevaluationen werden in der Regel in die Abschlussveranstaltungen dieser befristeten Aktivitäten einbezogen, auch wenn sie noch – weil das Projekt noch gar nicht abgeschlossen ist – selbst nicht abgeschlossen sind. Im Regelfall liegen sie dann in der Endfassung dem Endbericht des Programms bei und werden mit diesem publiziert und rezipiert.

Diejenigen Präsentationen, die im Evaluationsdesign enthalten sind, haben jeweils eine spezifische Funktion bei der Umsetzung der Ergebnisse. So gesehen sind sie genuiner Teil der Wirkung. In der Regel folgen sie – adressatenspezifisch – differenzierten Regeln methodisch-didaktischer Vermittlung und erschöpfen sich nicht im Referieren der Berichtstexte. Schriftmedien (Poster etc.), Workshops, Broschüren, Internetauftritte, Arbeitsgespräche, Szenariotechniken und sogar Rollenspiele und Zukunftswerkstätten werden zur Dissemination von Evaluationsergebnissen genutzt.

Gerade im Bildungsbereich, in dem Evaluationen einen großen Teil der Forschung ausmachen, wird auch gegenüber der Scientific Community präsentiert. Hier stellen sich völlig andere Anforderungen an die Präsentation, es geht um die Wissenschaftlichkeit der Evaluation und ihrer Ergebnisse, weniger um die Bedeutung und Umsetzbarkeit derselben und der Empfehlungen. Solche Präsentationen erfolgen fast ausschließlich im Eigeninteresse der Evaluatoren.

Bei größeren und politisch wichtigeren Evaluationen, vor allem dann, wenn die Ergebnisse bedeutsam waren, bestehen oft noch Jahre nach dem Abschluss der Evalua-

tion Nachfragen zu einer Präsentation. Dies sind ungeplante Präsentationen. Sie stellen dann die Evaluatoren, häufig ja nur befristet in der Evaluation beschäftigt, vor die Schwierigkeit – längst in anderen Arbeitszusammenhängen befindlich eine zusätzliche Arbeit zu leisten.

In besonderen Fällen gewinnen Evaluationsergebnisse im Nachhinein eine eigene Dynamik, die PISA – Diskussion in Deutschland ist dafür ein Beispiel. Das Zusammenwirken von Überraschung, Betroffenheit, Problembewusstsein und Verantwortung führte zu einer lang dauernden und nachhaltig wirkenden Diskussion, die sich in ihren Konsequenzen immer mehr von den eigentlichen Evaluationsergebnissen und deren Reichweite entfernte.

7.2 Interessen und Verfahren

Ohne Interessen an einer Beurteilung und Bewertung gibt es keine Evaluation, und diese Interessen werden jeweils im Kontext einer Evaluation von Akteuren vertreten; man nennt diese auch oft „Stakeholder“, also wichtige Repräsentanten einer Position, Institution, Organisation oder eines Interesses. Evaluation ist, wie wir gesehen haben, immer anwendungsorientiert und findet immer in einem (gesellschaftlichen) Feld statt. Dies bedeutet auch immer: die Akteurs- und Interessenstruktur ist komplex.

Um das Entstehen einer Evaluation (aus gegebenem oder geschaffenen Anlass) und das Umgehen mit ihren Ergebnissen zu verstehen, muss man zunächst eine „Landkarte“ der Akteure oder „Stakeholder“ zeichnen. Im (Weiter-)Bildungsbereich lassen sich hier folgende typischen Akteure (oder Akteursgruppen) erkennen (vgl. Nuissl 2007):

- *Lernende*

In der organisierten Weiterbildung werden sie oft auch die „Teilnehmenden“ genannt, es sind die erwachsenen Lerner und Lernerinnen. Für sie besteht der ganze Bereich, ihr gelingendes Lernen sollte das Ziel allen Bemühens in der Weiterbildung sein. Allerdings: Sie stellen nicht einmal die Hälfte der erwachsenen Bevölkerung in Deutschland, die andere Hälfte nimmt nicht teil. Es müsste also, von einem politisch normativen Ziel her betrachtet, nicht von den Lernenden, sondern schlicht von den Erwachsenen gesprochen werden, deren Lernen unterstützt werden soll. Die Lernenden werden in Deutschland immer gebildeter, stellen immer höhere Ansprüche an die Qualität von Bildungsmaßnahmen. Dies liegt auch daran, dass sie einen immer höheren Anteil an den Angeboten finanzieren. Die Lernenden haben einen immer höheren Bedarf an Information, Beratung und Qualitätskontrolle. Und ein immer präziseres Bewusstsein davon, was „gute“ Weiterbildung ist.

- *Lehrende*

Es handelt sich um viele Teilgruppen mit teilweise ganz unterschiedlichen Interessen und Kompetenzen. Einem Kern von zehn Prozent Hauptberuflichen stehen die neunzig Prozent der nebenberuflich Tätigen, hauptsächlich Lehrende, gegenüber. Die Beschäftigten in der Weiterbildung sind engagiert, wollen aber auch eine Verbesserung der Arbeitsbedingungen, des Ansehens und der beruflichen Entwicklungsmöglichkeiten. Oft verstehen sie sich gar nicht als Lehrende in der Erwachsenenbildung, sondern als Manager, Coach, Trainer, Museumsführer oder Dozent. Ihre Motive, sich in der Weiterbildung zu betätigen, sind außerordentlich heterogen, mit vielen Konsequenzen für ihre Fortbildung, ihr Lehrverhalten und ihren Bezug zu den Lernenden.
- *Politiker*

Anders als die Lernenden und die Lehrenden sind die Politiker relativ wenig erforscht. Dies liegt nahe, denn sie lassen eher forschen (und evaluieren) als dass sie selbst als Evaluationsgegenstände herhalten. Bei den Politikern muss man unterscheiden zwischen den „richtigen“ Politikern, also den gewählten Parlamentariern, und den Administratoren, die sich wiederum in „Politische Beamte“ und „normale“ Beamte untergliedern lassen. Die „richtigen“ Politiker müssen in kurzen Abständen (alle fünf Jahre) Erfolge nachweisen; die administrativ agierenden Politiker versuchen, eine bestimmte Linie über einen längeren Zeitraum hin umzusetzen. Im Bildungsbereich haben die Politikergruppen meist unattraktiv wenig Geld auszugeben, was umso schwerer wiegt, als Bildungsausgaben insgesamt in Deutschland sehr zersplittert sind – zwischen den Ressorts auf allen Ebenen von Bund und Ländern sowie zwischen Bund und Ländern (mit den jeweils unterschiedlichen Kompetenzen).
- *Wissenschaftler*

Eine marginale Größe, was sie als Adressaten von Evaluationsergebnissen angeht, aber eine respektable Größe, was ihren Wettbewerb um Evaluationsmittel betrifft.
- *Öffentlichkeit*

Ein sehr diffuser Stakeholder, der sich aus unterschiedlichen Stakeholdern zusammensetzt. Im Bildungsbereich sind dies eigentlich alle Menschen, denn sie sind alle von Bildung meist sehr direkt tangiert, auch wenn sie unterschiedlich aktiv agieren. Und natürlich die Medien als die Stimme – gelegentlich – der öffentlichen Meinung, die sogenannte veröffentlichte Meinung.

Die große Linie der Qualifizierung von Akteursgruppen ist natürlich wenig erklärend, wenn es um das konkrete Evaluationsvorhaben geht. Hier verschieben und gewichten sich die Dinge meist ganz anders – die geldgebenden Politiker sind die wichtigste, die Lernenden die am wenigsten einflussreiche Gruppe. Demokratietheoretisch ist das

legitimiert durch die Wahl der Politiker als Repräsentanten der Menschen in der Gemeinschaft, aber im konkreten Aushandeln einer Evaluation ist das nicht immer zu beobachten.

In der konkreten Auseinandersetzung um eine Evaluation (beim Auftrag, bei der Behandlung der Ergebnisse) spielen die konkreten Akteure eine Rolle. Das sind (mit den jeweils antizipierbaren Interessen):

- *Auftraggeber*
Er will eine vorzeigbare Evaluation (Qualität) mit vorzeigbaren Ergebnissen, sofern er für den Evaluationsgegenstand zuständig und verantwortlich ist. Vorzeigbar entweder zur Legitimation („gute Sache“) oder zur Rechtfertigung eigener Maßnahmen (Kürzungen, Änderungen etc.).
- *Untersuchungsgegenstand*
Er will eine faire Evaluation, die seine Bedingungen und Ziele respektiert, bestätigende Ergebnisse oder solche, die bei der Weiterentwicklung helfen, und eine Evaluationsprozedur, die seine Regelarbeit nicht über Gebühr belastet.
- *Evaluatoren*
Sie wollen eine qualitativ gute Evaluation, vorzeigbar auch in der Wissenschaft, eine nicht extern (z.B. durch den Auftraggeber) gesteuerte Arbeit, vorzeigbare Ergebnisse und nach Möglichkeit eigenes Lernen und verbesserte Reputation.
- *Öffentlichkeit*
Sie will Ergebnisse, die verständlich sind und die Situation beim Lernen Erwachsener dadurch verbessern, dass sie auch umgesetzt werden.

Es bedarf einer geregelten Transparenz (→Kap. 7.1), um überhaupt in der Lage zu sein, ein Verfahren zu finden, nach dem diese unterschiedlichen Interessen balanciert berücksichtigt werden können. Evaluationsergebnisse sind – dafür sind Evaluationen da – niemals nur positiv bestätigend, sondern zeigen immer auch Probleme und notwendige Entwicklungen auf. Gerade an diesen entzünden sich nicht selten die unterschiedlichen Interessen.

Im Verfahren sind daher – zumindest bei öffentlich beauftragten und geförderten Evaluationen – gestufte Bearbeitungsformen vorgesehen oder werden praktiziert. Meist sind diese an die Präsentationsformen gekoppelt (→Kap. 7.1), oft auch ohne direkte Präsentationen der Evaluatoren Gegenstand politischer Aushandlungen in anderen Kontexten. Dabei werden die Interessen jeweils deutlich benannt, auch wenn das nicht immer öffentlich geschieht. Im Zweifelsfall, bei nicht erfolgter Einigung, droht den Evaluationsberichten das, was schon mit vielen Evaluationsberichten geschah: Sie werden (im Schwyzerdütsch) „schubladisiert“.

7.3 Nutzung und Wirkung

Von bedeutendem Interesse ist es natürlich, besonders wenn man die weite Verbreitung von Evaluationen und die hohen Beträge bedenkt, die in sie investiert werden, ob sie genutzt werden und wie sie eigentlich wirken oder besser: Was sie bewirken. Es geht also um die Evaluation der Evaluationen, die „Meta-Evaluation“: Werden mit den in Evaluationen investierten Mitteln die gewünschten Ziele erreicht?

Zunächst geht es um die Nutzung, die ja einen wesentlichen Teil der Wirkung ausmacht. Mit Nutzung ist hier nicht die (bereits behandelte) Prozessnutzung (→ Kap. 4.4) gemeint, also die Nützlichkeit der Evaluation für den Evaluationsgegenstand, die etwa in Fokussierung, Selbstvergewisserung und Umweltanalyse liegt, sondern die Nutzung der Ergebnisse. Und auch hier ist zu differenzieren zwischen Nutzung, Dissemination und Rezeption. Evaluationsergebnisse müssen disseminiert werden, um transparent und nutzbar zu sein (zu Verfahren dazu → Kap. 7.1). Die Verbreitung der Ergebnisse ist aber noch kein Wert an sich, sondern die Voraussetzung für eine Nutzung. Eine bloße Rezeption, also zur Kenntnisnahme der Ergebnisse, ist ebenfalls noch keine Nutzung. Eine Nutzung beginnt erst dann, wenn das Evaluationsergebnis praktisch umgesetzt wird.

Beywl und Balzer unterscheiden hier zwischen einer konzeptionellen und einer instrumentellen Nutzung (o.J., S. 119). Eine konzeptionelle Nutzung liegt dann vor, wenn die Ergebnisse in das allgemeine Fachwissen einfließen, in Büchern, Artikeln etc. diskutiert und reflektiert werden, ohne dass konkrete Verbesserungen, Veränderungen oder Entscheidungen getroffen werden. Eine instrumentelle Nutzung liegt dann vor, wenn sich in der Realität etwas auf der Grundlage der Evaluationsergebnisse (oder deren Empfehlungen) verändert. Dies kann unterschiedlich weitreichend sein. So verzichtet etwa ein Lehrender nach einer Evaluation seiner Tätigkeit auf eine bestimmte Interaktionsstrategie, eine Einrichtung beginnt einen intensiven Prozess der Organisationsentwicklung oder eine Regierung erlässt ein neues Gesetz oder novelliert ein bestehendes.

Der empirische Kenntnisstand über die Nutzung von Evaluationsergebnissen ist gering, es gibt nur wenig Forschung im Bereich der Meta-Evaluation (wer sollte das auch mit welchem Interesse fördern?). Immerhin weiß man aus kleineren Studien und Nachfolgeevaluationen, welche Faktoren die Nutzung von Evaluationsergebnissen fördern:

- Die potentiellen Nutzer von Evaluationsergebnissen müssen von Anfang an in die Evaluation einbezogen werden, Sinn und Ziel mit diskutieren und beeinflussen können (was für viele Evaluationsforscher eine relativ ungewohnte Herangehensweise ist).
- Die Beteiligung der potentiellen Nutzer an der Evaluation darf nicht bedeuten, dass sie die Evaluationsarbeit machen müssen, das wird schnell als Ausbeutung verstanden.

- Die Ergebnisse müssen frühzeitig zur Verfügung stehen und bearbeitbar sein, die Empfehlungen müssen so operationalisiert sein, dass sie auch umsetzbar sind.
- Der Evaluator muss mit allen Hierarchiestufen und Sektionen des Evaluationsgegenstands gleichberechtigt und vertrauensvoll kommunizieren.
- Die Evaluation muss transparent und offen stattfinden.
- Die Ergebnisse sollten eher im Rahmen des Erwarteten liegen und ausgewogen sein zwischen positiver Bestärkung und kritischer Reflexion.
- Die Ergebnisse sollten mit den Erfahrungen und Werthaltungen der Vertreter des Evaluationsgegenstands harmonieren.

Betrachtet man den Aspekt der *Wirkungen* von Evaluationen, spielt die Nutzung der Ergebnisse natürlich die zentrale Rolle. Evaluationen, deren Ergebnisse weder konzeptionell noch instrumentell genutzt werden, bleiben in der Regel wirkungslos. Allerdings geht die Wirkung vielfach über die reine Nutzung der Ergebnisse hinaus. Vielfach kann man gar nicht genau feststellen, ob eine bestimmte Entwicklung noch eine Wirkung der Evaluation und ihrer Empfehlungen ist oder nicht; in der Wirkungsforschung sind die investigativen Instrumente – gerade im Bildungsbereiche – limitiert.

Bei der Wirkung („impact“) von Evaluationen sind folgende vier Ebenen zu berücksichtigen, wobei dort sowohl positive (intendiert und nicht intendiert) als auch negative (nicht intendiert) Wirkungen auftreten können:

- *Evaluationsgegenstand*
Schon im Prozess der Evaluation entstanden dort Effekte, die nicht explizit in der Evaluation vorgesehen waren, etwa das soziale Verhalten von Lehrenden und Lernenden oder ihre Beziehung zueinander, das Aufdecken von Störungen in einer Einrichtung, das Bewusstsein um die Stärken eines Systems. Im Nachgang können sich das Image verändern, der Bekanntheitsgrad, die Marktgängigkeit, das Selbstbewusstsein. Es kann aber auch Demotivation eintreten, Unklarheit und Unsicherheit.
- *Auftraggeber*
Es ist möglich, dass sie den Evaluationsgegenstand näher kennengelernt und schätzen gelernt haben, seine Bedeutung besser abschätzen können und ihn verstärkt in den Blick und in die Pflicht nehmen – es kann aber auch das Gegenteil eintreten.
- *Evaluatoren*
Mit Sicherheit haben sie einen Gegenstand genauer kennengelernt (gewissermaßen seine „Innereien“), zu dem sie sonst keinen Zugang gehabt hätten. Sie haben sich weiterqualifiziert, ihr Profil geschärft, eine gegenstandsbezogene Kompetenz aufgebaut und – in der Regel – eine feldbezogene Sensibilität entwickelt. Vielleicht sind sie aber auch „verbrannt“, festgelegt auf ein Image, im Negativen abgestempelt.

- *Öffentlichkeit*

Vielleicht ist ihr durch die Publikation des Evaluationsberichts die Bedeutung einer Sache klarer geworden, ein Problem bearbeitbarer, eine Gefahr wichtiger. Vielleicht beklagt sie den großen Aufwand für ein zu erwartendes Ergebnis, die schlechten Ergebnisse, die unklaren Perspektiven.

Für alle diese Impact-Faktoren lassen sich Beispiele finden – selten empirisch untersucht, aber im Alltagswissen verankert. Was sich zeigt: In der deutschen Evaluationsforschung ist noch einiges zu tun.

Will man ein Evaluationsprojekt insgesamt, also nicht nur von der Wirkung her, „evaluieren“, empfiehlt es sich, auf die differenzierten Kriterien der DeGEval zurückzugreifen. Mit den Kategorien der Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit hat die Gesellschaft 25 Standards (=Kriterien) formuliert, die eine sachgerechte Beurteilung jeder Evaluation ermöglichen (→ Kap. 2.2).

Schließlich bleibt, die Frage danach zu beantworten, ob sich das Lernen Erwachsener nun durch Evaluationen verbessert oder nicht. Vielleicht nähert man sich der Antwort am besten, indem man überprüft, ob Evaluationen schaden. Evaluationen kosten in der Regel (viel) Geld und zehren von den Ressourcen, die ansonsten in die Sache investiert worden wären – etwa Mittel für Lehre oder Verbesserung der Infrastruktur. Evaluationen belasten den Alltag, das Dauergeschäft, teilweise nicht unerheblich, haben – als Feldaktivität – etwa großen Einfluss auf Lehr-/Lernprozesse und die Arbeitsabläufe in Organisationen. Evaluationen binden viel Energie, die sonst in andere Tätigkeiten geflossen wäre – etwa in die Konzipierung eines Seminars, die Entwicklung eines strategischen Managements oder die eigene Fortbildung. Evaluationen können – im „worst case“ – die eigene Entwicklung in eine Richtung drängen, die weder gewünscht noch wünschenswert ist.

Andererseits ist der Nutzen von Evaluationen unübersehbar. In Lehr-/Lernprozessen haben Evaluationen auch traditionellen Typs (Hospitationen, „Sitting in the classroom“) immer zu Einsichten in pädagogische Prozesse und zur Verbesserung der pädagogischen Aktivitäten geführt, was auch bei den heutigen, methodisch weiter entwickelten Evaluationen der Fall ist.

Bei Einrichtungen sind Evaluationen unabdingbar im Rahmen von Verfahren des Qualitätsmanagements und immer nützlich zur Standortbestimmung und perspektivischen Weiterentwicklung. Gerade regelmäßige Gesamtevaluationen sind hier zukunftsweisend.

Bei Projekten und Programmen sind Evaluationen praktisch unverzichtbar, wenn es darum geht, die Sinnhaftigkeit und Zielerreichung zu überprüfen und zu dokumentieren. Gerade die Befristung von Projekten und Programmen (etwa bei der Implementation von Modellen) erschließt diese für weitergehende Erkenntnisse erst über Evaluationsberichte.

In Systemen machen Evaluationen auf Aspekte aufmerksam, die sonst nicht beachtet würden, für das zielgerichtete Funktionieren der Systeme aber wesentlich sind. Sie dienen auch dazu, die Konsequenzen des steten Wandlungsprozesses zu reflektieren und zu überprüfen.

Zusammenfassend kann gesagt werden: Evaluationen verbessern die Weiterbildung, wenn sie gewissenhaft und korrekt ausgeführt und mit den Betroffenen intensiv kommuniziert werden. Und dadurch und dann verbessern sie auch das Lernen Erwachsener.

ZUR REFLEXION

- In welchen Zusammenhang wurden Nutzung und Interessen bei Zielstellung und Realisierung der Evaluation gebracht – und welche Kommunikation erfolgte in diesem Kontext?
- Welche Gründe kann es geben, Ergebnisse und Empfehlungen von Evaluationen unberücksichtigt zu lassen?
- Welche Situationen können es möglich machen, dass das Befolgen von Evaluationsergebnissen und -empfehlungen für den Evaluationsgegenstand schädlich ist?
- Was in der Erwachsenenbildung legt überhaupt eine sinnvolle Evaluation nahe?



Lektüreprüfungen

- Tietgens, H. u.a. (1986): Aufgaben und Probleme der Evaluation in der Erwachsenenbildung. Bonn
- Wesseler, M. (2010): Evaluation und Evaluationsforschung. In: Tippelt, R./Hippel, A. v. (Hg.): Handbuch Erwachsenenbildung/Weiterbildung. 4. Aufl. Wiesbaden, S. 1031–1048

Glossar

Akkreditierung

Anerkennung einer Einrichtung oder eines Curriculums durch eine beauftragte Stelle.

Assessment

Überprüfung eines Standes oder Zustandes (meist von Kompetenzen) anhand eines multiplen Verfahrens.

Audit

Form einer „kleinen“ Evaluation, wörtlich: Anhörung, meist in Peer-Review-Verfahren verwandt.

Auftrag

Vergabe einer Evaluation an Personen oder Institutionen durch die fördernde und an den Evaluationsergebnissen interessierte Instanz; meist „Drittmittelauftrag“. Im Auftrag werden (mindestens) Ziel, Verfahren und Zuständigkeiten der Evaluation festgelegt.

Beschreibung

Grundlage in einem Evaluationsverfahren, erfasst den Evaluationsgegenstand anhand nachprüfbarer Daten noch ohne Bewertung. Die Beschreibung erfolgt anhand von Merkmalen und Indikatoren.

Daten

Informationen, die belastbar, reliabel und valide sind für weitergehende empirische Analysen.

Bewertung

Wertendes Urteil über den Evaluationsgegenstand auf der Grundlage der Beschreibung und anhand der vorliegende Daten und Informationen. Die Bewertung folgt explizierten und begründeten Kriterien auf der Grundlage von Zielsetzungen, Interessen, Normen oder Bedarfen.

Benchmark

Messdatum, das eine Zielleistungsgröße angibt (z.B. Teilnahmequote an Weiterbildung) und damit auch einen Vergleich ermöglicht.

Daten-Adäquanz

Bezeichnet das angemessene inhaltliche und mengenmäßige Verhältnis der in die Evaluation einbezogenen Daten für die Beantwortung der Evaluationsfragen.

Effizienz

Angemessener und sparsamer Umgang mit vorhandenen Ressourcen („Tue ich das Richtige richtig?“). Eines der wichtigsten Bewertungskriterien in allen Evaluationen.

Effektivität

Zielgenauer Umgang mit vorhandenen Ressourcen („Tue ich das Richtige?“). Meist das wichtigste Bewertungskriterium in Evaluationen.

Evaluation

Wertgutachten, das sich auf eine mit wissenschaftlichen Methoden erstellte Beschreibung stützt und hinsichtlich der angewandten Kriterien begründet ist. Im Grunde ist Evaluation eine wissenschaftliche Dienstleistung.

Evaluationsgegenstand

Dasjenige, was evaluiert wird, etwa ein Lehr-/Lernprozess, eine Einrichtung, ein Programm oder ein System. Hier auch synonym „Evaluationsobjekt“ genannt.

Externe Evaluation

Form der Evaluation, die von Personen oder Institutionen durchgeführt wird, die keinen Bezug zum evaluierten Objekt oder „Gegenstand“ hat, oft als „Peer-Review-Evaluationen“ durchgeführt.

Feed-Back

Wörtlich „Rückmeldung“, bezeichnet unterschiedliche Verfahren der unmittelbaren Rückmeldung insbesondere in Lernprozessen und bezieht sich meist auf Inputs von Lehrenden oder Teilnehmenden, welche einen Input übernommen haben.

Formative Evaluation

Evaluation, die einen Prozess, ein Verfahren, ein Projekt in seiner Entwicklung begleitet und deren Ergebnisse für die Optimierung derselben unmittelbar zurückgespiegelt werden.

Fremdevaluation

Im Wesentlichen synonym zu → „externe Evaluation“.

Indikatoren

„Anzeiger“ für Eigenschaften, die nicht unmittelbar ersichtlich oder messbar sind (z.B. bei „Intelligenz“).

Impact

Auswirkung der erzeugten Produkte im weiteren Kontext, etwa besseres gegenseitiges interkulturelles Verständnis nach Sprachlernprozessen. Impact kann beabsichtigt, aber auch unbeabsichtigt sein. Im letzteren Falle kann der Impact auch gegen die ursprünglichen Ziele gerichtet sein.

Interpretation

Analyse und Auslegung vorliegender oder erhobener Daten auf der Grundlage weitergehender Kenntnisse und Informationen sowie nach einem festgelegten und nachprüfbareren Verfahren.

Interne Evaluation

Form der Evaluation, die von Personen oder Abteilungen durchgeführt wird, die zum Objekt oder „Gegenstand“ der Evaluation gehören.

Qualität

Beschaffenheit eines Gegenstands, an sich noch wertneutral; die Alltagsbedeutung von Qualität = gute Qualität entstand erst in den letzten Jahrzehnten.

Kriterien

Aufmerksamkeits- und Bewertungspunkte, abgeleitet aus einem gemeinsamen Ziel- und Bewertungssystem.

Merkmale

Konkrete Eigenschaften eines Gegenstands, die erfassbar und messbar sind (z.B. Quantitäten).

Monitoring

Fortlaufende Kontrolle eines Arbeits- oder Entwicklungsprozesses.

Output

Direktes Ergebnis eines Produktionsprozesses, im pädagogischen Bereich etwa ein testbares Lernergebnis oder ein konkretes Seminarangebot.

Outcome

Indirektes Ergebnis eines Produktionsprozesses, im pädagogischen Bereich etwa die Motivation, das Gelernte anzuwenden, oder verbesserte Kommunikation zwischen Teilnehmergruppen aufgrund von Kursteilnahmen.

Programmevaluation

Eine sehr häufige und in der Realität meist verpflichtende Evaluation; Programme (oder Projekte), die in einem definierten Zeitraum ein definiertes Ziel verfolgen, werden fast immer evaluiert, meist danach, ob die angestrebten Ziele erreicht wurden.

Peer-Review-Evaluation

Der „Peer-Review“ ist eine weit verbreitete Experten-Evaluation, in der Fachleute den Gegenstand beurteilen; dabei handelt es sich um mehr als einen „Peer“, um persönliche Sichtweisen, Vorlieben etc. zu korrigieren (Intersubjektivität). Im Bildungsbereich ist die Peer-Review-Evaluation vor allem bei Fachzeitschriften und in Evaluationen von Einrichtungen zu finden. Bei Peer-Review-Evaluationen existieren strengste Regeln zur Kontrolle der Intersubjektivität.

Reliabilität

Begriff dafür, ob die erhobenen Daten auch verlässlich sind in ihrer Qualität und Quantität. Neben der Validität ein zentraler Begriff der empirischen Sozialforschung.

Responsive Evaluation

Evaluationsverfahren, bei dem die untersuchten und evaluierten Objekte in den Prozess kommunikativ eingebunden werden und zur Genauigkeit und Akzeptanz des Evaluationsergebnisses beitragen.

Rohdaten

Unbearbeitete, vor allem noch nicht ausgewertete Datenbestände aus Erhebungen oder Datenbanken, die eine ausgewiesene Qualität ausweisen und für Auswertungszwecke adäquat gruppiert werden können.

Sample

Bezeichnet wird damit die Auswahl aus einer Gesamtheit, deren Merkmale repräsentativ für die Gesamtheit sein sollen. Das Erstellen eines S. nennt man „Samplebildung“, die Verfahren reichen vom Zufallsprinzip bis hin zu einer detailliert erstellten und begründeten Einzelauswahl. Je nach Größe der Gesamtheit kann das Sample auch mit ganz niedrigen Anteilen aussagekräftig sein, etwa mit 0,1 Prozent der Gesamtbevölkerung (bei Wahlprognosen).

Selbstevaluation

Im Wesentlichen synonym zu → „interne Evaluation“

Summative Evaluation

Abschließende Evaluation eines Prozesses, Projekts oder Programms, die in der Regel dessen Qualität und Nützlichkeit bewertet; in der Abfolge von mehreren Projekten etc. nacheinander können Verfahren der summativen Evaluation auch den Charakter der formativen Evaluation erhalten.

Stakeholder

Personen (oder auch Institutionen), die für bestimmte Prozesse, Abläufe oder Entscheidungen wesentliche Bedeutung – und eigene Interessen – haben.

Triangulation

Verbindung/Verknüpfung von verschiedenen Methoden, Daten, Erhebungsinstrumenten, Evaluatoren und Ansätzen bei der Erfassung eines Evaluationsgegenstands.

Validierung

Bewertung eines erreichten Standes oder Zustands durch nachgewiesene Verfahren.

Validation

Sicherstellung des nachhaltigen Wertes von Projekten oder Programmen.

Validität

Begriff dafür, ob die erhobenen Daten auch das messen, was gemessen werden soll; so liefert etwa die Analyse, welche Informationen die Lehrperson gibt, keine validen Daten darüber, was die Teilnehmende lernen.

Wissenschaftliche Begleitung

Bei Projekten, Programmen und Modellversuchen werden die Evaluatoren oft auch „wissenschaftliche Begleitung“ genannt, das entspricht einer formativen Evaluation auf wissenschaftlicher Grundlage.

Annotierte Literatur

Deutsche Gesellschaft für Evaluation (DeGEval) (Hg.) (2002): Standards für Evaluation. Köln

Hier handelt es sich um die Grundregeln für Evaluation, wie sie die in Deutschland arbeitenden Evaluatoren kurz nach Gründung der „Deutschen Gesellschaft für Evaluation“ formuliert haben. Ein systematischer Zugang zur Evaluation, dessen positiv formulierte Regeln auch den Umkehrschluss auf Fehler und Probleme bei Evaluationen erlauben.

Kuper, H. (2005): Evaluation im Bildungssystem. Stuttgart

Eine sehr dichte, auf das gesamte Bildungssystem bezogene Darstellung der Evaluation. Besonders interessant ist die Einbindung in theoretische Kontexte und der Bezug zur erziehungswissenschaftlichen Diskussion.

Reischmann, J. (2006): Weiterbildung – Evaluation. Lernerfolge messbar machen. 2. Aufl. Neuwied

Ein Klassiker der Evaluation in der Erwachsenenbildung mit dem Schwerpunkt auf der Evaluation von Lehr-/Lernprozessen. Sehr praxisnah und informativ, mit Beispielen für Fragebogen und andere Auswertungsverfahren.

Stockmann, R./Meyer, W. (2010): Evaluation. Eine Einführung. Opladen

Standardwerk zur Evaluation, verfasst von zwei herausragenden Akteuren der deutschen Evaluationsdiskussion (auch in der DeGEval). Ein genereller Draufblick ohne explizite Differenzierung und Thematisierung erziehungswissenschaftlicher Kontexte.

Widmer, T./Beywl, W./Fabian, C. (Hg.) (2009): Evaluation. Ein systematisches Handbuch. Wiesbaden

Ein sehr lesbares Nachschlagewerk zu allen Fragen der Evaluation, an einigen Stellen auch Fundgrube für erziehungswissenschaftliche Aspekte.

Literatur

- Aeppli, J./Gasser, L./Gutzwiller, E./Tettenborn, A. (2011): Empirisches wissenschaftliches Arbeiten. Ein Studienbuch für die Bildungswissenschaften. Bad Heilbrunn
- Arnold, R. u.a. (2000): Forschungsmemorandum für die Erwachsenen- und Weiterbildung. URL: www.die-frankfurt.de/esprid/dokumente/doc-2000/arnold00_01.pdf (Stand: Mai 2013)
- Backhaus, K./Erichson, B./Plinke, W./Weiber, R. (2008): Multivariate Analysemethoden. 12. Aufl. Berlin
- Beywl, W. (1988): Zur Weiterentwicklung der Evaluationsmethodologie. Frankfurt a.M. u.a.
- Beywl, W./Widmer, T. (2000): Handbuch der Evaluationsstandards. Opladen
- Beywl, W./Balzer, L. (o.J.): Evaluation in der Weiterbildung, Studienbrief EB 0720 an der Technischen Universität Kaiserslautern
- Bortz, J./Döring, N. (2002): Forschungsmethoden und Evaluation für Sozialwissenschaftler. 3. Aufl. Berlin/Heidelberg/New York
- Brandt, T. (2009): Evaluation in Deutschland: Professionalisierungsstand und Perspektiven. Dissertation. Saarbrücken
- Brauwer, R./Rumpel, K.-D. (2008): Bildungscontrolling. Ansätze, Modelle und Kennzahlen. Aachen
- Bryman, A. (2004): Social Research Methods. Oxford
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1973): Strukturplan für das Bildungswesen. Bonn
- Calder, J. (1997): Programme Evaluation and Quality. A comprehensive guide to setting up an evaluation system. London
- Chelimsky, E./Shadish, W.R. (Hg.) (1997): Evaluation for the 21st Century – a handbook. Thousand Oaks
- Dahrendorf, R. (1965): Bildung ist Bürgerrecht. Plädoyer für eine aktive Bildungspolitik. Hamburg
- Deutsche Gesellschaft für Evaluation (DeGEval) (Hg.) (2001): Evaluation – Reformmotor oder Reformbremse? Köln
- Deutsche Gesellschaft für Evaluation (DeGEval) (Hg.) (2002): Standards für Evaluation. Köln
- Deutsches Institut für Erwachsenenbildung (DIE) (Hg.) (2008): Trends der Weiterbildung. DIE-Trendanalyse 2008, Bielefeld
- Deutsches Institut für Erwachsenenbildung (DIE) (Hg.) (2010): Trends der Weiterbildung. DIE-Trendanalyse 2010. Bielefeld
- Deutsches Institut für Erwachsenenbildung (DIE) (Hg.) (2011): Evaluation der Wirksamkeit des nordrhein-westfälischen Weiterbildungsgesetzes. Bonn
- Deutsches Institut für Erwachsenenbildung (DIE) (2011): Lernende fördern – Strukturen stützen. Evaluation der Wirksamkeit der Weiterbildungsmittel des Weiterbildungsgesetzes (WbG) Nordrhein-Westfalen im Auftrag des Ministeriums für Schule und Weiterbildung des Landes Nordrhein-Westfalen. Abschlussbericht des Deutschen Instituts für Erwachsenenbildung. URL: www.die-bonn.de/doks/2011-evaluation-weiterbildungsgesetz-nrw-01.pdf (Stand: April 2013)

- Dietrich, S./Schade, H.-J. (2008): Mehr Transparenz über die deutschen Weiterbildungsanbieter. URL: www.die-bonn.de/doks/dietrich0802.pdf (Stand: Mai 2013)
- Edding, F. (Hg.) (1987): *Zwanzig Jahre Bildungsforschung – Zwanzig Jahre AfeB*. Bad Heilbrunn
- Faulstich, P. (2003): *Weiterbildung. Begründungen lebensentfaltender Bildung*. München
- Fitzpatrick, J.L./Sanders, J.R./Worthen, B.R. (2004): *Programm Evaluation*. 3. Aufl. Boston u.a.
- Furubo, J.-E./Rist, R.C./Sandahl, R. (Hg.) (2002): *International Atlas of Evaluation*. New Brunswick
- Geißler, K.A. (2005): *Schlussituationen: Die Suche nach dem guten Ende*. 4. Aufl. Weinheim
- Gerl, H./Pehl, K. (1983): *Evaluation in der Erwachsenenbildung*. Bad Heilbrunn
- Gollwitzer, M./Jäger, R.S. (2007): *Evaluation. Workbook*. Weinheim/Basel
- Guba, E.G./Lincoln, Y.S. (1989): *Fourth Generation Evaluation*. London
- Hartz, S./Meisel, K. (2011): *Qualitätsmanagement*. 3. akt. und überarb. Aufl. Bielefeld
- Hartz, S./Schrader, J. (Hg.) (2008): *Steuerung und Organisation in der Weiterbildung. Analysen und Beiträge zur Aus- und Weiterbildung*. Bad Heilbrunn
- Hornbostel, S. (2008): Neue Evaluationsregime? Von der Inquisition zur Evaluation. In: Matthies, H./Simon, D. (Hg.): *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen*. Wiesbaden (S. 59–82)
- Huntemann, H./Reichart, E. (2012): *Volkshochschul-Statistik*. 50. Folge, Arbeitsjahr 2011. URL: www.die-bonn.de/doks/2012-volkshochschule-statistik-01.pdf
- Kejcz, Y./Nuissl, E. u.a. (1979–1981): *Das Bildungsurlaubs-Versuchs- und Entwicklungsprogramm*. 8 Bde. und 2 Sammelbände. Heidelberg
- Kolwe-Jung, G. u.a. (Hg.) (1980): *Modellversuche in der Weiterbildung*, 2 Bde. Heidelberg
- Kommission „Weiterbildung“ Baden-Württemberg (1984): *Weiterbildung – Herausforderung und Chance*. Stuttgart
- Kromrey, H. (2007): Begleitforschung und Evaluation – fast das Gleiche, und doch etwas Anderes! In: Glaser, M./Schuster, S. (Hg.): *Evaluation präventiver Praxis gegen Rechtsextremismus: Positionen, Konzepte und Erfahrungen*. Leipzig, S. 113–135
- Kuckartz, U./Dresing, T./Rädiker, S./Stefer, C. (2008): *Qualitative Evaluation – der Einstieg in die Praxis*. 2. Aufl. Wiesbaden
- Kuper, H. (2005): *Evaluation im Bildungssystem*. Stuttgart
- Kuper, H. (2011): Evaluation. In: Reinders, H. u.a. (Hg.): *Empirische Bildungsforschung – Strukturen und Methoden*. Wiesbaden, S. 131–144
- Lattke, S./Nuissl, E./Pätzold, H. (2010): *Erwachsenenbildung und Europa*. Bielefeld
- Mayring, P. (2007 [1983]): *Qualitative Inhaltsanalyse, Grundlagen und Techniken*. 9. Aufl. Weinheim
- Mertens, D.M. (1998): *Research Methods in Education and Psychology: Integrating diversity with quantitative & qualitative approaches*. Michigan
- Nuissl, E. (2007): Akteure der Weiterbildung. In: Heuer, U./Siebers, R. (Hg.): *Weiterbildung am Beginn des 21. Jahrhunderts*. Münster u.a., S. 371–384
- Nuissl, E. (2010a): *Empirisch forschen in der Weiterbildung*. Bielefeld
- Nuissl, E. (2010b): Ordnungsgrundsätze der Erwachsenenbildung in Deutschland. In: Tippelt, R./Hippel, A. v. (Hg.): *Handbuch Erwachsenenbildung/Weiterbildung*. Wiesbaden, S. 329–346

- Nuissl, E./Schlutz, E. (Hg.) (2001): Systemevaluation und Politikberatung. Gutachten und Analysen zum Weiterbildungssystem. Bielefeld
- Nuissl, E. u.a. (Hg.) (2006): Regionale Bildungsnetze. Ergebnisse zur Halbzeit des Programms „Lernende Regionen – Förderung von Netzwerken“. Bielefeld
- Nuissl, E./Brandt, P. (2008): Porträt Weiterbildung Deutschland. Bielefeld
- Nuissl, E./Siebert, H. (2013): Lehren an der VHS. Ein Leitfaden für Kursleitende. Bielefeld
- Picht, G. (1964): Die deutsche Bildungskatastrophe. München
- Posavac, E.J./Carey, R.G. (1997): Program Evaluation: Methods and case studies. New York
- Reischmann, J. (2006): Weiterbildung – Evaluation. Lernerfolge messbar machen. 2. Aufl. Neuwied
- Rosenblatt, B. v./Bilger, F. (2011): Weiterbildungsbeteiligung 2010. Trends und Analysen auf Basis des deutschen AES. Bielefeld
- Rossi, P.H./Lipsey, M.W./Freeman, H.E. (2004): Evaluation. A systematic approach. Thousand Oaks
- Royle, D. u.a. (2001): Program Evaluation. An introduction. Belmont
- Scriven, M. (1974): Evaluation Perspectives and Procedures. In: Popham, W.J. (Hg.): Evaluation in Education. Los Angeles
- Scriven, M. (2002): Evaluation Thesaurus. 4. Aufl. Newbury Park u.a.
- Scriven, M. (2007): Key Evaluation Checklist. Michigan
- Shaw, I.F./Greene, J.C./Melvin, M. (Hg.) (2006): The Sage Handbook of Evaluation. London
- Stake, R.E. (1995): The art of case study research. Thousand Oaks
- Steinert, B./Klieme, E. (2008): Evaluation im Bereich der empirischen Bildungsforschung. In: Faulstich-Wieland, H./Faulstich, P. (Hg.): Erziehungswissenschaft. Ein Grundkurs. Reinbek, S. 641–654
- Strauch, A./Jütten, S./Mania, E. (2009): Kompetenzerfassung in der Weiterbildung. Instrumente und Methoden situativ anwenden. Bielefeld
- Stockmann, R./Meyer, W. (2010): Evaluation. Eine Einführung. Opladen
- Stufflebeam, D.L. u.a. (1971): Educational Evaluation and Decision Making. Itasca/Illinois
- Stufflebeam, D.L./Shinkfield, A.J. (2007): Evaluation: Theory, models and applications. San Francisco
- Tietgens, H. u.a. (1986): Aufgaben und Probleme der Evaluation in der Erwachsenenbildung. Bonn
- Vedung, E. (2000): Evaluation Research and Fundamental Research. In: Stockmann, R. (Hg.): Evaluationsforschung. Opladen, S. 103–127
- Weber, M. (1985 [1922]): Gesammelte Aufsätze zur Wissenschaftslehre. Hg. von Johannes Winckelmann, 6., erneut durchges. Aufl. Tübingen
- Weiss, C. H. (1998): Evaluation – methods for studying programs and policies. Upper Saddle River
- Wesseler, M. (2010): Evaluation und Evaluationsforschung. In: Tippelt, R./Hippel, A. v. (Hg.): Handbuch Erwachsenenbildung/Weiterbildung. 4. Aufl. Wiesbaden, S. 1031–1048
- Widmer, T./Beywl, W./Fabian, C. (Hg.) (2009): Evaluation. Ein systematisches Handbuch. Wiesbaden
- Will, H./Winteler, A./Krapp, A. (Hg.) (1987): Evaluation in der beruflichen Aus- und Weiterbildung. Heidelberg
- Wottawa, H./Thierau, H. (1990): Evaluation. Bern
- Zech, R./Tödt, K. (2012): Gelungenes Lernen – Qualität und Qualitätsmanagement in der Weiterbildung, Studienbrief TU Kaiserslautern

Abbildungen und Tabellen

Abbildungen

| | |
|--|----|
| Abbildung 1: Ablaufplan Evaluation der 4. Generation | 29 |
| Abbildung 2: Acht Schritte der Evaluation | 29 |
| Abbildung 3: Leitfunktionen von Evaluation | 31 |
| Abbildung 4: Evaluationsgegenstände in der Weiterbildung | 36 |
| Abbildung 5: „Korridore“ der Ergebnisevaluation | 42 |
| Abbildung 6: Partizipativer Evaluationsansatz | 62 |
| Abbildung 7: Realisierung der Evaluation | 66 |
| Abbildung 8: Zeitstrahl Evaluation | 77 |
| Abbildung 9: Datenerhebungsmethoden | 89 |
| Abbildung 10: Datenaufbereitung | 93 |

Tabellen

| | |
|--|----|
| Tabelle 1: Ranking of Countries on the Indicators of an Evaluative Culture | 21 |
| Tabelle 2: Unterschiede zwischen Grundlagen- und Evaluationsforschung | 24 |

Autorenportrait

Ekkehard Nuissl von Rein ist aktuell Seniorprofessor an der Technischen Universität Kaiserslautern und Honorarprofessor an den Universitäten Florenz (Italien), Timisoara (Rumänien) und Torun (Polen). Er leitete vor seiner Emeritierung (2011) über zwanzig Jahre das Deutsche Institut für Erwachsenenbildung (DIE) in Bonn, nachdem er zuvor Leiter des Forschungsinstituts für Weiterbildung in Heidelberg (AfeB) und Direktor der Hamburger Volkshochschule war.

Mit den Fragen der Evaluation ist E. Nuissl seit langem intensiv befasst. Im Heidelberger Forschungsinstitut AfeB leitete er mehrere Jahre das Bildungsurlaubs- Versuchs- und -Entwicklungsprogramm der Bundesregierung, in dem es um die Evaluation unterschiedlicher Modelle der damals neuen Form des Bildungsurlaubs ging. Die Ergebnisse dieser Analyse von Lehr-/Lernprozessen liegen in acht Bänden vor (1979–1981). In zahlreichen Projekten der Folgezeit bearbeitete er Evaluationen im Bereich der Erwachsenenbildung, etwa die Implementation des „Sozialen Trainings“ im Strafvollzug (1986) oder die erste Generation von Projekten der Europäischen Union in der Erwachsenenbildung (2000).

Als Vizepräsident der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (2000–2006) war E. Nuissl zuständig für das dort eingeführte, mittlerweile beispielhafte Evaluationsverfahren der wissenschaftlichen Institute dieser Organisation – jedes Mitgliedsinstitut wird alle sieben Jahre im Peer-Review-Verfahren evaluiert. In diesen Jahren war E. Nuissl auch Leiter größerer Evaluierungsprojekte in der Weiterbildung; z.B. hat er für das Land Nordrhein-Westfalen die Evaluation der Wirksamkeit des Weiterbildungsgesetzes in NRW (2009 bis 2010) verantwortet (DIE 2011).

E. Nuissl verfiert Ansätze von Evaluation, die auf Partizipation, Offenheit, Diskurs und Entwicklung gerichtet sind. Evaluationen, denen es um Kontrolle, Prüfung und Regulierung geht, steht er kritisch gegenüber. Aktuell arbeitet E. Nuissl – neben seiner Lehrtätigkeit – an der Konzeption eines neuen Evaluationsprojekts zum Lernen im Museum.

Kontakt: nuissl@die-bonn.de

Zusammenfassung

Der Studientext zeigt, welche wissenschaftlichen Kriterien bei einer Evaluation in der Erwachsenenbildung zu beachten sind. Zu diesem Zweck werden Grundbegriffe und Theorien geklärt sowie Ziele, Aufgaben und Prinzipien der Evaluation beschrieben. Um die Besonderheiten einer Evaluation in der Erwachsenenbildung darzustellen, bezieht sich der Autor auf die für diesen Kontext klassischen „Evaluationsgegenstände“, also auf Lehr-/Lernprozesse, Bildungseinrichtungen, Bildungsprogramme sowie auf das Bildungssystem. In diesem Zusammenhang wird insbesondere der Frage nachgegangen, welche „gegenstandsbezogenen“ Daten nötig sind, wie diese erhoben und bewertet werden. Hinter allen Ausführungen steht die kritisch-konstruktive Frage, unter welchen Umständen Evaluationsergebnisse zur nachhaltigen Verbesserung der Erwachsenenbildung beitragen können.

Abstract

The study text points out relevant research criteria for an evaluation in the field of adult education. For that purpose, basic concepts and theories are discussed and objectives, tasks, and principles of evaluation are outlined. The author refers to classic targets of evaluation – teaching/learning processes, educational institutions, programmes as well as the education system – to present the features of adult education evaluation. In this context, the issue of required object-related data, as well as the collection and evaluation of data is pursued. All statements are based on the critical, yet constructive question under which circumstances evaluation results lead to a sustainable improvement of adult education.

Evaluation in der Erwachsenenbildung

Der Studientext zeigt, welche wissenschaftlichen Kriterien bei einer Evaluation in der Erwachsenenbildung zu beachten sind. Zu diesem Zweck werden Grundbegriffe und Theorien geklärt sowie Ziele, Aufgaben und Prinzipien der Evaluation beschrieben. Um die Besonderheiten einer Evaluation in der Erwachsenenbildung darzustellen, bezieht sich der Autor auf die für diesen Kontext klassischen „Evaluationsgegenstände“, also auf Lehr-/Lernprozesse, Bildungseinrichtungen, Projekte und Programme sowie auf das Bildungssystem. In diesem Zusammenhang wird insbesondere der Frage nachgegangen, welche „gegenstandsbezogenen“ Daten nötig sind, wie diese erhoben und bewertet werden. Hinter allen Ausführungen steht die kritisch-konstruktive Frage, unter welchen Umständen Evaluationsergebnisse zur nachhaltigen Verbesserung der Erwachsenenbildung beitragen können.