

Lars Balzer
Wolfgang Beywl

evaluiert

erweitertes Planungsbuch für Evaluationen im Bildungsbereich



2., überarbeitete Auflage

Inhaltverzeichnis

| | |
|--|------------|
| Vorwort | 7 |
| 1 Einstieg ins Thema | 11 |
| 2 Was ist Evaluation? | 14 |
| 2.1 Wissenschaftliche Evaluation statt Alltagsbewertung | 15 |
| 2.2 Evaluation als wissenschaftliche Dienstleistung statt (Grundlagen-)Forschung | 19 |
| 2.3 Evaluation als Qualitätsmanagement bzw. Controlling? | 25 |
| 2.4 Evaluation als Beitrag zur gesellschaftlichen Entwicklung | 26 |
| 3 Evaluationsprozess und Auftrag | 29 |
| 4 Bestimmung des Evaluationsgegenstandes | 35 |
| 5 Bestimmung der interessierten Akteure und die Rolle der Evaluierenden | 46 |
| 5.1 Geld- und Auftraggebende der Evaluation | 48 |
| 5.2 Finanzierende und Träger des Bildungsangebots | 49 |
| 5.3 Programmleitende und -mitarbeitende | 50 |
| 5.4 Teilnehmende | 50 |
| 5.5 Vorgesetzte, Arbeitgebende, Arbeitskollegen | 51 |
| 5.6 Sonstige Akteure | 51 |
| 5.7 Evaluierende | 52 |
| 5.8 Akteure im Zusammenspiel | 56 |
| 6 Bestimmung der Evaluationszwecke und -fragenstellungen | 59 |
| 6.1 Evaluationszwecke | 59 |
| 6.1.1 Evaluationszwecke nach Funktionen | 59 |
| 6.1.2 Evaluationszwecke nach Verwendungsabsichten | 62 |
| 6.1.3 Zusammenspiel von Zwecken nach Funktion, nach Verwendungsabsicht sowie Rolle der Evaluation | 67 |
| 6.1.4 Verdeckte Zwecksetzungen | 69 |
| 6.2 Evaluationsfragenstellungen | 72 |
| 6.2.1 Festlegen der Evaluationsfragenstellungen | 72 |
| 6.2.2 Formulierung von Evaluationsfragenstellungen | 74 |
| 6.2.3 Arten von Evaluationsfragenstellungen | 78 |
| 7 Festlegung der Bewertungskriterien | 81 |
| 8 Auswahl von Erhebungsdesign und -methoden | 92 |
| 8.1 Erhebungsdesigns | 92 |
| 8.2 Datenerhebungsmethoden | 103 |
| 8.3 Erhebungsdesign und Datenerhebungsmethoden im Zusammenspiel | 110 |
| 9 Durchführung der Erhebungen | 113 |
| 10 Datenauswertung, Interpretation und Bewertungssynthese | 119 |
| 11 Berichterstattung | 129 |
| 11.1 Adressierte | 129 |
| 11.2 Zeitpunkte | 131 |
| 11.3 Verantwortung und Zuständigkeiten | 132 |
| 11.4 Auswahl geeigneter Berichterstattungsformate | 134 |
| 11.4.1 Schriftliche Berichte | 135 |
| 11.4.2 Mündliche Berichterstattungsformate | 138 |

| | | |
|-----------|---|-----|
| 12 | Nutzung der Ergebnisse | 141 |
| 13 | Evaluation der Evaluation | 150 |
| 14 | Fragebogen erstellen | 156 |
| 14.1 | Inhalt von Fragebogen und Verbindung zum Evaluationszweck | 156 |
| 14.2 | Aufbau und Gestaltung von Fragebogen. | 158 |
| 14.3 | Zwei Grundformen der Frage. | 161 |
| 14.4 | Herausforderungen bei Antwortvorgaben. | 166 |
| 14.5 | Hinweise zur Frageformulierung | 169 |
| 14.6 | Checkliste für den Vorabtest | 171 |
| 15 | Text-Sortier-Technik (TST) | 173 |
| 16 | Fallbeispiele | 186 |
| 16.1 | Evaluation «Jugendprojekt LIFT» | 186 |
| 16.1.1 | Bestimmung des Evaluationsgegenstandes | 186 |
| 16.1.2 | Bestimmung der Interessierten Akteure und die Rolle der Evaluierenden | 188 |
| 16.1.3 | Bestimmung der Evaluationszwecke und -fragestellungen | 189 |
| 16.1.4 | Festlegung der Bewertungskriterien | 190 |
| 16.1.5 | Auswahl von Erhebungsdesign und -methoden | 191 |
| 16.1.6 | Durchführung der Erhebungen | 192 |
| 16.1.7 | Datenauswertung, Interpretation und Bewertungssynthese | 193 |
| 16.1.8 | Berichterstattung | 195 |
| 16.1.9 | Nutzung der Ergebnisse | 195 |
| 16.1.10 | Evaluation der Evaluation | 196 |
| 16.2 | Evaluation der Campusbibliothek Brugg-Windisch | 197 |
| 16.2.1 | Bestimmung des Evaluationsgegenstandes | 197 |
| 16.2.2 | Bestimmung der interessierten Akteure und die Rolle der Evaluierenden | 198 |
| 16.2.3 | Bestimmung der Evaluationszwecke und -fragestellungen | 198 |
| 16.2.4 | Festlegung der Bewertungskriterien | 199 |
| 16.2.5 | Auswahl von Erhebungsdesign und -methoden | 199 |
| 16.2.6 | Durchführung der Erhebungen | 201 |
| 16.2.7 | Datenauswertung, Interpretation und Bewertungssynthese | 201 |
| 16.2.8 | Berichterstattung | 203 |
| 16.2.9 | Nutzung der Ergebnisse | 204 |
| 16.2.10 | Evaluation der Evaluation | 205 |
| | Literaturverzeichnis | 206 |
| | Verzeichnis der Abbildungen | 223 |
| | Verzeichnis der Tabellen | 223 |
| | Verzeichnis der Übungsaufgaben | 224 |
| | Onlineressourcen für die Evaluation | 225 |
| | Musterlösungen zu den Übungsaufgaben | 227 |
| | Glossar | 251 |
| | Stichwortverzeichnis | 261 |

Vorwort

Evaluation hat in den letzten 25 Jahren auch im deutschsprachigen Raum eine steile Karriere gemacht. Ihre Wurzeln und großen Potenziale liegen insbesondere auch im Bildungsbereich. Evaluation kann helfen, Angebote vom Eltern-Kind-Turnen über den Computerkurs bis hin zur Facharzt-Weiterbildung und viele Hunderte mehr systematisch zu beschreiben und zu bewerten, meist mit dem Zweck, diese zu verbessern.

Diejenigen, die Bildungsprogramme finanzieren oder verantwortlich leiten, geben eine Evaluation in Auftrag, sei es an interne Mitarbeitende oder an Externe. Diese Evaluierenden müssen über einschlägiges Evaluationswissen und -können verfügen: wie sie zu relevanten und beantwortbaren Evaluationsfragestellungen kommen; wie sie Evaluationspläne an diese Fragestellungen anpassen; wie sie Evaluationskriterien festlegen; welche Rahmenbedingungen sie bei der Realisierung einer Evaluation zu beachten haben; welche methodischen Herausforderungen sie meistern müssen; wie sie gewonnene Daten verarbeiten; wie sie die Nutzung der erzielten Ergebnisse fördern können und vieles Weitere mehr.

Dieses Buch ist ein praxisnaher Leitfaden für die Planung, Durchführung und Bewertung einer Evaluation und bietet Ihnen eine systematische Anleitung zum Erwerb bzw. Ausbau der dafür erforderlichen Kompetenzen.

Wenn Sie sich mit Unterstützung dieses Buches erarbeiten, wie Sie nützliche und glaubwürdige Evaluationen planen und durchführen, dann nutzen Sie Erfahrungen aus vielen erfolgreichen und oft herausfordernden Lerngeschichten. Personen, die mit verschiedenen Vorgängerversionen dieses Buches autodidaktisch oder in einer der zahlreichen Aus- und Weiterbildungsveranstaltungen gearbeitet haben, hatten es mit unterschiedlichen Problemstellungen und Hindernissen der Evaluation zu tun, und konnten vielfach entsprechende Lösungen für diese finden.

Der vorliegende Text hat seine Wurzeln in Veranstaltungsskripten und Studienbriefen (z.B. Beywl & Balzer, 2009, Balzer & Beywl, 2014) und natürlich in seiner ersten Auflage (Balzer & Beywl, 2015), die mit dem vorliegenden Text grundlegend überarbeitet und erweitert worden ist: Dazu gehören einerseits Ergänzungen und Präzisierungen vorhandener Kapitel. So wird die einheitliche Begriffsverwendung weiter geschärft – z.B. wird nun konsequent die Bezeichnung «Evaluationsprojekt» verwendet, wenn eine geplante oder realisierte Evaluation gemeint ist. Das Wort «Projekt» als eine Klasse von Evaluationsgegenständen wird hingegen wenn immer möglich durch Bezeichnungen wie «Programm» oder «Maßnahme» ersetzt. Ebenso ist das Designkapitel grundlegend überarbeitet, und Evaluationszwecke werden nun konsistenter als verbesserungsorientiert, rechenschaftslegungsorientiert, grundsatzentscheidungsorientiert und wissensgenerierend unterschieden. Die aktuellen Revisionen der DeGEval- und SEVAL-Evaluationsstandards sind eingearbeitet. In der deutschsprachigen Fachliteratur bislang wenig ausgeführte Themen wie die Typisierung von Evaluationsfragestellungen, das Verständnis von «Daten» in der Evaluation, die Festlegung der Evaluationskriterien, die Bewertungssynthese sowie der Einfluss durch Evaluation sind weiter vertieft. Darüber hinaus kommen Kapitel zur Fragebogenkonstruktion und zur Text-Sortier-Technik sowie zwei ausführliche Fallbeispiele hinzu. Sowohl neue wie aktualisierte Literatur ist eingearbeitet.

Der Titel dieses Buches ist ein Vorgriff auf das, was Sie nach einer systematisch durchgeführten Evaluation erreicht haben, und markiert den Abschluss eines Dreischritts:

- **evaluiert?** – Ein legitimer, vielfach notwendiger Checkpunkt, der in der Lebensphase eines Programms anzuwenden ist.
- **evaluiert!** – Ein Planungsentschluss, der gefasst wird, wenn die Evaluation einer Maßnahme als möglich und nutzenversprechend eingeschätzt wird.
- **evaluiert.** – Der Schlusspunkt hinter der Durchführung einer Evaluation, an den sich gelegentlich ein neuer Dreischritt anschließt.

Mit dem im Untertitel genannten «Bildungsbereich» sind auch zahlreiche benachbarte Felder angesprochen. Zum einen sind dies alle Bildungsstufen von der frühen Bildung über die allgemeinbildenden Schulstufen und die grundständige Berufsbildung bis hin zur hochschulischen Bildung sowie der allgemeinen und beruflichen Weiterbildung. Zum anderen sind dies Handlungsfelder der Sozialen Arbeit, des Gesundheitsbereichs, der Verbraucherarbeit, der Beratung, des Coachings und viele andere mehr. Wichtig ist, dass es zumindest im weitesten Sinne um intentionales, also zielgerichtetes Lernen geht, das im Rahmen von Programmen stattfindet. Was hier unter einem Programm verstanden wird, schlagen Sie am besten gleich im → **Glossar** am Ende des Buches nach – das Wort bzw. der Wortbestandteil wird über 500-mal in diesem Buch genutzt. Sie werden das Buch auch nutzen können, wenn das Gegenstandsfeld ein anderes ist als Bildung – vieles lässt sich 1:1 übertragen. Die Erarbeitung der für Evaluationen kennzeichnenden Methodologie wird dadurch erleichtert, dass die Beispiele aus dem Bereich Bildung aus einem für viele bekannten Kontext stammen.

Adressierte dieses Buches sind alle, die sich die fachlichen Grundlagen der Evaluation erarbeiten wollen, um Evaluationen Schritt für Schritt planen und schließlich erfolgreich umsetzen zu können. Angesprochen sind zum einen Studierende in den grundständigen Bachelor-, besonders den Masterstudiengängen im Bereich der Bildungswissenschaften und auch der sozialen Arbeit sowie ihrer Bezugsdisziplinen, u. a. Psychologie, Soziologie und Wirtschaftswissenschaft. Zum anderen richtet sich das Buch an Weiterbildungsstudierende in Programmen, die ein substanzielles Element zur Evaluation enthalten, z. B. im Bereich des Projekt-, Change- oder Qualitätsmanagements. Schließlich gehören auch Fachpersonen in Bildungsorganisationen und -abteilungen, die ihr Wissen und Können im Prozess der Arbeit auf- und ausbauen wollen, zu den Hauptadressierten.

Erfahrene Evaluierende können dieses Buch im Rahmen eines kombinierten Schulungs- und Beratungsangebotes einsetzen. Präsenzzeiten der Teilnehmenden können – wie schon vielfach mit den Vorläufern dieses Buches umgesetzt – durch die vorbereitende Lektüre und evtl. in Kombination mit der Bearbeitung der Übungsaufgaben reduziert werden. Durch ein solches Konzept eines «inverted classroom» kann diese Präsenzzeit effektiver für kollegiale Beratung und Coaching durch Evaluationsexpertinnen und -experten genutzt werden. Dabei kann z. B. eine reale Evaluation über zehn – jeweils durch ein Buchkapitel erläuterte – Schritte geplant und gegebenenfalls danach tatsächlich umgesetzt werden.

Detaillierte Lernziele dieses Buches finden sich je zu Beginn der → **Kapitel 2–13**. Durch ein intensives Studium dieses Buches können Sie u. a.:

- wissenschaftliche Evaluation definieren,
- eine systematische Schrittfolge für eine für den Auftrag maßgeschneiderte Evaluation planen,
- den jeweiligen Evaluationsgegenstand differenziert beschreiben,

- die im Rahmen einer Evaluation zu berücksichtigenden Akteure identifizieren,
- Evaluationszwecke und -fragestellungen präzise formulieren,
- Kriterien für die systematische Bewertung klären und anwenden,
- geeignete Erhebungsdesigns auswählen und Erhebungsmethoden bestimmen,
- die Erhebungsarbeiten effizient und flüssig in den Bildungsalltag einbetten, mit einem Schwerpunkt bei Fragebogen,
- Auswertungs- und Bewertungsarbeiten konzipieren und abstimmen, mit einer Vertiefung bei qualitativen Daten,
- Evaluationsberichte auf die Adressierten zeitlich und inhaltlich abstimmen,
- die effektive Nutzung der Evaluation und ihrer Ergebnisse vorbereiten,
- sowie Voraussetzungen für eine Evaluation der Evaluation – die sogenannte Meta-Evaluation – schaffen.

Es empfiehlt sich, das Buch von vorne nach hinten durchzuarbeiten. → **Kapitel 1** enthält eine kleine Fallbeschreibung dazu, welche Herausforderungen sich schon ganz zu Beginn der Evaluation einer Bildungsmaßnahme stellen. Das Beispiel ist bewusst einfach gewählt. Dennoch entfaltet sich in seinen fiktiven Dialogen bereits die Komplexität, deren Bewältigung auch für erfahrene Evaluierende immer wieder eine große Herausforderung darstellt. In vielen der 39 Beispielkästen der nachfolgenden Kapitel wird auf dieses Eingangsszenario zurückgegriffen. Andere enthalten verdichtete kurze Fallbeispiele von Evaluationen, jeweils bezogen auf das im Kapitel behandelte Thema.

Die folgenden → **Kapitel 2–13** führen in die Methodologie und Fachsprache der Evaluation ein. Die Kompetenzen der Datenerhebung müssen Sie entweder bereits aus anderen Studien mitbringen oder durch Bearbeitung von einschlägigen Lehrbüchern zu empirischen Methoden evtl. in Kombination mit entsprechenden Studienangeboten erwerben. Gleiches gilt für die qualitativen und quantitativen Auswertungsverfahren. In den entsprechenden Kapiteln dieses Buches geben wir Literaturempfehlungen, welche durch Internetquellen ergänzt werden.

Für die zweite Auflage wurden alle Kapitel überarbeitet. Das gilt besonders für die → **Kapitel 7** (Bewertungskriterien) und **8** (Erhebungsdesign). In → **Kapitel 10** ist die Bewertungssynthese ausführlicher behandelt. Neu erarbeitet sind auch die Übungsaufgabe zur Unterscheidung von Fragestellungs-Arten sowie die Ausführungen zur Verbindung von Evaluationsfragestellungen und Erhebungsmethoden in → **Kapitel 6**. Von den knapp 320 Literaturangaben ist mehr als ein Viertel neu hinzugekommen oder aktualisiert.

Eine exemplarische Erweiterung und Vertiefung bei Erhebung und Auswertung bilden die neu aufgenommenen → **Kapitel 14** und **15**, welche unter einer evaluationsspezifischen Perspektive in die Konstruktion von Fragebogen sowie in die qualitative Auswertungsmethode Text-Sortier-Technik einführen. Ebenfalls neu ist → **Kapitel 16**, in dem zwei reale Fallbeispiele den zuvor geschilderten zehnschrittigen Evaluationsprozess illustrieren.

Vielfach wird in späteren Kapiteln das Wissen der vorangehenden vorausgesetzt. Eine Besonderheit besteht darin, dass alle Fachbegriffe nach bestimmten Regeln definiert sind und sie bei jeder Verwendung im Buch stets dieselbe Bedeutung haben. Sie finden die Begriffserklärungen im angehängten Glossar, das – wie einige andere im Literaturverzeichnis aufgeführte aktuelle Publikationen – auf dem noch umfangreicheren Online-Glossar des Evaluationsinstituts Univation basiert (vgl. Schobert, Beywl & Niestroj, 2016). Es ist unter http://eval-wiki.org/glossar/Eval-Wiki:_Glossar_der_Evaluation verfügbar.

Die meisten Kapitel des Evaluationsprozesses enthalten Übungsaufgaben. Diese ermöglichen es, Ziele unterschiedlicher Anspruchsniveaus zu verfolgen, von der Erarbeitung des Begriffsverständnisses über die Anwendung von Wissen auf vorgegebene Fallbeispiele bis hin zum Transfer auf reale Praxisbeispiele aus den Bildungsfeldern, in denen Sie tätig sind. Im Anhang des Buches finden Sie passende Musterlösungen. Für am Selbststudium Interessierte empfiehlt es sich, alle Übungsaufgaben durchzuarbeiten. Wenn Sie in Lerngruppen arbeiten, tauschen Sie sich zu Ihren Lösungen aus. Wenn Sie an einer Präsenzweiterbildung teilnehmen, in der Sie von Evaluationsexpertinnen und -experten begleitet werden, ist auf die Bearbeitung der Übungsaufgaben eventuell zu verzichten.

Zum Beleg von Originalaussagen werden in diesem Buch an verschiedenen Stellen fremdsprachige Textstellen zitiert. Zum leichteren Verständnis haben wir diese sinngemäß ins Deutsche übersetzt und die übersetzten Textteile in den Text integriert. In den Kapiteln zur Fragebogenerstellung und zur Text-Sortier-Technik sowie in den beiden Fallbeispielen werden Originaldokumente aus Evaluationen in der Deutschschweiz genutzt, mit leicht abweichender Rechtschreibung.

Sollten Sie bei der Lektüre des Buches auf Fehler oder unklare Passagen stoßen, schreiben Sie uns bitte eine E-Mail an: evaluiert@lars-balzer.info.

Lars Balzer und Wolfgang Beywl, Zollikofen und Bern, im Oktober 2018

Einstieg ins Thema

KAPITEL 1

Was ist eigentlich eine Evaluation?

Stellen Sie sich einen Weiterbildungskurs zum Thema «Wie erstelle ich meine erste Homepage?» vor, in dem an vier Abenden Basiswissen über die Erstellung von Internetseiten vermittelt wird.

Fallbeispiel

Nachdem dieser Kurs als Pilot gestartet ist, fragen sich die Weiterbildungsverantwortlichen in einem Teamgespräch, wie gut ihre Arbeit eigentlich ist. Nach Kursende gehen die Teilnehmenden ihrer Wege, man trifft sie nicht mehr. Ob mehr Wissen vorhanden ist als zuvor und ob tatsächlich eigene Internetseiten erstellt worden sind, ist den Weiterbildungsverantwortlichen normalerweise nicht bekannt. «Wir brauchen eine Evaluation, um zu überprüfen, ob wir gute Arbeit leisten», schlägt einer der Verantwortlichen vor. Bevor sie einen Evaluationsauftrag an eine interne oder eine externe Evaluationsfachperson formulieren, diskutieren sie die Möglichkeiten.

Herr Schmidt ist von dieser Idee sogleich begeistert: «Lasst uns doch am Ende des Kurses unsere Teilnehmenden fragen, ob sie mit uns und unserer Arbeit zufrieden sind. Wenn z.B. 85 Prozent von ihnen mit unseren Veranstaltungen zufrieden sind, gerne zu weiteren Veranstaltungen kommen würden und uns weiterempfehlen, können wir doch mit Recht sagen, dass wir Gutes geleistet haben.» – «Ja, das stimmt», ist man sich schnell einig, «doch hilfreicher und interessanter wäre zu wissen, was den restlichen 15 Prozent oder auch den Zufriedenen nicht gefallen hat, damit wir uns weiter verbessern können!»

Erfolg = Zufriedenheit?

«Darüber hinaus», stellt Frau Zbinden fest, «sollten wir mehr wissen als nur, wie zufrieden unsere Teilnehmenden sind. Ich glaube, bei uns geht es auch mal ganz lustig zu und her, und das gefällt allen. Zufriedenheit ist wichtig, denn das erhöht die Chance, dass unsere Kunden wiederkommen. Aber hauptsächlich sollen sie doch etwas lernen.» Kopfnicken macht die Runde. «Wir könnten unsere Teilnehmenden also ergänzend befragen, ob sie etwas gelernt haben.» – «Wobei: Können das unsere Teilnehmenden wirklich selbst beurteilen? Wie wäre es anstelle dessen mit einem Wissenstest am Ende der letzten Stunde? So könnten sie sich einerseits selbst überprüfen, und wir könnten andererseits erkennen, was von den Inhalten tatsächlich behalten worden ist!» – «Ja, das würde uns sicher helfen», ist die einhellige Meinung.

Erfolg = kurzfristiges Wissen?

«Aber was heißt eigentlich ‹behalten›?», wirft Frau Lavric ein. «Ein Wissenstest im Anschluss an eine intensive Fortbildung überprüft vielleicht das Kurzzeitgedächtnis oder den Fleiß, also wie intensiv sich jemand auf die Abschlussprüfung vorbereitet hat. Ist aber nicht vielmehr interessant, was längerfristig noch gewusst wird?» Ratlosigkeit macht sich breit. «Sollen wir unsere Teilnehmenden vielleicht zwei Monate später noch einmal testen?», fragt Herr Schmidt ungläubig, gar nicht mehr so begeistert wie noch zu Beginn der Diskussion. «Und wie sollen wir sie dazu motivieren? Der ganze Aufwand!» Ein leises Stöhnen in der Runde ist nicht zu überhören.

Erfolg = nachhaltiges Wissen?

«Es wird noch viel schwieriger, bei genauer Überlegung.» Frau Lavric spricht sich in Fahrt: «Selbst wenn nach zwei Monaten noch alles richtig gewusst wird, ist das zwar schön, aber unser eigentliches Ziel ist doch, dass die Teilnehmenden nicht nur bei einem Wissenstest gut abschneiden, sondern dieses Wissen auch tatsächlich anwenden. Denn selbst wenn der Wissenstest zu einem späteren Zeitpunkt erfolgreich ablaufen sollte: Wer sagt uns, dass jemand auch vernünftige In-

Erfolg = Anwendung?

1

ternetseiten erstellen kann? Und vielleicht auch tatsächlich solche erstellt hat? Wissensvermittlung ist nur ein Zwischenziel; schöne, funktionelle, sichere und suchmaschinenoptimierte Internetseiten sind hingegen das, was wir eigentlich erreichen wollen.» – «Stimmt, also lasst uns doch die Teilnehmenden zwei Monate nach Abschluss des Kurses anrufen und fragen, ob sie Internetseiten erstellt haben.» Damit ist Herr Malte überhaupt nicht einverstanden: «Das ist zwar eine gute Idee, aber zwei Monate sind eine zu kurze Zeitspanne, und einfach ein paar Internetseiten zu erstellen, ist keine große Herausforderung. Wir müssen die Internetseiten anschauen und selbst nach gewissen Usability-Kriterien überprüfen. Also müssen wir nach den Internetadressen fragen.» – «Das stimmt. Und vielleicht können wir uns schon ein wenig auf die Schultern klopfen, wenn die einfachsten Regeln der Internetseitenprogrammierung eingehalten worden sind. Und wer weiß, vielleicht haben sie nach einer gewissen Zeit auch eine sehr gute Platzierung in diversen Suchmaschinen?!» – «Das Problem ist nur: Wie bekommen wir das heraus?» – «Was wir also brauchen, ist eine Datenerhebung längere Zeit nach Ende unserer Veranstaltung.» Erneutes Kopfnicken: «Hiermit würden wir weiterkommen!»

Was genau löst den Erfolg aus?

«Aber was wüssten wir dann eigentlich genau?», meldet sich Frau Lavric wieder zu Wort: «Selbst wenn wir das alles in Erfahrung bringen können und feststellen sollten, dass 46 Prozent unserer Teilnehmenden gute Internetseiten erstellt haben: Was bedeutet ein solcher Prozentsatz? Sind das viele, sind das wenige? Wie viele Personen hätten auch ohne Weiterbildung Internetseiten erstellt? Ihr wisst, das Thema ist aktuell, und überall gibt es Do-it-yourself-Bausätze. Könnten wir also mit 46 Prozent zufrieden sein? Und hat unsere Weiterbildung eigentlich einen Anteil am möglichen Erfolg?» – «Wir brauchen einen Vergleichsmaßstab», ist die einstimmige Schlussfolgerung. Doch das ist leichter gesagt als getan. «Wie wäre es, wenn wir uns mit unseren Kolleginnen und Kollegen in Düsseldorf vergleichen würden? Die haben ein sehr ähnliches Konzept und eine ähnliche Klientel.» – «Das führt uns zwar weiter, aber ob wir absolut betrachtet Erfolg haben, wissen wir nicht. Vielleicht sind wir besser als das Team in Düsseldorf, aber was nutzt das, wenn wir alle eigentlich recht schlecht sind? Umgekehrt könnte ich gut damit leben, im Vergleich etwas schlechter zu sein, aber absolut betrachtet ein gutes Weiterbildungsprogramm anzubieten.»

Wie kann man Wirksamkeit messen?

Jemand hat eine Idee: «Wir brauchen eine andere Personengruppe, die unsere Weiterbildung nicht besucht hat. Dann vergleichen wir die Internetseiten unserer Teilnehmenden mit denjenigen besagter Gruppe. Fällt der Vergleich positiv aus, können wir doch sagen, dass unsere Weiterbildung Erfolg hatte, oder?» Zufriedenheit ist auf den Gesichtern abzulesen. «Und wen nehmen wir als Vergleichsgruppe?» – «Wie wäre es mit den Teilnehmenden des Englischkurses? Die erfahren dort sicher nichts über Internetseitenerstellung.» – «Das stimmt zwar», wirft einer ein, «aber ist das wirklich ein fairer Vergleich? Unsere Teilnehmenden sind motiviert, Internetseiten zu erstellen. Die Teilnehmenden des Englischkurses teilen dieses Interesse nicht. Also könnte es auch sein, dass ein möglicher Erfolg auf diese Motivation zurückzuführen ist und sich dieser auch ohne unseren Kurs eingestellt hätte.»

Systematische Evaluationsplanung ist erforderlich

Ein Kollege mit Wissen über sozialwissenschaftliche Forschung hat dazu einen Einfall: «Wir brauchen eine Vergleichsgruppe, die sich hinsichtlich aller für das Erreichen der Weiterbildungsziele wichtigen Bedingungen wie Motivation oder Vorwissen nicht von den Teilnehmenden unserer Kurse unterscheidet. Eine solche Gruppe können wir ganz leicht aufreiben. Wir haben doch ohnehin mehr Anmeldungen, als wir in einem Monat bedienen können. Wenn wir von den

nächsten 100 Anmeldungen 50 zufällig aussuchen, die wir sofort aufnehmen, und die anderen eben ein wenig später, haben wir die Gruppe bereits gebildet. Nach den Gesetzen der Wahrscheinlichkeit haben wir nämlich eine gute Chance, durch diese fast schon zufällige Zuordnung die relevanten Eigenschaften, die wir nicht einmal kennen müssen, gleich auf beide Gruppen zu verteilen. Beide Gruppen unterscheiden sich dann in nichts außer in der Tatsache, dass die eine an der Weiterbildung teilnimmt und die andere eben nicht. Ist unsere Gruppe dann besser, können wir recht sicher sein, dass wir mit unserer Weiterbildung einen Effekt erzielen.»

Doch der nächste Einwand folgt sofort: «Moment! Ich glaube nicht, dass wir lange genug warten können mit unserer Vergleichsgruppe, um längerfristige Resultate überprüfen zu können. Die wollen doch bald loslegen, und wer weiß, was sie in der Wartezeit unternehmen, um schnell ihre gewünschten Internetseiten erstellen zu können.»

«Eigentlich ist mir das zu kompliziert. Das Überprüfen des Erfolges unserer Weiterbildung ist schon wichtig, aber eigentlich hätte ich einfach gerne ein paar Informationen für mich und meine weiteren Planungen, sodass ich von den Teilnehmenden gerne gewusst hätte, welche konkreten Verbesserungsvorschläge sie haben. Es gibt schließlich keine Weiterbildungsmaßnahme, die so gut wäre, dass man sie nicht noch verbessern könnte!»

«Ach, worauf lassen wir uns mit einer solchen Evaluation nur ein? Wie soll eine solche Evaluation jemals abgeschlossen werden?»

Es bietet sich an, einen Evaluationsauftrag an hierfür qualifizierte interne Mitarbeitende oder Externe zu vergeben, die über Wissen und Können verfügen, wie sie zu beantwortbaren Evaluationsfragestellungen kommen, wie sie Evaluationspläne an diese Fragestellungen anpassen, welche Rahmenbedingungen sie bei der Realisierung einer Evaluation zu beachten haben, welche methodischen Klippen sie umschiffen müssen, wie sie gewonnene Daten verarbeiten, was mit den erzielten Ergebnissen zu geschehen hat und vieles Weitere mehr.

Dieses Buch bietet Ihnen eine systematische Anleitung zum Erwerb bzw. Ausbau der hierfür erforderlichen Kompetenzen.

KAPITEL 2

Was ist Evaluation?



Lernziele von Kapitel 2:

- Den Begriff der wissenschaftlichen Evaluation definieren können
- Evaluation von Forschung abgrenzen können
- Werte und Bewertung als Besonderheiten von Evaluation an Beispielen erläutern können

Auf dem Weg zur Definition

Auf die Frage, was Evaluation genau ist, gibt es keine einfache Antwort. Die Aussage von Franklin und Thrasher (1976), wonach es so viele Evaluationsdefinitionen wie Evaluierende gibt¹ («To say that there are as many definitions as there are evaluators is not too far from accurate», S. 20), hat an Gültigkeit eingebüßt. Theoriebildung und Professionalisierung haben in der vergangenen Jahrzehnten zu einer Schärfung des Begriffs geführt. Doch wird in der Bildungspraxis weiterhin vieles unter dem Begriff subsumiert, was in der Evaluationsgemeinschaft nicht oder allenfalls als schlechte Evaluation gelten würde. Daher ist es wichtig, ein fundiertes Verständnis davon zu entwickeln, was wissenschaftliche Evaluation ausmacht. Doch auch das ist nicht einfach, denn Evaluation wird nach Glass und Ellet (1980, S.211) sehr unterschiedlich definiert und abgegrenzt – und das gilt heute noch ebenso wie damals. Viele theoretische wie auch praktische Aktivitäten werden darunter gefasst, ohne dass ein allgemeingültiges Evaluationsparadigma vorherrschen würde: «Evaluation is a set of theoretical and practical activities without a widely accepted paradigm.» Und je nach Zuordnung zu einer bestimmten wissenschaftlichen Disziplin wird der Begriff zudem unterschiedlich akzentuiert (Götz, 1998, S.20). Erschwerend kommt hinzu, dass «allzu leicht [...] heute auch triviale Formen der Rückmeldung zu Evaluationen» werden (Böttcher, Kerlen, Maats, Schwab & Sheikh, 2014, S.7).

Diese beinahe beliebige Möglichkeit der Inanspruchnahme macht Evaluation für viele attraktiv, bietet aber auch schier unüberschaubare Ansatzpunkte für Kritik und Polemik. Unterschiedliche Ausdrücke und Aussagen werden in diesem Zusammenhang verwendet und diskutiert, wie «Evaluation als modernes Ritual» (Schwarz, 2006), Evaluation als notwendiges Übel (evaluation as a «necessary evil», Sullivan, 2011), «Evaluations Mania» (Frey, 2007a), oder auch «von der Inquisition zur Evaluation» (Hornbostel, 2008). Schon 2000 beschreibt Simon die «Evaluitis» als «eine fiebrige Erkrankung, die unversehens den Körper der Wissenschaft in seiner Gesamtheit erfasst hat» (S. 15). Umbach (2002) fragt wegen dramatisch ansteigender Evaluierungswünsche im Wissenschaftssystem, ob «Evaluitis» heilbar sei. Stoellger (2005a, 2005b, 2005c) spricht von einem akuten «Morbus Evaluitis» aufgrund einer epidemischen Breite der Evaluations-tätigkeit. Für qualitativ hochwertige Evaluationen und gegen eine «Evaluitis» im Sinne von reinen Befragungsroutinen spricht sich Döring (2005) im Bereich der Lehrevaluation aus, und Frey (2007b, 2008) greift die Krankheitsmetapher auf, indem er insbesondere der Wissenschaft die Diagnose «Evaluitis» für Evaluation als sich epidemisch ausbreitende, neue Krankheit stellt – dieser Begriff ist seitdem immer wieder in der Literatur anzutreffen (z. B. Burzan & Jahnke, 2010; Hornbostel, 2016; Munske, 2014; Niggli, 2011). Es werden auch Warnungen

1 Zum Beleg von Originalaussagen werden in diesem Buch an verschiedenen Stellen fremdsprachige Textstellen zitiert. Zum Verständnis ist die Kenntnis der anderen Sprache allerdings nicht erforderlich, da alle fremdsprachigen Zitate sinngemäß ins Deutsche übersetzt und die übersetzten Passagen in den erläuternden Text um das Zitat herum integriert sind.

ausgesprochen: «Vorsicht vor Evaluationismus!» (Kappler, 2010). Oder ist es gar so, dass wir uns zu Tode evaluieren, wie Preußler (2008) fragt – oder dass ein «Evaluationsnotstand» herrscht (Niedermaier, 2012, S.8)?

Für einen differenzierten Umgang mit diesem vermeintlichen Unwort ist es notwendig, einige Definitionen vorzustellen und eine Arbeitsdefinition für dieses Buch zu formulieren.

Auf der Suche nach einer adäquaten Definition gibt der Wortstamm einen ersten Hinweis. Auch wenn eine unmittelbare Herleitung aus dem Lateinischen nicht angebracht ist (entgegen der weitverbreiteten Meinung existiert das Wort «evaluare» im Lateinischen nicht), legt das Lateinlexikon erste Spuren. Wurzeln finden sich nämlich im lateinischen Wort «valor», das im Deutschen so viel bedeutet wie «bei Kräften sein», «wert sein» oder «gültig sein». Man beachte hierbei die eindeutig positive Konnotation.

Wortstamm

Den etymologischen Herleitungen des Begriffes von Karbach (1998) folgend, entwickelte sich daraus zunächst das französische «valoir», woraus die Substantivierung «valeur» (im Sinne vom «prix», also auch Wert) entstand. Daraus wurde wiederum das Verb «évaluer» abgeleitet und von diesem das Substantiv «évaluation» («Schätzung», «Ermittlung» oder «Wertbestimmung») gebildet.

Die daraus hervorgegangenen englischen Wörter «evaluate» («bewerten») sowie «evaluation» («Einschätzung», «Auswertung») bilden die Grundlage für die heute im Deutschen gebräuchliche Form des Begriffes.

Eine erste, vom Wortstamm ausgehende Umschreibung von Evaluation lautet also:

«Bestimmung des Wertes einer Sache»
(Bedeutung nach Wortstamm).

Folgt man dieser Bestimmung, so ist Evaluation – zumindest im französischen oder angelsächsischen Sprachraum – eine Bezeichnung für alltägliches menschliches Handeln. Denkend oder sprechend wird auf Basis eines Sinneseindrucks, z. B. des Blicks aus dem Fenster oder der herausgehaltenen Hand, ein Urteil – hier: über das Wetter – abgegeben. Es handelt sich um eine einfache Alltagsbewertung. Obwohl es sich auch in der deutschen Alltagssprache – z. B. in Tageszeitungen, Fernsehinterviews oder Talkrunden – seit einigen Jahren zu etablieren begonnen hat, solche subjektiven Ad-hoc-Bewertungen als «Evaluationen» zu bezeichnen, wird an dieser Stelle dafür plädiert, «Evaluation» und «evaluieren» für das wissenschaftlich abgestützte, systematische Beschreiben und Bewerten zu reservieren. Was darunter zu verstehen ist, wird nachfolgend präzisiert.

2.1 Wissenschaftliche Evaluation statt Alltagsbewertung

Kromrey (2001) unterscheidet den alltäglichen und den wissenschaftlichen Sprachgebrauch von Evaluation danach, was von wem wie und nach welchen Kriterien bewertet wird. Auf dem Weg zu unserer Definition von Evaluation für dieses Buch gilt es nun, diese und andere relevante Dimensionen zu konkretisieren.

Zunächst ist zu klären, wer oder was evaluiert werden soll. Dies wird als Evaluationsgegenstand bezeichnet.

Wer oder was? – Der Evaluationsgegenstand

2

Die Menge an potenziellen Evaluationsgegenständen ist beinahe unüberschaubar. Wottawa und Thierau (2003, S. 59) nennen «Personen, Umwelt-/Umgebungsfaktoren, Produkte, Techniken/Methoden, Zielvorgaben, Programme, Projekte, Systeme/Strukturen, Forschungsergebnisse/Evaluationsstudien» und haben der Evaluation damit ein bereits sehr breites Tätigkeitsfeld eröffnet. Scriven (1981, S. 4) weitet dieses noch aus, indem er jedes Hauptwort eines Wörterbuches zu einem möglichen Evaluationsgegenstand macht: «One can begin at the beginning of a dictionary and go through to the end, and every noun, common or proper, readily calls to mind a context in which evaluation would be appropriate.» Cook und Matt (1990, S. 15) bringen es auf den Punkt: «Alles kann evaluiert werden.»

Wie? – Die Evaluationsmethoden

Entscheidend für die Abgrenzung zur Alltagsbewertung ist, dass die Bestimmung des Wertes (Güte und Tauglichkeit) eines Evaluationsgegenstandes systematisch, umfassend und objektiv durchgeführt werden soll, was folgende Definitionen festhalten:

«Evaluation: The systematic investigation of the worth or merit of an object»
(Joint Committee on Standards for Educational Evaluation, 1994, S.3).

«Good evaluation is the systematic, comprehensive, objective determination of merit or worth»
(Scriven, 1974, S.23).

Andere Autorinnen und Autoren gehen einen Schritt weiter und fordern explizit sozialwissenschaftliche Methoden:

«wissenschaftliche Evaluation nutzt sozialwissenschaftliche Methoden [...]»
(Döring & Bortz, 2016, S. 979).

Evaluation, evaluative research und Evaluationsforschung

Schon früh gab es aber auch Positionen, die mit dem Begriff «Evaluation» einen eher alltäglichen Bewertungsvorgang bezeichneten, ohne die Notwendigkeit, systematisch vorzugehen: «While it implies some logical or rational basis for making such judgments, it does not require any systematic procedures for marshaling and presenting objective evidence to support such judgment. Thus, we retain the term «evaluation» in its more common-sense usage as referring to the general process of assessment or appraisal of value» (Suchman, 1967, S.7). Der Begriff «evaluative research» wurde hingegen reserviert für eine Bewertung, die auf wissenschaftlichen Forschungsmethoden basiert: «Evaluative research» on the other hand, will be restricted to the utilization of scientific research methods and techniques for the purpose of making an evaluation» (Suchman, 1967, S.7).

Dieses Begriffsverständnis wurde auch im deutschsprachigen Raum diskutiert, allerdings eher unter Verwendung des Begriffspaares «Evaluation – Evaluationsforschung» (Wottawa & Thierau, S. 13). Eine strikt akademisch wissenschaftliche Vorgehensweise als Evaluationsforschung zu bezeichnen, ist insbesondere in methodisch orientierten Texten anzutreffen (z.B. Döring & Bortz, 2016; Gollwitzer & Jäger, 2014; Häder, 2010). Stufflebeam und Coryn (2014, S.133–172) zählen solche Ansätze zu den «Quasi-Evaluationen», wegen ihrer Verengung bei Evaluationsfragestellungen bzw. methodischen Zugängen: «A quasi-evaluation approach provides direction for performing a high-quality study that is narrow in terms of the scope of questions addressed, the methods employed, or both» (S.133).

Manches spricht gegen diese Art der Begriffsverwendung. So legen semantisch ähnliche Wortkonstruktionen wie Sozialforschung, Genforschung oder Bildungsforschung keine soziale, genetische oder gebildete Forschung nahe, sondern eine Forschung *über* Soziales, Gene oder Bildung (vgl. Beywl, 1991; Hense, 2006, S.26). Vermutlich hatte Suchman Ähnliches im Sinn, als er mit «evaluative re-

search» ebenfalls eine Adjektiv-Konstruktion verwendete: «In this sense, «evaluative» becomes an adjective specifying a type of research» (Suchman, 1967, S.7).

Eine explizite Differenzierung zwischen Evaluation und Evaluationsforschung im beschriebenen Sinn ist eher rückläufig, wie an den letzten Auflagen des international meistverkauften Lehrbuchs zur Evaluation exemplarisch aufgezeigt werden kann. So wird in der fünften Auflage aus dem Jahr 1993 Evaluationsforschung als eine systematische Anwendung von sozialwissenschaftlichen Verfahren zur Einschätzung/Bewertung der Konzeption, Gestaltung, Umsetzung und Nützlichkeit sozialer Interventionsprogramme definiert:

«Evaluation research is the systematic application of social research procedures for assessing the conceptualization, design, implementation, and utility of social intervention programs»

(Rossi & Freeman, 1993, S.5).

Demgegenüber lassen die Autoren seit der sechsten Auflage den Forschungszusatz «research» bei ihren zentralen Definitionen weg und definieren Programmevaluation als den Einsatz sozialwissenschaftlicher Verfahren zur systematischen Untersuchung der Wirksamkeit sozialer Interventionsprogramme:

«Program evaluation is the use of social research procedures to systematically investigate the effectiveness of social intervention programs [...]»

(Rossi, Freeman & Lipsey, 1999, S.4).

In der aktuellen siebten Auflage schreiben die Autoren explizit, dass die Begriffe «Evaluation», «Programmevaluation» und «Evaluationsforschung» beliebig austauschbar verwendet werden: «Note that throughout this book we use the terms evaluation, program evaluation, and evaluation research interchangeably» (Rossi, Lipsey & Freeman, 2004, S.6).

Auch andernorts ist diese Gleichsetzung zu beobachten: «Furthermore, I will make no distinction between evaluation research and evaluation» (Vedung, 2004, S.112). Ähnliches wird ebenfalls in vielen deutschsprachigen Lehrtexten zum Ausdruck gebracht. So schreibt Stockmann (2004, S.13): «Die Begriffe «Evaluierung», «Evaluation» und «Evaluationsforschung» werden hier synonym verwendet», Döring (2014, S.167) beginnt ihren Text mit den Worten «Mit «Evaluationsforschung», «wissenschaftlicher Evaluation» oder kurz «Evaluation» [...]» und Döring und Bortz (2016, S.977) halten fest: «Wir verwenden den Begriff der Evaluationsforschung synonym mit wissenschaftlicher Evaluation (kurz: Evaluation) [...]».

Was beim Lesen dieses Textes vielleicht als Wortklauberei erscheint, markiert einen tiefen Umbruch im Verständnis von Evaluation. Dieser nahm Ende der 1960er-Jahre im Bildungsbereich in den USA seinen Anfang und hat dazu geführt, dass sich Evaluation zu einem eigenständigen, von der Alltagspraxis, von der Forschung und von weiteren Verfahren unterscheidbaren Ansatz der wissenschaftlichen Beschreibung und Bewertung pädagogischer und anderer komplexer Gegenstände entwickelt hat (vgl. Beywl, 1988, S.127–135). Dies geht mit eigenen Lehrbüchern, Fachzeitschriften, Berufsverbänden, Ausbildungsgängen sowie berufsethischen Grundlagen einher. Auf einen Ausschnitt davon wird in diesem Buch eingegangen.

Im vorliegenden Buch wird der Terminus «Evaluation» verwendet und von dem der «Forschung» abgegrenzt (→ Kapitel 2.2). Der Begriff «Evaluationsforschung» bleibt der Forschung *über* Bedingungen, Praxis, Methoden, Nutzung und Auswirkungen von Evaluation vorbehalten.

2

Wer? – Die Evaluierenden

Wodurch genau sich das systematische, methodische Vorgehen einer Evaluation auszeichnet, ist Kernthema dieses Buches. Die hierfür notwendigen Kompetenzen gehen über eine korrekte Anwendung der empirischen Forschungsmethoden deutlich hinaus und schließen auch persönliche und soziale Qualifikationen ein, die sich Evaluierende aneignen müssen, sodass Evaluation mehr ist als nur Sozialwissenschaft (Scriven, 2006).

Im Gegensatz zur Alltagsevaluation erfordert wissenschaftliche Evaluation spezielle Kenntnisse und Fertigkeiten: «Ein weiteres Unterscheidungsmerkmal besteht darin, dass wissenschaftliche Evaluationen von «Experten» durchgeführt werden» (Stockmann, 2004, S. 14), die für die Anforderungen besonders ausgebildet sind. Die DeGEval – Gesellschaft für Evaluation e. V. hat Anforderungen und Kompetenzen zusammengestellt, «die für die angemessene Durchführung von Evaluationen unerlässlich sind» (2008, S. 7). Der Vergleich mit einem Kompetenzprofil für Evaluation aus den USA zeigt übereinstimmende Schwerpunkte. Neben der Befähigung zur Durchführung empirischer Untersuchungen werden namentlich Kompetenzen zur Analyse von Situation und Kontext, zum Projektmanagement, zur reflexiven Praxis und im Bereich der interpersonellen und interkulturellen Kommunikation gefordert (Beywl, 2006a, S. 333–335). Eine Liste der Schweizerischen Evaluationsgesellschaft (2012), die an Evaluationsteams und nicht unbedingt an jeden einzelnen Evaluierenden gerichtet ist, formuliert ähnliche Anforderungen: Grundkenntnisse (Grundeigenschaften der Evaluation, verschiedene Evaluationsansätze, Evaluationsstandards), methodische Kenntnisse (Konzepte und Designs von Evaluationen, Methoden und Instrumenten der Datenerhebung und -analyse, Projektmanagement, Kommunikationsstrategien und -fähigkeiten) sowie soziale und diverse persönliche Kompetenzen.



→ Lösung auf Seite 227

Übungsaufgabe 1:**«Merkmale wissenschaftlicher Evaluation»**

Lösen Sie nun die Übungsaufgabe 1: Stellen Sie fest, bei welchen der folgenden Vorgehensweisen es sich zumindest näherungsweise um wissenschaftliche Evaluation handelt, also nicht um Alltagsevaluation, und begründen Sie Ihre Zuordnung kurz. Zu bewerten ist in diesem Fall ein Kurs, in dem Informatikstudierende lernen sollen, vor einem Fachpublikum zu präsentieren.

- a) Die zwanzigminütigen Präsentationen der Studierenden werden von drei verschiedenen Beobachtenden entlang eines einheitlichen, strukturierten Beobachtungsschemas protokolliert. Nachfolgend werden die festgehaltenen Beobachtungsdaten miteinander verglichen. Als Zusammenfassung werden für jede Beobachtungsdimension ein numerisch abgestuftes Urteil und eine kurze Begründung angegeben.
- b) Ein besonders erfahrener Projektmanager eines mittelgroßen Softwareanbieters hält zu jeder Studierendenpräsentation seine persönlichen Eindrücke in Stichworten fest. Er notiert hauptsächlich aussagekräftige Adjektive (z. B. «exzellent», «unvollständig»), um die Besonderheiten der jeweiligen Präsentationen zu kennzeichnen.
- c) Die PowerPoint-Präsentationen werden als Dokumente nach bestimmten Kriterien ausgewertet, z. B. Strukturiertheit des Aufbaus der gesamten Präsentation, erkennbarer Fokus der Präsentation, deutliche Botschaften an das Zielpublikum, korrekte Verwendung von Fachbegriffen. Falls möglich werden Zitate als Ankerbeispiele festgehalten. Diese werden als eine Basis für die Bestimmung des Lernerfolgs der Präsentierenden genutzt.

- d) Die zwanzigminütigen Präsentationen werden von den Zuschauenden bewertet. Dabei wird jeder/jede gefragt, wie es ihm/ihr gefallen habe, was es zu verbessern gebe und welche Anmerkungen er/sie außerdem habe.
- e) Zu jeder der zwanzigminütigen Präsentationen werden an die Zuschauenden Fragebogen ausgeteilt, die Dimensionen wie «inhaltliche Angemessenheit», «Verständlichkeit» oder «Präsentationstechnik» behandeln, bestehend aus je fünf Items mit sechs Antwortalternativen von «trifft gar nicht zu» bis «trifft voll und ganz zu». Die zusammengefassten Daten aus den eingesammelten Fragebogen werden für eine Bewertung der Präsentation genutzt.

2

2.2 Evaluation als wissenschaftliche Dienstleistung statt (Grundlagen-)Forschung

Bei aller Ähnlichkeit und auch angesichts gleicher Wurzeln unterscheidet sich Evaluation substanziell in ihrer typischen Zwecksetzung von Forschung.

Patton (2008, S. 40) bringt dies auf den Punkt. Forschung suche nach Wissen und Wahrheit, nützliche Evaluation unterstütze praktisches Handeln: «Research aims to produce knowledge and truth. Useful evaluation supports action.»

Auch Forschung verfolgt nicht selten soziale Anwendungsziele (vgl. Mertens, 2015, S. 52). Ihr zentrales Interesse besteht jedoch darin, («wahres») Wissen zu generieren und darauf aufbauend Theorien aufzustellen bzw. zu testen sowie auf möglichst unterschiedliche Personengruppen, Zeitpunkte, Situationen und geografische Räume zu verallgemeinern. Grundlagenforschung kann praktisch nützlich sein, aber ihre Nutzung ist zufällig und ungeplant: «Basic research may be useful, but its use is accidental and unplanned» (Vedung, 2004, S. 118).

Bei Evaluationen geht es hingegen selten um Theoriebildung. Es geht ihnen primär, wie auch Ansätzen der anwendungsbezogenen Forschung, um instrumentellen Nutzen für bestimmte Stakeholder in einem konkreten Kontext (vgl. Beywl, Künzli, Messmer & Streit, 2015). Das durch Informationen gesicherte Handeln in der Praxis steht im Vordergrund. Die Ergebnisse sollen binnen kurzer Frist handlungsrelevant und verwertbar sein:

«Zur Evaluation wird empirische Wissenschaft [...] durch ein spezifisches Erkenntnis- und Verwertungsinteresse»

(Kromrey, 2001, S. 112).

Generalisierbarkeit von Ergebnissen spielt bei Evaluationen selten eine Rolle, denn Nutzen soll für konkrete Maßnahmen oder Programme erzielt werden. Evaluation kann analog zur Erwachsenenpädagogik als «Handlungswissenschaft» bezeichnet werden.

SCHLÜSSELAUSSAGE

Forschung im Bereich Bildung will möglichst verallgemeinerbare Erkenntnisse über Lernendentypen, Mechanismen des Lernens, Zusammenhänge von Bildungsinstitutionen, Lernarrangements und Lernmethoden, Strategien der Bildungsfinanzierung etc. erzielen, will diese auf Theorien und Begriffssysteme abstützen und gesicherte verallgemeinerte Erkenntnis in der wissenschaftlichen Gemeinschaft der Bildungsforschenden vorantreiben.



Wofür? –
Die Evaluationszwecke

2

Bildungsevaluation will für konkrete, räumlich meist begrenzte Bildungsprogramme, -studiengänge oder -kurse bestimmter Bildungsanbieter oder -träger und auf deren Informationsbedarfe zeitgenau abgestimmte Daten, Schlussfolgerungen und Bewertungen bereitstellen, sodass jene Akteure Grundsatzentscheidungen treffen, Rechenschaft ablegen oder Optimierungsschritte einleiten können.

Im zweiten Fall ist präzise vorausgedacht, wozu und wann sowie über welche Kommunikationskanäle und Schnittstellen die Ergebnisse der Evaluation genutzt werden sollen: Leviton und Hughes (1981, S.528) nennen diese Art der vorgesehenen Nutzung «instrumentell». Auf die damit verbundenen Evaluationszwecke wird später ausführlich eingegangen (→ [Kapitel 6.1](#)).

Evaluation am Beispiel des Bildungsbereiches ist die Sammlung und Nutzung von Informationen, um über ein Bildungsprogramm Entscheidungen zu treffen, wie einer der Pioniere der Bildungsevaluation schreibt:

«[...] collection and use of information to make decisions about an educational program»
(Cronbach, 1963, S.672).

Dass die Nutzung von Evaluationsergebnissen zwar theoretisch gefordert, aber praktisch nicht immer vollzogen wird, unterstreicht Patton, indem er der allgemeinen Programmevaluation ein Konzept gegenüberstellt, welches einen vorgesehenen Evaluationsnutzen für vorgesehene Nutzende zugrunde legt:

«Utilization-focused program evaluation (as opposed to program evaluation in general) is evaluation done for and with specific, intended primary users for specific, intended uses»
(Patton, 2008, S.37).

Auf welcher Basis? – Die Evaluationskriterien

Das nachfolgende Zitat von Weiss, die zu den herausragenden Evaluationstheoretikerinnen des letzten Jahrhunderts zählt, betont die Wichtigkeit des Bewertens auf Basis von Bewertungskriterien. Sie definiert Evaluation als eine systematische Bewertung der Durchführung und/oder der Resultate eines Programms oder einer Politik auf der Basis einer Reihe von expliziten oder impliziten Normen – und dies als Mittel zur Förderung der Verbesserung des Programms oder der Politik:

«Evaluation is the systematic assessment of the operation and/or the outcomes of a program or policy, compared to a set of explicit or implicit standards, as a means of contributing to the improvement of the program or policy»
(Weiss, 1998a, S.4).

Indem Evaluation als nicht direkt zu einer Bewertung führend, sondern zunächst als die Identifikation, Klärung und Anwendung von belastbaren Kriterien definiert wird, um daraufhin den Wert (Güte und Tauglichkeit) eines Gegenstandes in Bezug auf diese Kriterien zu bestimmen, werden die Bewertungskriterien stark betont:

«[...] we define evaluation as the identification, clarification, and application of defensible criteria to determine an evaluations object's value (worth and merit) in relation to these criteria»
(Fitzpatrick, Sanders & Worthen, 2012, S.7).

Stufflebeam und Coryn definieren Evaluation als einen systematischen Prozess der Bestimmung, Beschaffung, Berichterstattung und Nutzung von beschreibenden und bewertenden Informationen. Dabei benennen sie die sieben Kriterien-dimensionen *Güte, Tauglichkeit, Integrität* (Rechtschaffenheit/Redlichkeit), *Um-*

setzbarkeit, Sicherheit, Bedeutsamkeit und/oder Gerechtigkeit, auf deren Basis die Beschreibung und Bewertung des Evaluationsgegenstandes erfolgt:

«[...] evaluation is the systematic process of delineating, obtaining, reporting and applying descriptive and judgemental information about some object's merit, worth, probity, feasibility, safety, significance and/or equity»
(Stufflebeam & Coryn, 2014, S.12).

Patton argumentiert, dass Evaluation als Wissenschaft angesehen werden kann. Das Ziel von Wissenschaft sei es zu verstehen und zu erklären, wie die Welt funktioniert. Die Besonderheit der Evaluation bestehe darin, klären zu wollen, wie und wie gut Programme, Maßnahmen oder Interventionen funktionieren, die Veränderungen – zu ergänzen wäre: Stabilisierungen – auslösen sollen. Der Evaluation als Wissenschaft liegt eine systematische Vorgehensweise zur Bestimmung von Güte, Wert, Tauglichkeit, Nutzen und Bedeutsamkeit des Evaluationsgegenstandes zugrunde, die sich an wissenschaftliche Normen hält, zu denen die Anwendung von Logik, die Verwendung transparenter Methoden, die Überprüfung der Ergebnisse und die Bereitstellung von Belegen und expliziten Begründungen gehören, um vernunftgemäße Interpretation, Bewertung und Beurteilung zu unterstützen.

«Evaluation science is systematic inquiry into how, and how well, interventions aimed at changing the world work. Evaluation science involves systematic inquiry into the merit, worth, utility, and significance of whatever is being evaluated by adhering to scientific norms that include employing logic, using transparent methods, subjecting findings to review, and providing evidence and explicit rationales to support reason-based interpretation, valuing, and judgment»
(Patton, 2018a, S. 187).

Im Vergleich von Evaluation und Forschung ergibt sich darüber hinaus ein für die Praxis höchst relevanter, geradezu dramatischer Punkt: «Grundlagenforschung darf sich «irren». Damit ist gemeint: Hypothesen, die als Ausgangspunkt gewählt werden, dürfen sich im Verlauf der Forschung als falsch erweisen. [...] Deren informationsreiches Scheitern ist nicht selten der Startpunkt für grundlegende Erkenntnisse, die eine neue Forschungslinie begründen» (Kromrey, 2003, S.98).

Eine solche, die Fehlbarkeit preisende Haltung gefährdet hingegen Legitimität und Akzeptanz von Evaluation: «Bei der Konzipierung des [...] Designs ist große Sorgfalt darauf zu verwenden, dass die zugrunde liegenden Annahmen und Hypothesen einen hohen Grad empirischer Bewährung aufweisen und dass der Prozess der Gewinnung, Auswertung und Interpretation aller Informationen methodisch abgesichert und mit begleitender Qualitätskontrolle abläuft. Jede falsche Schlussfolgerung im Verwertungskontext, die wegen fehlerhafter [...] Daten gezogen wird, hat Konsequenzen für einen nicht absehbaren Kreis von Betroffenen» (Kromrey, 2003, S.98).

Wenn also grundlegende Skepsis, Misstrauen gegenüber jedem sicheren Befund sowie ständiges und wiederholendes Infragestellen einer Vorannahme hohe Tugenden der Grundlagenforschenden sind (gegen die allerdings nicht selten in unethischer Weise verstoßen wird), würde eine solche Haltung – zumindest, wenn in übertriebenem Maße praktiziert – die Akzeptanz und das wirtschaftliche Überleben eines Evaluationsbüros gefährden. Hier deutet sich ein erstes der vielen Dilemmata an, mit denen in Evaluationen umgegangen werden muss.

Für die angemessene Planung und Durchführung einer Untersuchung ist es unverzichtbar, eine klare Priorität auf Forschung oder Evaluation zu setzen: «Eine

2

Auf wessen Initiative –
autonom oder beauftragt?

(unklare) Mischung schadet zumeist in beide Richtungen» (Reischmann, 2006, S.30).

Der Weg zu Fragestellungen ist bei Evaluation und Forschung oft sehr unterschiedlich: In der Forschung bestimmen – zumindest vom Ideal her – die Forschenden die Fragestellungen und die wissenschaftlichen Hypothesen. Auf der Suche nach Erkenntnis sollen sie allein sich selbst und den Ansprüchen ihrer Disziplin gegenüber verantwortlich sein. In der Ausrichtung von Forschungsthemen sind sie dabei nicht selten intuitiv, durch biografische oder zeitgeschichtliche Besonderheiten geleitet. Friedrichs (1973, S.50–55) nennt dies in seiner, für die sozial- und erziehungswissenschaftliche Forschung prägenden Dreigliederung den «Entdeckungszusammenhang». Diesem widmen die Forschenden relativ wenig Aufmerksamkeit, ebenso wie dem «Verwertungszusammenhang». Für sie zentral ist hingegen der «Begründungszusammenhang» mit seinen Theorien, Hypothesen, Begriffen und dem gesamten forschungsmethodischen Inventar. Was in der Forschung eine nebensächliche Aufgabe ist – die genaue Festlegung des Evaluationsgegenstandes (→ Kapitel 4) inklusive der Fragestellungen –, ist in der Evaluation die erste Kernleistung, für die genügend Ressourcen zur Verfügung stehen müssen.



→ Lösung auf Seite 228

Übungsaufgabe 2:

«Risiken pragmatischer Evaluationen»

Lösen Sie nun die Übungsaufgabe 2:

- a) Notieren Sie Gefahren und Risiken, wenn Sie sich – z. B. wegen Zeitknappheit – von wissenschaftlichen Anforderungen an Evaluation lösen und auf «Evaluationen» ausweichen, die allein pragmatischen Überlegungen folgen.
- b) Machen Sie einige Stichworte dazu, wie Sie sich – in Ihrer aktuellen oder künftigen Arbeitssituation – gegen solche Risiken schützen können.

Bedeutung des Kontextes

Bei Evaluationen läuft es oft anders ab: Sie sind zumeist Auftragsarbeiten, und ihre Zwecke und Fragestellungen sind zumindest grob vorgegeben. Sie werden nicht durch die Evaluierenden in eigener Regie nach ihrer Neugier und Kreativität konzipiert, sondern in enger Abstimmung mit den Auftraggebern und weiteren Beteiligten aus dem Bereich des Evaluationsgegenstandes. Evaluationsvorhaben finden in einem kontrollierten Rahmen als Auftrag von Entscheidungsträgerinnen und -trägern statt und sind oft in politische Settings eingebunden. Mertens hebt als Besonderheit von Evaluation hervor, dass in ihr Politik und Wissenschaft von Natur aus miteinander verflochten seien. Evaluationen werden zur Bestimmung der Güte und Tauglichkeit öffentlicher Programme durchgeführt, die wiederum selbst Antworten auf – durch politische Entscheidungen bevorzugte – individuelle und gemeinschaftliche Bedürfnisse sind: «what distinguishes evaluation from other forms of social inquiry is its political inherency; that is, in evaluation, politics and science are inherently intertwined. Evaluations are conducted on the merit or worth of programs in the public domain, which are themselves responses to prioritized individual and community needs that resulted from political decisions» (Mertens, 2015, S. 52, auf Basis von Greene, 2000). Auch die nachfolgend geschilderte Möglichkeit, eine Vielzahl an Perspektiven auf den Evaluationsgegenstand einzunehmen, indem man Beteiligte und Betroffene einbezieht, kann der Vereinseitigung einer Evaluation vorbeugen, deren Vermeidung aber nicht garantieren.

Beteiligte und Betroffene – Stakeholder – im Bildungsbereich sind z. B. Bildungsverantwortliche, Kursleitende, Bildungsteilnehmende und ihre Angehörigen/Vorgesetzten. Ihnen soll ein hoher Einfluss auf die Gegenstandsbestimmung und Themenfindung in der Evaluation eingeräumt werden. Im Vergleich zu deren Informationsinteressen spielen diejenigen der Evaluierenden keine oder höchstens eine untergeordnete Rolle.

Bereits in der ersten Evaluationsphase wird in den Blick genommen, wer wann was konkret mit den erzeugten Evaluationsergebnissen tun soll – Pattons vorgesehener Evaluationsnutzen für vorgesehene Nutzende: «intended use for intended users» (2008, S. 59). Die Ergebnisverwendung wird vorbereitet und eingeleitet in Abstimmung zwischen Evaluierenden, Auftraggebenden und Akteuren aus dem pädagogischen Feld.

Die darauffolgende Phase der Evaluation ist die Domäne der Evaluationsspezialistinnen und -spezialisten. Hier erfolgen die Auswahl der passenden Erhebungs- und Auswertungsmethoden, die Entwicklung und der Einsatz von empirischen Datenerhebungsinstrumenten, die Anwendung qualitativer und quantitativer Auswertungstechniken etc. Allerdings handeln die Evaluierenden auch hier nicht abgeschottet und isoliert von den Beteiligten und Betroffenen. Gelingende Nutzung von Evaluationsergebnissen setzt voraus, dass die Datengebenden und späteren Ergebnisnutzenden zumindest in groben Zügen verstehen, wie die Daten entstehen, und dass Transparenz und Nachvollziehbarkeit geschaffen werden.

Was zunächst unproblematisch scheint, erweist sich in der Konfrontation mit dem klassischen Verständnis von Forschung, wie es in der verbreiteten analytisch-nomologischen Wissenschaftstheorie vorliegt, als höchst problematisch: Hier beschränkt sich das «Wissenschaftliche» auf den Begründungszusammenhang, also auf Theorie und daran anschließende Empirie. Werturteile sind in diesem Verständnis aus dem Kern des Forschungsprozesses fernzuhalten. Letzterer verläuft objektiviert, ganz unabhängig von Vorlieben, Interessen sowie sozialen oder moralischen Überzeugungen der Forschenden und weiterer Mitglieder des Forschungsfeldes. Werte spielen nur im Entstehungskontext – also: «Was wird geforscht und welche Forschung wird gefördert?» – und dem Verwertungskontext – «Wer macht mit Ergebnissen von Forschung was?» – eine Rolle. Für deren Behandlung hält Forschung keine systematischen Verfahren bereit, sie gelten im Forschungsprozess als unberührbar, bleiben folglich dem Spiel politisch-gesellschaftlicher Kräfte überlassen und werden dem Bereich wissenschaftsethischer Reflexion zugewiesen.

Wenn Evaluation darauf abzielt, Gegenstände der sozialen Realität zu bewerten (also sie als mehr oder weniger «gut» einzustufen), hierfür – bereits in der Phase der Gegenstandsbestimmung (Entstehungszusammenhang) – Kriterien festzulegen (z. B. Bildungsziele, Bildungsbedarfe, vermiedene unerwünschte Nebenfolgen, Kosten-Nutzen-Relationen, Gerechtigkeit bzw. eine Kombination aus diesen Kriterien) und im Verwertungskontext dazu beizutragen, dass die erzeugten Evaluationsergebnisse in Richtung der zugrunde liegenden Werte genutzt werden, verstößt sie damit nicht diametral gegen Grundüberzeugungen wissenschaftlicher Forschung (vgl. ausführlich Kromrey, 2007a)?

Es zeigt sich ein weiterer Grundkonflikt: Evaluation will, soll und muss bewerten (sonst ist sie keine Evaluation). Sie überschreitet dabei bewusst das sichere Terrain des Begründungszusammenhangs und verwendet systematische Verfahren auch für die Festlegung von Evaluationsfragestellungen und Bewertungskriterien. Sie will die Erarbeitung von Interpretationen und Bewertungen transparent

2

Aktive Stakeholder als Partner

Wie verwenden? – Nutzung der Evaluationsergebnisse

Wertfreiheit versus Bewerten als zentrale Aufgabe

Keine Evaluation ohne Bewertung!

2

und nachvollziehbar, auf Daten basierend und die verschiedenen Perspektiven der Beteiligten und Betroffenen einbeziehend gestalten. Der Anspruch des systematischen, regelgeleiteten, wissenschaftlichen Vorgehens wird ausgedehnt vom Begründungszusammenhang auf den vorgelagerten Entdeckungszusammenhang und den nachgelagerten Verwertungszusammenhang. Dabei droht, dass Evaluation an distanzierter Objektivität verliert. Ein weiteres Dilemma, dem sich Evaluation stellen muss.



→ Lösung auf Seite 229

Übungsaufgabe 3:

«Bewertungsmaßstäbe verschiedener Beteiligter»

Lösen Sie nun die Übungsaufgabe 3: Versetzen Sie sich in die folgende Situation und schreiben Sie auf, nach welchen Bewertungsmaßstäben aus der Sicht der verschiedenen Beteiligten diese Bildungsmaßnahme zu bewerten sein könnte. Ergänzen Sie z. B. folgenden Satz: «Für Person x ist besonders wichtig, dass ...»

Die Volkshochschule in Musterstadt bietet einen Kurs «Internet für Einsteiger» an. Die Teilnehmerschaft, die sich zur ersten Unterrichtsstunde im Kursraum einfindet, ist sehr heterogen. Da sitzt Frau Wunder, die mit ihren 70 Jahren einfach einmal wissen möchte, was es mit diesem ominösen Internet auf sich hat. Sie hat schon viel darüber gelesen, aber wirklich vorstellen kann sie sich nichts darunter.

Mitgebracht hat sie ihren Enkel Lukas, der schon toll mit seinem Tablet umgehen kann, dessen Eltern aber wünschen, dass er sich einmal systematisch mit diesem Thema auseinandersetzt.

Herr Mazzini wurde von seinem Chef geschickt. Er soll über kurz oder lang die Firmenhomepage betreuen, die schon eine ganze Weile nicht mehr aktualisiert worden ist.

Frau Schulze hat ein eigenes Geschäft, und sie sucht nach Möglichkeiten, neue Kundenkreise zu erschließen. Ihr wurde geraten, sich einmal intensiv mit den Möglichkeiten des Internets auseinanderzusetzen.

Herr Scheid ist ebenfalls Senior, und dieser neumodische Kram war ihm schon immer suspekt. Da will er doch einmal sehen, was das für Leute sind, die sich stundenlang vor den Computer setzen wollen, und welche vielleicht unlaute- ren Beweggründe dahinterstecken.

Frau Özoğuz ist die Dozentin im Kurs. Sie hat viele Jahre Erfahrung im IT-Unterricht und hat auch schon Senioren- und Kinderkurse gegeben. Am meisten Spaß macht es ihr aber, wenn alle zusammen bei ihr im Unterricht sitzen. Diese Heterogenität sei eine ganz besondere Herausforderung, sagt sie.

Herr Felix hat nichts direkt mit dem Unterrichtsgeschehen zu tun, aber als IT-Verantwortlicher in der Volkshochschule Musterstadt ist er für ein kundengerechtes IT-Angebot verantwortlich. Für ihn sind es schwierige Zeiten, denn die Belegungszahlen sind schon seit einigen Semestern rückläufig. Die goldenen Computerzeiten im Schulungsbereich scheinen vorbei zu sein.