

Teil V

Ausgewählte Tätigkeitsfelder

19 Diagnostik und Evaluation

Martin Brunner • Petra Stanat • Hans Anand Pant

19.1 Diagnostik

- 19.1.1 Begriffe, Anwendungsfelder und Definition
- 19.1.2 Statistische Kennwerte
- 19.1.3 Prozessablauf
- 19.1.4 Gütekriterien von Testwerten und Auswahlkriterien von diagnostischen Instrumenten
- 19.1.5 Diagnostik im schulischen Alltag von Lehrkräften

19.2 Evaluation

- 19.2.1 Begriffe, Funktionen und Beispiele
- 19.2.2 Ablauf
- 19.2.3 Qualitätsstandards zur Durchführung von Evaluationsstudien
- 19.2.4 Die Konzeption von Evaluationsstudien
- 19.2.5 Interpretation und Kommunikation der Ergebnisse

Die Ergebnisse von PISA 2000, die im Jahr 2001 publiziert wurden (Baumert et al., 2001), haben in Deutschland für Aufregung gesorgt. Wider Erwarten konnten sehr viele Jugendliche in Deutschland nicht gut verstehend lesen und hatten Schwierigkeiten, die in der Studie eingesetzten Aufgaben in Mathematik und in den Naturwissenschaften zu lösen. PISA 2000 zeigte auch, dass in Deutschland die Leistungsschere zwischen Jugendlichen aus sozioökonomisch gut gestellten und sozioökonomisch weniger privilegierten Familien besonders weit aufging. So war der Zusammenhang zwischen dem erreichten Kompetenzniveau und dem sozioökonomischen Status der Familie in kaum einem PISA-Teilnehmerstaat enger als in Deutschland. Schließlich wies PISA 2000 darauf hin, dass schwache Lesekompetenz bei Jugendlichen in Hauptschulbildungsgängen häufig unerkannt blieb. Dieses Ergebnis warf die Frage auf, wie gut Lehrkräfte die Leistungen ihrer Schüler diagnostizieren können.

Die Ergebnisse groß angelegter Schulleistungsstudien wie PISA ermöglichen es, Bildungssysteme quasi aus der Vogelperspektive zu betrachten und ganze Staaten hinsichtlich ihrer Bildungserträge miteinander zu vergleichen. Um jedoch mögliche Ursachen der Ergebnisse eines solchen Bildungsmonitorings zu bestimmen und identifizierten Schwächen zu begegnen, ist es erforderlich, Lehr-Lern-Prozesse und deren Effekte auch auf anderen Ebenen zu untersuchen. Wie können Lehrkräfte die Fähigkeiten ihrer Schüler diagnostizieren, um Förderbedarfe oder besondere Begabungen zu identifizieren? Welche Programme zur Leseförderung sind tatsächlich wirksam? Ist die Ganztagschule als Organisationsform besser geeignet, um gute Schülerleistungen zu erzielen, als die Halbtagschule? Um diese und zahlreiche andere praxisrelevante Fragen beantworten zu können, sind professionelle pädagogisch-psychologische Diagnostik und wissenschaftliche Evaluationen in pädagogischen Handlungsfeldern unerlässlich.



Titelbild des SPIEGEL am 10. 12. 2001 zur Bekanntgabe der Ergebnisse von PISA 2000

19.1 Diagnostik

19.1.1 Begriffe, Anwendungsfelder und Definition

In diesem Kapitel beschäftigen wir uns mit zentralen Konzepten und Methoden pädagogischer Diagnostik, die wir häufig mit Rückgriff auf das nachfolgende Beispiel erläutern.

Beispiel

Stellen Sie sich vor, Sie würden im psychologischen Dienst für Grundschulen eines Stadtbezirks arbeiten. Ein Elternpaar kommt zu Ihnen mit seinem 8-jährigen Sohn Karl, der die 2. Klasse besucht. Die Eltern sind sehr besorgt, denn Karl hat große Schwierigkeiten beim Lesen und ist allgemein oft unkonzentriert. Ausgehend vom Erfahrungsbericht seiner Eltern, aber auch nach Rücksprache mit seiner Klassenlehrerin, ziehen Sie als eine mögliche Erklärung für die Lese-schwierigkeiten von Karl in Betracht, dass er eine Lese-Rechtschreib-Schwäche (LRS) haben könnte.

Neben der Diagnose von Lernschwächen (wie z. B. LRS oder spezifische Lernschwierigkeiten in Mathematik) spielt pädagogisch-psychologische Diagnostik in vielen anderen Bereichen eine wichtige Rolle. Dazu zählen u. a. die Diagnose von Verhaltensauffälligkeiten oder auch von Hochbegabung, die Hochschulzulassung und Studierendenauswahl sowie die Schul- und Berufslaufbahnberatung oder sozialpädagogische Beratung (z. B. in der Jugendhilfe). Diagnostische Tätigkeiten in diesen Kontexten, die mit wissenschaftlich fundierten Verfahren arbeiten, lassen sich durch eine gemeinsame Definition charakterisieren (vgl. Hasselhorn & Gold, 2009; Leutner, 2006). Die diagnostischen Tätigkeiten von Lehrkräften in der Schule (z. B. bei der Notengebung oder bei der Empfehlung für eine weiterführende Schule) betrachten wir gesondert (s. Abschn. 19.1.5), da sie nicht in vollem Umfang dieser Definition entsprechen.

Definition

Pädagogisch-psychologische Diagnostik nutzt Konzepte und Methoden der psychologischen Diagnostik, um in systematischer und möglichst akkurater Weise Informationen über Personen zu sammeln, aufzubereiten und zu bewerten. Diese Informationen dienen als Grundlage für Entscheidungen in Bezug auf Einzelpersonen in pädagogischen Handlungsfeldern, in denen es um das Lernen, Lehren und Erziehen geht.

Einige Bestimmungsstücke und Konsequenzen dieser Definition heben wir nochmals hervor. Pädagogisch-psychologische Diagnostik ...

- ▶ findet im Kontext pädagogischer Handlungsfelder statt. Zu diesen Handlungsfeldern gehören u. a. Familien, Kindergärten, Schulen, Universitäten, Beratungsstellen (z. B. in der Jugendhilfe) und Institutionen der beruflichen Aus-, Fort- und Weiterbildung sowie Erwachsenenpädagogik.
- ▶ bezieht sich auf Einzelpersonen. Die Erfassung verschiedener Merkmale von Lernumwelten, von pädagogischen Institutionen oder Bildungssystemen betrachten wir als Gegenstand von Evaluation.
- ▶ nutzt die Methoden der Psychologie. Hierzu gehören Tests, Fragebögen, Verhaltens- und Beobachtungsinventare, psychophysische Apparate, neurowissenschaftliche Verfahren (z. B. funktionelle Magnetresonanztomografie, fMRT) sowie die systematische Analyse biografischer Daten.
- ▶ sammelt Informationen über Personen, nicht nur um deren Verhalten und Erleben zu beschreiben, sondern auch um es zu erklären und vorherzusagen. Hierzu ist es in der Regel erforderlich, die relevanten Bedingungen und Situationen zu identifizieren, in denen eine bestimmte Person ein bestimmtes Verhalten zeigt. So ist es z. B. bei der Erklärung von aggressivem Verhalten eines Kindes wichtig festzustellen, gegen wen oder was sich die Aggression richtet, welche Situationen Anlass zum aggressiven Verhalten geben und wie betroffene Personen auf das aggressive Verhalten reagieren.

- ▶ hilft bei der Entscheidungsfindung, und zwar bezogen auf praktische Probleme in pädagogischen Handlungsfeldern (z. B. hinsichtlich der Frage, ob Karl zusätzliche Förderung in den Bereichen Lesen und Rechtschreibung benötigt). Pädagogisch-psychologische Diagnostik kann weitreichende Auswirkungen haben (z. B. auf die schulische und berufliche Zukunft von Karl). Sie wird grundsätzlich nicht als Selbstzweck angewendet, sondern im Dienste der Optimierung von wichtigen Entscheidungen unter Unsicherheit.
- ▶ unterscheidet sich in mehreren Aspekten von typischer Alltagsdiagnostik, wie sie z. B. häufig von Lehrkräften im schulischen Alltag praktiziert wird (vgl. Wild & Krapp, 2006). Zu den Unterscheidungsmerkmalen gehören der Rückgriff auf wissenschaftlich verankerte Konstrukte bei der Beschreibung und Erfassung von Personenmerkmalen, die Anwendung von wissenschaftlich fundierten Messverfahren (zur Verringerung von Mess- und Prognosefehlern), das systematische Vorgehen im diagnostischen Prozess (z. B. die Systematik der Datengewinnung und das Prüfen von Hypothesen) sowie die Transparenz des diagnostischen Prozesses (durch die Berichtlegung in Form von Gutachten und die Offenlegung potenzieller Einschränkungen bei der Interpretation der Daten).

Diagnostische Zielsetzungen

Ausgehend von der Definition pädagogisch-psychologischer Diagnostik können in enger Anlehnung an die einflussreiche Arbeit von Pawlik (1976) verschiedene diagnostische Zielsetzungen unterschieden werden. Diese lassen sich anhand von zwei Dimensionen beschreiben:

- (1) Selektions- vs. Modifikationsdiagnostik und
- (2) Status- vs. Prozessdiagnostik.

Selektions- vs. Modifikationsdiagnostik. Selektionsdiagnostik in pädagogischen Handlungsfeldern bildet die Grundlage für Entscheidungen, die auf die Auswahl (a) der richtigen Person für eine bestimmte Lernumwelt oder (b) der richtigen Lernumwelt für eine bestimmte Person abzielen. Ein Beispiel für die Personenselektion ist die Hochschulzulassung, bei der es darum geht, die am besten geeigneten Bewerber für eine begrenzte Anzahl von Studienplätzen in einem bestimmten Fach auszuwählen, sofern die Anzahl der Bewerber die Zahl der verfügbaren Plätze übersteigt (ansonsten würde sich der finanzielle, zeitliche und personelle Aufwand für die

Selektionsdiagnostik nicht lohnen). Ein Beispiel für die Auswahl einer Lernumwelt ist die Auswahl einer bestimmten Berufsausbildung aus einer Vielzahl von Möglichkeiten im Rahmen der Berufslaufbahnberatung. Ein weiteres Beispiel für die Wahl einer Lernumwelt ist die Wahl einer Schule, die sich auf die Förderung von Kindern mit LRS spezialisiert hat.

Modifikationsdiagnostik in pädagogischen Handlungsfeldern bildet die Grundlage für Entscheidungen, die auf Veränderungen (Modifikation) von (a) Verhalten oder psychologischen Merkmalen einer Person oder (b) pädagogischen Lernumwelten bzw. Lernsituationen abzielen. Ein Beispiel für die Modifikation von Personen ist die systematische Förderung der Leseleistung von Karl. Eine Fördermaßnahme könnte etwa darauf abzielen, Karl Strategien zu vermitteln, die es ihm ermöglichen, Texte zu dekodieren und verstehend zu erfassen. Eine Veränderung einer Lernumwelt wäre z. B. die Einrichtung zusätzlicher Förderkurse für lese-/rechtsschreibschwache Schüler in einer Schule.

Status- vs. Prozessdiagnostik. Eng verbunden mit der Unterscheidung von Selektions- und Modifikationsdiagnostik ist die Unterscheidung von Status- und Prozessdiagnostik. Die Aufgabe von Statusdiagnostik besteht darin, die Ausprägung des individuellen Verhaltens oder Erlebens einer Person zu einem bestimmten Zeitpunkt möglichst präzise zu erfassen. Ein Beispiel hierfür ist die Bestimmung des allgemeinen Intelligenzniveaus von Karl, um auszuschließen, dass seine Lese-schwierigkeiten auf gering ausgeprägte allgemeine kognitive Grundfähigkeiten zurückzuführen sind. Prozessdiagnostik hingegen zielt darauf ab, spontane oder gezielt herbeigeführte Veränderungen individuellen Verhaltens oder Erlebens über die Zeit hinweg zu messen, um diese Veränderungen sichtbar zu machen. So würde man Karls Leseleistung wiederholt erfassen, um den Erfolg der gewählten Fördermaßnahme zu bestimmen und die Förderstrategie ggf. anzupassen.

Die Beschreibung der verschiedenen diagnostischen Strategien verdeutlicht, dass mit ihrer Anwendung eine Annahme darüber verbunden ist, inwieweit das jeweilige Verhalten oder Erleben zeitlich (weitgehend) stabil ist oder nicht. Zeitlich relativ stabile Verhaltens- und Erlebensweisen werden in der pädagogisch-psychologischen Forschung als Dispositionen oder Eigenschaften (Traits) bezeichnet. Selektions- und Statusdiagnostik setzen häufig (weitgehende) zeitliche Stabilität der gemessenen Eigenschaften voraus und nehmen an, dass

anhand dieser Eigenschaften zukünftiges Verhalten vorhergesagt werden kann. Im Unterschied hierzu gehen Modifikations- und Prozessdiagnostik von der grundsätzlichen Veränderbarkeit des individuellen Verhaltens und Erlebens durch Lern-, Bildungs- oder Erziehungsprozesse aus. Ein Schwerpunkt dieser Diagnostik liegt daher auch auf der Bestimmung von (situativen) Faktoren, die ein bestimmtes Verhalten oder Erleben bedingen, um Ansatzpunkte für konkrete Fördermaßnahmen zu erhalten.

Mischformen. Die Unterscheidung von Selektions- vs. Modifikationsdiagnostik bzw. Status- vs. Prozessdiagnostik ist hilfreich, um den Fokus einer diagnostischen Situation zu charakterisieren. In pädagogischen Handlungsfeldern lassen sich jedoch diagnostische Zielsetzungen nicht immer trennscharf den jeweiligen Dimensionen zuweisen, sondern es handelt sich meist um Mischformen (Leutner, 2006). Die Ursache hierfür liegt in der Kernannahme pädagogischen Handelns, dass individuelles Verhalten und Erleben durch Lern-, Bildungs- oder Erziehungsprozesse veränderbar ist. Beispielsweise kann die Diagnose einer LRS für Karl die Entscheidungsgrundlage dafür bilden, eine geeignete Lernumwelt auszuwählen (z.B. die Auswahl einer Schule, die sich auf die Förderung von Kindern mit Lese-Rechtschreib-Schwierigkeiten spezialisiert hat). Gleichzeitig sollten diagnostische Informationen dazu genutzt werden, die Fördermaßnahmen auf Karls spezifische Schwierigkeiten beim Lesen wie auch auf seinen Lernfortschritt abzustimmen, um ihn so optimal zu fördern. Die Lernumwelt wird hierbei also kontinuierlich aufgrund dieser Informationen modifiziert.

In diesem Zusammenhang ist es auch wichtig zu unterstreichen, dass dasselbe Merkmal aus unterschiedlichen Perspektiven betrachtet werden kann und nicht per se einer bestimmten diagnostischen Zielsetzung zuzuordnen ist. Die Leseleistung ist hierfür ein gutes Beispiel, denn die Fähigkeit zum Lesen basiert auf einem langfristigen, kumulativen Lernprozess. Ab einem bestimmten Alter kann man von einer weitgehenden Stabilität der Lesekompetenz ausgehen. Im Rahmen ei-

ner Statusdiagnostik für Schullaufbahnnempfehlungen am Ende der Primarschule stellt die Leseleistung damit ein potenziell geeignetes Kriterium dar, da eine solch weitgehende Entscheidung eine (zumindest mittelfristige) zeitliche Stabilität des Merkmals voraussetzt. Andererseits kann die Lesekompetenz durch Bildungsprozesse aber auch gefördert werden. Wird etwa im Fall von Karl entschieden, eine gezielte Förderung der Lesekompetenz durchzuführen, wäre eine kontinuierliche Prozessdiagnostik der Leseleistung erforderlich, um die Maßnahmen auf seine spezifischen Stärken und Schwächen abzustimmen. Die Unterscheidung von Status- vs. Prozessdiagnostik bezieht sich also auf die Ziele, die mit einer diagnostischen Entscheidung verbunden sind; sie ist nicht an das zu diagnostizierende Merkmal (wie etwa die Leseleistung) gebunden.

19.1.2 Statistische Kennwerte

Um die Logik pädagogisch-psychologischer Diagnostik und Evaluation verstehen zu können, ist die Kenntnis

a. Mittelwert

$$M_L = \frac{\sum_{i=1}^n L_i}{n} = \frac{451}{15} = 30,1$$

b. Standardabweichung

$$S_L = \sqrt{\frac{\sum_{i=1}^n (L_i - M_L)^2}{n}} = \sqrt{\frac{1402,9}{15}} = 9,7$$

c. z-Wert

$$z_{L,1} = \frac{L_i - M_L}{S_L} = \frac{35 - 30,1}{10,0} = 0,49$$

d. Korrelation

$$r_{L,L} = \frac{\sum_{i=1}^n z_{L_i} \cdot z_{L_i}}{n - 1} = \frac{9,89}{14} = 0,706$$

e. 95%-Konfidenzintervall

Untergrenze:

$$95\% - KI_i = X_i - 1,96 \cdot S_x \cdot \sqrt{1 - r_{tt}}$$

Obergrenze:

$$95\% - KI_i = X_i + 1,96 \cdot S_x \cdot \sqrt{1 - r_{tt}}$$

f. Cohens d

$$d = \frac{M_1 - M_2}{\sqrt{\frac{n_1 \cdot S_1^2 + n_2 \cdot S_2^2}{n_1 + n_2}}}$$

Abbildung 19.1 Formeln zur Berechnung wichtiger statistischer Kenngrößen. M = Mittelwert; S = Standardabweichung; z = z -standardisierter Testwert; r = Korrelationskoeffizient; i = Personenindex; Σ = Summe; X_i = Testwert von Person i ; r_{tt} = Reliabilität der Testwerte eines bestimmten Tests; d = Cohens d als standardisiertes Effektstärkemaß für Mittelwertsunterschiede für ein bestimmtes Maß zwischen Gruppe 1 und Gruppe 2 mit den Gruppengrößen n_1 und n_2 . Die Zahlenbeispiele beziehen sich auf Tabelle 19.1

einiger statistischer Konzepte und Kenngrößen erforderlich, die wir an einem Beispiel erläutern. Angenommen eine Grundschullehrkraft hat einen Intelligenztest (mit den Testwerten I) und einen Lesetest (mit den Testwerten L) in ihrer Klasse mit 15 Grundschulkindern eingesetzt. Tabelle 19.1 führt die in den Tests erreichten Punkte der Kinder auf.

Die Lehrkraft möchte nun die Verteilungen der Testwerte beschreiben und wissen, ob ein Zusammenhang zwischen den Testwerten im Intelligenz- und Lesetest besteht. In Abbildung 19.1 werden die mathematischen Formeln zur Berechnung der statistischen Kennwerte aufgeführt, die sie hierfür benötigt. (Das Konfidenzintervall bzw. das Effektstärkemaß Cohens d , die ebenfalls in Abbildung 19.1 aufgelistet sind, besprechen wir weiter unten in Abschn. 19.1.3 bzw. Abschn. 19.2.5). Als Maß zur Beschreibung der zentralen Tendenz der Verteilung

von I und L berechnet die Lehrkraft die Mittelwerte $M_I = 20,1$ und $M_L = 30,1$. Als Maße für die Leistungsheterogenität der Kinder berechnet sie die Streuung der Testwerte I bzw. L in Form der Standardabweichung S_I bzw. S_L . Die Standardabweichung gibt an, wie stark die Testwerte im »Mittel« um den Mittelwert der Testwertverteilung streuen. Zu beachten ist hierbei (daher steht Mittel in Anführungszeichen), dass die Testwerte nicht mit gleichem Gewicht in die Berechnung der Standardabweichung eingehen: Testwerte, die weiter entfernt vom Mittelwert der Verteilung liegen, gehen durch das Quadrieren der Differenzen stärker gewichtet ein als Testwerte, die näher am Mittelwert der Verteilung liegen (s. Abb. 19.1 b). Im Beispiel liegen die Streuungswerte im Intelligenztest bei etwa $S_I = 9,6$ Punkten und im Lesetest in etwa bei $S_L = 9,7$ Punkten.

Tabelle 19.1 Datenbeispiel für Testwerte in einem Intelligenztest (I) und in einem Lesetest (L)

Kind	Punkte im Intelligenztest (I)	Punkte im Lesetest (L)	$(I_i - M_I)^2$	$(L_i - M_L)^2$	Z_I	Z_L	$Z_I \times Z_L$
1	29	35	79,8	24,3	0,90	0,49	0,44
2	16	18	16,5	145,6	-0,41	-1,21	0,50
3	15	26	25,7	16,5	-0,51	-0,41	0,21
4	25	41	24,3	119,5	0,50	1,09	0,54
5	25	30	24,3	0,0	0,50	-0,01	0,00
6	15	42	25,7	142,4	-0,51	1,19	-0,61
7	34	41	194,1	119,5	1,41	1,09	1,54
8	12	24	65,1	36,8	-0,82	-0,61	0,49
9	5	15	227,0	227,0	-1,52	-1,51	2,29
10	1	17	363,5	170,7	-1,93	-1,31	2,52
11	33	46	167,3	253,9	1,31	1,59	2,08
12	25	19	24,3	122,5	0,50	-1,11	-0,55
13	26	34	35,2	15,5	0,60	0,39	0,24
14	27	33	48,1	8,6	0,70	0,29	0,21
15	13	30	49,9	0,0	-0,71	-0,01	0,00
Summe	301	451	1370,9	1402,9	0,00	0,00	9,89
M	20,1	30,1			0,0	0,0	
S	9,6	9,7			1,0	1,0	

M = Mittelwert; S = Standardabweichung; z = z-standardisierter Testwert; i = Personenindex

Wenn die Lehrkraft daran interessiert ist, die Testwerte der einzelnen Kinder in Beziehung zur jeweiligen Verteilung der Testwerte in ihrer Klasse zu setzen, kann sie hierfür z -Werte berechnen. Ein z -Wert gibt an, um wie viele Standardabweichungen und in welcher Richtung ein Messwert vom Mittelwert der Vergleichsstichprobe – in diesem Fall der Klasse – abweicht. Ein Vorzug der z -Werte ist, dass dadurch Testwerte aus Verteilungen mit unterschiedlichen Mittelwerten und Streuungen in Bezug auf ihre relative Abweichung vom Mittelwert vergleichbar werden; die z -Werte weisen immer einen Mittelwert von 0 und eine Standardabweichung von 1 auf. Ist der z -Wert 0, so entspricht der Wert der Person genau dem Mittelwert der Vergleichsstichprobe. Bei einem z -Wert größer 0 ist ihre Leistung überdurchschnittlich, bei einem z -Wert kleiner 0 unterdurchschnittlich ausgeprägt. Kind 2 z.B. hat einen z -Wert von $-1,21$ im Lesetest erreicht. Dies bedeutet, dass seine Leistung um 1,21 Standardabweichungen unter dem Durchschnitt der Grundschulkinder in dieser Klasse liegt.

Um zu erfahren, ob ein Zusammenhang zwischen den Testwerten im Intelligenz- und Lesetest besteht, erstellt die Lehrkraft zunächst ein Streudiagramm (s. Abb. 19.2). Man sieht, dass Kinder mit besserer Leistung im Intelligenztest tendenziell auch bessere Leistungen im Lesetest erzielen.

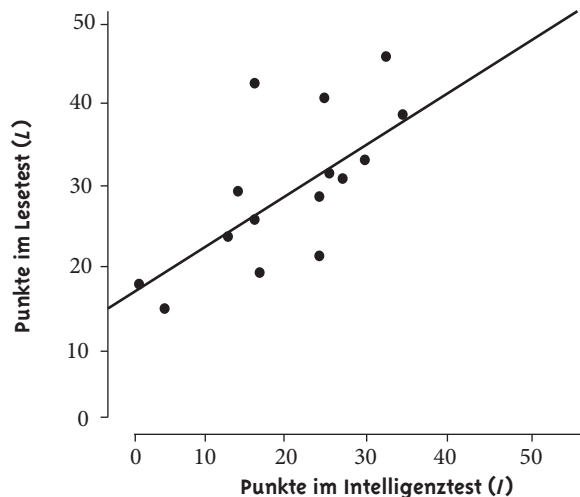


Abbildung 19.2 Streudiagramm zur Darstellung des Zusammenhangs zwischen den Testwerten im Intelligenztest und im Lesetest, die in Tabelle 19.1 aufgeführt sind. Die Gerade stellt den linearen Zusammenhang zwischen den beiden Variablen dar

Um den linearen Zusammenhang zwischen den Testwerten im Intelligenz- und Lesetest zu quantifizieren, berechnet man deren (Pearson-Produkt-Moment-) Korrelation. Die Korrelation ist ein statistisches Zusammenhangsmaß, das auch bei der Bestimmung der Gütekriterien diagnostischer Verfahren eine zentrale Rolle spielt (s. Abschn. 19.1.4). Der Korrelationskoeffizient r kann zwischen -1 und $+1$ variieren. Beispielsweise kennzeichnet ein Korrelationskoeffizient von $r=0,71$ einen hohen positiven linearen Zusammenhang zwischen I und L . Hoch bedeutet hierbei, dass man aus der Kenntnis von I bei einem Kind mit hoher »Treffsicherheit« auf seine Ausprägung von L schließen kann. Positiv bedeutet hier, dass mit höheren Werten im Intelligenztest tendenziell höhere Werte im Lesetest einhergehen (ein negativer Korrelationskoeffizient würde anzeigen, dass mit höheren Werten im Intelligenztest tendenziell niedrigere Werte im Lesetest korrespondieren). Bei einer Korrelation von $r=0$ besteht kein Zusammenhang zwischen den Merkmalen. Linear schließlich bedeutet, dass der statistische Zusammenhang in Form einer Geraden abgebildet werden kann (vgl. Abb. 19.2); nicht-lineare (z.B. U-förmige) Zusammenhangsmuster zwischen I und L , die möglicherweise in den Daten vorliegen, werden durch den Korrelationskoeffizienten nicht erfasst.

19.1.3 Prozessablauf

Pädagogisch-psychologische Diagnostik ist in einen Prozess eingebunden, der mehrere Phasen umfasst (s. Abb. 19.3). Auswahlkriterien für die diagnostischen Instrumente werden in Abschnitt 19.1.4 besprochen.

Definition des Entscheidungsproblems

Um Entscheidungen im Rahmen pädagogisch-psychologischer Diagnostik zu treffen, ist es zunächst erforderlich, das Entscheidungsproblem und die damit verbundenen diagnostischen Zielsetzungen zu definieren: Geht es um Selektion oder Modifikation von Personen? Ist es eine Status- oder eine Prozessdiagnostik angezeigt? Müssen mehrere Strategien kombiniert werden? Hierbei ist es besonders wichtig, die Fragestellungen in Form von Hypothesen so zu konkretisieren, dass sie im Rahmen des diagnostischen Prozesses untersucht und beantwortet werden können. Im Beispiel von Karl wird die allgemeine Fragestellung zunächst von den Eltern definiert: Warum hat Karl solche Probleme mit dem Lesen

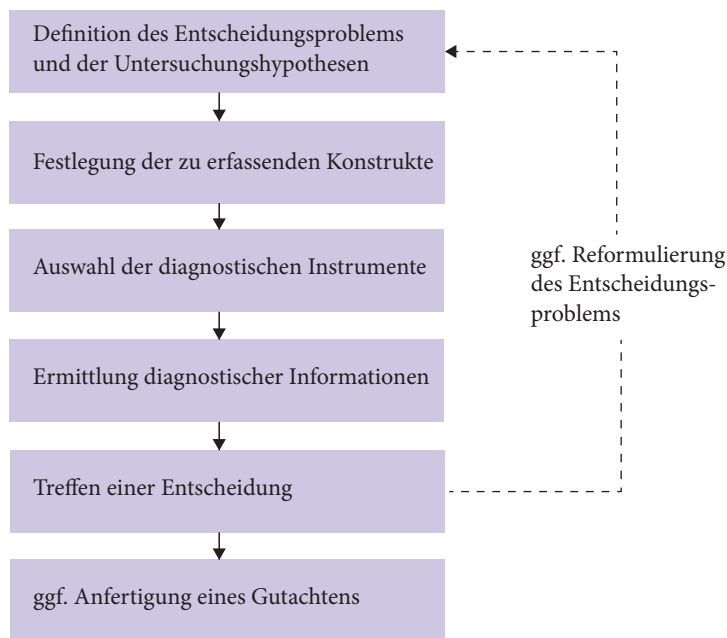


Abbildung 19.3 Prozessablauf pädagogisch-psychologischer Diagnostik

und warum ist er oft unkonzentriert? Diese Fragestellung muss im diagnostischen Prozess zunächst anhand von Informationen zur Vorgeschichte spezifiziert werden, die zu der Vermutung führen könnten, dass bei Karl möglicherweise eine LRS vorliegt. Wenn eine LRS vorliegt, ist die Entwicklung der Lesefertigkeiten und – damit sehr häufig verbunden – die Entwicklung der Rechtschreibung beeinträchtigt.

Für die Diagnose einer Lese-Rechtschreib-Schwäche müssen gemäß den »Leitlinien zu Diagnostik und Therapie von psychischen Störungen im Säuglings-, Kindes- und Jugendalter«, die von der Deutschen Gesellschaft für Kinder- und Jugendpsychiatrie, Psychosomatik und Psychotherapie (DGKJP) und Ärzteverbänden herausgegeben wurden (2007), drei Kriterien erfüllt sein:

- (1) Die Leistung im Lesekompetenztest fällt in der Vergleichsgruppe in den Bereich der unteren 10 % (etwa 90 % der Kinder in der Vergleichsgruppe sollten also eine bessere Leistung als Karl erzielen).
- (2) Das Intelligenzniveau liegt nicht im Bereich der geistigen Behinderung (d.h., der Intelligenzquotient [IQ] liegt über 70).

- (3) Die Lese- oder Rechtschreibleistung sollte deutlich schlechter sein, als dies nach der allgemeinen Intelligenzentwicklung zu erwarten ist.

Nach diesen Leitlinien wäre ab einem bestimmten Schweregrad – gemäß dem internationalen Klassifikationssystem medizinischer Diagnosen ICD-10 – eine »Lese- und Rechtschreibstörung« (F81.0) zu diagnostizieren. Dieser Begriffswahl folgen wir hier nicht, da der Begriff »Störung« eng mit dem Begriff »Krankheit« verbunden und umstritten ist.

Um festzustellen, ob bei Karl eine LRS vorliegt, müssen die drei genannten Kriterien in Form von Hypothesen überprüft werden: Karls Leseleistung gehört zu den unteren 10 % (Hypothese 1) und Karls IQ liegt über 70 (Hypothese 2). Die dritte Hypothese zur Diskrepanz zwischen Leseleistung und Intelligenzentwicklung klammern wir hier aus, da dieses Diskrepanzkriterium sehr kritisch diskutiert wird, u. a. weil es für die Frage, wie gefördert werden sollte, kaum relevant ist

(Stuebing et al., 2002). Wichtig für die Diagnosestellung ist ferner, neben kognitiven Beeinträchtigungen weitere mögliche Alternativerklärungen auszuschließen, wie etwa Störungen der Hör- und Sehfunktionen sowie neurologische oder allgemeine psychische Störungen (s. hierzu ausführlich Weber & Marx, 2008).

Festlegung der Konstrukte

Bei der Definition und Konkretisierung der Fragestellung wird gleichzeitig festgelegt, welche Konstrukte (z. B. die Lesekompetenz oder die allgemeine Intelligenz) zu erfassen sind.

Definition

Im Kontext der psychologischen Diagnostik bezeichnen **Konstrukte** einen klar definierten Gegenstandsbereich zeitlich stabiler oder zeitlich veränderlicher Merkmale des individuellen Verhaltens und Erlebens.

Sie unterscheiden sich in ihrer Breite, also der Anzahl und Vielfalt der Verhaltens- und Erlebensweisen, die einem Konstrukt zugeordnet werden. Konstrukte unterscheiden

sich auch in der Organisation ihres Gegenstandsbereichs, also darin, ob verschiedene Facetten des Konstrukts differenziert werden können, sowie ob und in welcher Form diese Facetten hierarchisch organisiert sind.

Beispielsweise ist Intelligenz ein sehr breites kognitives Konstrukt, das viele Facetten umfasst und hierarchisch organisiert ist (s. z. B. Deary, 2001 zur Einführung in die Intelligenzforschung; vgl. Abschn. 7.3.1). An der Spitze der Hierarchie steht die allgemeine Intelligenz. Facetten der Intelligenz, die der allgemeinen Intelligenz untergeordnet sind, sind kognitive Operationen wie das schlussfolgernde Denken, die Merkfähigkeit oder die mentale Geschwindigkeit. Ein im Vergleich dazu einfacher strukturiertes Konstrukt ist das Leseverstehen, das sich auf das Verstehen von Wörtern und Sätzen sowie kurzen und längeren Texten bezieht (vgl. z. B. Schiefele, Artelt, Schneider & Stanat, 2004).

Im Fall von Karl besteht das diagnostische Ziel darin zu untersuchen, ob eine LRS vorliegt. Entsprechend der oben beschriebenen Definition müssen dazu sowohl Karls Niveau der allgemeinen Intelligenz als auch seine Lesekompetenz erfasst werden.

Auswahl der diagnostischen Instrumente

Konstrukte (wie die Intelligenz oder die Lesekompetenz) sind nicht direkt beobachtbar, sondern müssen mittels diagnostischer Messinstrumente erfasst und damit »sichtbar« gemacht werden (vgl. American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999). Zur Erfassung von Konstrukten ist es deshalb notwendig, geeignete diagnostische Instrumente auszuwählen (z. B. Tests, Fragebögen, Verhaltens- und Beobachtungsinventare, psychophysische Apparate, Leitfäden für biografische Interviews; vgl. Abschn. 4.4), um die für die Entscheidung erforderlichen diagnostischen Informationen zu gewinnen. Im Idealfall handelt es sich um quantitative Aussagen im Sinne von Testwerten (*test scores*). Hierzu ist anzumerken, dass in der pädagogisch-psychologischen Diagnostik der Begriff »Testwert« zumeist allgemein verwendet wird, um Messwerte zu bezeichnen, die mit diagnostischen Instrumenten erhoben werden, auch wenn es sich bei diesen Instrumenten nicht um Tests im engeren Sinn handelt, sondern z. B. um einen Fragebogen zur Erfassung schulischer Interessen oder um die Ergebnisse eines systematischen Beobachtungsverfahrens. Zur Feststellung einer LRS ist jedoch v. a. der Einsatz von

Tests im engeren Sinne erforderlich, wobei aus der Vielzahl der verfügbaren Intelligenz- und Lesetests bestimmte Instrumente ausgewählt werden müssen.

Zum Beispiel könnte man den häufig eingesetzten Test »Coloured Progressive Matrices« (CPM; Bulheller & Häcker, 2002) nutzen, um die Intelligenz von Karl zu messen. Bei diesem Verfahren müssen die Kinder unvollständige farbige Figuren oder Muster ergänzen, indem sie die richtige Antwort aus einer von sechs vorgegebenen Antwortalternativen auswählen. Der CPM hat den Vorteil, dass er durch das figurale Aufgabenmaterial praktisch nur beim Verständnis der Instruktion, die beschreibt, wie die Testaufgaben zu bearbeiten sind, Anforderungen an sprachliche Fähigkeiten stellt.

Als Maß für Karls Leseleistung könnte z. B. der »ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler« (Lenhard & Schneider, 2006), verwendet werden. Dieses Instrument enthält Subtests zur Messung des Verständnisses von einzelnen Wörtern, von einzelnen Sätzen und von kurzen Texten. Welche Kriterien für die Auswahl diagnostischer Instrumente herangezogen werden sollten, wird in Abschnitt 19.1.4 diskutiert.

Bestimmung und Einordnung der Testwerte einer Person (Bezugsnormen)

Die mit einem diagnostischen Instrument erhobenen Antworten werden üblicherweise zu einem Testwert zusammengefasst. Beispielsweise wird beim CPM oder beim ELFE 1-6 die Anzahl der richtig gelösten Aufgaben jeweils zu einem Testwert addiert. Die Testwerte werden als Indikator für die Ausprägung des zu erfassenden Konstrukts bei der jeweiligen Person betrachtet. Zur Interpretation der Testwerte können verschiedene Bezugsnormen als Maßstäbe herangezogen werden: die soziale, die kriteriale und die individuelle Bezugsnorm. Die Einordnung und Beurteilung der Testwerte erfolgt dabei mit Bezug auf

- ▶ die Verteilung der Testwerte in einer Vergleichsgruppe (sozialnormorientierte Diagnostik oder Vergleichsdiagnostik),
- ▶ ein sachlich definiertes Kriterium (kriteriumsnormorientierte Diagnostik oder Diagnostik der Zielerreichung) und/oder
- ▶ die individuelle Entwicklung der Testperson sowie individuelle Profile (individualnormorientierte oder ipsative Diagnostik).

Sozialnormorientierte Diagnostik. Die sozialnormorientierte Diagnostik legt das Augenmerk auf interindividuelle Unterschiede, also darauf, wie sich Personen

hinsichtlich ihrer Testwerte unterscheiden. Hierfür werden die Ergebnisse aus sog. Normierungsstudien benötigt, bei denen repräsentative Stichproben von Personen das diagnostische Verfahren (z. B. einen Test) bearbeiten. Die Verteilung der Testwerte in der Normierungsstudie gibt Aufschluss über die Verteilung der Testwerte in der Referenzpopulation. Zur Einordnung des Ergebnisses einer bestimmten Person werden die Testwerte in Normwerte umgewandelt, die über die relative Position einer Testperson in Bezug auf die Verteilung von Testwerten in der Referenzpopulation informieren (soziale Bezugsnorm; vgl. Abschn. 4.2.3).

Die grundlegende Idee dieser Art von Normwerten veranschaulicht Abbildung 19.4, in der die Verteilung von (Intelligenz-)Testwerten einer Normalverteilung entspricht. Dies muss aber nicht immer so sein, denn die Verteilung der Testwerte kann (z. T. deutlich) von einer Normalverteilung abweichen (Micceri, 1989). Zur Beschreibung der interindividuellen Unterschiede kann man auf die Prozentrangskala oder auf diverse Skalen zurückgreifen, die sich an den Abweichungen der individuellen Testwerte vom Mittelwert der Verteilung in der Normierungsstichprobe orientieren.

Der Prozentrang gibt an, welcher Anteil von Personen in der Referenzgruppe einen gleich hohen oder einen geringeren Testwert erreicht hat. So bedeutet beispielsweise ein Prozentrang von 10, dass in der Referenzgruppe 10 % der Personen einen niedrigeren oder gleich hohen Testwert und 90 % der Personen in der Referenzgruppe einen höheren Testwert erzielt haben. Wie aus Abbildung 19.4 ersichtlich wird, hat die Prozentrangskala den Nachteil, dass die Messwerte über die gesamte Spannweite der Messwerte sehr ungleich verteilt sind. Eine bessere Differenzierung ermöglichen die sog. Abweichungsnormen, die sich am Mittelwert und der Standardabweichung der Verteilung orientieren. Dazu zählen insbesondere

- ▶ z-Werte (vgl. Abschn. 19.1.2),
- ▶ T-Werte und
- ▶ IQ-Werte.

Die Abweichungsnormen können dabei durch einfache Transformationen ineinander überführt werden. Prozentränge können allerdings nur dann in Abweichungsnormen überführt werden, wenn – wie in dieser Abbildung – die Verteilung der Testwerte einer Normalverteilung folgt.

Allgemein gilt bei Häufigkeitsverteilungen mit einer (angenäherten) Normalverteilungsform, dass ca. zwei

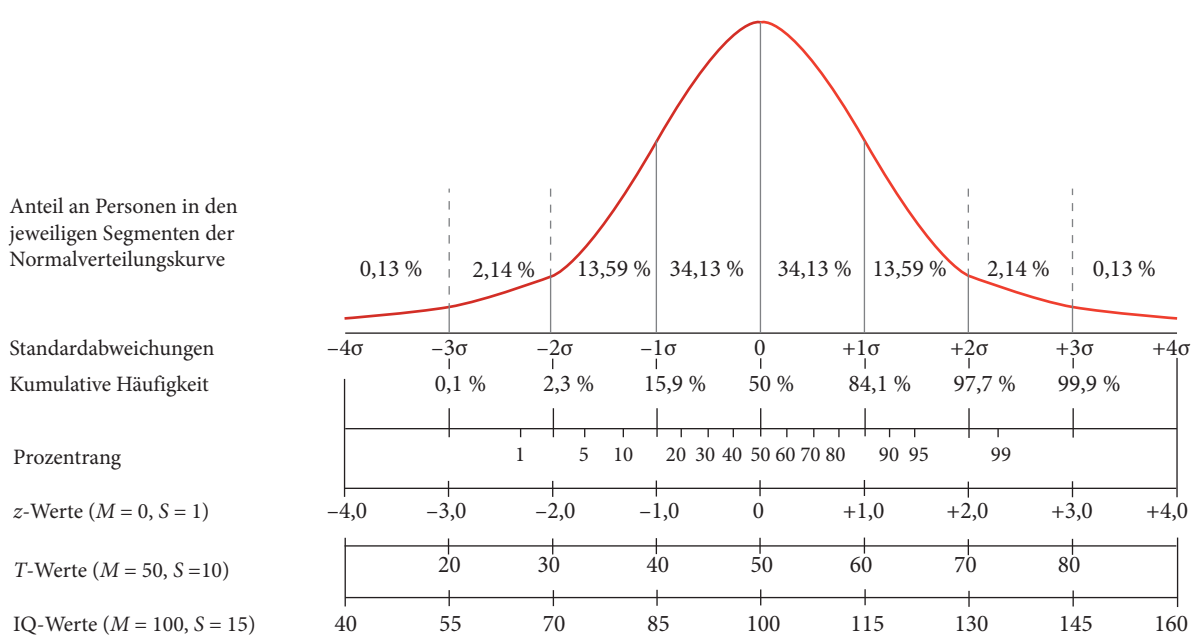


Abbildung 19.4 Normalverteilungskurve von Testwerten sowie darauf basierende Normwerte und kumulative Häufigkeiten. M = Mittelwert, S = Standardabweichung

Drittel (genauer: 68,26 %) der Fälle einer Stichprobe im Messwertbereich von $M \pm 1S$ und ca. 95 % der Fälle (genauer: 95,44 %) im Messwertbereich von $M \pm 2S$ liegen (s. Abb. 19.1 a und b in Abschn. 19.1.2 zur Definition von M und S). Die wechselseitige Überführung von Abweichungsnormen in Prozentränge ist aber umso stärker verzerrt, je mehr die tatsächliche Verteilung der Testwerte von einer Normalverteilung abweicht.

Karl beispielsweise hat 28 von 36 Aufgaben im CPM richtig gelöst. Anhand der im Testmanual des Verfahrens angegebenen Transformationsregeln lässt sich ermitteln, dass dies einem IQ-Wert von 100 und damit genau dem Mittelwert der Referenzpopulation und auch dem mittleren Rangplatz von 50 entspricht. Demnach erzielten in der Referenzgruppe Gleichaltriger 50 % der Kinder niedrigere oder gleich hohe IQ-Werte und 50 % der Kinder höhere IQ-Werte als Karl.

Kriteriumsnormorientierte Diagnostik. Da der individuelle Testwert bei der sozialen Bezugsnorm ausschließlich in Bezug auf die Verteilung der Testwerte in einer Referenzgruppe beurteilt wird, erhält man lediglich eine Information über den individuellen Rangplatz innerhalb der Bezugsgruppe, z. B. den Rangplatz im Leistungsspektrum bei einem schulischen Kompetenztest. Wenn allerdings alle anderen in der Gruppe schwache Leistungen erzielen, ist auch der Spitzenplatz keine Garantie für eine objektiv gute Leistung. Hier setzt die kriteriumsnormorientierte Diagnostik an. Testwerte, die sich an einer kriteriumsorientierten Bezugsnorm orientieren, geben Auskunft darüber, ob eine Person ein bestimmtes Kriterium erreicht hat oder nicht, und zwar unabhängig davon, wie viele Personen dieses Kriterium insgesamt erreicht haben. Beim kriteriumsnormorientierten Vorgehen spielen interindividuelle Unterschiede in den Testwerten also eine untergeordnete Rolle.

Ein wichtiges Beispiel hierfür sind Tests zur Erfassung schulischer Kompetenzen, wie etwa die Tests zur Überprüfung des Erreichens der länderübergreifenden Bildungsstandards, die von der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Kultusministerkonferenz, KMK) verabschiedet wurden. Bildungsstandards sind Kompetenzerwartungen. Sie legen fest, über welche fachspezifischen Kompetenzen Schüler in Deutsch, in der ersten Fremdsprache (Englisch, Französisch), in Mathematik und in den naturwissenschaftlichen Fächern bis zu einer bestimmten Jahrgangsstufe verfügen sollen. Abbildung 19.5 stellt beispielhaft das Kompetenzstufenmodell für

den Bereich Lesen vor, das zu den Bildungsstandards für den Mittleren Schulabschluss im Fach Deutsch entwickelt worden ist (Bremerich-Vos et al., 2010).

Zur Bestimmung sachlich definierter Kriterien wird die Verteilung der Testwerte mithilfe sog. Standardsetting-Prozeduren in mehrere Kompetenzstufen eingeteilt, indem Experten des jeweiligen Faches entsprechende Schwellenwerte (*cut scores*) bestimmen (Pant, Tiffin-Richards & Köller, 2010). Die Grundlage dieser Entscheidungen bilden die Aufgaben, die Schüler mit bestimmten Testwerten mit einiger Wahrscheinlichkeit lösen können. Für ihre Einschätzung erhalten die Experten zu diesen Aufgaben einerseits empirische Daten (z. B. zu den Lösungshäufigkeiten) und andererseits qualitative Informationen zu lösungsrelevanten kognitiven Prozessen oder erforderlichen Wissensstrukturen. Auf diese Weise wird es möglich, die Kompetenzstufen in Bezug auf die Bewältigung inhaltlicher Anforderungen zu beschreiben.

Um eine bestimmte Kompetenzstufe bei einer Schülerin oder einem Schüler diagnostizieren zu können, müssen beobachtbare Kriterien erfüllt worden sein. Zentral ist hierbei der Gedanke, dass Schüler, die eine bestimmte Kompetenzstufe erreicht haben, Aufgaben der jeweiligen Kompetenzstufe und Aufgaben der niedrigeren Kompetenzstufe mit einer festgelegten Mindestwahrscheinlichkeit (z. B. 65 %) lösen können, Aufgaben der höheren Kompetenzstufe aber nur mit geringerer Wahrscheinlichkeit. Wenn beispielsweise eine Schülerin Kompetenzstufe IV im Bereich Lesen im Fach Deutsch erreicht hat, dann kann sie mit relativ hoher Wahrscheinlichkeit die Aufgaben 1, 2, 3 und 4 in Abbildung 19.5 lösen, die den Kompetenzstufen I, II, III und IV zugeordnet sind. Die Wahrscheinlichkeit, dass sie Aufgabe 5 auf Kompetenzstufe V lösen kann, ist dagegen deutlich geringer. Weitere Aufgabenbeispiele für die Erfassung von Kompetenzen auf der Grundlage der KMK-Bildungsstandards finden sich auf den Internetseiten des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB; z. B. www.iqb.hu-berlin.de/laendervergleich).

Kompetenzstufen können auch darüber informieren, ob ein bestimmter Leistungsstandard, d. h. eine definierte Leistungserwartung, erreicht wurde oder nicht (Klieme et al., 2007). So legte beispielsweise die KMK für die Kompetenzstufenmodelle, die auf der Grundlage der Bildungsstandards entwickelt wurden, als Mindeststandard allgemein die Stufe II fest. Jugendliche, deren Leistungen unter Kompetenzstufe II liegen, erreichen demnach nicht das von der KMK definierte Bildungs-








<p>Aufgabe: Mann über Bord</p> <p>Stell dir vor, die Leute auf der Insel würden die wahre Geschichte des Schiffbrüchigen erfahren. Einige halten ihn für schuldig, einige für unschuldig am Tod der anderen Seeleute. Finde Argumente aus dem Text, was man dem Matrosen vorwerfen bzw. zugutehalten könnte.</p> <table><tr><th>»Schuldig«</th><th>»Unschuldig«</th></tr><tr><td> ----- ----- -----</td><td> ----- ----- -----</td></tr></table>	»Schuldig«	»Unschuldig«	 ----- ----- -----	 ----- ----- -----	<p>V</p> <p>Optimalstandard erreicht</p> <p>640</p>	<ul style="list-style-type: none">▶ im Text weit verstreute Informationen identifizieren und verknüpfen▶ selbstständig einen mehr oder weniger umfangreichen Text produzieren
»Schuldig«	»Unschuldig«					
 ----- ----- -----	 ----- ----- -----					
<p>Aufgabe: Mann über Bord</p> <p>Was könnte einer der Inselbewohner über den Überlebenden berichten? Notiere stichpunktartig.</p> <p> ----- ----- -----</p>	<p>IV</p> <p>Regelstandard plus erreicht</p> <p>560</p>	<ul style="list-style-type: none">▶ im Text nicht explizit genannte, sondern zu erschließende Wissensbestände von Figuren sowie das Motiv eines Erzählers erkennen▶ Aufgaben lösen, die Kombinationen von Grafik und kontinuierlichem Text erfordern				
<p>Aufgabe: Mann über Bord</p> <p>Die Geschichte handelt hauptsächlich von</p> <ul style="list-style-type: none"><input type="checkbox"/> einem Schiff, das bei einem Sturm kentert.<input type="checkbox"/> von der Ungeschicklichkeit eines Matrosen.<input type="checkbox"/> von den Gefahren auf hoher See.<input type="checkbox"/> einer Verkettung von Ereignissen, nachdem jemand ins Meer gestürzt ist.	<p>III</p> <p>Regelstandard erreicht</p> <p>480</p>	<ul style="list-style-type: none">▶ Aufgaben meistern, in deren Rahmen bei längeren Texten aus vorgegebenen Alternativen das zutreffende Textthema auswählen▶ Aufgaben lösen, bei denen die Bedeutung von Wörtern kontextuell erschlossen werden muss				
<p>Aufgabe: Mann über Bord</p> <p>Die Hauptfigur ...</p> <ul style="list-style-type: none"><input type="checkbox"/> schwimmt schnell zur rettenden Insel.<input type="checkbox"/> kann sich in ein Boot retten.<input type="checkbox"/> wird an Land gespült.<input type="checkbox"/> wird aus dem offenen Meer gefischt.	<p>II</p> <p>Mindeststandard erreicht</p> <p>400</p>	<ul style="list-style-type: none">▶ Aufgaben bewältigen, bei denen die Verbindung von Informationen verlangt ist, die über mehrere Abschnitte verteilt sind▶ Aufgaben lösen, bei denen ein mehrfaches Durchsuchen von Texten nach einzelnen Informationen gefordert ist				
<p>Aufgabe: Bewerbung</p> <p>Wann fängt die Lehre an?</p> <ul style="list-style-type: none"><input type="checkbox"/> zum Jahresbeginn<input type="checkbox"/> zum Schuljahresende<input type="checkbox"/> zum 1. August<input type="checkbox"/> zum 1. September	<p>I</p> <p>Mindeststandard nicht erreicht</p>	<ul style="list-style-type: none">▶ einzelne Informationen im Text lokalisieren, vor allem dann, wenn sie auffällig platziert sind▶ Aufgaben lösen, bei denen angesichts eines strukturell einfachen und kurzen Textes das Textthema identifiziert werden muss				

Abbildung 19.5 Kompetenzstufenmodell und illustrierende Aufgaben für den Mittleren Schulabschluss für den Bereich Lesen im Fach Deutsch (nach Bremerich-Vos et al., 2010). Die Zahlen in Klammern geben jeweils den Testwert an, den Jugendliche mindestens erzielen müssen, um eine bestimmte Kompetenzstufe zu erreichen. Die Metrik ist hier so bestimmt, dass ein Mittelwert von 500 der durchschnittlichen Leistung von Jugendlichen in der 9. Klassenstufe entspricht. Die Standardabweichung wurde auf 100 fixiert

minimum. Jugendliche, deren Leistungen auf der Kompetenzstufe III liegen, erreichen die Regelstandards der KMK, während die Leistungen von Jugendlichen auf den Kompetenzstufen IV oder V dem sog. Regelstandard plus bzw. dem Optimalstandard entsprechen (Bremerich-Vos et al., 2010).

Ipsative Diagnostik. Bei der ipsativen Diagnostik werden die individuellen Testwerte derselben Person als Vergleichsmaßstab herangezogen, d. h., die Messung basiert auf einer (intra-)individuellen Bezugsnorm. Hierzu können (bei Verwendung desselben Instruments) die individuellen Testwerte über die Zeit hinweg zu einer Zeitreihe zusammengefügt werden, um die individuelle Entwicklung abzubilden. Diese Form der Prozessdiagnostik kommt v. a. bei der Anwendung von Modifikationsstrategien zum Tragen, da hier die Veränderung des Erlebens oder des Verhaltens einer Person über die Zeit im Mittelpunkt steht. Im Fall von Karl kann z. B. sein individueller Lernfortschritt erfasst werden, indem die Leseflüssigkeit bei wiederholter Vorgabe ähnlich schwieriger Texte gemessen wird. Eine verhaltensnahe ipsative Norm kann im Idealfall selbst geringfügige individuelle Entwicklungen abbilden, wohingegen sozialnorm- oder kriteriumsnormorientierte Ansätze meist deutlich weniger sensitiv gegenüber individuellen Veränderungen sind. Damit kann die Verwendung ipsativer Normen einen wesentlichen Beitrag zum Erfolg von Fördermaßnahmen leisten, da die Erfassung des Lernfortschritts ein wichtiger Bestandteil effektiver Förderstrategien ist.

Ein zweites Einsatzgebiet von ipsativen Normen stellen Profilanalysen von Konstrukten dar, die mehrere Facetten umfassen. Hier werden die Ausprägungen auf einer Konstruktfacetten (z. B. die Leseleistung bei kontinuierlichen Texten, wie etwa Kurzgeschichten) mit den Ausprägungen auf anderen Konstruktfacetten (z. B. die Leseleistung bei diskontinuierlichen Texten, wie etwa einem TV-Programmheft) verglichen. Solche Profilanalysen sind wiederum bei der Modifikationsdiagnostik wichtig, um die Fördermaßnahmen an das individuelle Leistungsprofil bestmöglich anzupassen. Des Weiteren spielen Profilanalysen im Kontext der Selektionsdiagnostik eine wichtige Rolle, etwa wenn es bei der Berufs- oder Studierendenberatung darum geht, Berufsausbildungen oder Studiengänge zu identifizieren, die an das Interessen- und Begabungsprofil der jeweiligen Person angepasst sind.

Messfehler und Konfidenzintervall. Grundsätzlich ist davon auszugehen, dass Testwerte mit einem Messfehler

behaftet sind (vgl. unter »Reliabilität« in Abschn. 19.1.4). Ein Testwert für eine Person spiegelt nicht nur die individuelle Ausprägung des Zielkonstrukts (z. B. Leseverstehen oder Intelligenz) wider, sondern kann darüber hinaus auch durch das Wirken vielfältiger anderer Faktoren zustande gekommen sein (z. B. Flüchtigkeitsfehler, Unaufmerksamkeit, Verständnisschwierigkeiten in Bezug auf die Aufgabeninstruktionen, Unruhe im Untersuchungszimmer, körperliches Unwohlsein). Der Anteil am Testwert, der auf das Wirken solcher Störfaktoren zurückgeht, wird in psychometrischen Modellen allgemein mit dem Begriff »Messfehler« beschrieben.

Um den Messfehler zu berücksichtigen, werden üblicherweise Konfidenzintervalle (»Vertrauensintervalle«) um den jeweiligen Testwert angegeben. Die Unter- und Obergrenzen für das Konfidenzintervall markieren dabei eine Spannbreite plausibler Werte für die Merkmalsausprägung einer Person (vgl. Cumming & Finch, 2005, für eine verständliche Erklärung der statistischen Bedeutung von Konfidenzintervallen). Die Berechnung der Konfidenzintervalle von (sozialnormorientierten) Testwerten X_i ist in Abbildung 19.1 e dargestellt (s. Fisseni, 1997, zur Berechnung des Konfidenzintervalls für kriteriumsnormorientierte Testwerte). Um den Messfehler zu quantifizieren, benötigt man Informationen zur Messgenauigkeit eines Testwerts. Diese lässt sich mithilfe von Reliabilitätsschätzungen bestimmen (s. Abschn. 19.1.4). Anhand des Testwerts einer Person und der Kenntnis der Reliabilität des verwendeten Testinstruments lässt sich dann das Konfidenzintervall berechnen.

In der pädagogisch-psychologischen Diagnostik sind Konfidenzintervalle bei der Prüfung der zu untersuchenden diagnostischen Hypothesen wichtig. Dies wollen wir am Beispiel von Karl erläutern. Im Test zum Leseverstehen erzielte Karl einen T -Wert von 20 Punkten. Die Untergrenze für das 95%-Konfidenzintervall (bei $S_{\text{Lesen}} = 10$ und einer Reliabilität des Tests von $r_{tt} = 0,90$) liegt bei $20 - 1,96 \cdot 10 \cdot \sqrt{(1 - 0,9)} = 20 - 6,2 = 13,8$. Die Obergrenze liegt bei 26,2. Dies bedeutet, dass die Werte 13,8 und 26,2 plausible Unter- und Obergrenzen für Karls »wahren« Wert im Leseverstehen markieren. Die Informationen zu Unter- und Obergrenzen des Konfidenzintervalls können dann formal z. B. so zusammengefasst werden: 95%-Konfidenzintervall im Lesetest = [13,8; 26,2].

Ein T -Wert von 20 entspricht in etwa einem Prozentrang von 0,1 % (s. Abb. 19.4). Selbst wenn Karls »wahrer« Wert im Leseverstehen an der Obergrenze des Konfidenz-

intervalls läge (also bei 26,2), würden die Testleistungen von über 98 % der Kinder in der Referenzgruppe über der Testleistung von Karl liegen. Dieses Ergebnis stützt die erste Hypothese, dass Karls Prozentrang im Leseverstehen bei höchstens 10 % liegt. Karls Leistung im Leseverstehen ist also als deutlich unterdurchschnittlich zu bewerten. Im Intelligenztest CPM (mit $S = 15$ und einer angenommenen Reliabilität von $r_{tt} = 0,90$) erzielte Karl einen IQ-Wert von 100. Die Untergrenze seines 95%-Konfidenzintervalls liegt somit bei $100 - 1,96 \cdot 15 \cdot \sqrt{(1 - 0,9)} = 100 - 9,3 = 90,7$ und die Obergrenze bei 109,3. Selbst die Untergrenze des Konfidenzintervalls von Karls IQ-Wert befindet sich somit klar über dem Grenzwert von 70, ab dem generell eine geistige Behinderung angenommen wird. Dieses Ergebnis stützt damit die zweite Hypothese, dass Karls IQ über 70 liegt.

Treffen einer Entscheidung

Im Beispiel von Karl legen die im vorherigen Abschnitt dargestellten Ergebnisse die Diagnose einer LRS nahe: Karls Intelligenz ist durchschnittlich, es liegt also keine geistige Behinderung vor, und seine Leseleistung liegt unter dem

10. Prozentrang. Zudem konnten – ohne dass dies hier dargestellt wurde – mögliche Alternativerklärungen, wie z. B. eine Sehestörung oder eine Konzentrationsstörung, ausgeschlossen werden. Hierbei ist es aber wichtig zu beachten, dass, selbst wenn (wie im Beispiel von Karl) die eingesetzten Messinstrumente relativ präzise sind, im Kontext pädagogisch-psychologischer Diagnostik immer die Möglichkeit einer Fehlentscheidung besteht (s. Kasten).

Bei einer Entscheidung sind stets die Konsequenzen verschiedener Entscheidungsfehler gegeneinander abzuwägen: Welche Konsequenzen hätte es einerseits für Karl, wenn Sie eine LRS diagnostizierten, er aber in Wirklichkeit keine hat? Und welche Konsequenzen hätte es andererseits für Karl, wenn Sie keine LRS diagnostizierten, tatsächlich aber eine solche Schwäche vorliegt? Karls Leseleistungen sind deutlich unterdurchschnittlich. Daher überwiegt (unseres Erachtens) in diesem Fall der potenzielle Nutzen einer gezielten Förderung von Karls Leseverstehen die potenziellen Nachteile, die etwa aus einer möglichen Stigmatisierung durch die Diagnose einer LRS entstehen könnten.

Unter der Lupe

Wie Schwellenwerte bei einem Aufnahmetest diagnostische Entscheidungsfehler beeinflussen

In vielen Ländern und mittlerweile auch an einigen deutschen Universitäten ist es Praxis, dass Studierende u. a. mithilfe eines Aufnahmetests ausgewählt werden. Der Aufnahmetest dient dazu, Studienbewerber zu identifizieren, die mit großer Wahrscheinlichkeit das Studium erfolgreich abschließen werden. Hierzu benötigt man einen Aufnahmetest, der eine möglichst gute Vorhersage der Abschlussleistung ermöglicht. Im (hier stark vereinfachten) Beispiel betrachten wir eine Selektionsstrategie, in der zur Studierendenauswahl nur ein Test eingesetzt wird, der mit der Abschlussleistung mit $r = 0,70$ korreliert.

Nachfolgend erklären wir anhand von Abbildung 19.6, wie Entscheidungsfehler und die verwendeten Schwellenwerte zusammenhängen. Entscheidend für die Zulassung zum Studium ist, dass eine Person eine bestimmte Mindestpunktzahl im Aufnahmetest (von maximal 100 erreichbaren Punkten) erzielen muss. Diese Mindestpunktzahl definiert den Schwellenwert. Die Abschlussleistung wird im Beispiel als Summe über 10 Einzelprüfungsleistungen berechnet, in denen jeweils maximal 10 Punkte erreicht werden können. Das Studium gilt als bestanden, wenn man mindestens

60 der maximal 100 erzielbaren Punkte erreicht. Somit betreffen in Abbildung 19.6 a und 19.6 b die diagnostischen Entscheidungsfehler jeweils zwei Felder:

- ▶ Personen in Feld FP sind »falsch positiv«, d. h., diese Personen werden zum Studium zugelassen, erreichen aber nicht die erforderliche Mindestpunktzahl zum erfolgreichen Abschluss des Studiums.
- ▶ Personen in Feld FN sind »falsch negativ«, d. h., diese Personen werden nicht zum Studium zugelassen, hätten aber (sofern sie zugelassen würden) die erforderliche Mindestpunktzahl zum Abschluss des Studiums erreicht.

Die beiden Auswahlverfahren mit jeweils 300 Probanden, die in Abbildung 19.6 dargestellt sind, unterscheiden sich im Schwellenwert, der zur Zulassung herangezogen wird. Dies hat zur Folge, dass bei einem Schwellenwert von 50 der Anteil von FP bei $87/300 = 29\%$ und der Anteil von FN bei $13/300 = 4\%$ liegt (s. Abb. 19.6 a). Bei einem Schwellenwert von 70 liegt der Anteil von FP deutlich niedriger, und zwar bei $17/300 = 6\%$, jedoch steigt gleichzeitig der Anteil von FN auf $43/300 = 14\%$ (s. Abb. 19.6 b). Je restriktiver die Auswahl erfolgt, desto höher ist

zwangsläufig die Quote der zu Unrecht Zurückgewiesenen. Die Zahl der falsch Positiven sinkt also auf Kosten der falsch Negativen. Anders herum gilt auch, dass die Zahl der falsch Positiven steigt, wenn die Zahl der falsch Negativen sinkt.

Abbildung 19.6 verdeutlicht, dass im Idealfall zur Bewertung von diagnostischen Selektionsentscheidungen beide Fehlerarten herangezogen werden sollten. Jedoch liegen bei realen selektiven Entscheidungen in pädagogischen Kontexten häufig nicht genügend Informationen vor, um das Ausmaß beider Fehlentscheidungen abschätzen zu können. Da beispielsweise beim Entwicklungsprozess des Selektions-

verfahrens in der Regel nicht alle Personen, die sich um einen Studienplatz beworben haben, zugelassen werden, weiß man nicht, wer unter den abgewiesenen Bewerbern erfolgreich gewesen wäre. In diesem Fall kennt man nur die Gesamtanzahl der abgelehnten Personen, nicht aber die Anzahl der Personen in den Feldern RN bzw. FN. In solchen Fällen kann die Qualität des Selektionsverfahrens lediglich anhand der zu Unrecht Aufgenommenen (Feld FP) bestimmt werden. Jedoch ist eine solche Art der Qualitätsbeurteilung irreführend, da die Fehlerrate der falsch negativen Entscheidungen unbekannt ist.

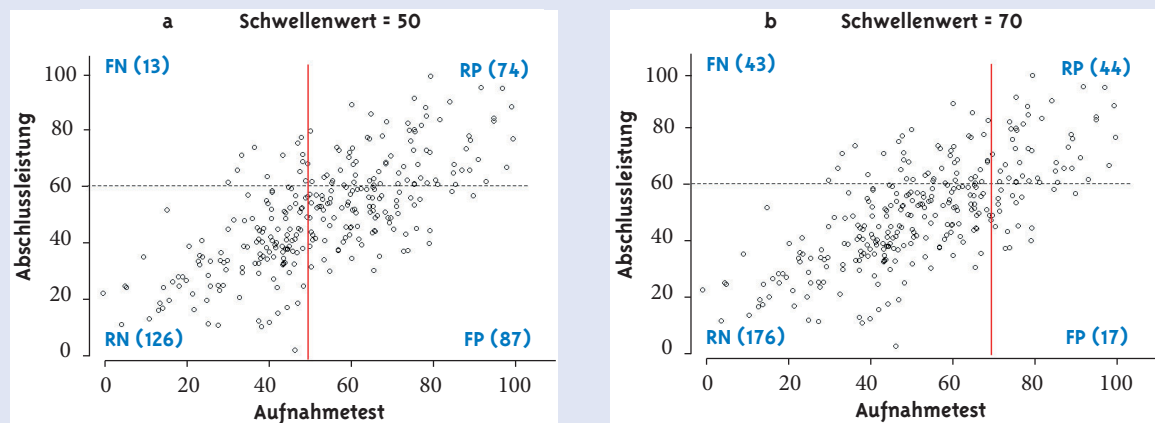


Abbildung 19.6 Zusammenhang zwischen den Anteilen von richtigen bzw. falschen Diagnoseentscheidungen und Schwellenwerten am Beispiel einer Studienzulassungsentscheidung über 300 Personen. RP = richtig positiv; FP = falsch positiv; RN = richtig negativ; FN = falsch negativ. In beiden Teilabbildungen korreliert die Leistung im Aufnahmetest mit der Abschlussleistung mit $r = 0,70$. Die Zahlen in Klammern geben die Anzahl der Personen in den jeweiligen Fallgruppen an. Der Schwellenwert ist jeweils rot dargestellt

Anfertigung eines Gutachtens

Die Entscheidung bzw. Empfehlung für eine Entscheidung wird häufig in Form schriftlicher Gutachten festgehalten. Diese folgen in der Regel einer Gliederung, die sich am diagnostischen Prozess orientiert. Bei der Formulierung des Gutachtens sind einige Qualitätsmerkmale zu beachten, zu denen insbesondere die folgenden gehören (Fisseni, 1997; Kubinger, 2009):

- Erklärung von Fachbegriffen (z. B. LRS), wenn davon auszugehen ist, dass einige Empfänger des Gutachtens keine psychologische Ausbildung haben

- konkrete Darstellung der diagnostischen Fragestellung und der Untersuchungszeitpunkte
- vollständige Auflistung der verwendeten Instrumente und Methoden
- klare Trennung zwischen Ergebnissen und Interpretation von Ergebnissen
- thematische Integration der verschiedenen Testwerte mit Bezug auf die Fragestellung

Viele ausführliche Beispiele und Empfehlungen für das Verfassen psychologischer Gutachten finden sich in Kubinger (2009) und Fisseni (1997).

19.1.4 Gütekriterien von Testwerten und Auswahlkriterien von diagnostischen Instrumenten

Zur Erfassung von Konstrukten steht häufig eine Vielzahl von alternativen Messinstrumenten zur Verfügung. Einen guten Überblick über die im deutschsprachigen Markt verfügbaren Instrumente geben das »Brickenkamp Handbuch psychologischer und pädagogischer Tests« (Brähler, Holling, Leutner & Petermann, 2002), die elektronische Datenbank »PSYNDEXplus Tests« sowie die Reihen »Deutsche Schultests« und »Tests und Trends: Jahrbuch der pädagogisch-psychologischen Diagnostik« (Hasselhorn & Gold, 2009). Um ein Instrument auszuwählen, ist es erforderlich, die Gütekriterien der Testwerte zu bewerten, die mit dem Instrument erhoben werden (American Educational Research Association et al., 1999; Testkuratorium, 2010). Die zentralen Gütekriterien von Testwerten sind: Objektivität, Reliabilität und Validität. Zusätzlich ist im Rahmen pädagogisch-psychologischer Diagnostik auch die Qualität der Normen sehr wichtig. Professionell entwickelte Instrumente zeichnen sich dadurch aus, dass die Informationen zur Beurteilung von Testwerten detailliert im Manual des Instruments dokumentiert sind. Wegweisend für die Auswahl eines Instruments können darüber hinaus auch die in Fachzeitschriften veröffentlichten Testkritiken sein (Testkuratorium, 2010).

Objektivität

Fallen die Testwerte einer Testanwendung immer gleich aus, egal wer den Test (a) durchführt, (b) auswertet und (c) interpretiert? Ist die Antwort auf diese Teilfragen jeweils »ja«, dann sind Durchführungs-, Auswertungs- und Interpretationsobjektivität für die Testwerte gegeben. Um die Objektivität der Testwerte zu gewährleisten, müssen die Anwendungsbedingungen für die Durchführung, Auswertung und Interpretation stan-

dardisiert werden. Hierfür sollte das Manual u. a. auf die folgenden Aspekte eingehen (Testkuratorium, 2010):

- Es sollte vorschreiben, was die durchführende Person als Testinstruktion wortwörtlich sagen soll und was sie auf keinen Fall sagen sollte, welche Handlungen zu vollziehen sind (z. B. für eine ruhige und ungestörte Durchführung des Tests sorgen) und welche Antworten auf Rückfragen oder bei Unklarheiten gegeben werden sollen.
- Es sollte definieren, wie fehlende Antworten auf Testfragen zu kodieren sind und wie viele gültige Antworten vorliegen müssen, damit eine Interpretation der Testwerte überhaupt sinnvoll ist.

Einige diagnostische Instrumente erfordern spezifische Vorkenntnisse und Erfahrungen, um das Instrument fachgerecht einzusetzen, auszuwerten und die gewonnenen Testwerte zu interpretieren. In diesen Fällen ist es erforderlich, Anwender anhand des Manuals zu schulen.

Reliabilität

Wie messgenau erfassen die Testwerte das Zielkonstrukt? Die Frage nach der Messpräzision betrifft die Reliabilität (oder Zuverlässigkeit) von Testwerten (Brunner, Nagy & Wilhelm, 2012). Für die Bestimmung der Reliabilität stehen verschiedene methodische Zugänge zur Verfügung (s. Kasten), die darauf abzielen, die Messfehler der Testwerte zu schätzen. Die Grundidee ist, dass das jeweilige Konstrukt mehrfach (wiederholt mit demselben Instrument, mit parallelen Fassungen des Instruments etc.) gemessen wird. Sind die Testwerte präzise, so sollten Personen, die in der einen Messung hohe Werte erzielt haben, auch in der anderen Messung hohe Werte erzielen. Die Korrelation zwischen den Messwerten sollte also hoch sein. Abweichungen von einem perfekten Zusammenhang werden als Messfehler bewertet.

Übersicht

Methoden zur Reliabilitätsbestimmung

Zur Bestimmung der Reliabilität stehen mehrere »klassische«, aber auch moderne Methoden zur Verfügung:

- **Retest-Methode:** Dabei wird dasselbe Instrument bei derselben Personenstichprobe zweimal eingesetzt. Die Korrelation der Testwerte zwischen erstem und zweitem Messzeitpunkt ist ein Maß der Messgenauigkeit des Instruments. Die Wahl des

Zeitintervalls zwischen den beiden Messzeitpunkten sollte von Annahmen zur zeitlichen Stabilität des Konstrukts abhängig gemacht werden: Je stabiler das Konstrukt ist, desto länger kann das Zeitintervall gewählt werden (Schermelleh-Engel & Werner, 2012).

- ▶ **Paralleltest-Methode:** Dabei werden zwei nach den gleichen Kriterien konstruierte Tests in derselben Personenstichprobe eingesetzt. In diesem Fall ist die Korrelation zwischen den Testwerten der beiden Paralleltests das Maß für die Messgenauigkeit des Instruments.
- ▶ **Testhalbierungs-Methode:** Dabei werden die einzelnen Indikatoren eines Messinstruments (z.B. die Testaufgaben eines Lesetests) per Zufall in zwei Hälften eingeteilt und für jede Hälfte jeweils ein Testwert berechnet. Das Maß für die Messgenauigkeit des Instruments ist die Korrelation dieser beiden Testwerte.
- ▶ **Interne Konsistenz:** Reliabilitätskoeffizienten zur Bestimmung der internen Konsistenz basieren auf den Interkorrelationen zwischen den Werten aller Einzelindikatoren (z.B. Aufgaben eines Lesetests) untereinander. Das am häufigsten eingesetzte Maß zur Bestimmung der internen Konsistenz ist Cronbachs Alpha (Schmitt, 1996).
- ▶ **Faktoranalytische Modelle:** Über die vier bereits genannten »klassischen« Verfahren zur Bestimmung

der Reliabilität hinaus wurden in jüngerer Zeit Reliabilitätskoeffizienten entwickelt, die auf faktoranalytischen Modellen basieren (z. B. Omega; s. McDonald, 1999). Ihr besonderer Mehrwert liegt darin, dass sie explizit die hierarchische Struktur des Zielkonstrukts bei der Berechnung der Reliabilität berücksichtigen können (Brunner et al., 2012). So kann beispielsweise untersucht werden, wie messgenau Testwerte Intelligenz auf einem übergeordneten Niveau (allgemeine Intelligenz) und auf dem Niveau untergeordneter, spezifischerer Facetten wie schlussfolgerndes Denken erfassen.

- ▶ **Modelle der Item-Response-Theorie:** Weiterhin sind die statistischen Modelle der Item-Response-Theorie (IRT) interessant, da diese es ermöglichen, die Messgenauigkeit eines Instruments in Abhängigkeit von der Ausprägung der Testwerte zu berechnen (McDonald, 1999; Rost, 2004). Damit lässt sich beispielsweise bestimmen, ob das Instrument sehr schwache und sehr gute Leistungen ebenso präzise erfasst wie Leistungen im mittleren Bereich.

Auswahlkriterien des Reliabilitätskoeffizienten. Welchen Reliabilitätskoeffizienten soll man nun zur Bewertung der Messpräzision von Testwerten heranziehen? Die Retest-Reliabilität ist insbesondere dann relevant, wenn die zeitliche Stabilität einer Eigenschaft für das jeweilige Entscheidungsproblem eine wichtige Rolle spielt (z.B. bei Selektionsstrategien). Wenn hingegen die zeitliche Stabilität weniger wichtig ist, sind die anderen Reliabilitätskoeffizienten vorzuziehen, da diese eine Aussage über die Messgenauigkeit der Testwerte für einen bestimmten Zeitpunkt machen. Beispielsweise ist die zeitliche Stabilität von Testwerten zur Bestimmung der Lesekompetenz von nachrangiger Bedeutung, wenn Veränderungen des Leseverstehens von Karl im Laufe einer Fördermaßnahme erfasst werden sollen.

Bewertung der Reliabilitätskoeffizienten. Unabhängig von der Methode der Reliabilitätsbestimmung können Reliabilitätskoeffizienten Werte zwischen 0 (keine Reliabilität) und 1 (perfekte Reliabilität) annehmen. Wie wir oben bereits dargestellt haben, ist die Reliabilität der Testwerte als Maß der Messgenauigkeit ein wichtiger Faktor, der die Breite der Konfidenzintervalle um gemessene Testwerte und damit die Sicherheit der Prüfung diagnostischer Hypothesen beeinflusst. In diesem

Zusammenhang stellt sich die Frage, was ein guter Wert für die Reliabilität von Testwerten ist. Generell ist es bei der Beantwortung dieser Frage wichtig, zwei Aspekte zu beachten (vgl. Testkuratorium, 2010):

- ▶ den Anwendungskontext des Tests und
- ▶ die Referenzpopulation, für die der Reliabilitätskoeffizient bestimmt wurde.

Reliabilitätskoeffizienten können auf einen aktuellen Anwendungskontext generalisiert werden, (1) wenn der vorliegende Anwendungskontext mit dem bei der Konstruktion des Tests ins Auge gefassten Anwendungskontext möglichst übereinstimmt und (2) wenn die zu untersuchende Person Teil derjenigen Referenzpopulation ist, für welche die Reliabilitätsschätzung des Messverfahrens ursprünglich erfolgte. Sind diese Bedingungen (weitgehend) erfüllt, so werden bei Verwendung von Testwerten im Kontext von individualdiagnostischen Entscheidungsproblemen mit weitreichenden individuellen Konsequenzen (z.B. die Diagnose einer LRS) häufig die folgenden Bewertungskonventionen herangezogen: Reliabilitäten größer 0,90 gelten als gut, Werte zwischen 0,80 und 0,90 als ausreichend und Werte kleiner 0,80 als unzureichend (Evers, 2001). Hier- von abzugrenzen ist der Forschungskontext, in dem

auch niedrigere Reliabilitätswerte von Testwerten akzeptabel sein können, wenn das Instrument z. B. nur für Aussagen auf Gruppenebene verwendet werden soll (für eine ausführlichere Diskussion s. Schermelleh-Engel & Werner, 2012; Schmitt, 1996).

Validität

Messen die Testwerte das Konstrukt, das sie messen sollen? Diese fundamentale Frage zur Validität von Testwerten stellt sich, da die individuelle Konstruktausprägung nicht direkt beobachtbar ist, sondern (indirekt) anhand von Testwerten abgebildet wird (s. o.). Um zu beurteilen, wie gut diese Abbildung mit einem bestimmten Instrument gelingt, muss die Validität von Testwerten analysiert werden. Da die Testwerte die empirische Grundlage bei der diagnostischen Entscheidungsfindung bilden, schließt sich eine zweite, praxisorientierte Frage zur Validität von Testwerten an: Inwieweit stützen Theorie und empirische Befunde die Interpretation von Testwerten im Hinblick auf die vorgesehene Testanwendung? Zur Beantwortung beider Fragen sollten systematisch verschiedene Formen empirischer Evidenz herangezogen werden (American Educational Research Association et al., 1999). Hierzu gehört die Analyse

- ▶ der Inhaltsvalidität,
- ▶ von Antwortprozessen,
- ▶ der internen Struktur sowie
- ▶ von konvergenten und diskriminanten Zusammenhängen.

Inhaltsvalidität. Bei der Inhaltsvalidität geht es darum, ob die einzelnen Indikatoren (Testaufgaben, Fragen im Fragebogen u. Ä.), die zu einem Testwert zusammengefasst werden, das jeweilige Zielkonstrukt repräsentativ abbilden. Die inhaltliche Validität wird häufig anhand von Experteneinschätzungen überprüft. Es wird beurteilt, inwieweit die Indikatoren das definierte Konstrukt inhaltlich und in seiner Breite und Struktur angemessen abdecken. Hierbei ist insbesondere auch darauf zu achten, dass Konstruktfacetten, die zentral für die Definition des Konstrukts sind, durch eine größere Anzahl von Indikatoren abgebildet werden (Haynes, Richard & Kubany, 1995), wie etwa bei einem Lesekompetenztest verschiedene Ziele des Lesens (z. B. Heraussuchen von Informationen, textimmanente Interpretation, Reflexion und Bewertung) in Bezug auf verschiedene Textsorten (Artelt, Stanat, Schneider & Schiefele, 2001). Ein üblicherweise verwendeter Indikator und Gradmesser

für die inhaltliche Validität stellt die Übereinstimmung der Expertenurteile dar (s. Wirtz & Caspar, 2002, zu Methoden zur Bestimmung der Beurteilerübereinstimmung).

Antwortprozesse. Die Analyse individueller Antwortprozesse untersucht die Passung zwischen Zielkonstrukt und den (kognitiven) Prozessen, die Personen bei der Bearbeitung des diagnostischen Instruments tatsächlich zeigen. Zur Erfassung der Bearbeitungsprozesse können vielfältige Methoden herangezogen werden. Neben »klassischen Methoden«, wie etwa dem lauten Denken, sind insbesondere technologiegestützte Verfahren (z. B. Erfassung des Leseverstehens am Computer) sehr vielversprechend. Denn mit solchen Verfahren können detailgenaue Bearbeitungsprotokolle aufgezeichnet werden. Die Abfolge und Dauer der Bearbeitungsschritte dienen dabei als Indikatoren für individuelle Antwortprozesse beim Lösen von Testaufgaben (für Leseaufgaben vgl. z. B. Organisation for Economic Co-operation and Development [OECD], 2011).

Interne Struktur. Die Analyse der internen Struktur eines Instruments beschäftigt sich mit dem Zusammenhangsmuster zwischen den Konstruktindikatoren. Es wird der Frage nachgegangen, ob das empirisch vorgefundene Zusammenhangsmuster die angenommene Unterscheidung von Facetten oder die angenommene hierarchische Struktur des Konstrukts stützt. So wird beispielsweise im Rahmen von PISA zwischen Lesekompetenz bei verschiedenen Textformaten unterschieden (OECD, 2010). Ein sehr wichtiges Unterscheidungsmerkmal ist hierbei, ob es sich um kontinuierliche Texte (z. B. Erzählungen) oder um nicht-kontinuierliche Texte (z. B. Zeitungsartikel, die Texte und Bilder oder Tabellen enthalten) handelt. Statistische Methoden zur Prüfung von Konstruktstrukturen sind fortgeschrittene multivariate statistische Verfahren, wie etwa die exploratorische und die konfirmatorische Faktorenanalyse (McDonald, 1999), Modelle der Item-Response-Theorie sowie Mischverteilungsmodelle (z. B. Rost, 2004).

Konvergente und diskriminante Validität. Bei Analysen konvergenter und diskriminanter Validität werden Zusammenhangsmuster zwischen den Testwerten des zu validierenden Instruments mit externen Kriterien untersucht. Zu diesen Kriterien gehören Messwerte, die mit anderen Testinstrumenten gewonnen werden, aber auch einschlägige Verhaltensmaße und vielfältige andere praktisch relevante Indikatoren (z. B. Bildungszertifikate wie das Abitur oder Studienabschlüsse, Urteile

von Lehrkräften oder Eltern, Vorgesetztenurteile). Konvergente Validität lässt sich dadurch belegen, dass die Testwerte substanziell mit Indikatoren für dasselbe oder ein eng verwandtes Konstrukt bzw. inhaltlich relevanten Indikatoren zusammenhängen, etwa wenn die Leistungen in einem Lesetest mit den Noten im Fach Deutsch oder Lehrereinschätzungen der Lesekompetenz korrelieren. Diskriminante Validität bedeutet, dass keine oder nur geringe Zusammenhänge zu inhaltlich nicht oder nur entfernt verwandten Indikatoren bestehen, etwa wenn die Korrelation zwischen der Leistung in einem Lesetest und der Sportnote gering ist oder um 0 liegt.

Bei Fragen zur konvergenten Validität von Testwerten trifft man in der Fachliteratur häufig auf zwei weitere verwandte Begriffe:

- **Kriteriumsvalidität:** Handelt es sich bei Befunden zur konvergenten Validität um Zusammenhänge mit praktisch relevanten Kriterien (z.B. den oben genannten Indikatoren), dann bezeichnet man dies häufig auch als Nachweis der Kriteriumsvalidität der Testwerte.
- **Prognostische Validität:** Wenn Testwerte ein Kriterium vorhersagen können, das zeitlich später auftritt bzw. erfasst wird, wird dies auch als Nachweis für deren prognostische Validität betrachtet, etwa wenn die Leistungen 15-jähriger Schüler mit ihren später erworbenen Berufsabschlüssen korrelieren.

Normen und Schwellenwerte

Wie sind die resultierenden Testwerte zu beurteilen? Wie oben beschrieben, können zur Beantwortung dieser Frage Testwerte einer Person (a) in Bezug zur Verteilung der Testwerte in einer Referenzgruppe (soziale Bezugsnorm) oder (b) in Bezug zu einem definierten Kriterium (kriteriale Bezugsnorm) gesetzt werden. Welche Qualitätskriterien sind hier zu beachten? Bei der sozialnormorientierten Diagnostik sind die Aktualität, Repräsentativität und Größe der Vergleichsstichprobe, anhand derer die Normwerte bestimmt wurden, von zentraler Bedeutung (Testkuratorium, 2010). Darüber hinaus gilt es zu beurteilen, wie genau die Population definiert wurde und wie die Stichprobe gezogen wurde, um diese Population abzubilden. Bei der kriteriumsnormorientierten Diagnostik spielt die Validität der Schwellenwerte eine wichtige Rolle, weil diese Auskunft darüber geben, ob eine Person ein bestimmtes (Lernziel-)Kriterium erfüllt oder nicht und daraus ggf. handlungsrelevante Schlussfolgerungen gezogen werden können (z.B.

Sprachförderbedarf). Wichtig ist hierbei, dass der Prozess der Schwellenwertbestimmung hinreichend wissenschaftlich abgesichert und dokumentiert ist (z.B. durch Anwendung einer etablierten Standardsetting-Prozedur; s. hierzu Pant et al., 2010) und die Schwellenwerte durch ausgewiesene Experten bestimmt wurden.

19.1.5 Diagnostik im schulischen Alltag von Lehrkräften

Die bislang diskutierten Konzepte, Begriffe und Methoden der pädagogisch-psychologischen Diagnostik sind allgemein anwendbar. Insbesondere die Konzepte und Begriffe sind auch auf den schulischen Alltag von Lehrkräften anwendbar, der in vielerlei Hinsicht durch diagnostische Aufgaben geprägt ist. Das methodische Vorgehen lässt sich jedoch nur sehr eingeschränkt auf den schulischen Alltag übertragen.

Diagnose von Lernausgangslagen und Lernfortschritt.

Eine wichtige diagnostische Aufgabe von Lehrkräften ist die Erfassung von Lernausgangslagen (z.B. Vorwissen oder Interesse) und des Lernfortschritts einzelner Schüler. Diese Arten der Diagnostik bilden wichtige Komponenten im Handeln von Lehrkräften, um die Unterrichtsanforderungen optimal an die Lernausgangslagen der Schüler anzupassen und so deren Lernfortschritt effektiv zu unterstützen (Helmke, 2006). Viele Studien haben gezeigt, dass Lehrkräfte generell sehr gut darin sind, die Schüler in ihren Klassen in eine Leistungsrangreihe zu bringen (Südkamp, Kaiser & Möller, 2012). Sie haben aber häufig große Probleme, das Leistungsniveau ihrer Klasse genau einzuschätzen (Artelt et al., 2001; Brunner, Anders, Hachfeld & Krauss, 2011). Deshalb fällt es ihnen auch schwer, das absolute Leistungsniveau von einzelnen Schülern exakt zu beurteilen.

Bewertung durch Schulnoten. Eine zweite wichtige diagnostische Aufgabe von Lehrkräften ist die Bewertung von Schülerleistungen anhand von Noten. Schulnoten sind wichtig für die Bildungskarriere und damit den Lebensweg der Schüler, denn sie bilden die Grundlage von Empfehlungen für weiterführende Schulformen. Weiterhin sind Noten entscheidend für das Erreichen von Schulabschlüssen und erlauben oder verwehren somit den Zugang zu bestimmten Berufen oder Studiengängen. Die Notengebung sollte auf einem Vergleich der Schülerleistung mit einem Leistungsmaßstab basieren bzw. die Passung zwischen Leistung und Anforderungen abbilden.

Allerdings sind die Leistungsmaßstäbe und Anforderungen (z. B. in den Lehrplänen), mit denen individuelle Leistungen verglichen werden, in aller Regel nicht genau genug definiert. Dies macht es für die Lehrkräfte besonders schwierig, die individuellen Leistungen ihrer Schüler sachgerecht (»kriterial«) zu beurteilen. Daraus folgt fast zwangsläufig, dass Lehrkräfte sich (zumindest zu einem gewissen Maß) darin unterscheiden, wie sie die »Anforderungen« im Unterricht und damit bei der Notenvergabe definieren. Zum anderen ist es empirisch gut gesichert, dass sich Lehrkräfte auch darin unterscheiden, in welchem Ausmaß sie bei der Notenvergabe die Leistung in Bezug setzen (a) zur Leistung der anderen Schüler im Klassenverband (soziale Bezugsnorm) oder (b) zur individuellen Leistungsentwicklung einer Schülerin oder eines Schülers (individuelle Bezugsnorm).

Dementsprechend sind die empirischen Befunde zur Objektivität von Schulnoten ernüchternd. Beispielsweise zeigte eine Untersuchung mit den Daten aus PISA-2000, dass Jugendliche mit ähnlichen Mathematikleistungen im PISA-Test im selben Bundesland an unterschiedlichen Gymnasien sehr unterschiedliche Mathematiknoten erhielten (Baumert, Trautwein & Artelt, 2003). Ebenso ergaben zahlreiche Studien, dass dieselbe Schülerarbeit (z. B. ein Aufsatz oder eine Mathematikarbeit) von unterschiedlichen Lehrkräften oft sehr unterschiedlich benotet wird und dass selbst ein und dieselbe Lehrkraft bei wiederholter Benotung derselben Schülerarbeit häufig nicht mit ihren früheren Benotungen übereinstimmt (zsf. Ingenkamp & Lissmann, 2008).

Trotz dieser Mängel konnte in vielen empirischen Studien auch gezeigt werden, dass Abschlussnoten weiterführende Studien- oder Ausbildungserfolge in substantiellem Maße vorhersagen können (Richardson, Abraham & Bond, 2012; Trapmann, Hell, Weigand & Schuler, 2007). Dieser breiten Befundlage zufolge weist also die schulische Leistungsbewertung konvergente, prognostische Validität in Bezug auf spätere Bildungserfolge auf.

Bildungsstandards zur Qualitätssicherung schulischer Diagnostik. Die professionelle Entwicklung von diagnostischen Instrumenten, wie auch professionelle Diagnostik selbst, ist mit einem hohen zeitlichen, finanziellen und personellen Aufwand verbunden. Zudem macht die Diagnostik zwar einen wichtigen, aber eben

doch nur einen Teil der vielfältigen Arbeit von Lehrkräften aus. Daher ist klar, dass von Lehrkräften entwickelte Verfahren der Leistungsüberprüfung (z. B. Klassenarbeiten) und ihre diagnostischen Urteile nicht in vollem Umfang dem Anspruch an professionelle Diagnostik gerecht werden können. Verschiedene Befunde zur Güte von Diagnosen im schulischen Alltag durch Lehrkräfte weisen darauf hin, dass die Güte dieser Urteile oft eingeschränkt ist, was vor dem Hintergrund der oft weitreichenden Konsequenzen für das spätere Leben von Schülern ein Problem darstellt (vgl. Ingenkamp & Lissmann, 2008). Dies ist ein Grund dafür, dass in Deutschland Bildungsstandards eingeführt wurden. Sie sollen dazu beitragen, die diagnostische Kompetenz von Lehrkräften zu verbessern. Bildungsstandards definieren die inhaltlichen Anforderungen in verschiedenen Schulfächern und Kompetenzbereichen (vgl. Abb. 19.5) und konkretisieren diese Anforderungen durch zahlreiche Aufgabenbeispiele und Unterrichtsmaterialien. Zudem erhalten Lehrkräfte für Schüler der 3. und 8. Jahrgangsstufe jährlich Rückmeldungen aus landesweiten Lernstandserhebungen bzw. Vergleichsarbeiten, die auf den Bildungsstandards basieren (Pant, Emmrich, Harych & Kuhl, 2011; s. Abschn. 19.2.1). Diese informieren die Lehrkräfte über die Leistungen ihrer Schüler (z. B. in Form von erreichten Kompetenzstufen) wie auch über den relativen Leistungsstand ihrer Klassen im Vergleich zu anderen Klassen (z. B. Klassen derselben Schule oder allen Klassen derselben Schulart im Land). Auf dieser Datengrundlage können Lehrkräfte ihre klassenspezifischen Leistungsniveaus mit den Anforderungen der Bildungsstandards abgleichen, die Leistungserwartungen mit Kollegen an ihren Schulen abstimmen und Förderbedarf für ihre Klasse identifizieren.

Als weitere Maßnahme zur Qualitätssicherung der Diagnostik durch Lehrkräfte wurden in vielen Bundesländern Konzepte und Methoden pädagogisch-psychologischer Diagnostik als feste Bestandteile in die Lehreraus- und -fortbildung integriert. Hierzu gehört v. a. die Vermittlung methodischer Kenntnisse darüber, wie man Klassenarbeiten konstruiert oder mündliche Prüfungen durchführt, um die Objektivität, Reliabilität und Validität daraus resultierender Messwerte zu verbessern. Informationen hierzu finden sich auch auf den Internetseiten des Projekts UDiKom (www.kmk-udikom.de), das von der KMK gefördert wurde.

19.2 Evaluation

19.2.1 Begriffe, Funktionen und Beispiele

Im allgemeinen Sprachgebrauch bezeichnet der Begriff Evaluation eine »Beurteilung« oder »Bewertung«. Diese Bewertungen zielen nach Kuper (2005) darauf ab, anwendungsbezogene Entscheidungen zu treffen (vgl. Sechrest & Figuerado, 1993). Gegenstand von pädagogisch-psychologischen Evaluationen sind z. B. Interventionsprogramme (s. Abschn. 20.2.2) sowie größere organisatorische Einheiten wie einzelne Schulen, Schularten oder ganze Bildungssysteme, die im Hinblick auf ihre Input-, Prozess- oder Ergebnisqualität beurteilt werden. Die Evaluation von Bildungssystemen im Rahmen großer internationaler Schulleistungsstudien wie PISA, TIMSS (Trends in International Mathematics and Science Study) und IGLU (Internationale Grundschul-Lese-Untersuchung) sowie im Rahmen des Ländervergleichs zur Überprüfung des Erreichens der Bildungsstandards durch das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) wird auch Bildungsmonitoring genannt (s. u.). Am Ende dieses Abschnitts erläutern wir zentrale Konzepte und Methoden der Evaluation im pädagogischen Kontext am Beispiel einer Evaluation von Maßnahmen zur Leseförderung (s. Kasten).

Die auf der Basis von Evaluationen getroffenen Entscheidungen sollen dazu beitragen, die Qualität von pädagogischen Maßnahmen, Einrichtungen oder Bildungssystemen zu sichern oder zu verbessern. Die Entscheidungen haben eine normative Komponente, eine technologische Komponente oder beides (Kuper, 2005). Die normative Komponente betrifft die Bewertung, inwieweit eine bestimmte pädagogische Intervention, eine pädagogische Einrichtung oder ein Bildungssystem vorgegebene Zielkriterien erreicht. Die technologische Komponente hingegen bewertet alternative pädagogische Einrichtungen oder Maßnahmen im Hinblick darauf, inwieweit sie diese Zielkriterien in unterschiedlichem Ausmaß erreichen.

Zentral für Evaluationen im pädagogischen Kontext ist dabei die Anwendung sozialwissenschaftlicher Methoden, um die Entscheidungen evidenzbasiert zu treffen, d. h. auf der Grundlage von Analysen empirischer Daten (Kuper, 2005). Empirische Daten können mithilfe von qualitativen oder quantitativen Methoden oder auch durch Kombination beider Methoden gewonnen werden (Mayring, 2002; Sechrest & Figuerado, 1993, vgl. Abschn. 4.1.2). Dieser Beitrag fokussiert auf Evaluationen, die auf quan-

titativen Methoden beruhen. Ausgehend von diesen Bestimmungsstücken kann pädagogisch-psychologische Evaluation wie folgt definiert werden.

Definition

Evaluationen im pädagogischen Kontext zielen darauf ab, pädagogische Maßnahmen oder Prozesse, Einrichtungen oder Systeme im Bildungswesen zu bewerten. Diese Bewertung dient dazu, evidenzbasierte Entscheidungen zu treffen, um die Qualität im Bildungswesen zu sichern und zu verbessern. Als Entscheidungsgrundlage werden empirische Daten herangezogen, die mit sozialwissenschaftlichen Methoden gewonnen und ausgewertet werden.

Bildungsmonitoring und Qualitätssicherung im Bildungswesen

Input- vs. Output-orientierte Steuerung des Bildungswesens. Mit den internationalen Schulleistungsstudien wie TIMSS und v. a. PISA wurde in Deutschland eine grundsätzliche Wende der Steuerung des Bildungswesens durch Bildungspolitik und Bildungsverwaltung eingeleitet (Klieme et al., 2007). Vor Durchführung dieser Studien erfolgte Steuerung primär über den Input, der solche Elemente wie Lehrpläne und Ausbildungsbestimmungen für Lehrpersonen oder Prüfungsregelungen umfasst. Bei einer solchen Input-orientierten Steuerung sind der Anteil der Bildungsausgaben am Bruttosozialprodukt, die Größe von Schulklassen oder die Anzahl qualifizierter Lehrpersonen wichtige Qualitätskriterien für die Evaluation des Bildungssystems. Eine Output-orientierte Steuerung des Bildungswesens stellt hingegen die Erträge von Bildungsprozessen stärker in den Vordergrund. Bei diesem Ansatz bemisst sich die Qualität eines Bildungssystems v. a. daran, inwieweit Schüler fachspezifische und fächerübergreifende Kompetenzen, motivationale Orientierungen, Werthaltungen und persönliche Eigenschaften erworben haben, die wichtig für lebenslanges Lernen sowie die erfolgreiche Beteiligung in der Berufswelt und an einer demokratischen Gesellschaft sind (Klieme et al., 2007). Steuerung, die sich am Output orientiert, benötigt regelmäßige Rückmeldungen darüber, inwieweit die angestrebten Ziele erreicht werden.

Gesamtstrategie zum Bildungsmonitoring. Um die Datengrundlage für solche Rückmeldungen zu schaffen,

hat die KMK eine Gesamtstrategie zum Bildungsmonitoring verabschiedet, die vier Säulen umfasst (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2006):

- (1) internationale Schulleistungsstudien (PISA, IGLU, TIMSS),
- (2) zentrale Überprüfungen des Erreichens der Bildungsstandards im Ländervergleich,
- (3) auf den Bildungsstandards beruhende flächendeckende Vergleichsarbeiten (VERA) zur Überprüfung der Leistungsfähigkeit aller Klassen und Schulen in der 3. Jahrgangsstufe (VERA 3) und in der 8. Jahrgangsstufe (VERA 8) sowie
- (4) gemeinsame Bildungsberichterstattung von Bund und Ländern (Autorengruppe Bildungsberichterstattung, 2012).

Die Analysen dieser Datenquellen liefern kontinuierlich, aus unterschiedlichen Perspektiven und auf unterschiedlichen Ebenen Informationen über Bildungsprozesse und Bildungserträge sowie deren Rahmenbedingungen.

Schulleistungsstudien. So bieten die internationalen und nationalen Schulleistungsstudien eine Vielzahl an Informationen, die zur Evaluation des deutschen Bildungssystems bzw. der Systeme der einzelnen Bundesländer herangezogen werden können. Solche Standortbestimmungen erscheinen vor dem Hintergrund eines nationalen und internationalen Wettbewerbs notwendig, da die Bildungsqualität in vielfacher Hinsicht eine wichtige Rolle spielt, etwa aufgrund ihrer Bedeutung für die Ausbildung von Fachkräften (Bos, Postlethwaite & Gebauer, 2009). Ein datengestützter Vergleich mit anderen Bildungssystemen kann zudem Aufschluss geben über wichtige bildungsplanerische Fragen (z. B. »Welche Rolle spielt die Klassengröße für den Kompetenzerwerb?«) sowie Hinweise geben, ob politisch heftig umkämpfte Themen, wie die Frage nach der besten Schulstruktur (z. B. gegliedertes Schulsystem vs. Gesamtschulsystem), überhaupt die unterstellte Relevanz für die Bildungserträge haben (Bos et al., 2009). Weiterhin liefern Leistungserhebungen auf der Grundlage

von Bildungsstandards bzw. Vergleiche mit den Ergebnissen aus früheren Jahren (Trendanalyse) Hinweise darauf, ob politische Zielsetzungen zur Standardsicherung tatsächlich erreicht wurden und bildungspolitische Entscheidungen zielführend waren.

Vergleichsarbeiten. Einen wesentlichen Beitrag zur Qualitätsentwicklung im Bildungswesen sollen weiterhin die Vergleichsarbeiten in den Klassenstufen 3 (VERA 3) und 8 (VERA 8) leisten (vgl. Pant et al., 2011). In VERA 3 und VERA 8 bearbeiten alle Schüler der jeweiligen Bundesländer Testaufgaben, die auf den Bildungsstandards und den auf der Grundlage der Bildungsstandards entwickelten Kompetenzstufenmodellen basieren. In der Regel führen die Lehrkräfte dieser Schüler die Tests in ihren Klassen durch und werten die Antworten auf die Testaufgaben nach vorgegebenen Korrekturanweisungen aus. Die Lehrkräfte erhalten anschließend eine externe Rückmeldung zum Kompetenzstand der Schüler in ihren Klassen, die Schulleitungen ein Feedback zum Leistungsstand ihrer Schule insgesamt sowie zur schulinternen Leistungshomogenität in den Parallelklassen und die Eltern eine lehrkraftunabhängige Information über den Leistungsstand ihrer Kinder.

Weiterhin werden in vielen Bundesländern sog. »faire Vergleiche« durchgeführt, bei denen in der Leistungsrückmeldung die Ergebnisse der einzelnen Klasse oder Schule in Bezug gesetzt werden zur Leistung von Klassen oder Schulen, deren Schülerschaft hinsichtlich leistungsrelevanter Merkmale (z. B. sozioökonomischer Hintergrund und Bildungshintergrund der Eltern, Migrationsstatus) ähnlich zusammengesetzt ist (Kuhl, Lenkeit, Wendt & Pant, 2011). Insgesamt sollen damit die Ergebnisse von VERA 3 und VERA 8 eine inhaltliche Auseinandersetzung mit den kompetenzorientierten Bildungsstandards durch Lehrkräfte, Schulleitungen und Eltern fördern. Diese Informationen können Lehrkräfte und Schulleitungen als wertvolle Impulse für die Sicherung der Qualität von diagnostischen Urteilen sowie die Unterrichts- und Schulentwicklung nutzen, wie sie in Konzepten einer eigenverantwortlichen Schule eingefordert werden.

Studie

Interventionsstudie zur Evaluation eines Programms der Leseförderung

Die Frage, wie sich die Lesekompetenz von Kindern und Jugendlichen gezielt fördern lässt, ist in zahlreichen Studien untersucht worden. Als Beispiel für eine solche Untersuchung soll eine neuere Studie von Streblow, Schiefele und Riedel (2012) herangezogen werden, in der die Effekte eines Trainings zur Förderung von Lesekompetenz und Lesemotivation (»LekoLemo«) überprüft wurden. Die Trainingsaktivitäten zur Förderung der Lesekompetenz bezogen sich auf die Vermittlung und das Anwenden von vier Lesestrategien: Aktivierung des Vorwissens, Wichtiges unterstreichen, Methoden zum Umgang mit Verständnisschwierigkeiten und Wichtiges zusammenfassen. Um die Lesemotivation zu fördern, wurden für die in der Förderung verwendeten Sachtexte Themen ausgewählt, die auf die Interessen der Zielgruppe (Schüler der 7. Jahrgangsstufe) abgestimmt waren. Die Jugendlichen konnten aus dem Angebot dieser Texte frei

wählen. Weiterhin übernahmen die Schüler in der Förderung eine Expertenrolle bei der Bearbeitung der Texte zu den von ihnen gewählten Themen, und es wurde ihnen der Nutzen des Trainings ausführlich erläutert. Als motivierend wurde zudem die Arbeit in Kleingruppen erachtet und der häufige Einsatz positiver Rückmeldungen durch die Personen, die das Training durchführten.

Am Training nahmen Schüler aus 13 Klassen an vier Realschulen und an jeweils einer Haupt- und Gesamtschule teil. In Schulen, in denen sich zwei oder drei Klassen für die Teilnahme an der Studie zur Verfügung gestellt hatten, wurde jeweils eine Klasse der Kontrollgruppe zugewiesen. Nach Ausschluss von Jugendlichen mit unzureichenden Daten umfasste die Trainingsgruppe $n = 150$ und die Kontrollgruppe $n = 85$ Schüler. Die Förderung erfolgte in Kleingruppen von 4 bis 6 Schülern durch Studierende der Psychologie, der

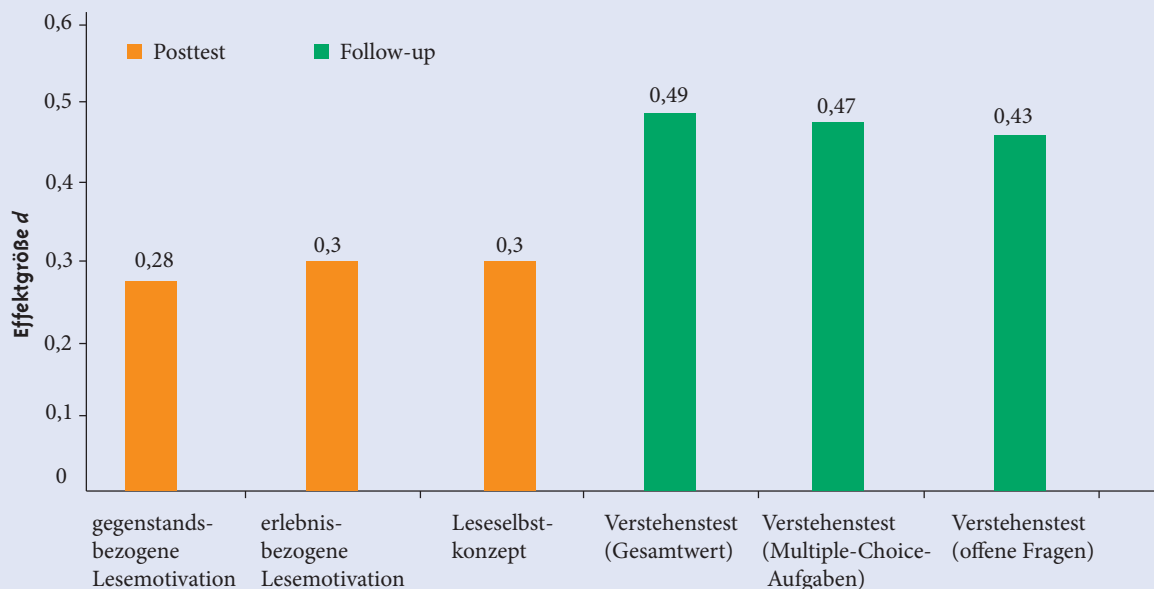


Abbildung 19.7 Ergebnisse der Evaluation des LekoLemo-Trainings (Streblow et al., 2012). Die Trainingswirkung auf die Lesekompetenz wurde ausgewiesen in Bezug auf den Gesamtwert im Verstehenstest sowie auf die Leistungen bei offenen Fragen und Multiple-Choice-Aufgaben. Streblow et al. berechneten für alle Ergebnisse eine standardisierte Effektgröße d_{korrt} , die im Prätest bestehende Leistungsunterschiede zwischen den Gruppen berücksichtigt. Die Autoren interpretieren Werte von d_{korrt} zwischen 0,20 und 0,50 als »klein« und zwischen 0,50 und 0,80 als »mittel«. Einige Werte von d_{korrt} fehlen in dieser Abbildung, da Streblow et al. nur Effektgrößen berichteten, die statistisch bedeutsam waren

Pädagogik oder des Lehramts. In den Förderstunden wurde jeweils ein Text anhand eines vorgegebenen Strategiefahrplans bearbeitet. Das Training umfasste acht Sitzungen von jeweils 90 Minuten Dauer über einen Zeitraum von bis zu drei Wochen. Die Kontrollgruppe nahm während der Förderung am regulären Unterricht teil.

In der Studie wurde ein sog. Prä-Post-Kontrollgruppensdesign mit Follow-up umgesetzt. Da die Zuweisung zur Trainings- und Kontrollgruppe nicht per Zufall erfolgte, handelt es sich um eine quasi-experimentelle Studie. Die Erfassung von Lesekompetenz, Lesemotivation und lesebezogenem Selbstkonzept er-

folgte unmittelbar vor (Prätest) und direkt nach der Förderung (Posttest) sowie nochmals zwei Monate nach Abschluss des Trainings (Follow-up-Test). Die statistischen Analysen, in denen die Testergebnisse der Trainingsgruppe mit den Testergebnissen der Kontrollgruppe verglichen wurden, zeigten, dass das Training statistisch bedeutsam positive Effekte auf die Lesekompetenz hatte, die auch noch zwei Monate nach der Förderung nachgewiesen werden konnten. Für die Lesemotivation und das Leseselbstkonzept konnten dagegen zwar kurzfristige positive Effekte der Förderung identifiziert werden, diese waren jedoch nicht nachhaltig (s. Abb. 19.7).

19.2.2 Ablauf

Unabhängig von ihrer Zielsetzung befinden sich Evaluationen generell im Schnittbereich von grundlagenwissenschaftlicher und anwendungsbezogener Forschung. Der Ablauf von Evaluationen (s. Abb. 19.8) orientiert sich daher einerseits am typischen Ablauf von sozialwissenschaftlichen Studien (s. Punkte 2 bis 6), berücksichtigt dabei aber andererseits die praktischen Rahmenbedingungen der Evaluation und das Ziel, anwendungsbezogene Empfehlungen abzuleiten oder gar Entscheidungen zu treffen (s. Punkte 1 und 7).

19.2.3 Qualitätsstandards zur Durchführung von Evaluationsstudien

Worauf ist bei der Durchführung von Evaluationen zu achten? Unabhängig davon, ob qualitative oder quantitative Methoden eingesetzt werden, zeichnen sich professionell durchgeführte Evaluationen dadurch aus, dass mehrere Qualitätsstandards berücksichtigt werden, die wegweisend für die Planung einer Evaluationsstudie sind (s. Abb. 19.8). Diese Standards zielen auf vier zentrale Aspekte ab (Deutsche Gesellschaft für Evaluation, 2008):

- (1) Nützlichkeit
- (2) Genauigkeit
- (3) Durchführung
- (4) Fairness

Nützlichkeitsstandards. Damit ist gemeint, dass sich die Evaluation an ihren jeweils vorab definierten Zielen und Fragestellungen orientiert sowie den Informationsbedarf der Auftraggeber erfüllt. Typische Zielkriterien

im pädagogischen Kontext sind zum Beispiel die Förderung von

- ▶ schulfachbezogenen Kompetenzen (z. B. im Bereich Lesen, in Mathematik oder in den Naturwissenschaften),
- ▶ allgemeinen oder spezifischen kognitiven Fähigkeiten (z. B. allgemeine Problemlösefähigkeit; kognitive Lernstrategien),
- ▶ sozialen Kompetenzen (z. B. Bewältigung sozialer Konfliktsituationen),
- ▶ motivationalen oder affektiven Merkmalen (z. B. schulische und berufliche Interessen, schulbezogene Selbstkonzepte oder Reduktion von Leistungsängstlichkeit) oder
- ▶ pädagogisch wünschenswerten Persönlichkeitsmerkmalen (z. B. Gewissenhaftigkeit).

Genauigkeitsstandards. Informationen, die im Rahmen einer Evaluation zur Beantwortung der Fragestellungen erhoben werden, sollen reliabel und valide sein. Für die professionelle Auswahl von geeigneten Messinstrumenten ist daher die Berücksichtigung der Gütekriterien maßgeblich (s. Abschn. 19.1.4). Oftmals sind jedoch die Fragestellungen einer Evaluation sehr spezifisch und es sind keine geeigneten Messinstrumente verfügbar, die zur Erhebung der relevanten Informationen eingesetzt werden könnten. In solchen Fällen ist es erforderlich, neue Messinstrumente zu entwickeln und deren Gütekriterien vor Durchführung der Evaluation in sog. Pilotstudien empirisch zu überprüfen.

Durchführbarkeitsstandards. Es ist darauf zu achten, dass die Belastung, die mit der Datensammlung verbunden ist, für alle Beteiligten in einem angemessenen Verhältnis zum Nutzen der Evaluation steht.

Fairnessstandards. In einer Evaluation soll mit allen Akteuren respektvoll und fair umgegangen werden. Dies erfordert, dass die Datensammlung den ethischen Richtlinien für Forschung mit Menschen entsprechend erfolgt und relevante Datenschutzaufgaben beachtet werden (Deutsche Gesellschaft für Psychologie & Berufsverband Deutscher Psychologinnen und Psychologen, 2004). Ebenso soll der Fairnessstandard sicherstellen, dass pädagogische Interventionen und Einrichtungen möglichst vollständig, unvoreingenommen und

ergebnisoffen überprüft werden, damit ihre Stärken weiter ausgebaut und ihre Schwachpunkte reduziert werden können.

19.2.4 Die Konzeption von Evaluationsstudien

Die empirischen Ergebnisse von Evaluationsstudien im pädagogischen Kontext dienen dazu, evidenzbasierte Entscheidungen zu treffen, um die Qualität im Bildungswesen zu sichern und zu verbessern. Um die Aus-

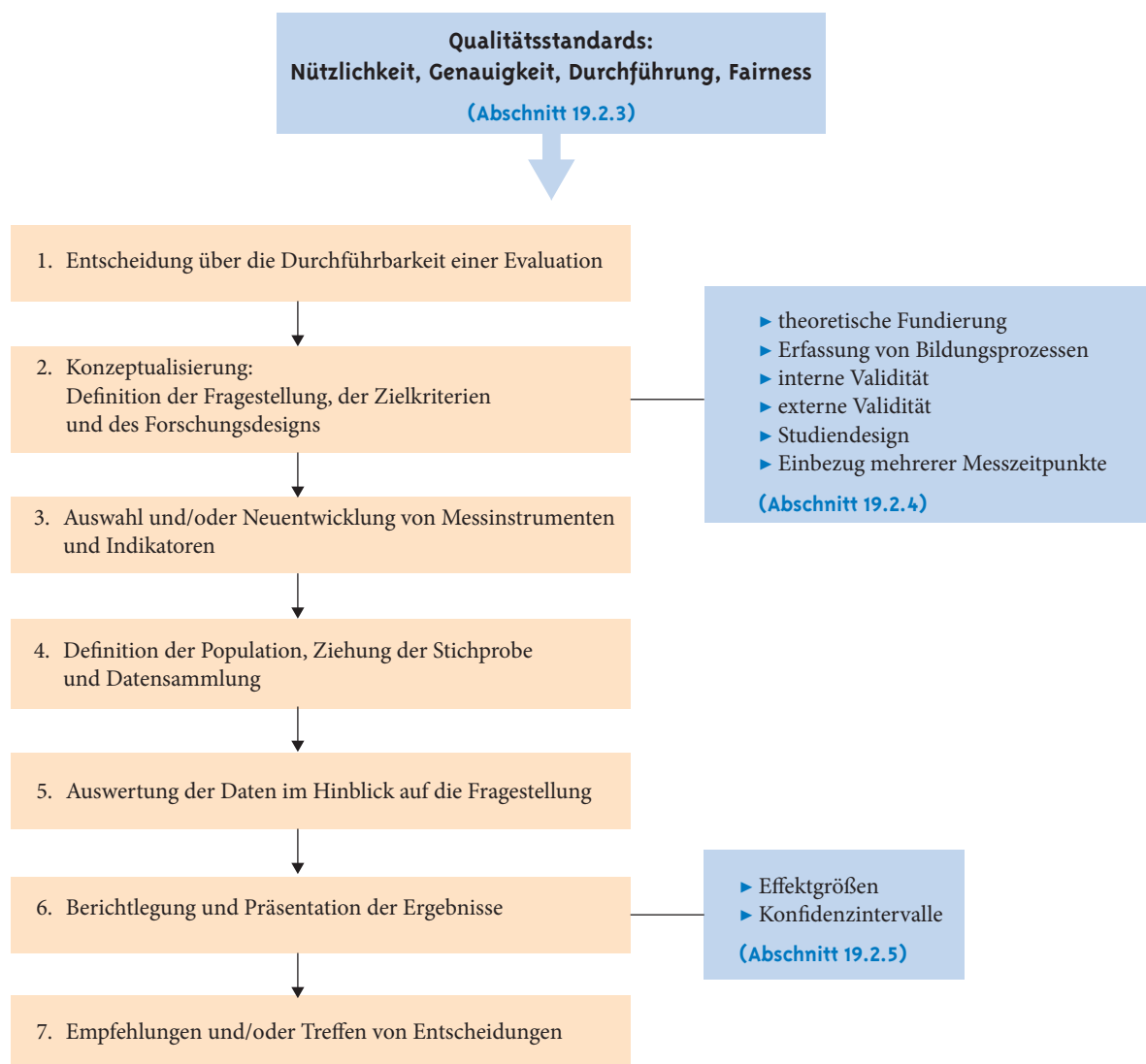


Abbildung 19.8 Ablauf von Evaluationen und zentrale Konzepte und Methoden. In Klammern sind die Abschnitte genannt, in denen die jeweiligen Begriffe etc. beschrieben werden

sagekraft der empirischen Daten zu gewährleisten, sind bei der Planung von Evaluationsstudien mehrere theoretische und methodische Konzepte zu berücksichtigen, die sich (teilweise) inhaltlich überlappen und wechselseitig bedingen:

- ▶ theoretische Fundierung
- ▶ Erfassung von Bildungsprozessen
- ▶ interne Validität
- ▶ externe Validität
- ▶ Studiendesign
- ▶ Einbezug mehrerer Messzeitpunkte

Theoretische Fundierung

In wissenschaftlich fundierten Evaluationen werden pädagogische Interventionen auf der Basis von Theorien konzipiert oder es wird bei bereits bestehenden pädagogischen Programmen auf Theorien zurückgegriffen, um Annahmen über ihre Wirkmechanismen zu generieren. In die Konzeption des LekoLemo-Trainings beispielsweise wurde ein Bündel von Maßnahmen integriert, für die in früheren, theoretisch fundierten Studien empirisch nachgewiesen werden konnte, dass sie sich positiv auf die Entwicklung der Leseleistung und -motivation auswirken. Hierzu gehören die Vermittlung von effektiven Lesestrategien, die Berücksichtigung bereits vorhandener thematischer Interessen der Jugendlichen, die Übernahme der Expertenrolle durch die teilnehmenden Kinder, die Arbeit in Kleingruppen sowie positive Rückmeldungen. Weiterhin betrifft die theoretische Fundierung der Evaluation die Auswahl der Zielkriterien, die zur Bewertung der pädagogischen Intervention bzw. der Einrichtungen oder Systeme im Bildungswesen herangezogen werden sollen. Da das LekoLemo-Training auf eine Förderung der Leseleistung und -motivation abzielt, wurden für die Evaluation des Trainings sowohl kognitive Maße (Lesekompetenz und Lesestrategien) als auch motivationale Maße (Lesemotivation, Leseselbstkonzept) eingesetzt.

Erfassung von Bildungsprozessen

Um die Wirkung einer pädagogischen Intervention zu evaluieren, vergleicht man häufig die Ergebnisse der Interventionsgruppe mit den Ergebnissen einer Kontrollgruppe. Für die Interpretation der Ergebnisse ist es dabei wichtig, dass für alle experimentell manipulierten Bedingungen (also Experimentalgruppe und Kontrollgruppe) bzw. für alle untersuchten pädagogischen Interventionen detailliert erfasst wird, in welcher Weise und in welchem

Ausmaß die theoretisch postulierten Bildungsprozesse tatsächlich implementiert wurden (Cook & Shadish, 1994; Lipsey & Cordray, 2000). Dies wird als Grad der Implementationstreue bezeichnet. Durch die sorgfältige Analyse der tatsächlich implementierten Bildungsprozesse in der Interventionsgruppe (z. B. durch Videoanalysen) kann beispielsweise festgestellt werden, ob das Ausbleiben von Effekten der pädagogischen Intervention darauf zurückzuführen ist, dass in der Interventionsgruppe pädagogisch wirksame Bildungsprozesse nicht stattfanden, obwohl sie im Rahmen der Evaluation geplant waren. Solche Analysen können damit auch wichtige Hinweise darauf geben, welche der umgesetzten Bildungsprozesse wirksam waren.

Auch wenn eine pädagogische Intervention theoriekonform implementiert wurde, bleibt es ohne empirische Überprüfung Wunschdenken, dass für alle Personen das Ausmaß und die Intensität der Bildungsprozesse identisch waren, also alle Personen die gleiche »Interventionsdosis« erhalten haben (vgl. Lipsey & Cordray, 2000). Beispielsweise haben sicherlich einige Jugendliche aus verschiedenen Gründen nicht an allen Sitzungen des LekoLemo-Trainings teilnehmen können. Es ist daher eine offene Frage, ob für diese Jugendlichen der positive Effekt des Trainings aufgrund des geringeren Umfangs des Trainings geringer ausgefallen ist. Bei Analysen der Wirksamkeit von pädagogischen Interventionen sollten daher zusätzlich zu ihrer generellen Wirksamkeit auch Effekte in Abhängigkeit vom Ausmaß und von der Intensität der Bildungsprozesse analysiert werden.

Interne Validität

Um pädagogische Maßnahmen und pädagogische Einrichtungen und deren Wirkung detailliert zu bewerten, ist es wichtig, die Frage der Kausalität zu klären. In der Regel wird davon ausgegangen, dass die Zielkonstrukte, die im Rahmen einer Evaluation erfasst worden sind (z. B. die Lesekompetenz), von der untersuchten Intervention (z. B. dem LekoLemo-Training), pädagogischen Einrichtung o. Ä. maßgeblich beeinflusst wurden. Nur wenn dies tatsächlich der Fall ist, lassen sich die Ergebnisse der Evaluation auch tatsächlich der jeweiligen Maßnahme zuschreiben und für ihre Optimierung nutzen. Diese Feststellung scheint auf den ersten Blick trivial zu sein, sie zählt jedoch zu den größten Herausforderungen von Evaluationen im pädagogischen Kontext und wird in der Praxis häufig unzureichend berücksichtigt.

Übersicht

Bedingungen für kausale Schlüsse

Nach Cohen, Cohen, Aiken und West (2003) müssen für die Aussage »X hat eine kausale Wirkung auf Y (z.B. das LekoLemo-Training auf die Lesekompetenz)« vier Bedingungen erfüllt sein:

- (1) Die Ursache X geht der Wirkung Y zeitlich voraus.
- (2) Es gibt eine theoretische Erklärung dafür, warum X und Y statistisch miteinander zusammenhängen sollen.
- (3) X und Y sind empirisch nachweisbar statistisch assoziiert (z.B. korreliert).
- (4) Die Wirkung von X auf Y bleibt auch dann bestehen, wenn andere Ursachen von Y berücksichtigt werden.

Die Untersuchung von kausalen Mechanismen erfordert eine theoretische Fundierung dieser Mechanismen, aber auch theoretische Überlegungen zu möglichen Alternativerklärungen. Ebenso erfordert sie den Nachweis eines statistischen Zusammenhangs von X und Y sowie die Anwendung von statistischen Modellen, die Alternativerklärungen explizit berücksichtigen und so weit wie möglich ausschließen. Vor allem aber erfordern kausale Schlussfolgerungen ein methodisch gut fundiertes Design der Evaluationsstudie (Shadish, Cook & Campbell, 2002). Ein zentrales Konzept ist dabei die interne Validität.

Definition

Interne Validität von Evaluationsstudien: Eine Evaluationsstudie wird dann als intern valide bezeichnet, wenn ihre Ergebnisse die eindeutige Schlussfolgerung zulassen, dass X eine kausale Wirkung auf Y hat.

Gefährdungen der internen Validität

Bei der Planung des Designs von Evaluationsstudien ist es wichtig, Faktoren zu berücksichtigen, welche die interne Validität gefährden können (für eine ausführliche Diskussion s. Shadish et al., 2002). Hierzu gehören u. a.

- ▶ Selektionseffekte,
- ▶ Retest-Effekte und
- ▶ Stichprobenausfall.

Selektionseffekte. Ein Selektionseffekt liegt vor, wenn sich Personen in einer pädagogischen Maßnahme, die in der Evaluation bewertet werden soll, in wichtigen Merkmalen systematisch von Personen unterscheiden, die diese Maßnahme nicht erhalten. Dann besteht die Gefahr, dass nicht die pädagogische Maßnahme, sondern diese Merkmale den beobachteten Effekt bewirkt haben. Beispielsweise unterschieden sich die Jugendlichen, die am LekoLemo-Training teilnahmen, auch schon vor Trainingsbeginn von Jugendlichen, die regulären Unterricht erhielten, in leistungsrelevanten Merkmalen (z.B. der Lesekompetenz). Der beobachtete Kompetenzunterschied im Lesen, der im Posttest beobachtet wurde, dürfte damit also nicht allein auf das Training, sondern auch auf Kompetenzunterschiede vor Beginn des Trainings zurückgehen. Das beste Mittel, um solchen Selektionseffekten zu begegnen, ist die zufällige Zuweisung (Randomisierung) von Personen auf die Untersuchungsbedingungen (z.B. die zufällige Zuweisung von Schülern zum LekoLemo-Training und zu regulärem Unterricht). Ist dies nicht möglich, kann (wie in der Studie von Streblow et al., 2012) versucht werden, durch geeignete statistische Modellierungen eine Abschätzung der kausalen Wirkung zu erreichen (Cook & Shadish, 1994).

Retest-Effekte. Eine weitere Gefährdung der internen Validität ist potenziell mit der wiederholten Bearbeitung eines Tests (Retest) verbunden, die dazu führen kann, dass nachfolgend erhobene Testwerte beeinflusst werden, z.B. aufgrund von Übungseffekten, Ermüdungseffekten oder zunehmender Vertrautheit mit dem Messinstrument. Wurde also etwa die Lesekompetenz von Schülern vor und nach Durchführung des Förderprogramms mit demselben Test gemessen, so ist es möglich, dass die beobachteten Veränderungen nicht primär auf die Intervention, sondern auf Retest-Effekte durch die zunehmende Vertrautheit mit dem Instrument (z.B. Erinnerung daran, wie bestimmte Leseaufgaben zu lösen sind) zurückzuführen sind.

Stichprobenausfall. Der Verlust von Teilnehmern im Verlauf einer Evaluation kann zu einer Über- oder Unterschätzung der kausalen Effekte führen, wenn der Stichprobenausfall systematisch mit der Zugehörigkeit zu experimentellen Bedingungen zusammenhängt. Nehmen z.B. Kinder der Fördergruppe, die im Prätest einer Evaluation eher schwache Leistungen aufweisen, nach der Förderung nicht mehr am Posttest teil, wird bei unzureichender Berücksichtigung dieses Ausfalls die

mittlere Leistung der Fördergruppe im Posttest überschätzt. Dies hätte zur Folge, dass auch der Effekt der pädagogischen Förderung überschätzt würde.

Externe Validität

Schlüsse aus Evaluationsstudien, die eine hohe interne Validität aufweisen, basieren meist auf experimentellen Designs (vgl. Abschn. 4.3.3). In diesen Studien wird sichergestellt und streng kontrolliert, dass nur diejenigen Faktoren implementiert werden, von denen angenommen wird, dass sie wirksam sind, und dass die Studienbedingungen dem experimentellen Protokoll entsprechen. Das damit verbundene Ziel, kausale Effekte zu isolieren, steht in einem gewissen Spannungsverhältnis mit der Generalisierbarkeit der Ergebnisse, die als externe Validität bezeichnet wird (Shadish et al., 2002).

Definition

Externe Validität von Evaluationsstudien: Eine Evaluationsstudie wird dann als extern valide bezeichnet, wenn die auf ihren Ergebnissen basierenden Schlussfolgerungen nicht auf die spezifischen Bedingungen der Untersuchung begrenzt sind, sondern sich auf andere Personengruppen, andere vergleichbare pädagogische Maßnahmen oder Einrichtungen, andere Zielkriterien und zukünftige Zeitpunkte generalisieren lassen.

Gefährdungen der externen Validität. Gefährdungen der externen Validität betreffen letztlich immer Wechselwirkungen zwischen dem untersuchten Effekt und den Randbedingungen der Untersuchung. Welchen Effekt eine pädagogische Maßnahme hat, kann u. a. abhängig sein von bestimmten Merkmalen der Stichprobe, von der Durchführung der Studie in bestimmten pädagogischen Einrichtungen, von der Verwendung bestimmter Zielkriterien, der Anwendung bestimmter Methoden, mit denen die untersuchten Konstrukte erfasst wurden, oder vom historischen Zeitpunkt der Untersuchung. Einige der Gefährdungen externer Validität überschneiden sich auch mit Gefährdungen der internen Validität.

So blieben beispielsweise in der Studie von Streblow et al. (2012) einige Fragen zur Generalisierbarkeit der Befunde zum LekoLemo-Training offen. Am Training nahmen überwiegend Jugendliche aus Realschulen teil, sodass unklar ist, ob das Training auch für sehr schwa-

che Leser geeignet ist, die sich eher in Haupt- und Gesamtschulen befinden. Zudem wurde die Förderung von intensiv geschulten Studierenden durchgeführt. Dies wirft die Frage auf, ob das LekoLemo-Training auch für die breite Umsetzung im schulischen Alltag geeignet ist, wenn Lehrkräfte deutlich weniger umfassend in der Durchführung des Trainings geschult werden können.

Um solchen Bedrohungen der externen Validität effektiv zu begegnen, sind möglichst für die Zielpopulation der Evaluation repräsentative Stichproben, der Einbezug einer größeren Zahl unterschiedlicher pädagogischer Einrichtungen, die Erfassung einer möglichst großen Anzahl an Zielkriterien, die Anwendung möglichst vielfältiger Erhebungsmethoden und der Einbezug mehrerer Messzeitpunkte notwendig. Da dies oft nicht innerhalb einer Studie gewährleistet werden kann, sollten im Idealfall Replikationsstudien durchgeführt werden, die solche Aspekte der Durchführungsbedingungen variieren.

Wenn zu einer Fragestellung (z. B. zu Effekten der systematischen Förderung von Lesestrategien auf Lesekompetenz) bereits mehrere Evaluationsstudien vorliegen, kann mit einer quantitativen Synthese dieser Einzelstudien in Form einer Metaanalyse die Stabilität der Effekte über verschiedene Variationen der Designelemente hinweg untersucht werden (vgl. Abschn. 4.5.2). Dabei lässt sich auch prüfen, inwieweit bestimmte Designelemente, wie etwa das Alter der Schüler oder die Dauer der Förderung, die Größe der Effekte beeinflussen (Cooper, Hedges & Valentine, 2009; Cortina & Pant, 2009). Metaanalysen sind also besonders hilfreich, um Fragen zur externen Validität von Befunden zu analysieren.

Studiendesign

Um intern und extern valide Schlussfolgerungen auf der Grundlage von Evaluationsstudien treffen zu können, ist es unerlässlich, ein für die Beantwortung der jeweiligen Fragestellung geeignetes Forschungs- oder Studiendesign umzusetzen (vgl. Abschn. 4.3). Allgemein werden drei wichtige Typen von Studiendesigns unterschieden (Shadish et al., 2002):

- ▶ experimentelle Designs
- ▶ quasi-experimentelle Designs und
- ▶ korrelative Designs.

Experimentelle Designs. In experimentellen Designs werden die Interventionen experimentell manipuliert. Weiterhin zeichnen sich diese Designs dadurch aus, dass

Personen zufällig zu den experimentell manipulierten Bedingungen (z. B. Experimental- vs. Kontrollgruppe) bzw. zu den pädagogischen Interventionen, die verglichen werden sollen, zugewiesen werden (Randomisierung). Experimentelle Designs gelten als »Königsweg« in der Evaluationsforschung, denn bei erfolgreicher Randomisierung unterscheiden sich die experimentellen Gruppen nur darin, ob bzw. welcher pädagogischen Intervention sie ausgesetzt werden. Viele Gefährdungen der internen Validität (z. B. Selektionseffekte) können so (weitgehend) ausgeschlossen werden.

Quasi-experimentelle Designs. In quasi-experimentellen Designs werden zwar auch die Interventionen experimentell manipuliert, jedoch erfolgt die Zuweisung von Studienteilnehmern nicht nach dem Zufallsprinzip. Dieses Design ist immer dann von Interesse, wenn aus pragmatischen oder ethischen Gründen keine randomisierte Zuweisung von Personen zu den Untersuchungsbedingungen möglich ist. So konnten beispielsweise in der quasi-experimentellen Studie von Streblow et al. (2012) die Schüler nicht zufällig den Trainings- und Kontrollgruppen zugewiesen werden. Deshalb ließ sich anhand des quasi-experimentellen Untersuchungsdesigns dieser Studie nicht mit Sicherheit die Frage beantworten, ob der beobachtete Leistungsunterschied zwischen Trainings- und Kontrollgruppe allein auf das LekoLemo-Training zurückzuführen war.

In quasi-experimentellen Studien ist der Zuweisungsmechanismus, der die Zugehörigkeit zu den experimentellen Gruppen bestimmt, nicht vollständig bekannt. Daher sollten zumindest diejenigen individuellen Merkmale der Studienteilnehmer erfasst werden, die systematisch mit der Gruppenzugehörigkeit variieren könnten (z. B. beim LekoLemo-Training die Lesekompetenz und -motivation sowie der sozioökonomische Hintergrund der Schüler vor Trainingsbeginn). Damit wird es möglich, den Zuweisungsmechanismus (nachträglich) statistisch zu modellieren, um auf diese Weise die Wirkung der pädagogischen Intervention selbst abzuschätzen zu können. So wurden zur Evaluation der Effekte des LekoLemo-Trainings mehrere leistungsrelevante Merkmale in einem komplexen statistischen Analysemodell berücksichtigt. Trotz des empirisch nachgewiesenen Effekts bleibt jedoch bei einer nachträglichen statistischen Modellierung immer die Unsicherheit bestehen, ob tatsächlich alle relevanten Merkmale, die für den Effekt verantwortlich sein könnten, erfasst und kontrolliert wurden.

Korrelative Designs. In korrelativen Designs werden weder die Bedingungen experimentell manipuliert noch erfolgt die Zuweisung von Personen zu Bedingungen randomisiert. Diese Designs sind von Interesse, wenn die Untersuchungsbedingungen nicht beeinflusst werden können, sondern natürlich vorzufindende Bedingungen miteinander verglichen werden sollen. Dies ist z. B. der Fall bei der Evaluation einer bestimmten Einzelschule (Watermann, Thurn, Tillmann & Stanat, 2005) oder bei Studien zur Wirkung schulischen Unterrichts (Clausen, 2002). Zur Auswertung solcher Designs werden die (im weiteren Sinne) korrelativen Zusammenhänge (daher auch der Name des Studiendesigns) zwischen den natürlich auftretenden Bedingungen und den Zielkriterien analysiert. Obwohl diese Studien die methodisch schwächsten sind, ist das korrelative Design relativ häufig die Methode der Wahl, wenn ökologisch valide Lernumwelten untersucht werden sollen.

Definition

Ökologische Validität bedeutet, dass die gesamte Forschungsanlage (inklusive Personen, eingesetzte Erhebungsmaterialien, pädagogische Interventionen usw.) die realen Bedingungen bestmöglich repräsentiert bzw. die Studie unter realen Bedingungen stattfindet.

Trotz der potenziellen Einschränkungen der internen Validität haben methodisch fundierte korrelative Studien (welche den Zuweisungsmechanismus zu Bedingungen statistisch modellieren) den Vorteil, dass ihre Studienanlage die Generalisierbarkeit der Ergebnisse stützt und einigen Gefährdungen der externen Validität effektiv begegnet.

Längsschnittdesign mit mehreren Messzeitpunkten

Zur Untersuchung der meisten Fragestellungen im Kontext pädagogisch-psychologischer Evaluationsstudien ist der Einbezug mehrerer Messzeitpunkte in einem Längsschnittdesign erforderlich, um intern bzw. extern valide Interpretationen der Ergebnisse zu ermöglichen. Dies wollen wir in Anlehnung an die Arbeit von Raubenbush (2001) erläutern, indem wir die Ergebnisse aus fiktiven weiteren quasi-experimentellen Evaluationsstudien des LekoLemo-Trainings diskutieren (s. Abb. 19.9). Wie in der Originalstudie von Streblow et al. (2012) erfolgte in diesen fiktiven Studien die Zuweisung der Jugendlichen zu Trainings- und Kontrollgruppe

nicht randomisiert, was dazu führte, dass die Lesekompetenz der Jugendlichen der Trainingsgruppe bereits im Prätest (Messzeitpunkt T_1) im Mittel stärker ausgeprägt war als die Lesekompetenz der Jugendlichen in der Kontrollgruppe. Ebenso zeigte sich (ebenfalls wie in der Originalstudie), dass das mittlere Niveau der Lesekompetenz der Trainingsgruppe im Posttest (T_2) über dem der Kontrollgruppe lag.

Um zu bestimmen, wie groß der potenzielle Effekt des LekoLemo-Trainings war, wurde in allen (fiktiven) Beispiel-Studien zuerst für die Trainings- und Kontrollgruppe die mittlere Leistungsentwicklung als Differenz zwischen Prä- und Posttest berechnet. Im nächsten Schritt wurde der Unterschied in der Leistungsentwicklung (Δ) zwischen Trainings- und Kontrollgruppe bestimmt. In allen Studien resultierte der gleiche Unter-

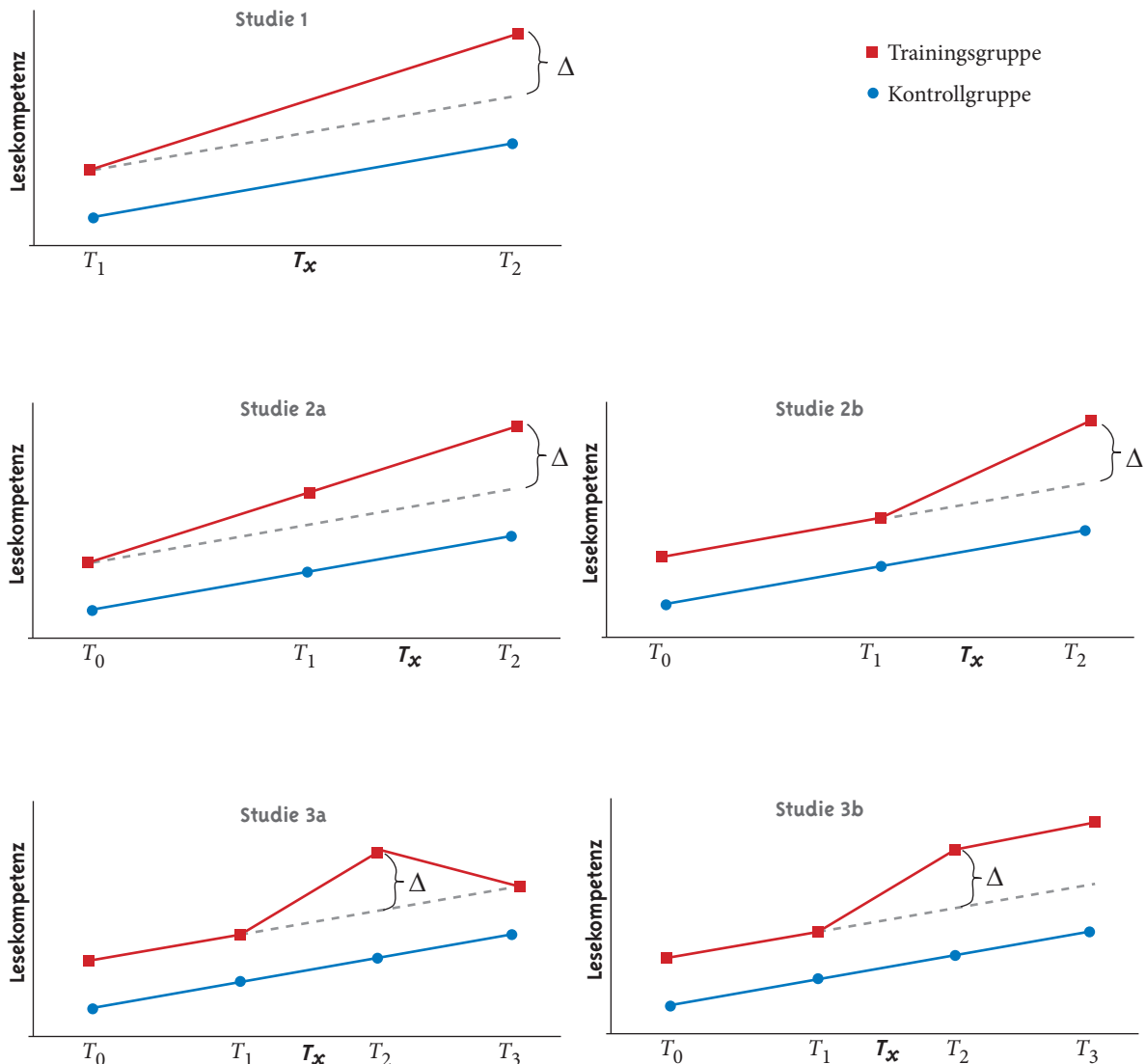


Abbildung 19.9 Fünf fiktive quasi-experimentelle Studien zur Evaluation eines Trainings zur Förderung der Lesekompetenz. T_x = Das Training fand zwischen T_1 und T_2 statt. Δ = Unterschied in der Leistungsentwicklung zwischen Trainings- und Kontrollgruppe. Δ ist in allen Studien gleich groß, aber im Hinblick auf die Wirkung des Trainings anders zu interpretieren

schied Δ in der Leistungsentwicklung. Zwei wichtige Fragen sind nun zu klären:

- (1) War der Unterschied in der Leistungsentwicklung kausal auf das LekoLemo-Training zurückzuführen?
- (2) War der (potenzielle) Effekt des Trainings nachhaltig? Die Studien unterschieden sich darin, wie gut sie diese Fragen empirisch beantworten konnten.

Prä-Post-Design. In Studie 1 (s. Abb. 19.9) wurde die Lesekompetenz nur im Rahmen eines Prätests (T_1) und eines Posttests (T_2) erfasst; das LekoLemo-Training wurde zwischen T_1 und T_2 durchgeführt. Ohne Randomisierung von Jugendlichen auf die beiden Gruppen muss man davon ausgehen, dass sich die Leistungsentwicklung der beiden Gruppen vor Durchführung des Trainings nicht voneinander unterschieden hatte, damit man den Unterschied in der Leistungsentwicklung Δ kausal auf das LekoLemo-Training zurückführen kann (Raudenbush, 2001). Diese Annahme kann auf der Grundlage des in Studie 1 verwendeten quasi-experimentellen Designs jedoch nicht empirisch überprüft werden.

Prä-Post-Design mit zusätzlichem Prätest. Um dieser potenziellen Gefährdung der internen Validität effektiv zu begegnen, wurde daher in den Studien 2a und 2b jeweils ein zusätzlicher Prätest (Messzeitpunkt T_0 ; s. Abb. 19.9) durchgeführt. Auf diese Weise kann die Entwicklung der Lesekompetenz untersucht werden, die nicht vom LekoLemo-Training beeinflusst wurde, denn anhand der zusätzlich gewonnenen Daten kann man überprüfen, ob sich bereits vor Beginn des Trainings die durchschnittlichen Entwicklungsverläufe der Lesekompetenz zwischen der Trainingsgruppe und der Kontrollgruppe voneinander unterschieden. Dies war in Studie 2a der Fall. Daraus lässt sich die Schlussfolgerung ziehen, dass der Unterschied in der Leistungsentwicklung Δ wahrscheinlich nicht durch das Training bewirkt wurde. In Studie 2b hingegen verliefen die Leistungsentwicklungen von Trainings- und Kontrollgruppe vor Beginn des Trainings parallel. Die Ergebnisse dieser Studie weisen somit darauf hin, dass der positive Effekt auf die Leistungsentwicklung durch das Training verursacht wurde. Doch war dieser Effekt nachhaltig?

Prä-Post-Design mit zusätzlichem Prätest und Follow-up-Test. Um diese Frage zu beantworten, wurde in den Studien 3a und 3b jeweils ein weiterer Posttest (ein sog. Follow-up-Test) zum Zeitpunkt T_3 durchgeführt (s.

Abb. 19.9). In Studie 3a zeigte sich, dass das Training einen positiven Effekt auf die Leistung zum Zeitpunkt T_2 hatte. Dieser Effekt war aber nicht nachhaltig, da die Lesekompetenz zum Zeitpunkt T_3 wieder abfiel, und zwar auf das Niveau der Lesekompetenz, das ohne Lesetraining zu erwarten war. In Studie 3b hingegen hatte das Training einen nachhaltigen positiven Effekt auf die Leistungsentwicklung. Der Leistungsunterschied, den das Training zum Posttest (T_2) bewirkte, blieb auch beim Follow-up-Test (T_3) bestehen.

Generell gilt, dass der Einbezug von Follow-up-Tests ein Qualitätsmerkmal von Evaluationsstudien im pädagogischen Kontext darstellt. Insbesondere ein zeitlich deutlich versetzter Follow-up-Messzeitpunkt (z.B. mehrere Monate nach dem ersten Posttest) gibt Aufschluss darüber, ob die festgestellten Verbesserungen in den Zielkriterien über die Zeit hinweg stabil bleiben und die Befunde (im Sinne der externen Validität) zeitlich generalisiert werden können.

19.2.5 Interpretation und Kommunikation der Ergebnisse

Neben einer verständlichen, adressatengerechten, aber dennoch genauen Berichtlegung sind bei der Ergebniskommunikation von Evaluationsstudien zwei Aspekte wichtig (vgl. Kap. 4):

- (1) die Bewertung der Größe des Effekts einer pädagogischen Intervention (Effektgröße) und
- (2) die statistische Unsicherheit, die mit dieser Effektgröße verbunden ist (Konfidenzintervall).

Effektgröße

Effektgrößenmaß. Zur Bewertung der Ergebnisse in Evaluationsstudien ist die Metrik (d.h. die mit der Quantifizierung der Ergebnisse verbundene Messeinheit) der Zielkriterien zentral. Im Idealfall ist die Metrik selbsterklärend (z.B. der Effekt einer beruflichen Fortbildung auf das persönliche Netto-Einkommen in Euro; die Metrik ist hier Euro). Im Kontext der Evaluation pädagogischer Interventionen sind in dieser Hinsicht kriteriumsnormorientierte Kompetenzstufenmodelle und damit verbundene Leistungsstandards von besonderer Bedeutung (s. Abschn. 19.1.3). So dürfte etwa die Relevanz solcher Aussagen wie »ein Drittel der 15-Jährigen erreicht nicht den Mindeststandard in der Lesekompetenz« allen Adressaten eines Evaluationsberichts relativ leicht vermittelbar sein.

Für viele Zielkriterien im pädagogisch-psychologischen Kontext (z. B. Leistungsmaße oder motivational-emotionale Merkmale) ist die Metrik jedoch weniger klar interpretierbar. In manchen Anwendungsfällen ist zudem eine kriteriumsnormorientierte Rückmeldung zu ungenau. In solchen Fällen ist es erforderlich, standardisierte Effektgrößen anzugeben. Ein häufig verwendetes Effektgrößenmaß ist Cohens d für Mittelwertsunterschiede (Cohen, 1992). Die mathematische Formel zur Berechnung von Cohens d ist in Abbildung 19.1f (Abschn. 19.1.2) dargestellt. Ein großer Vorzug von standardisierten Effektgrößen ist, dass ihre Metrik unabhängig von der Metrik der Messinstrumente ist, die in einer bestimmten Studie eingesetzt wurden. Dies ermöglicht es, die Effektgrößen studienübergreifend zu vergleichen, wie man es z. B. in Metaanalysen macht (Cooper et al., 2009; vgl. Abschn. 4.5.2).

Interpretation von Effektgrößen. Wie interpretiert man standardisierte Effektgrößen? In der Studie von Strelow et al. (2012) wurde ein standardisierter Effekt des LekoLemo-Trainings auf die Leseleistung im Follow-up-Test von $d = 0,22$ gefunden. Dies bedeutet, dass die

Lesekompetenz von Jugendlichen, die am LekoLemo-Training teilnahmen, beim Follow-up-Test im Mittel um 0,22 Standardabweichungen über der Leseleistung von Jugendlichen lag, die am regulären Unterricht teilnahmen. Ist dies ein großer oder ein kleiner Effekt? Zur Beantwortung dieser Frage können mehrere Strategien angewendet und miteinander kombiniert werden (Cohen et al., 2003; Hill, Bloom, Black & Lipsey, 2008):

- (1) Zur Bewertung der Effektgrößen für Leistungstests kann es sehr aufschlussreich sein, die gefundenen Mittelwertsunterschiede mit den zu erwartenden Lernzugewinnen für eine bestimmte Altersgruppe zu vergleichen. So zeigten beispielsweise mehrere PISA-Zyklen, dass die Lernzuwächse über ein Schuljahr bei 15-jährigen Jugendlichen in allen drei Leistungsdomänen (Lesen, Mathematik und Naturwissenschaften) im internationalen Durchschnitt bei etwa $d = 0,40$ liegen (OECD, 2010). Ähnliche, wenn auch etwas geringere Lernzugewinne wurden für diese Jahrgangsstufen auf Grundlage von US-amerikanischen Studien in einer Arbeit von Hill et al. (2008) gefunden. In dieser

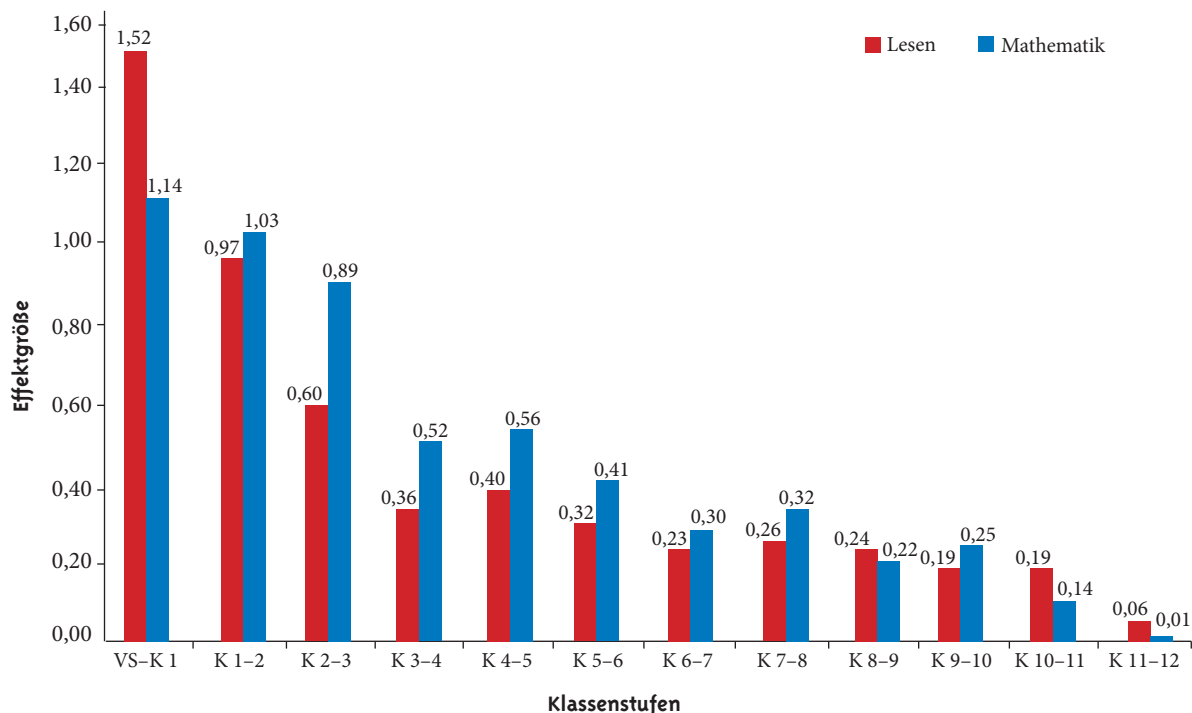


Abbildung 19.10 Mittlerer Lernzugewinn (in Form von standardisierten Mittelwertsdifferenzen) über ein Schuljahr in Lesen und Mathematik in Abhängigkeit von den besuchten Klassenstufen (nach Hill et al., 2008). VS = Vorschule; K = Klassenstufe

Arbeit zeigte sich weiterhin, dass die Lernzuwächse mit zunehmenden Jahrgangsstufen tendenziell abnehmen (s. Abb. 19.10). Bezogen auf die Befunde bei PISA entspricht demnach der oben genannte Effekt des LekoLemo-Trainings auf die Leseleistung etwa dem zu erwartenden Lerngewinn von einem halben Schuljahr; bezogen auf den Lernzugewinn von der 7. zur 8. Klasse (s. Abb. 19.10) entspricht er in etwa dem Lerngewinn von einem Schuljahr. In Bezug auf beide Vergleichsmaßstäbe ist der Effekt des LekoLemo-Trainings also als praktisch bedeutsam zu interpretieren.

- (2) Die Effektgrößen in der jeweiligen Evaluationsstudie können mit Effektgrößen vergleichbarer Studien oder auch anderer pädagogischer Interventionen verglichen werden. Hierzu bieten sich v. a. die gemittelten Effektgrößen an, die in thematisch relevanten Metaanalysen berichtet werden. Auf diese Weise lässt sich feststellen, wie stark die Effekte der untersuchten Intervention im Vergleich zu denen ähnlicher, bereits evaluierter Maßnahmen ausfallen. Ein Vergleich mit anderen pädagogischen Interventionen ist auch im Hinblick auf die Bewertung der Kosten aufschlussreich. Sind die Kosten für die Durchführung der verschiedenen Maßnahmen bekannt und können diese mit den Wirkungen in Beziehung gesetzt werden, so lassen sich nicht nur Aussagen über die relative Effektivität, sondern auch über die relative Effizienz der evaluierten Intervention ableiten.
- (3) Für Situationen, in denen wenig Vorwissen darüber vorliegt, wie die Effektgrößen für ein bestimmtes Maß oder eine bestimmte pädagogische Intervention zu bewerten sind, können die von Cohen (1992) vorgeschlagenen Richtlinien angewendet werden. Demnach sind Werte von $d = 0,20$ als »kleine Effekte«, Werte von $d = 0,50$ als »mittlere Effekte« und Werte von $d = 0,80$ und größer als

»große Effekte« zu bewerten. Solche Richtlinien liegen auch für andere Effektgrößenmaße als d (wie z. B. für den Korrelationskoeffizient r) vor (s. hierzu Cohen, 1992).

Konfidenzintervalle

Im Kontext von Evaluationsstudien entsteht statistische Unsicherheit dadurch, dass bei solchen Untersuchungen in der Regel nur Daten für Stichproben zur Verfügung stehen. Die Frage, ob die Intervention wirkt oder inwiefern die Schülerschaft in einem Bildungssystem bestimmte Zielkriterien erreicht, soll jedoch nicht nur für diese spezifische Stichprobe beantwortet, sondern auf die jeweilige Zielpopulation (z. B. alle Schüler in einem bestimmten Bundesland) generalisiert werden. Die Schätzung von statistischen Kennwerten für die Population anhand von Stichprobendaten ist mit Unsicherheit behaftet, die sich quantifizieren lässt. Mittlerweile ist es üblich, diese Unsicherheit in Form von Konfidenz- bzw. Vertrauensintervallen anzugeben. Hier ist es wichtig anzumerken, dass sich die Konfidenzintervalle bei der Individualdiagnostik auf die Präzision von Messwerten für eine einzelne Person beziehen, während Konfidenzintervalle in Evaluationsstudien die Präzision bei der Schätzung von Populationskennwerten betreffen. Die Breite der Konfidenzintervalle in Evaluationsstudien ist damit ein Maß für die Präzision, mit der Effektgrößen geschätzt wurden. Sie markieren einen Bereich plausibler Werte für den Effekt einer pädagogischen Intervention oder das Erreichen von bestimmten Zielkriterien in der jeweiligen Population (Cumming & Finch, 2005). Konfidenzintervalle geben damit beispielsweise Schätzwerte für eine konservative Untergrenze des zu erwartenden Interventionseffekts an. Gleichzeitig dient die Obergrenze des Konfidenzintervalls dazu, den maximalen Effekt z. B. eines Trainings abzuschätzen.

Zusammenfassung

- ▶ Pädagogisch-psychologische Diagnostik zielt darauf ab, evidenzbasierte Entscheidungen in Bezug auf Einzelpersonen zu treffen. Diese Entscheidungen können mit Entscheidungsfehlern verbunden sein.
- ▶ Professionelle Diagnostik zeichnet sich durch die Auswahl und Anwendung objektiver, reliabler und valider Messinstrumente sowie durch die fachgerechte Interpretation und Kommunikation der damit erfassten Testwerte aus.
- ▶ Im Rahmen diagnostischer Entscheidungen können Testwerte im Vergleich zu früheren Testwerten derselben Person (individualnormorientiert), im Vergleich zu einem inhaltlich definierten Kriterium (kriteriumsnormorientiert) oder im Vergleich zu den Testwerten anderer Personen (sozialnormorientiert) bewertet werden.
- ▶ Evaluationen im pädagogischen Kontext zielen darauf ab, empirische Informationen zur Bewertung pädagogischer Interventionsprogramme oder größerer organisatorischer Einheiten (z. B. Schulen, Schularten, Bildungssysteme) zu liefern. Diese Bewertungen sollen zur evidenzbasierten Qualitätssicherung im Bildungswesen beitragen.
- ▶ Außer mit quantitativen Methoden können Evaluationen auch mit qualitativen Methoden oder dem kombinierten Einsatz beider Methodenansätze durchgeführt werden. Unabhängig von den verwendeten Methoden sollen sich Evaluationen an den von Fachgesellschaften formulierten Standards für Nützlichkeit, Durchführung, Fairness und Genauigkeit orientieren.
- ▶ Bei der Konzeption einer Evaluationsstudie ist eine theoretische Fundierung wichtig und die Erfassung von Bildungsprozessen nützlich. Das Studiendesign sollte so angelegt sein, dass es möglichen Gefährdungen der internen und externen Validität effektiv begegnen kann. Hierbei ist in der Regel ein Studiendesign mit mehreren Messzeitpunkten hilfreich (Längsschnittdesign). Für evidenzbasierte Entscheidungen ist es zudem auch im Rahmen von Evaluationen unerlässlich, objektive, reliable und valide Messinstrumente einzusetzen und diese fachgerecht auszuwerten und zu interpretieren.
- ▶ Zur Bewertung und Kommunikation von Evaluationsergebnissen sind Zielkriterien mit intuitiver Metrik, Kompetenzstufenmodelle und Leistungsstandards sowie standardisierte Effektgrößen wichtige Hilfsmittel.

20 Beratung, Intervention, Supervision

Mara Gerich • Simone Bruder • Silke Hertel • Tina Hascher • Bernhard Schmitz

20.1 Beratung

- 20.1.1 Relevanz in pädagogischen Handlungsfeldern
- 20.1.2 Begriffsbestimmung und theoretischer Hintergrund
- 20.1.3 Anwendungsfelder im pädagogischen Bereich

20.2 Intervention

- 20.2.1 Relevanz und Ansatzpunkte in pädagogischen Handlungsfeldern
- 20.2.2 Begriffsbestimmung, Aspekte und Evaluation
- 20.2.3 Anwendungsfelder im pädagogischen Bereich

20.3 Supervision

- 20.3.1 Relevanz in pädagogischen Handlungsfeldern
- 20.3.2 Begriffsbestimmung und theoretischer Hintergrund
- 20.3.3 Formen der Supervision
- 20.3.4 Anwendungsfelder im pädagogischen Bereich

Frau M. ist 30 Jahre alt und seit zwei Jahren als Lehrerin an einem Gymnasium beschäftigt. Sie unterrichtet die Fächer Deutsch und Geschichte. Ihren Beruf mag sie sehr gerne und bisher hatte sie noch nie Schwierigkeiten mit ihren Klassen. Sie fühlt sich an ihrer Schule auch sehr wohl und hat eine unterstützende und engagierte Schulleiterin. Am liebsten unterrichtet Frau M. in den höheren Klassenstufen, aufgrund des längeren Ausfalls einer Kollegin musste sie aber die Leitung einer 6. Klasse übernehmen. Sie ist nun erst seit zwei Wochen in der neuen Klasse, aber mit den Nerven schon völlig am Ende. Die Klasse zeigt sich ihr gegenüber wenig respektvoll, im Unterricht ist es extrem laut und auch unter den Schülern scheint es große Schwierigkeiten zu geben; Frau M. hat mitbekommen, dass mindestens zwei Schüler aus der Klasse gemobbt werden.

Nachdem Frau M. mit ihrer Schulleiterin darüber gesprochen hat, ist diese bereit, sie zu unterstützen. Sie wusste auch von der letzten Klassenlehrerin, dass diese Klasse nicht einfach zu führen ist. Die Schulleiterin stellt Frau M. verschiedene Möglichkeiten vor: Frau M. könnte sich von der zuständigen Schulpsychologin beraten lassen, wie sie mit der Situation umgehen soll und welche Schritte sie als Erstes unternehmen könnte. Oder sie könnte die Situation mit einem Supervisor ausführlich besprechen und gemeinsam Lösungen erarbeiten. Außerdem schlägt die Schulleiterin Frau M. vor, an einer Fortbildung (Intervention) zum Klassenmanagement teilzunehmen, weil sie weiß, dass dies ein zentrales Thema ist und häufig in der pädagogischen Ausbildung zu kurz kommt. Auch eine Kombination der verschiedenen Maßnahmen wäre für die Schulleiterin denkbar.

Frau M. ist sehr zufrieden mit den Angeboten der Schulleiterin und entscheidet sich, zunächst als schnellste Lösung die Schulpsychologin zu konsultieren und sich dann supervidieren zu lassen. Eine passende Fortbildung hat sie auf die Schnelle leider nicht gefunden, sich aber fest vorgenommen, sich innerhalb des nächsten Jahres für Fortbildungen zu den Themen Klassenmanagement und Mobbing anzumelden.

Wie das Fallbeispiel zeigt, gibt es allein im schulischen Bereich eine Vielzahl von Belastungen und schwierigen Situationen, mit denen Lehrkräfte im Schulalltag umgehen müssen. Wie auch in anderen pädagogischen Ar-

beitsfeldern treten immer wieder schwierige Situationen mit Kindern, Eltern und Kollegen auf. Beratung, Intervention und Supervision bieten aber nicht nur Möglichkeiten der Unterstützung der im pädagogischen Bereich

tätigen Personen, sondern werden insbesondere von diesen selbst im Zusammenhang mit der Gestaltung von Entwicklungs-, Bildungs- und Bewältigungsprozessen verschiedener Personengruppen angeboten. Wie genau Beratung, Intervention und Supervision unterstützend eingesetzt werden können, was sie gemeinsam haben und was sie unterscheidet, wird in diesem Kapitel geschildert.

20.1 Beratung

20.1.1 Relevanz in pädagogischen Handlungsfeldern

Die zunehmende Komplexität unserer Gesellschaft und der beruflichen Tätigkeitsfelder erfordert vom Einzelnen ein hohes Maß an Orientierungs-, Umstellungs- und Lernfähigkeit, aber auch die Fähigkeit zur Lösung von Konflikten oder zur Bewältigung ungewohnter Aufgaben und Schwierigkeiten. In vielen Situationen fühlen sich Handelnde überfordert und bedürfen der Unterstützung von außen, z. B. durch eine fachkundige Beratung. Tatsächlich gewinnt derzeit die professionelle Beratung in vielen Handlungsfeldern zunehmend an Bedeutung. Sie stellt heute eines der am besten entwickelten und vielfältigsten professionellen Hilfsangebote dar und hat sich besonders in pädagogischen und psychologischen Arbeitsfeldern etabliert.

In pädagogischen Kontexten wird Beratung als ein zentraler Handlungsbereich verstanden, der sich in den letzten Jahren zu einem breiten Praxisfeld entwickelt hat. Auf der Grundlage unterschiedlicher theoretischer Ansätze und Konzepte kommt sie in verschiedenen pädagogischen Bereichen, wie z. B. der Schule, der Aus- und Weiterbildung oder der Erziehungsberatung, zur Anwendung. Das generelle Ziel pädagogisch-psychologischer Beratung ist die Erarbeitung von Lösungen für spezifische Problemfälle und das Aufdecken individueller Ressourcen im Zusammenhang mit Erziehung, Unterricht und Ausbildung. In den Bereichen Schule, Ausbildung und Hochschule geht es häufig um die Optimierung von Lehr-Lern-Prozessen.

20.1.2 Begriffsbestimmung und theoretischer Hintergrund

Definition der professionellen Beratung

Im alltäglichen Sprachgebrauch bezeichnet man mit Beratung die Hilfe und Unterstützung bei der Bewältigung persönlicher Probleme. Professionelle Beratung geht jedoch über das Ratgeben oder Sich-Beraten hinaus. Sie kann als eine Sonderform der sozialen Interaktion aufgefasst werden, die im Gegensatz zur Alltagsberatung planvoll und von einem fachkundigen und methodisch geschulten Berater durchgeführt wird und die auf einer beidseitigen Verbindlichkeit, Verantwortung sowie auf einem beidseitigen Vertrauensverhältnis beruht.

Definition

Beratung ist eine zumeist kurzfristige soziale Interaktion zwischen Ratsuchenden und Beratenden, bei der dem Ratsuchenden Unterstützung zur Bewältigung seines Problems angeboten wird. Sowohl bei lebenspraktischen Fragen als auch in psychosozialen Krisen erarbeiten Ratsuchende und Beratende gemeinsam kognitive, emotionale und praktische Problemlösungen.

Beratung wird dabei als Hilfe zur Selbsthilfe verstanden, bei der ein professioneller Berater eine kooperative und vertrauensvolle Beziehung zu einem (Individualberatung) oder mehreren Klienten (Gruppenberatung, Systemberatung) eingeht und den Beratungsprozess planvoll und zielgerichtet gestaltet. Hierbei unterstützt er die Eigenbemühungen des Ratsuchenden bei der Verbesserung seiner Fähigkeiten zur Bewältigung eines anstehenden Problems unter Berücksichtigung der individuellen Eigenschaften des Klienten, dessen aktuellen und potenziellen Kompetenzen und dessen Lebenssituation. In Anlehnung an Honai und Schlegel (2002) können folgende vier Aspekte als grundlegende Charakteristika professioneller Beratung aufgefasst werden (s. Übersicht).

Übersicht

Kennzeichen professioneller Beratung

- ▶ **Freiwilligkeit:** Der Klient wählt den Berater eigenständig aus und entscheidet über Beginn, Beendigung oder Fortführung des Beratungsprozesses.
- ▶ **Vertraulichkeit:** Das Vertrauensverhältnis und die Vertraulichkeit der Informationen und Angaben im Beratungsgespräch sind wichtige Voraussetzungen für den Erfolg der Beratung. Nur wenn ein stabiles Vertrauensverhältnis besteht, kann der Berater alle für den Beratungsprozess wichtigen Informationen seitens des Klienten erhalten.
- ▶ **Wissenschaftliche Fundierung:** Der Berater gestaltet den Beratungsprozess mithilfe wissenschaftlicher Theorien und Methoden und unterstützt damit einen erfolgreichen Verlauf der Beratung.
- ▶ **Unabhängigkeit:** Der Beratungsprozess findet möglichst unabhängig von externen Einflüssen statt.

Beratung als (Problemlöse-)Prozess

Da professionelle Beratung auf die Lösung umschriebener, subjektiv bedeutsamer Probleme von Individuen, Gruppen oder Institutionen hinarbeitet, kann der Beratungsprozess auch als Problemlöseprozess aufgefasst werden. Das Problem stellt dabei im Allgemeinen eine von dem oder den Ratsuchenden wahrgenommene Diskrepanz zwischen einem unerwünschten Ist-Zustand und einem anzustrebenden Soll-Zustand dar, welche im Verlauf des Beratungsprozesses verringert oder überwunden werden soll.

Der Beratungsprozess kann als eine Sequenz aufeinanderfolgender Phasen oder Ablaufschritte dargestellt werden (vgl. Abb. 20.1), die jedoch nicht als starre und lineare Abfolge, sondern vielmehr als Phasen eng miteinander vernetzter Rückkopplungsschleifen zu verstehen sind. Zur Bewältigung der im Verlauf einer Beratung anstehenden Aufgaben stehen dem Berater jeweils eine Vielzahl an Strategien und Methoden zur Verfügung. Durch deren gezielte Anwendung versucht der Berater das jeweilige Problem genauer zu analysieren, die Zielvorstellungen des Ratsuchenden zu klären und

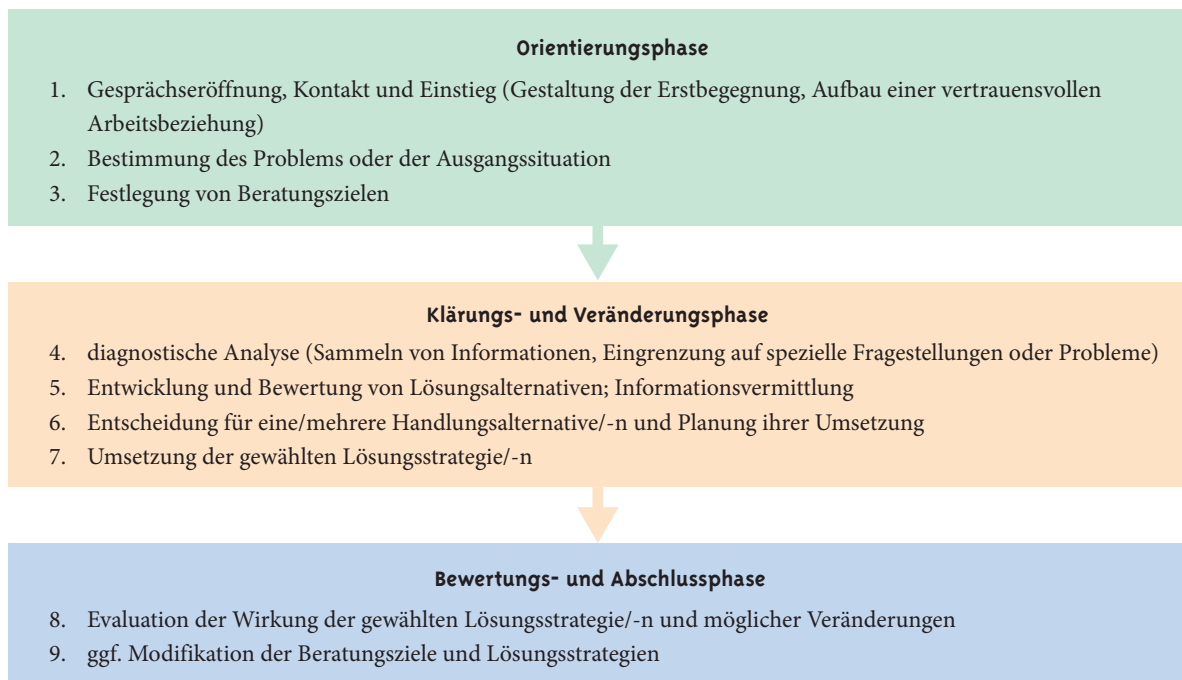


Abbildung 20.1 Phasenverlauf des Beratungsprozesses

ggf. die für ihn realisierbaren Handlungsalternativen zu erweitern, um so dessen Kompetenzen zur eigenständigen Bewältigung seines Problems zu stärken.

Zielsetzung von Beratung

Neben der Bewältigung von Problemsituationen besteht das Ziel professioneller Beratung auch in der Prävention antizipierter Problemzustände unter optimaler Ausschöpfung vorhandener Potenziale. Dies soll durch die Verbesserung der Bewältigungskompetenz, der Selbsthilfbereitschaft und Selbststeuerungsfähigkeit des Klienten erreicht werden. Das heißt, Beratung versteht sich stets als Hilfe zur Selbsthilfe. Durch die Anregung aktiver Lernprozesse sollen Kompetenzen für eine erfolgreichere Auseinandersetzung mit Problemen und

Schwierigkeiten erworben werden. Dies betrifft auch die Fähigkeit, selbstständig Entscheidungen zu treffen und bestimmte Vorhaben zielorientiert zu realisieren. Um dies zu erreichen, kann der Berater für die Problemlösung oder Entscheidung wichtige Aspekte erfragen und/oder allgemeine Voraussetzungen herbeiführen, welche den Klienten bei der Lösung seines Problems unterstützen. Dies geschieht in der Regel durch die Vermittlung neuer Informationen und/oder durch die Analyse, Neustrukturierung und Neubewertung vorhandener Informationen (Schwarzer & Posse, 2005). Im pädagogischen Kontext liegt der Fokus von Beratung insbesondere auf der Optimierung der Entwicklungsprozesse von Lernenden (Brem-Gräser, 1993).

Übersicht

Abgrenzung von Beratung zur Psychotherapie

Viele der im pädagogischen Bereich eingesetzten Beratungsansätze und -techniken haben ihren Ursprung in der Psychotherapie. Das ist einer der Gründe, warum die beiden Interventionsformen nur schwer voneinander abzugrenzen sind. Wesentliche Unterschiede bestehen in der Regel jedoch in den folgenden Merkmalen:

- **Dauer der Intervention:** Beratung erstreckt sich in der Regel über einen kürzeren Zeitraum als Therapie bzw. beinhaltet weniger Sitzungen.
- **Schwere des Problems:** Therapie setzt meist an Störungen mit Krankheitswert an, während in der

Beratung hauptsächlich nicht-pathologische Problemfälle thematisiert werden.

- **Grad der aktiven Einflussnahme:** Im Gegensatz zu einer professionellen Therapie, deren Verlauf von einem Therapeuten geplant und kontrolliert wird, erhält ein Klient von einem Berater in der Regel nur Anregungen für eine Veränderung des Verhaltens. Bei der Beratung steht die Selbststeuerungsfähigkeit des Klienten stärker im Vordergrund.
- **Einbezug des Unbewussten:** Beratung verzichtet auf die Thematisierung unbewusster Elemente des psychischen Geschehens, die in der klassischen Psychotherapie oft eine wichtige Rolle spielen.

Theoretische Ansätze der Beratung

In der Literatur findet sich eine Vielzahl von Beratungsansätzen, die in der Praxis nicht unbedingt konkurrierend nebeneinander stehen. Besonders in den letzten Jahren gibt es immer mehr Versuche einer integrativen Beratung, bei der sich die einzelnen Ansätze und Techniken je nach Art der Problemstellung in ihrer Anwendung gegenseitig ergänzen (Rausch, Hinz & Wagner, 2008).

Klientenzentrierte Beratung. Grundlage dieses Beratungskonzepts bildet die klientenzentrierte Gesprächs-

therapie nach Rogers (1973). Ein wesentliches Kennzeichen dieses Konzepts ist die starke Fokussierung des Beratungsprozesses auf die Individualität des Ratsuchenden. Sie gilt als besonders geeignet, um dem Ratsuchenden mehr Vertrauen in die eigenen Handlungs- und Entscheidungsmöglichkeiten zu vermitteln (Weinberger, 2011). Das Verhalten des Beraters gegenüber dem Klienten orientiert sich an bestimmten allgemeinen Prinzipien, welche nach Rogers grundlegende Bedingungen für ein wachstumsförderliches Interaktionsklima im Beratungsprozess darstellen.

Übersicht

Grundhaltungen des klientenzentrierten Beraters

- **Einfühlsames Verstehen:** die Fähigkeit und Bereitschaft des Beraters, sich in den Klienten hineinzuversetzen, um so dessen Erlebnisse und Gefühle nachvollziehen zu können, dabei jedoch auch eine kritische Distanz zu bewahren. Neben den verbalen Aussagen des Klienten sollte der Berater auch die nonverbalen Äußerungsformen von Gedanken und Gefühlen berücksichtigen und die von ihm wahrgenommenen emotionalen Erlebnisinhalte des Klienten an diesen zurückmelden.
- **Wertschätzung:** die grundlegend positive Einstellung des Beraters gegenüber seinen Klienten; er soll sie in ihrer individuellen Erlebniswelt vorbehaltlos und vorurteilsfrei annehmen.
- **Echtheit:** die Unverfälschtheit und Transparenz des Beraterverhaltens gegenüber dem Klienten sowie seine Fähigkeit, sich seiner selbst und seiner Gefühle so weit wie möglich bewusst zu sein.

Verhaltensbezogene Beratung. Die an klassischen lern- und verhaltenstheoretischen Ansätzen orientierte Beratung basiert auf der Annahme, dass (fast) alle Verhaltensweisen – auch fehlangepasste – im Laufe der Zeit erlernt und deshalb auch wieder umge- bzw. verlernt werden können (vgl. Abschn. 6.4). Hierzu wird das spezifische Problemverhalten im Rahmen einer differenzierten Verhaltensanalyse zunächst genau untersucht. Dabei steht insbesondere der Zusammenhang zwischen Verhalten und den situationsspezifischen Kontextbedingungen – also die einem beobachtbaren Verhalten vorausgehenden und nachfolgenden Ereignisse – im Fokus. Um das Problemverhalten in erwünschter Weise zu modifizieren, empfiehlt der Berater die Anwendung von Strategien und Handlungsmustern, die sich auch im verhaltenstherapeutischen Kontext bewährt haben, wie z. B. Techniken der Handlungskontrolle durch gezielt eingesetzte Formen der Verstärkung oder Bestrafung, systematische Desensibilisierung oder andere erprobte Maßnahmen der Verhaltensänderung (Margraf & Schneider, 2009). Seit der »kognitiven Wende« in den 1960er-Jahren (vgl. Abschn. 2.2.6 und 7.1) werden in der verhaltensbezogenen Beratung auch kognitive Aspekte berücksichtigt, wie die subjektive Bewertung äußerer Ereignisse und die damit einhergehenden Gedanken und Gefühle sowie die Planung, Bewertung und Überprüfung des eigenen Handelns durch den Ratsuchenden (Hautzinger, 2011). Verhaltensbezogene Beratungsansätze werden z. B. in der Sozialarbeit, im Elternt raining sowie im schulischen und erwachsenenpädagogischen Bereich eingesetzt.

Systemische Beratung. Die systemische Beratung geht davon aus, dass das Problemverhalten eines Individuums weniger durch innerpsychische Prozesse als viel-

mehr durch Störungen im sozialen Gefüge der betreffenden Person hervorgerufen und aufrechterhalten wird. Das individuelle Problemverhalten erweist sich aus dieser Sicht als Symptom für Störungen im sozialen Netzwerk und betrifft damit auch alle anderen Mitglieder der Gruppe. Bei einer systemischen Beratung werden nicht allein die Verhaltensmuster des primären »Symptomträgers«, sondern insbesondere die Kommunikation innerhalb des sozialen Systems und der Umgang der Mitglieder untereinander betrachtet. Die Beratung zielt sowohl auf Veränderungen der Person als auch auf Veränderungen ihres sozialen Lebensraumes ab. Dabei geht es v. a. um die Identifizierung und Förderung der Ressourcen und Kompetenzen des Ratsuchenden und seines sozialen Systems. Die systemischen Ansätze werden derzeit v. a. in der Familienberatung, aber auch in der Organisations- und Einzelberatung eingesetzt (Barthelmess, 2005).

Lösungsorientierte Beratung. Ein weiteres Konzept ist der auf De Shazer (1985) und Berg (1993) zurückgehende lösungsorientierte Ansatz, der sich aus der systemischen Beratung entwickelt hat. Im Vergleich zu klassischen Ansätzen werden Probleme und Konflikte nicht vertieft auf ihre Ursachen hin untersucht; vielmehr wird der Fokus der Beratung auf die Entwicklung möglichst rasch realisierbarer Lösungen gelegt. Zudem findet die lösungsorientierte Beratung meist in Form einer Kurzzeitberatung (mit nur einem oder wenigen Terminen) statt, in der häufig nur Anregungen und Anstöße gegeben werden für die eigentlichen Veränderungsprozesse, die sich im Alltag des Ratsuchenden vollziehen müssen. Ratsuchende werden im Rahmen dieses Beratungskonzepts aufgefordert, sich gleich zu Beginn selbst aktiv an der Lösungsfindung zu beteiligen.

Kompetenzen des Beraters

Unabhängig von der Orientierung an spezifischen Beratungsansätzen lassen sich allgemeine Kompetenzen eines professionellen Beraters benennen, die generell für einen erfolgreichen Beratungsverlauf erforderlich sind. Die meisten Konzepte zur Aufschlüsselung der verschiedenen Komponenten der Beratungskompetenz integrieren inhaltliches Wissen und (praktische) Handlungskompetenzen (vgl. Honal & Schlegel, 2002). Schwarzer und Buchwald (2006) weisen darauf hin, dass ein kompetenter Berater auch personale Ressourcen benötigt, die eine reflektierte Auseinandersetzung mit dem eigenen professionellen Beraterhandeln erlauben.

Trotz der Vielzahl unterschiedlicher Ansätze zur Beschreibung der Beratungskompetenz besteht weitgehend Konsens über die Relevanz der im Folgenden dargestellten Kompetenzbereiche eines professionellen Beraters (vgl. Abb. 20.2).

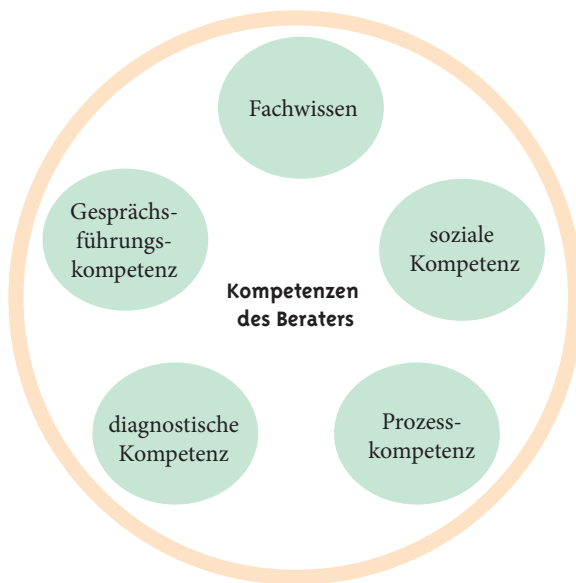


Abbildung 20.2 Allgemeine Kompetenzen des Beraters

Fachwissen. Um ihre Funktion adäquat wahrnehmen zu können, benötigen professionelle Berater ein hinreichend differenziertes und wissenschaftlich fundiertes pädagogisch-psychologisches Fachwissen, das dem Beratungsprozess zugrunde liegt. Diese professionelle Wissensbasis setzt sich zusammen aus Theorien, Modellen und empi-

risch gesicherten Erkenntnissen zur Beschreibung, Erklärung und Modifikation problematischer Phänomene. Da Beratung immer unter bestimmten gesellschaftlichen Rahmenbedingungen stattfindet, benötigt der Berater je nach Beratungsanlass außerdem Wissen über die aktuelle Gesetzeslage (z.B. Schulpflichtgesetz, Sozialgesetzgebung) oder andere Rahmenbedingungen in der Lebenswelt der Klienten, die für deren Probleme bedeutsam sein können, wie z.B. die Beschäftigungslage oder die speziellen Lern- und Arbeitsbedingungen in einem bestimmten pädagogischen Setting.

Soziale Kompetenz. Eine zentrale Komponente ist die Fähigkeit, eine vertrauensvolle und offene Beziehung und Atmosphäre aufzubauen, die es dem Ratsuchenden erleichtert, sich zu öffnen und auch persönlich belastende Sachverhalte zur Sprache zu bringen. Hierzu ist ein hohes Maß an Einfühlung und Akzeptanz gegenüber dem Klienten, soziales Fingerspitzengefühl und kommunikative Sensibilität für verbale sowie nonverbale Signale notwendig. In diesem Zusammenhang besitzen die von Rogers (1973) hervorgehobenen Merkmale eines positiven persönlichen Interaktionsstils im Rahmen einer klientenzentrierten Beratung (s.o.) einen besonderen Stellenwert.

Prozesskompetenz. Damit ist gemeint, dass der Berater in der Lage sein muss, den Ablauf der Beratung angemessen zu steuern. So ist es z.B. erforderlich, den Beratungsprozess sinnvoll zu strukturieren, die notwendigen Schritte zu identifizieren und in eine adäquate Reihenfolge mit richtigem zeitlichen Verhältnis zueinander zu bringen. Hierbei sind seine Kompetenzen zur Beschreibung und Analyse von Problemsituationen, zur Identifizierung und Reflexion von Zielen, zur Suche nach alternativen Lösungswegen, zur Planung ihrer Umsetzung, zur Steuerung der Durchführung und zur Überprüfung ihrer Wirkung von Bedeutung. Dies erfordert kognitive Flexibilität und Sensibilität sowie die Fähigkeit, die vom Ratsuchenden vermittelten Informationen in ein Gesamtbild einzuordnen.

Diagnostische Kompetenz. Der professionelle Berater sollte weiterhin über hinreichende diagnostische Kompetenzen verfügen, um je nach Beratungsanlass die Situation oder die Problematik des Ratsuchenden richtig erfassen und verstehen zu können. Dazu ist u. U. auch der Einsatz spezieller diagnostischer Instrumente und Strategien erforderlich, mit denen der Berater hinreichend vertraut sein muss (vgl. Abschn. 19.1). Neben einer Analyse der »inneren Welt« zielt die Diagnostik auch auf die

Identifikation der auslösenden Faktoren von Problemen oder die Einschätzung vorhandener Ressourcen des Ratsuchenden einschließlich seines sozialen Netzwerks. Eine sorgfältige Diagnose ist Voraussetzung für die Präzisierung des Problems und der Beratungsziele. Sie bildet auch die Grundlage für die Entscheidungen des Beraters, welche Beratungsstrategien, -techniken und -methoden im Beratungsprozess anzuwenden sind und welche pädagogischen Interventionen ggf. infrage kommen. Zur diagnostischen Kompetenz gehört nicht nur die fachgerechte Auswahl und Verwendung geeigneter diagnostischer Strategien und Methoden (wie Interviews, Fragebögen, Tests u. Ä.), sondern ebenso die Fähigkeit, auf der Basis dieser Daten und dem im Verlauf der

Beratung gewonnen Gesamteindruck handlungsrelevante Hypothesen zu entwickeln.

Gesprächsführungskompetenz. Um ein Beratungsgespräch zielführend gestalten zu können, benötigt man außerdem spezielle Kompetenzen der Gesprächsführung. Zu diesen zählt z. B. die Fähigkeit, durch gezielte Fragen die jeweils vorliegende Problematik möglichst exakt herauszuarbeiten oder die Erwartungen der am Beratungsprozess beteiligten Personen sowie das Ziel der Beratung zu klären. Darüber hinaus sollte der professionelle Berater über Fertigkeiten zur optimalen Strukturierung von Gesprächen verfügen sowie verschiedene Gesprächstechniken beherrschen und variabel einsetzen können.

Unter der Lupe

Gesprächstechniken

Zwei grundlegende Techniken, die zu den Basisfertigkeiten professioneller Berater zählen, stellen das Paraphrasieren und das aktive Zuhören dar. Ihre Anwendung kann dem Berater dabei helfen, dem Ratsuchenden Wertschätzung zu vermitteln, diesem zu zeigen, dass er das Gespräch interessiert und aufmerksam verfolgt und dass er bereit ist, sich in die Perspektive seines Klienten hineinzuversetzen.

Paraphrasieren. Beim Paraphrasieren spiegelt der Berater die sachlichen Aussagen des Klienten in seinen eigenen Worten wider. So signalisiert er dem Ratsuchenden, dass er das Wesentliche der Aussage erfasst hat und gibt ihm gleichzeitig die Möglichkeit, sich mit seinen angesprochenen Anliegen und Problemen auseinanderzusetzen und die paraphrasierten Inhalte im weiteren Gesprächsverlauf zu konkretisieren. Weiterhin kann der Berater durch den Einsatz dieser Technik überprüfen, ob er das vom Klienten Gesagte richtig verstanden hat, und somit Missverständnisse im Beratungsprozess frühzeitig vermeiden oder ausräumen. Dabei wiederholt der Berater nicht nur die Worte des Klienten, sondern bringt durch die Zusammenfassung und Schwerpunktsetzung des Gesagten in seinen eige-

nen Worten die Problembearbeitung voran. Mögliche Einstiegsformulierungen für das Paraphrasieren sind z. B.: »Wenn ich Sie richtig verstehe, meinen Sie ...«, »Ihnen ist also wichtig, dass ...«.

Aktives Zuhören. Es ist wichtig, dass der Berater nicht nur auf den sachlichen Inhalt der Äußerungen des Ratsuchenden achtet, sondern ebenso die begleitenden emotionalen Anteile des Gesagten wahrnimmt und diese dem Klienten ggf. widerspiegelt. Denn Gefühle, Hoffnungen, latente Wünsche und dergleichen werden oft nicht direkt vom Klienten formuliert, schwingen aber in fast jeder Äußerung mit. Der Berater denkt und fühlt sich in den Ratsuchenden hinein und zeigt ihm durch aktives Zuhören, dass er die Gefühle, die mit dessen Äußerungen verknüpft sind, erfasst hat. Auf diese Weise wird einerseits der Aufbau einer vertrauensvollen Beziehung zwischen Berater und Ratsuchendem unterstützt und andererseits – ähnlich dem Paraphrasieren – eine Konkretisierung und ggf. Korrektur der angesprochenen gefühlsmäßigen Anteile erzielt. Typische Einstiegsformulierungen für das Aktive Zuhören lauten z. B.: »Sie sind verärgert, weil ...«, »Sie hoffen nun, dass ...«.

Forschungsansätze zur Beratungskompetenz

Die in Abbildung 20.2 genannten allgemeinen Kompetenzen des Beraters sind aus theoretischen Überlegungen hervorgegangen. Ob diese Klassifikation auch einer empirischen Überprüfung standhält, wurde bislang kaum untersucht. Erste Ansätze zur Entwicklung

empirisch überprüfter Modelle der Beratungskompetenz von Lehrkräften bei der Lernberatung – einem der zentralsten Beratungsfelder von Lehrkräften – finden sich z. B. bei Hertel (2009), Bruder, Keller, Klug und Schmitz (2011) sowie Gerich, Bruder, Hertel, Trittel und Schmitz (submitted).

Studie

Beratungskompetenz von Lehrkräften

Das Ziel der Studie von Gerich et al. (submitted) bestand in der Identifikation der wichtigsten Komponenten der Beratungskompetenz von Lehrkräften und ihrer anschließenden Zuordnung zu übergeordneten Kompetenzbereichen in einem Kompetenzmodell. In einem ersten Schritt wurde anhand von Vorarbeiten von Hertel (2009) und Bruder et al. (2011) sowie auf der Grundlage der aktuellen Literatur zum Thema Beratung ein Modell mit vier zentralen Kompetenzbereichen bzw. Dimensionen theoretisch formuliert und anschließend im Rahmen einer Querschnittsuntersuchung (s. Abschn. 4.3.1) mit Lehrkräften der Primar- und Sekundarstufe empirisch überprüft. In einer schriftlichen Befragung bearbeiteten die teilnehmenden Lehrkräfte hierzu ein handlungsnahes Fallszenario, das sich aus einem Fallbeispiel einer Schülerin mit Lernschwierigkeiten sowie verschiedenen offenen Fragen zur Durchführung eines Beratungsgesprächs mit deren Mutter zusammensetzte.

Auf der Grundlage dieser Daten konnte mittels konfirmatorischer Faktorenanalysen (s. Abschn. 4.5.1) das postulierte vierdimensionale Modell bestätigt werden. Abbildung 20.3 zeigt die vier Kompetenzdimensionen mit ihren jeweils untergeordneten Teilkompetenzen. Im Vergleich zur Benennung allgemeiner Kompetenzbereiche von Beratern legt das Modell zum einen die Grundlage für die Entwicklung zielgruppenspezifischer Instrumente zur validen, reliablen und insbesondere differenzierten Messung der zentralen Beratungskompetenzen im Lehrberuf. Darüber hinaus liefert es konkrete inhaltliche Hinweise für die Konzeption und Evaluation von Aus- und Weiterbildungsmaßnahmen im Bereich der Beratung für (angehende) Lehrkräfte. Die detaillierte Aufgliederung der Beratungskompetenz in ihre einzelnen Teilbereiche ermöglicht dabei sowohl die Entwicklung von Programmen zur generellen Förderung der Beratungskompetenz als auch zur gezielten Entwicklung einzelner Teilkompetenzen.

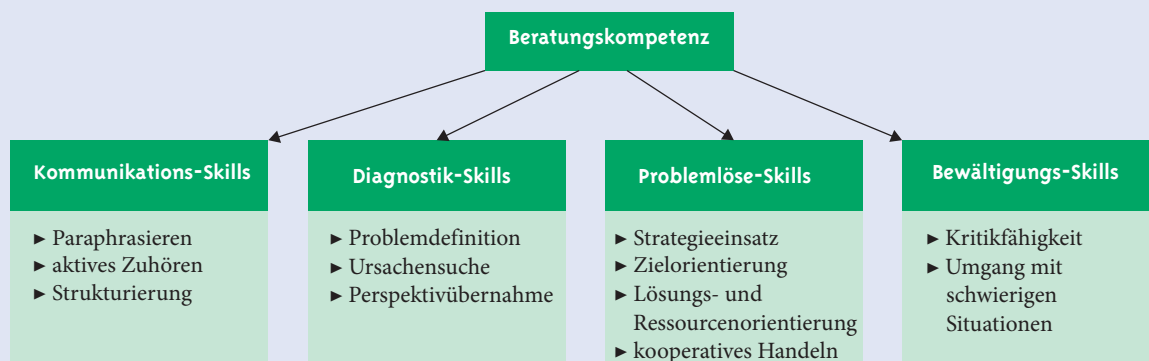


Abbildung 20.3 Modell der Beratungskompetenz von Lehrkräften in Bezug auf Lernberatung (Gerich et al., submitted)

20.1.3 Anwendungsfelder im pädagogischen Bereich

Erziehungs- und Familienberatung

Eines der wichtigsten Anwendungsfelder pädagogischer Beratung ist die Erziehungs- und Familienberatung. Sie zielt darauf ab, Erziehungsberechtigte in ihrer Erziehungsverantwortung zu unterstützen, um eine dem Wohl des Kindes oder Jugendlichen entsprechende Erziehung sicherzustellen (§ 27, SGB VIII, 1990). Nach dem Kinder- und Jugendhilfegesetz sollen Erziehungs-

und Familienberatungsstellen Kinder, Jugendliche, Eltern und andere Erziehungsberechtigte bei der Klärung und Bewältigung individueller und familienbezogener Probleme und der zugrunde liegenden Faktoren, bei der Lösung von Erziehungsfragen sowie bei Trennung und Scheidung unterstützen (§ 28, SGB VIII, 1990). Obwohl die Zahl der Kinder in Deutschland abnimmt, steigt die Inanspruchnahme von Erziehungs- und Familienberatung kontinuierlich an. Wurden 1993 noch insgesamt 197.955 (beendete) Beratungen registriert, waren es im Jahr 2006 bereits 310.561. Somit hat die Zahl der Bera-

tungen in diesem Zeitraum um 56,9% zugenommen (Gerth & Menne, 2009).

Die Fragestellungen, mit denen sich Ratsuchende an die Erziehungs- und Familienberatung wenden, sind sehr vielfältig und können sich sowohl auf akute Krisen als auch auf länger anhaltende Probleme beziehen. Typische Anlässe für die Inanspruchnahme von Beratung sind z. B. soziale Verhaltensauffälligkeiten (Aggressivität, Geschwisterrivalität), Probleme im Schul- und Leistungsbereich (Leistungsversagen, Schulschwänzen), Schwierigkeiten in der familiären Interaktion (Gesprächsverweigerung, Ablösung vom Elternhaus) oder psychosomatische Auffälligkeiten (Einnässen, Ess- und Sprachstörungen). Tabelle 20.1 gibt einen Überblick über die häufigsten Beratungsanlässe in der Erziehungs- und Familienberatung.

Die im Bereich der Erziehungs- und Familienberatung beschäftigten Berufsgruppen sind z. T. sehr unterschiedlich ausgebildet. Neben Psychologen und (Sozial-)Pädagogen führen auch Ärzte, Psychotherapeuten und gelegentlich auch Geistliche oder Vertreterinnen von Frauenverbänden die Beratungen durch. Auch die Art der Beratungsziele variiert sehr stark: Viele Ratsuchende begnügen sich mit einer rein informatischen Beratung; andere möchten mithilfe intensiver

Beratungsgespräche schwerwiegende Probleme lösen. Eine wichtige Aufgabe ist oft die psychodiagnostische Abklärung einer möglichen Störung, die einer professionellen Intervention (z. B. Therapie) bedarf. In Kapitel 19 wird dies exemplarisch anhand der Diagnose einer Lese-Rechtschreib-Schwäche (LRS) erläutert (Fallbeispiel »Karl«).

Beratung in der Schule

Sowohl auf Schüler-, Eltern- und Lehrerseite als auch aufseiten der Schule als Organisation ist in den letzten Jahrzehnten ein stetig steigender Beratungsbedarf festzustellen. Mögliche Gründe dafür sind die sich verändernden gesellschaftlichen Rahmenbedingungen (vgl. Abschn. 13.2) und Familienkonstellationen (vgl. Abschn. 14.2), bildungspolitische Reformen sowie die starke Ausdifferenzierung der bestehenden Bildungsangebote und der damit verbundenen Notwendigkeit, individuelle Entscheidungen über mögliche Bildungswege bzw. Schullaufbahnen für jeden einzelnen Schüler zu treffen.

Die von Lehrkräften, Beratungslehrern oder Schulpsychologischen Diensten praktizierten Formen schulischer Beratung und Unterstützung erfüllen unterschiedliche Funktionen: Sie unterstützen z. B. Schüler

Tabelle 20.1 Anzahl beendeter Beratungen im Jahr 2006 nach Geschlecht und Anlässen der Beratung (Statistisches Bundesamt, 2007)

Beratungsanlässe	Männlich	Weiblich	Insgesamt
Beziehungsprobleme	67.030	57.487	124.517
Entwicklungsauffälligkeiten	48.173	30.724	78.897
Schul-/Ausbildungsprobleme	50.670	26.682	77.352
Trennung/Scheidung der Eltern	39.315	34.782	74.097
Sonstige Probleme in und mit der Familie	32.295	31.157	63.452
Anzeichen für sexuellen Missbrauch	2.344	6.983	9.327
Suchtprobleme	3.872	2.101	5.973
Straftat des Jugendlichen/jungen Volljährigen	3.963	1.097	5.060
Anzeichen für Misshandlung	2.228	2.752	4.980
Wohnungsprobleme	1.163	1.300	2.463
Insgesamt	173.780	136.781	310.561

Anmerkung: Für jeden jungen Menschen konnten bis zu zwei Anlässe der Beratung angegeben werden.

Tabelle 20.2 Beratungsanlässe in der Schule

	Themen	Beratende
Beratung von Schülern und Eltern		
Lernberatung	Lernstrategien, Gestaltung der häuslichen Lernumgebung, Unterstützung beim Lernen durch die Eltern, spezifische Teilleistungsschwächen (z. B. LRS, Dyskalkulie)	Lehrkräfte, Beratungslehrer, Schulpsychologen, externe Beratungsdienste
(Schul-)Laufbahnberatung	Einschulung, Versetzung, Kurswahl, Übergangsentscheidungen, Praktika, Studien- und Berufswahl	Lehrkräfte, Beratungslehrer, Berufsberater
Beratung bei psychosozialen und Suchtproblematiken; Erziehungsberatung	Schul- und Prüfungsangst, Verhaltensauffälligkeiten, Mobbing, persönliche/familiäre Probleme, Suchtproblematiken	Lehrkräfte, Beratungslehrer, externe Beratungsdienste
Beratung von Lehrkräften		
Beratung in fachlichen Fragen	Unterrichtsplanung und -gestaltung, Nutzung neuer Medien, Disziplinprobleme	Lehrkräfte, Beratungslehrer, Schulpsychologen, Supervisoren
Beratung bei persönlichen Problemen	Burn-out, Depressionen, Lehrerangst, soziale Probleme im Lehrerkollegium	Schulpsychologen, Beratungslehrer, Supervisoren, externe Beratungsdienste
Beratung von Schulen und Schulleitungen		
Personalentwicklung	Entwicklung und Umsetzung von Fortbildungsplänen	Schulaufsicht, Schulpsychologen, Personalentwickler
Organisationsentwicklung	Verbesserung des Lehrangebots, Umsetzung von Bildungsstandards, Entwicklung und Umsetzung innovativer Konzepte	Schulaufsicht, Organisationsentwickler

und deren Eltern bei der Entscheidung über eine angemessene Schullaufbahn oder geben Anregungen und Hinweise zur Verbesserung des Lernverhaltens im häuslichen Umfeld. Auf diese Weise trägt das Beratungsangebot auch zur Verbesserung der Bildungschancen sowie der individuellen Förderung von Schülern bei. Pädagogische Beratung im schulischen Kontext unterstützt auch die Zusammenarbeit von Schule und Elternhaus und ist oft eine Hilfestellung für Eltern bei allgemeinen Erziehungs- und Bildungsfragen (vgl. Abschn. 14.5). Bei psychologisch schwierigeren Problemen (z. B. Suchtproblematiken, Verhaltensauffälligkeiten) arbeiten Schulen mit externen Beratungsdiensten zusammen. Auch die kollegiale Beratung von Lehrkräften untereinander stellt ein wichtiges Beratungsfeld im schulischen Alltag dar. Sie dient der gegenseitigen

Unterstützung z. B. bei der Planung und Gestaltung von neuen Unterrichtsmethoden oder einzelner Unterrichtseinheiten, beim Umgang mit schwierigen Schülern aber auch bei der Bewältigung persönlicher Probleme der Lehrkräfte. Beratung unterstützt auch Schulen bzw. Schulleitungen bei der Verbesserung und Individualisierung des Lehrangebots, der Umsetzung von Bildungsreformen und der Entwicklung und Verwirklichung von Innovationen. Die Beratung von Schulen im Sinne einer Organisationsberatung geschieht dabei – im Gegensatz zur Beratung von Schülern, Eltern und Lehrkräften – in der Regel durch externe Berater, die nicht aus der eigenen Schule kommen (z. B. Schulaufsicht). Tabelle 20.2 gibt einen Überblick über Beratungsanlässe, Adressaten und beratend Tätige im schulischen Kontext.

Beratung in der Hochschule

Beratung in der Hochschule kann nach ihren verschiedenen Adressaten in unterschiedliche Aufgabenbereiche unterteilt werden, welche je nach Anlass von verschiedenen Personen oder Institutionen (Allgemeine Studienberatung, Studienfachberatung, Lehrende, hochschuldidaktische Zentren, externe Berater) wahrgenommen werden.

Den Bereich mit der größten Nachfrage stellt die Beratung von Studierenden dar. Laut der Sozialerhebung des Deutschen Studentenwerks im Jahr 2009 bestand bei über 60 % der Studierenden ein Beratungs-

und Informationsbedarf bei mindestens einem Themenbereich (Isserstedt, Middendorff, Kandulla, Borchert & Leszczensky, 2010). Im Vordergrund stehen dabei finanzielle Themen, studienbezogene Leistungsprobleme und private Probleme der Studierenden im Hinblick auf die eigene Persönlichkeitsentwicklung bzw. ihre sozialen Beziehungen (s. Abb. 20.4).

Zwei weitere wichtige Bereiche stellen die Beratung von Lehrenden und der Hochschule als Institution dar. Tabelle 20.3 gibt einen Überblick über häufige Beratungsanlässe in diesen beiden Beratungsfeldern im Kontext der Hochschule.

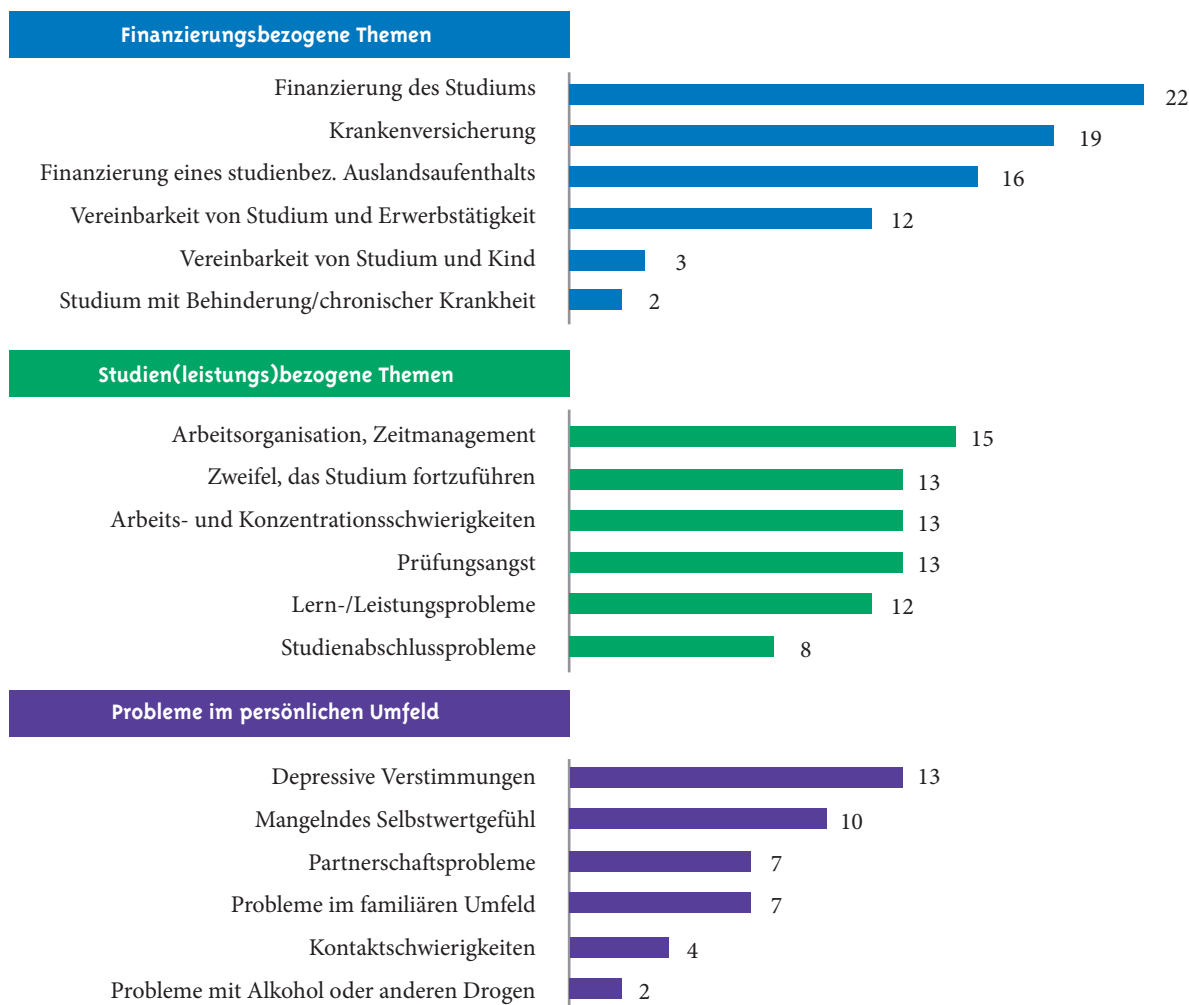


Abbildung 20.4 Beratungs- und Informationsbedarf von Studierenden im Erststudium, in Prozent (Mehrfachnennungen möglich; Isserstedt et al., 2010)

Tabelle 20.3 Anlässe von Beratung der Lehrenden und der Institution Hochschule

Beratung der Lehrenden

Beratung zur Qualität der Lehre

- ▶ Hochschuldidaktik: Verbesserung der Lehrkompetenzen der Dozenten
- ▶ Unterstützung der Lehrenden in Form von Supervision bei der Reflexion ihrer Lehrpraxis

Beratung bei psychosozialen Problemen

- ▶ Konflikte mit Vorgesetzten oder Mitarbeitern, Mobbing
- ▶ Depressionen, Burn-out

Beratung der Institution Hochschule

Beratung bei der Reformierung des Hochschulsystems

- ▶ Einführung neuer Studiengänge und Studienabschlüsse
- ▶ Festlegung von Eingangsbedingungen für bestimmte Studiengänge
- ▶ Neuorganisation des Lehrangebots durch Nutzung neuer Informations- und Kommunikationstechniken
- ▶ Ermittlung und Einhaltung von Qualitätsstandards in der Lehre
- ▶ Auswahl und Einsatz von Selektions- und Evaluationsinstrumenten

Beratung in der Aus- und Weiterbildung

Zwei der wichtigsten Beratungsfelder in diesem Kontext stellen die Berufsberatung sowie die berufliche Laufbahnberatung dar. Sie werden in Deutschland in der Regel durch die Bundesagentur für Arbeit und ihre Dienststellen angeboten, die laut Gesetz (§ 29, SGB III, 1997) den Auftrag haben, Jugendliche und Erwachsene vor Eintritt in das Berufsleben und während des Berufslebens in allen Fragen der Berufswahl und des beruflichen Fortkommens zu beraten. Wie in jedem Beratungsprozess geht es auch bei der Berufs- und Laufbahnberatung darum, Prozesse und Lösungsstrategien beim Ratsuchenden anzustoßen, die es ihm ermöglichen, autonom und reflektiert das zukünftige Berufs- und Arbeitsleben zu gestalten. Neben der Vermittlung von Informationen bezüglich verschiedener Berufs- und Studienmöglichkeiten bietet die Berufs- und Laufbahnberatung den Ratsuchenden Unterstützung bei der Realisierung eines selbstständigen Laufbahnmanagements. So werden im Rahmen des Beratungsprozesses individuelle Kompetenzen, Ziele und Wünsche des Klienten herausgearbeitet sowie die Verankerung der beruflichen Planung in seine persönliche Lebensplanung angestrebt. Im Bereich der Laufbahn- und Personalentwicklung kommen neben der Beratung häufig auch Coaching und Supervision (s. Abschn. 20.3.2) zur Anwendung (vgl. Sonntag, 1999).

Übersicht

Themen der Berufsberatung

Laut § 30, SGB III, 1997 befasst sich die Berufsberatung mit Fragen

- ▶ zur Berufswahl, zur beruflichen Entwicklung und zum Berufswechsel,
- ▶ zur Lage und Entwicklung des Arbeitsmarktes und der Berufe,
- ▶ zu den Möglichkeiten der beruflichen Bildung,
- ▶ zur Ausbildungs- und Arbeitsstellensuche,
- ▶ zu Leistungen der Arbeitsförderung,
- ▶ zu Möglichkeiten der Ausbildungsförderung und der schulischen Bildung, soweit sie für die Berufswahl und die berufliche Bildung von Bedeutung sind.

20.2 Intervention

20.2.1 Relevanz und Ansatzpunkte in pädagogischen Handlungsfeldern

Relevanz und Zielsetzungen von Interventionen

Das zielgerichtete Eingreifen in Entwicklungs- und Bildungsprozesse ist ein grundlegender Aspekt pädagogischen Handelns. Dies ist immer dann notwendig, wenn Entwicklungs- oder Bildungsprozesse ungünstig bzw. anders als erwartet verlaufen. Mithilfe von speziellen Maßnahmen sollen pädagogisch erwünschte Entwick-

lungen angeregt und Bildungsprozesse optimiert werden. Dies kann durch einen Abbau ungünstiger Verhaltensweisen oder durch einen Aufbau erwünschter Verhaltensmuster erfolgen; manchmal werden beide Änderungsrichtungen gleichzeitig verfolgt (Hager & Hasselhorn, 2000). Die gezielte Auswahl und Umsetzung von Maßnahmen zur Unterstützung günstiger Entwicklungs- und Bildungsprozesse kann unter dem Begriff »Intervention« zusammengefasst werden. Pädagogische Interventionen erfüllen mindestens drei wichtige Zwecke (s. Hascher & Schmitz, 2010):

- (1) Sie unterstützen erwünschte Entwicklungs- und Bildungsprozesse in der pädagogischen Praxis,
- (2) tragen zur Theorieentwicklung (z.B. im Bereich der Lehr- und Lernforschung) bei und
- (3) liefern empirisch basiertes Entscheidungswissen bei der Planung, Realisierung und Veränderung pädagogischer Maßnahmen (z.B. Einführung von Förderprogrammen).

Ansatzpunkte von Interventionen

Die psychologisch orientierten Interventionen können auf unterschiedlichen Ebenen pädagogischer Handlungsfelder angesiedelt sein (vgl. Abschn. 1.1.1). Sie können das Individuum betreffen und sich z.B. auf kognitive, motivationale und emotionale Aspekte der Lernprozesse der Person beziehen (Mikroebene). Außerdem können sie auf Veränderungen von Prozessen in Schulklassen, Schulen oder anderen (Bildungs-)Institutionen abzielen, z.B. durch die Einführung innovativer Unterrichtskonzepte oder die Umsetzung eines Beratungskonzepts an einer Schule (Mesoebene). Darüber hinaus können sich Interventionen auf Aspekte des Bildungssystems beziehen wie beispielsweise die Einführung von kompetenzorientiertem Unterricht oder die Umsetzung eines bundesweiten Bildungsmonitorings (Makroebene; vgl. Abschn. 19.2.1).

20.2.2 Begriffsbestimmung, Aspekte und Evaluation

Definition von Intervention

Der Begriff Intervention geht zurück auf das lateinische Verb »intervenire«, welches sich mit »dazwischentreten« oder »sich einschalten« übersetzen lässt. Bei einer Intervention in pädagogischen Kontexten kann es sich um eine einzelne Maßnahme handeln oder um ein Set von mehreren Maßnahmen. Die folgende Definition von

Hager und Hasselhorn (2000, S. 41) nennt die wichtigsten Bestimmungsstücke einer Interventionsmaßnahme.

Definition

Unter Interventionsmaßnahmen verstehen wir jede Art von außengesteuerter, zielorientierter und systematischer Beeinflussung von Personen- und/oder Systemmerkmalen. Jede Interventionsmaßnahme besteht mindestens aus einer Menge zu bearbeitender Aufgaben bzw. Probleme und mindestens einer Handlungsstrategie (z.B. Instruktion, Training).

Ein zentrales Unterscheidungsmerkmal von Interventionen gegenüber Beratung und Supervision ist die stärkere Außensteuerung. Bei professionellen Formen der Intervention werden die erforderlichen Maßnahmen von »außen«, d.h. von Fachleuten, geplant und umgesetzt. Beratung und Supervision hingegen erfolgen unter dem Leitsatz »Hilfe zur Selbsthilfe« – die Verantwortung für die Entwicklung und Umsetzung von Maßnahmen, die Veränderungen bewirken sollen, liegt bei dem Ratsuchenden bzw. dem »Supervisanden«. Auch bei der Intervention ist der Anteil der Partizipation durch die Teilnehmer ein wichtiges Kriterium für den Erfolg der Maßnahme, allerdings ist die Maßnahme vorstrukturiert bzw. vorgegeben.

Die Intervention sollte auf einer fundierten Diagnostik der Ausgangssituation beruhen und auf einen erwünschten Zielzustand hinarbeiten. Das heißt, eine Intervention erfolgt in der Regel, wenn die Diagnostik der Ausgangssituation Problembereiche aufdeckt oder ungünstige Entwicklungen sichtbar macht bzw. wenn ein Defizit identifiziert wurde (vgl. Abschn. 19.1.2). Die Diagnostik sollte dabei auf die Sachverhalte jener Ebene abzielen, die durch die Intervention beeinflusst werden soll (Mikro-, Meso-, Makroebene), d.h., sie kann entweder am Individuum ansetzen, an Unterrichtsbedingungen oder institutionellen Gegebenheiten oder an der (Bildungs-)Systemebene (vgl. Abschn. 6.3.1).

Der angestrebte Zielzustand sowie die Maßnahmen, die zur systematischen Beeinflussung eingesetzt werden, sollten möglichst aus empirisch hinreichend fundierten pädagogisch-psychologischen Theorien abgeleitet werden. Die theoretische Fundierung ist neben der wissenschaftlichen Wirksamkeitsüberprüfung (Evaluation; vgl. Abschn. 19.2) ein zentrales Qualitätskriterium für pädagogische Interventionen.

Kriterien für die Beschreibung von Interventionen

Um die Vielzahl an Interventionen in pädagogischen Handlungsfeldern zu beschreiben und zu unterscheiden, ist es hilfreich, sich an Kriterien zu orientieren. Interventionen lassen sich nach den folgenden Kriterien differenzieren:

- ▶ Zielsetzung
- ▶ Programmkomponenten
- ▶ Zugangswege
- ▶ Dosis und Zeitraum der Intervention
- ▶ Wirkung und Effekte

Zielsetzung. Im Hinblick auf die Zielsetzung lassen sich präventive, kurative und rehabilitative Interventionen unterscheiden.

- (1) Präventive Interventionen zielen darauf ab, im Anschluss an die Diagnose eines Gefährdungspotenzials Maßnahmen einzuleiten, um die antizipierten unerwünschten Folgen abzuwenden. Ein Beispiel sind Deutschkurse vor der Einschulung. Es ist ein gut gesicherter Befund, dass mangelnde Sprachkenntnisse eine Gefährdung für schulischen Erfolg darstellen. Wird bei der Einschulungsuntersuchung ein sprachliches Defizit erkannt (= Gefährdungspotenzial), erfolgt die Empfehlung für eine Teilnahme an einem Sprachkurs.
- (2) Kurative Interventionen umfassen Maßnahmen, die darauf abzielen, einen unerwünschten Zustand zu verändern oder eine bereits eingetretene Fehlentwicklung rückgängig zu machen. Ein Beispiel für kurative Interventionen sind verhaltenstherapeutische Programme zur Reduktion von Verhaltensauffälligkeiten in der Schule, etwa dem fortlaufenden Stören im Unterricht durch eine Schülerin. Auf der Grundlage des Verstärkungsprinzips könnte die Schülerin z.B. Bonuspunkte erhalten, wenn sie innerhalb einer bestimmten Zeitspanne nicht gestört hat. Diese Punkte können später gegen eine Belohnung eingetauscht werden.
- (3) Rehabilitative Interventionen beziehen sich auf Maßnahmen, die eingesetzt werden, um Rückfälle in unerwünschte Verhaltensmuster zu verhindern bzw. diesen vorzubeugen. Ein Beispiel für rehabilitative Interventionen sind Trainings oder Trainingsprogramme zum Umgang mit beruflichen Belastungen für Lehrkräfte, die nach einem Burn-out und dessen erfolgreicher Behandlung in den Berufsalltag zurückkehren.

Programmkomponenten. Interventionen können aus einer Kombination von Maßnahmen oder Komponenten eines umfassenden Programms bestehen: Sie können Bausteine zur Wissensvermittlung und/oder zur Vermittlung und Einübung von Handlungsstrategien enthalten oder auf die Änderung von Verhalten abzielen; zusätzlich können z.B. Komponenten zur Unterstützung der Reflexion und des Transfers sowie zur Rückmeldung von (Lern-)Fortschritten eingesetzt werden. Ein Interventionsprogramm zur Förderung der Beratungskompetenz von Lehrkräften könnte z.B. die folgenden drei Komponenten umfassen: (1) Vermittlung von Wissen, (2) Unterstützung der Selbstreflexion, (3) Rückmeldung zur Qualität der Beratungsleistung in Rollenspielen durch die Beurteilung von Peers. Diese drei Programmkomponenten können im Rahmen der Intervention gemeinsam eingesetzt werden, sie können aber auch einzeln durchgeführt bzw. in unterschiedlichen Kombinationen umgesetzt werden. Durch die Kombination der Programmkomponenten ergeben sich unterschiedliche Arten der Intervention.

Zugangswege. Bezogen auf die Zugangswege lassen sich direkte und indirekte Interventionen unterscheiden. Bei einer direkten Intervention setzen die Maßnahmen z.B. direkt an den Individuen bzw. der Lerngruppe (z.B. Schulklasse) oder dem Bildungssystem an. Ein Beispiel ist ein Lernstrategietraining für Schüler mit ungünstigem Lernverhalten. Im Rahmen des Trainings werden den Schülern konkrete Lernstrategien vermittelt. Indirekte Interventionen setzen hingegen im Umfeld der Individuen oder Lerngruppen an. Dabei wird erwartet, dass durch die Veränderung des Umfeldes auch Veränderungen bei der eigentlichen Zielgruppe erreicht werden können. Für das Beispiel des Lernstrategietrainings würde dies bedeuten, dass hier z.B. die Eltern der Schüler darin geschult werden, wie sie ihr Kind beim Lernen unterstützen können. Dahinter steht die Hypothese, dass sich das Training der Eltern auch positiv auf das Lernverhalten der Schüler auswirkt.

Dosis und Zeitraum der Intervention. Diese Kategorien bestimmen die Intensität und den Umfang der Maßnahme. Die Dosis kennzeichnet die Anzahl der Interventionseinheiten, z.B. die Anzahl der Trainingseinheiten bei einem Trainingsprogramm gegen Prüfungsängstlichkeit. Der Zeitraum der Intervention beschreibt die Zeitspanne, über die die Intervention durchgeführt wird, z.B. die Anzahl der Wochen, die das Trainingsprogramm gegen Prüfungsängstlichkeit dauert. Eine

hohe Intensität liegt vor, wenn eine hohe Dosierung (= viele Trainingseinheiten) in einem kurzen Interventionszeitraum (= geringe Zeitspanne) umgesetzt wird.

Wirkung und Effekte der Intervention. Beobachtbare Effekte können sofort einsetzen oder erst nach einiger Zeit auftreten (Sleeper-Effekt). Ein sofortiger Effekt ist z. B. das Ansteigen der Behaltensleistung nach einem Gedächtnistraining. Ein zeitverzögerter Effekt ist z. B. die Veränderung von Einstellungen und Überzeugungen durch eine Fortbildung. Zudem können kurzfristige, mittelfristige und langfristige Effekte von Interventionen betrachtet werden. Der kurzfristige Effekt eines Beratungstrainings für Lehrkräfte könnte die souveränere Anwendung von Gesprächsführungsstrategien in Beratungsgesprächen mit Eltern sein. Mittelfristig könnte die Intervention zu einem größeren Beratungserfolg (z. B. einer höheren Zufriedenheit der Eltern sowie der Schüler) führen, welcher sich langfristig in einer erhöhten Selbstwirksamkeit der Lehrkraft in der Beratung von Eltern auswirkt. Auf der Grundlage konzeptioneller und theoretischer Überlegungen lassen sich Annahmen über

die Wirkung einer Intervention ableiten. Diese können dann als Hypothesen formuliert und empirisch überprüft werden (vgl. Abschn. 4.3 und 19.1.2).

Vorgehen bei der Konzeption und Durchführung von Interventionen

Ein Rahmenmodell für die Konzeption und Durchführung von Interventionen findet sich bei Landmann, Schmidt und Schmitz (2010). In dem Modell werden drei Phasen unterschieden (vgl. Abb. 20.5):

- (1) die Phase vor der Intervention (= Konzeptionsphase), in der die Zielsetzung der Intervention formuliert wird und Inhalte und Vermittlungsstrategien ausgewählt werden,
- (2) die Phase während der Intervention (= Durchführungsphase), in der die Maßnahme(n) vorbereitet und durchgeführt wird/werden, und
- (3) die Phase nach der Intervention (= Evaluations-/Reflexionsphase), in der die Evaluation und Reflexion der Intervention und ggf. eine Optimierung der Intervention erfolgen.

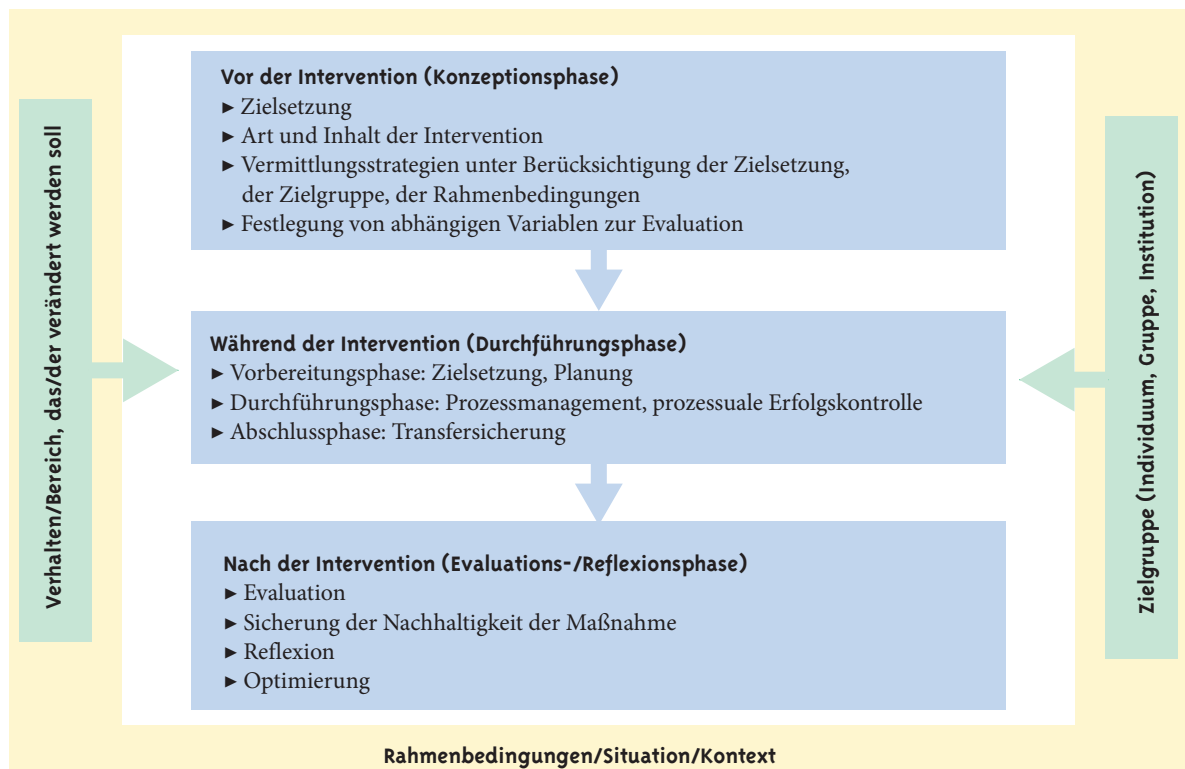


Abbildung 20.5 Modell zur Konzeption und Durchführung von Interventionen (Landmann et al., 2010)

Aspekte der Wirksamkeit von Interventionen

Aus wissenschaftlicher Perspektive sind insbesondere Beiträge zur Weiterentwicklung von Theorien sowie zur Bereitstellung von empirisch basiertem Entscheidungswissen für die Umsetzung von Innovationen und die Veränderung von pädagogischem Handeln von Bedeutung. Deshalb nimmt die Überprüfung der Wirksamkeit von Interventionen einen zentralen Stellenwert ein. Im Hinblick auf die Weiterentwicklung von Theorien stellen sich beispielsweise Fragen danach, welche konkrete Maßnahmen geeignet sind, um die erwünschten Veränderungen zu erzielen, oder welche zusätzlichen Programmkomponenten (z. B. Reflexionsunterstützung, Feedback) die Veränderungen unterstützen. Bezogen auf die Bereitstellung von Entscheidungswissen ist zu klären, unter welchen Bedingungen eine Intervention erfolgreich sein kann (z. B. Vorwissen, Veränderungsbereitschaft, Überzeugungen) oder welchen Nutzen die Intervention im Vergleich zu den Kosten aufweist.

Interventionsforschung. Diese Fragen werden im Rahmen der Interventionsforschung behandelt. Die Übergänge zwischen Grundlagenforschung und angewandter Forschung (z. B. im Bereich der Evaluation) sind dabei häufig fließend (vgl. Leutner, 2010a). Die Interventionsforschung zielt darauf ab, die Wirkungen einer Intervention unter Einsatz wissenschaftlicher Methoden zu evaluieren und dabei systematische Variationen von Maßnahmen vorzunehmen sowie Eingangsbedingungen der Teilnehmer zu kontrollieren (vgl. Hascher & Schmitz, 2010; Leutner, 2010). Hierfür wird häufig auf

die vier Ebenen der Evaluation nach Kirkpatrick (1998; s. u.) und die in Abschnitt 19.2 beschriebenen Strategien der Evaluationsforschung zurückgegriffen.

Im Rahmen (quasi-)experimenteller Studien (vgl. Abschn. 4.3.4) sollen kausale Schlussfolgerungen über die Wirksamkeit der Intervention abgeleitet werden. Meistens werden Daten zu mindestens zwei Zeitpunkten (vor und nach der Intervention) in zwei Gruppen erhoben, nämlich einer Experimentalgruppe, in welcher die zu überprüfende Intervention realisiert wurde, und einer Kontrollgruppe ohne Intervention bzw. mit alternativem Treatment. Die gewonnenen Daten werden dann mithilfe statistischer Verfahren daraufhin analysiert, ob sich signifikante Unterschiede zwischen den Veränderungen in den beiden Gruppen nachweisen lassen (vgl. Abb. 20.7). Um die Wirksamkeit eines Lernstrategietrainings für Schüler zu untersuchen, könnten z. B. die Zuwächse in der Anwendung von Lernstrategien bei Schülern, die ein Lernstrategietraining erhalten haben (Experimentalgruppe), mit denen von Schülern, die ein Training zur Rechtschreibung erhalten haben (Kontrollgruppe), verglichen werden. Eine besondere Herausforderung der Interventionsforschung in realen Kontexten (z. B. in Schulklassen) ist, dass viele Bedingungen und Einflussgrößen nicht wie im Labor kontrolliert werden können. Dies muss bei der Planung und Durchführung der Intervention, aber auch der Datenerhebung (z. B. Auswahl der Instrumente, Messzeitpunkte) sowie bei der Interpretation der Ergebnisse berücksichtigt werden (s. Abschn. 19.2.4).

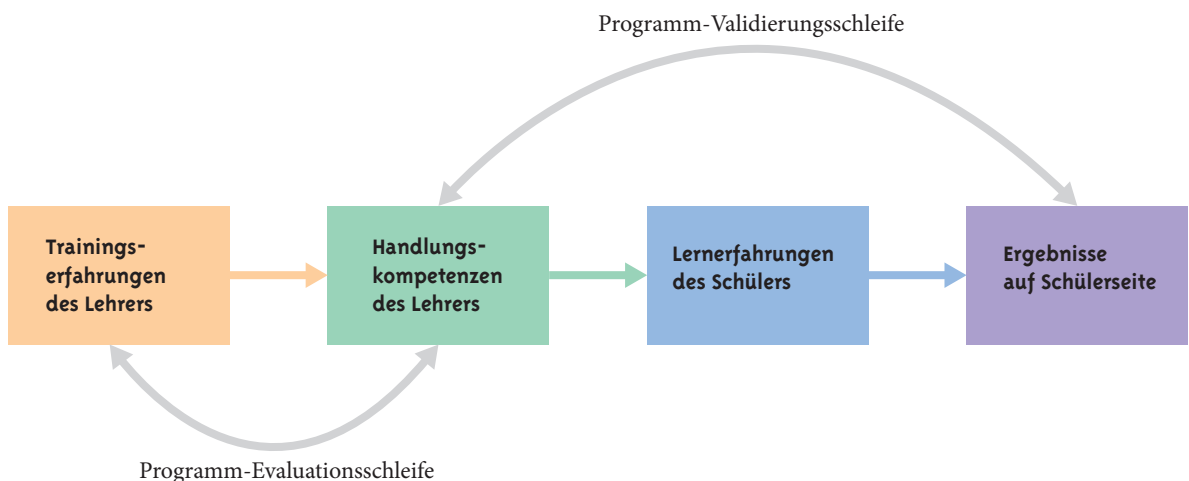


Abbildung 20.6 Modell zur Planung der Evaluation von Interventionen (Galluzzo & Craig, 1990)

Ein hilfreiches Modell zur Planung der Evaluation von Interventionen wurde von Galluzzo und Craig (1990) vorgeschlagen. Ihr Modell (vgl. Abb. 20.6) berücksichtigt sowohl direkte als auch indirekte Interventionseffekte: Während sich die »Programm-Evaluationschleife« auf die Veränderung von Handlungskompetenzen bezieht (direkte Effekte), richtet sich die »Programm-Validierungsschleife« auf nachfolgende, durch die gesteigerten Handlungskompetenzen ausgelöste indirekte Effekte.

Vier Ebenen der Evaluation nach Kirkpatrick. Die Wirkungen von Interventionen sollten stets systematisch überprüft bzw. evaluiert werden. Kirkpatrick (1998) nennt hier vier Ebenen der Evaluation von Interventionen:

- (1) Reaktion: Erfassung von Akzeptanz/Zufriedenheit mit der Maßnahme
- (2) Lernen: Erfassung von Wissens- und Kompetenzzuwachs, Veränderungen in Einstellungen und Überzeugungen
- (3) Verhalten: Erfassung von Verhaltensänderungen und Transfer in den Alltag
- (4) Ergebnisse: Erfassung von Ergebnissen/Veränderungen auf Organisations- und Systemebene

Diese vier Ebenen bauen aufeinander auf, d. h., wenn eine Intervention nicht positiv von den Teilnehmern bewertet wird, sind auch keine Effekte auf den folgenden Ebenen (Lernen, Verhalten, Ergebnisse) zu erwarten. Eine gewisse Zufriedenheit mit der Intervention ist demnach eine notwendige – aber nicht hinreichende – Voraussetzung für das Lernen, Verhaltensänderungen und Ergebnisse auf der Meso- und Makroebene. Die Aufschlüsselung und Betrachtung der Ergebnisse einer Intervention auf diesen vier Ebenen ermöglicht eine systematischere Interpretation und Diskussion der Befunde. Wenn eine Intervention nicht die gewünschten bzw. erwarteten Erfolge erzielt, kann eine entsprechend differenzierte Betrachtung dazu beitragen, die Intervention zu optimieren und ihre Wirksamkeit zu steigern. Folgende Fragen können gestellt werden: Waren die Teilnehmer mit der Intervention zufrieden? Kam es durch die Intervention zu einem Wissens-/Kompetenzzuwachs? Wurden durch die Intervention Verhaltensänderungen bewirkt? Welche Wirkungen konnten auf Systemebene gefunden werden?

Elf Schritte der Evaluation von Interventionen. In wissenschaftlichen Evaluationen können unterschiedliche Kriterien zur Beurteilung der Wirksamkeit von Interventionen herangezogen werden, z. B. Art und Ausmaß

des Leistungszuwachses (Performanzsteigerung) bei den Teilnehmern der Interventionsmaßnahme, Vergleich verschiedener Maßnahmen mit ähnlicher Zielstellung, detaillierte Untersuchungen über die Wirkungsweise (»Ursachen«) für Erfolg und Misserfolg. Hager und Hasselhorn (1995) unterscheiden in diesem Zusammenhang elf Schritte der Evaluation; diese umfassen sowohl die Betrachtung trainingsnaher Aspekte als auch weiterführende und vertiefende Analysen zur differenziellen Wirksamkeit von Interventionen:

- (1) trainingsnahe Performanzsteigerung
- (2) Performanzsteigerung trainingsferner Variablen
- (3) Kompetenzverbesserung (für trainingsferne abhängige Variablen)
- (4) Vergleich mit Konkurrenzprodukten
- (5) weiter Transfer auf verwandte Bereiche
- (6) Ausbleiben von Transfer auf nicht verwandte Bereiche
- (7) Transfer in den Alltag
- (8) Trainingswirkung in Abhängigkeit von Persönlichkeitsvariablen
- (9) Analyse der Wirkmechanismen
- (10) vertiefende Fragestellungen
- (11) Metaevaluation

Diese elf Schritte sind zwar als eine Abfolge von evaluativen Bewertungsmaßnahmen gedacht, aber sie bauen nicht zwingend aufeinander auf. Der Nachweis eines Transfers auf verwandte Bereiche (Punkt 5) ist z. B. unabhängig von einem Vergleich mit Konkurrenzprodukten (Punkt 4). Hier ist im Vorfeld der Intervention genau zu planen, welche Evaluationsschritte im Rahmen der Studie durchgeführt werden sollen. In Abhängigkeit vom Anlass und der konkreten Zielstellung der Evaluation können die mit diesen Schritten verbundenen Evaluationskriterien als mehr oder weniger verbindlich (obligatorisch oder fakultativ) eingeschätzt werden.

20.2.3 Anwendungsfelder im pädagogischen Bereich

Zentrale Anwendungsfelder

Interventionen in pädagogischen Handlungsfeldern beziehen sich in der Regel entweder auf die Unterstützung von Lern- und Bildungsprozessen (z. B. in Kindertageseinrichtungen; vgl. Abschn. 15.5) oder auf spezifische Verhaltensweisen, die aus pädagogisch-psychologischer Perspektive als interventionsbedürftig eingestuft werden (z. B. Verhaltensauffälligkeiten, Prüfungsangst, um-

schriebene Lernstörungen). Die Interventionen können wie dargestellt auf Mikro-, Meso- oder Makroebene ansetzen und sich auf unterschiedliche Entwicklungsphasen (Säuglings- und Kleinkindalter bis höheres Erwachsenenalter) oder Bildungsphasen (Vorschule, Schule, Ausbildung/Hochschule, Fort- und Weiterbildung) der Individuen beziehen. Entsprechend weit ist das Spektrum pädagogisch-psychologischer Interventionen; es reicht von Trainingsprogrammen zu Motivation oder sozialer Kompetenz über Trainings zur Optimierung selbstregulierten Lernens bis hin zu Aspekten der Werteerziehung oder der Umsetzung von Bildungsstandards. Interventionen erfolgen in der Regel in Form von Trainingsprogrammen oder im Rahmen von Fort- und Weiterbildungsveranstaltungen (s. Hascher & Schmitz, 2010; Spiel, Schober, Wagner & Reimann, 2010).

Besonderheiten bei der Durchführung von Interventionen in Schule und Unterricht

Bei der Durchführung von Interventionen in Schule und Unterricht sind einige Besonderheiten zu beachten. Zunächst ist es wichtig zu entscheiden, ob eine direkte Intervention erfolgen sollte, d. h. die Intervention bei der Zielgruppe direkt – also den Lehrkräften oder den Schülern – ansetzt oder ob eine indirekte Intervention geplant ist, z. B. eine gezielte Fortbildung der Lehrkräfte, um bei den Schülern bestimmte Effekte zu erzielen. Bei indirekten Interventionen besteht eine besondere Schwierigkeit darin sicherzustellen, dass die Lehrkräfte die intendierten Interventionen im Unterricht tatsächlich so wie geplant umsetzen. Hier kann es zu Abweichungen kommen, aus denen sich systematische Unterschiede in der Umsetzung der geplanten Maßnahmen ergeben können (z. B. eine Lehrkraft, die sich genau an die Vorgaben hält, und eine Lehrkraft, die immer nur

Teile der Vorgaben umsetzt). Zudem ist mit sog. Moderationseffekten zu rechnen: So hat sich z. B. gezeigt, dass die Einstellungen und Überzeugungen einer Lehrkraft zu bestimmten Aspekten der Intervention (z. B. Erwartungen bezüglich deren Wirksamkeit oder Passung mit den eigenen pädagogischen Zielen) einen großen Einfluss auf die Umsetzung der Inhalte und den Transfer in den Schulalltag besitzen. Um Abweichungen von der intendierten Intervention möglichst gering zu halten, werden die Interventionsschritte in Trainingsmanualen für die Lehrkräfte möglichst konkret beschrieben, die erforderlichen Materialien werden bereitgestellt und es werden Foren oder Sprechstunden eingerichtet, in denen die Lehrkräfte noch einmal Fragen stellen können. Insgesamt sind bei indirekten Interventionen schwächere Effekte zu erwarten als bei direkten Interventionen.

Weiterhin ist zu beachten, ob die Intervention in der Unterrichtszeit oder außerhalb des Unterrichts (z. B. am Nachmittag) stattfindet. Aufgrund curricularer Vorgaben ist es nicht immer ohne Weiteres möglich, Interventionen während der Unterrichtszeit durchzuführen. Wichtig ist in diesem Kontext, ob es sich um ein in den Fachunterricht integriertes bzw. auf den Fachunterricht bezogenes Programm (z. B. Programme zum Lesenlernen) oder um fächerübergreifende Aspekte (wie das selbstregulierte Lernen oder soziale Kompetenzen) handelt. An manchen Schulen werden bereits eigene Stundenkontingente für die Weiterentwicklung fächerübergreifender Kompetenzen bereitgestellt.

Grundsätzlich müssen bei Interventionen für geplante Befragungen oder Beobachtungen von Schülern Einverständniserklärungen der Erziehungsberechtigten, der Schulleitung (manchmal auch der Schulkonferenz) und der zuständigen Behörden (z. B. Schulaufsicht, Ministerien) eingeholt werden.

Beispiel

Lehrertraining zum selbstregulierten Lernen

Das Training »So unterstütze ich meine Schüler beim Lernen lernen« (Hertel, 2007) richtet sich an Mathematiklehrkräfte im Grundschulzweig, die in der 3. Jahrgangsstufe unterrichten. Ziel des Trainings ist es, Strategien zu vermitteln, mit denen Lehrkräfte das selbstregulierte Lernen ihrer Schüler fördern können. Die Konzeption erfolgte nach dem Prozessmodell der Selbstregulation (Schmitz & Schmidt, 2007). Die Intervention ist für Gruppen von 12 bis 20 Lehrkräften konzipiert und umfasst drei Einheiten. Es handelt sich

um eine direkte Intervention für die Lehrkräfte und eine indirekte Intervention für die Schüler.

In der ersten Trainingseinheit werden allgemeine Grundlagen des selbstregulierten Lernens sowie die Aspekte Motivation, Zielsetzung und Gewohnheiten der präaktionalen Phase (vor dem Lernen) vermittelt. Die Themen Aufschieben von Aufgaben, Lernen und Bewegung, Lernen und Entspannung sowie Konzentration sind Inhalt der zweiten Trainingseinheit. Hierbei werden Strategien vermittelt, die Schüler während

des Lernens bei der Bearbeitung von Aufgaben einsetzen können (aktionale Phase). Zudem werden mathematische Problemlösestrategien (Skizze, Überschlag, Probe) vermittelt. In der letzten Einheit des Trainings werden Inhalte behandelt, die nach dem Lernen (post-aktionale Phase) eine Rolle spielen, z. B. Ursachenzuschreibung, Bezugsnormorientierung und informatives Feedback.

Die Evaluation der Fortbildung bezog Lehrkräfte sowie Schüler mit ein. Als Kontrollgruppe wurden Lehrkräfte, die nicht an der Fortbildung teilgenommen hatten, und ihre Schüler herangezogen. Im Prä- und Posttest wurde erfragt, wie häufig die Lehrkräfte Strategien zur Förderung des selbstregulierten Lernens im

Unterricht einsetzen (Handlungskompetenz der Lehrkräfte). Die Schüler wurden vor und nach der Fortbildung ihrer Lehrkräfte zum Lernen und zu ihrem Mathematikunterricht befragt. Außerdem bearbeiteten sie einen mathematischen Problemlösetest.

Es zeigte sich, dass die Lehrkräfte, die an dem Training teilgenommen hatten (Experimentalgruppe), die beschriebenen Methoden im regulären Fachunterricht häufiger einsetzten als die Lehrkräfte in der Kontrollgruppe. Auf Schülerebene zeigte sich, dass Schüler, deren Lehrkraft an der Fortbildung teilgenommen hatte, häufiger mathematische Problemlösestrategien einsetzten als Schüler in der Kontrollgruppe.

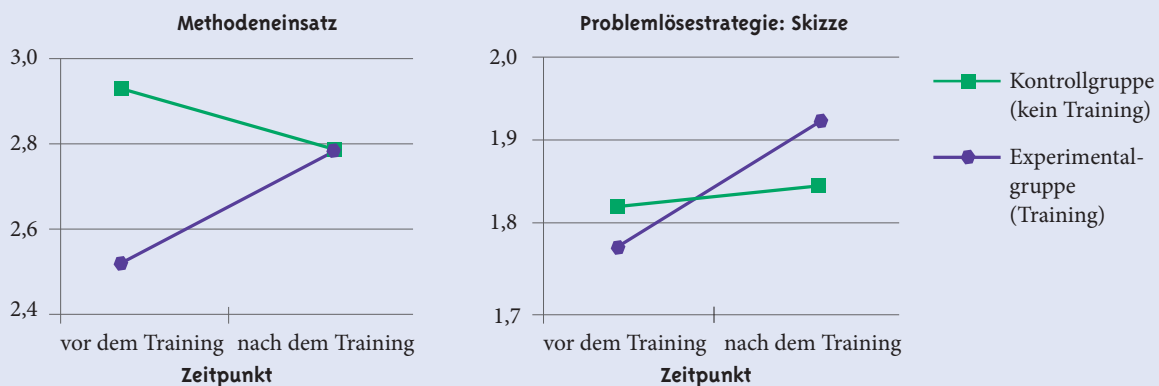


Abbildung 20.7 Methodeneinsatz im Mathematikunterricht durch die Lehrkräfte und Anwendung der Problemlösestrategie »Skizze« durch die Schüler

20.3 Supervision

20.3.1 Relevanz in pädagogischen Handlungsfeldern

Supervision hat nach Petermann (1995) insbesondere für beratende, soziale, erzieherische, therapeutische und pflegende Berufe eine herausragende Bedeutung. Dies liegt u. a. daran, dass Menschen, die mit anderen Menschen eng zusammenarbeiten und diese professionell beraten oder therapieren, hohen Belastungsfaktoren ausgesetzt sind. Bei ihnen besteht auch eine höhere Gefahr an Burn-out zu erkranken als bei anderen Berufsgruppen (Bauer, Häfner, Kächele, Wirsching & Dahlbender, 2003).

Supervision kann dazu beitragen, diese Belastungssituationen besser zu verarbeiten und das Risiko für Burn-out zu vermindern (Petermann, 1995; Ehinger & Hennig, 1994). Während Supervision in therapeutischen Berufen seit vielen Jahren zum Standard gehört und auch in den sozialen und erzieherischen Einrichtungen schon vermehrt Eingang gefunden hat, wird sie im schulischen Alltag bislang nur relativ selten genutzt und oft auch nicht vorbehaltlos angenommen (Palzkill, 1995). Angebote für Supervision in pädagogischen Kontexten richten sich nicht nur an Lehrkräfte, Ausbilder oder Trainer, sondern auch an Schulleiter, Schulpsychologen, Erziehungsberater oder Erzieher (Petermann, 1995).

Übersicht

Adressaten von Supervision in pädagogischen Handlungsfeldern

Die Zielgruppen von Supervision in pädagogischen Handlungsfeldern umfassen alle Personengruppen, bei denen professionelles Handeln auf zwischenmenschliche Beziehungen gerichtet ist und die beratend, unterrichtend, menschenführend, pflegend oder helfend tätig sind (Petermann, 1995). Dazu zählen v. a. die folgenden Personengruppen:

- ▶ Schulleiter/Heimleiter
- ▶ Lehrkräfte
- ▶ Schulpsychologen
- ▶ Erziehungsberater
- ▶ Kindergartenpädagogen
- ▶ Heilpädagogen
- ▶ Sozialarbeiter
- ▶ Hochschulmitarbeiter

Die Möglichkeiten zur Nutzung von Supervision hängen für die verschiedenen Adressaten u. a. von den institutionellen Bedingungen und der Einstellung der jeweiligen Arbeitgeber ab. In bestimmten Einrichtungen (z. B. Heime oder Erziehungsberatungsstellen) gibt es feste Pläne für Supervisionen, die regelmäßig stattfinden (z. B. Fallsupervision in vierwöchigem Rhythmus oder Teamsupervision einmal im Halbjahr). In vielen Fällen wird Supervision allerdings erst auf Nachfrage (z. B. in schwierigen Situationen im Team oder mit besonderen Eltern/Kindern) organisiert und nur kurzfristig eingesetzt.

20.3.2 Begriffsbestimmung und theoretischer Hintergrund

Supervision kann als eine personenbezogene Form der Beratung beschrieben werden, bei der berufliche Zusammenhänge besprochen und bearbeitet werden. Rappe-Giesecke (2009, S. 3) definiert Supervision als »personenbezogene berufliche Beratung für Professionals. Ihre Aufgabe ist es, Einzelne, Gruppen oder Teams von Professionals zu individueller und sozialer Selbstreflexion zu befähigen. Ziel dieser Reflexion ist die Über-

prüfung und Optimierung des beruflichen und methodischen Handelns«.

Im Gegensatz zur Intervention hat Supervision nicht das Ziel, Veränderungen explizit herbeizuführen, sondern entsprechende Prozesse anzustoßen, die eventuell zu Veränderungen führen. Von pädagogischer Supervision wird dann gesprochen, wenn diese im pädagogischen Arbeitsfeld, also in schulischen, vorschulischen und hochschulischen, aber auch außerschulischen Bildungseinrichtungen stattfindet (Jugert, 1998).

Historische Entwicklung der Supervision. Die Supervision hat ihre Wurzeln in der amerikanischen Sozialarbeit zu Beginn des 20. Jahrhunderts und später in der Psychotherapie. Zu Beginn des 20. Jahrhunderts waren in amerikanischen Großstädten viele Berufsanfänger und ehrenamtlich arbeitende Personen zur Unterstützung von Einwanderern und anderen hilfsbedürftigen Personen tätig. Zur Koordination und fachlichen Begleitung dieser ehrenamtlich engagierten Mitarbeiter haben die Organisationen daher fest angestellte Supervisoren auf der Ebene eines mittleren Managements eingestellt. Die Supervision wurde zu Beginn gewissermaßen von Vorgesetzten durchgeführt. Sie beschränkte sich zunächst sowohl in der Sozialarbeit als auch in der Psychotherapie auf Fallbesprechungen im Einzelsetting (Jugert, 1998). Erst seit den 1970er-Jahren lassen sich auch verschiedene Formen der Gruppensupervision oder der kollegialen Beratung finden. Im Verlauf der 1980er-Jahre setzte sich dann die »Teamsupervision« durch, bei der Menschen aus kooperierenden Arbeitsgruppen gemeinsam beraten werden (Schreyögg, 2000). Psychologen und Psychotherapeuten wurden u. a. systematisch zu Supervisoren ausgebildet und auch in der Supervision der Sozialarbeit und in der Psychotherapie eingesetzt (Pallasch, 1991).

Verbindung von »clinical supervision« und »educational supervision«. Eine wichtige theoretische Neuorientierung war die Verknüpfung von zwei traditionell getrennt durchgeführten Maßnahmen der beruflichen Weiterbildung bzw. Supervision, nämlich Selbsterfahrung und -reflexion (*clinical supervision*) und Instruktion (*educational supervision*), die bis heute ein zentrales Kennzeichen der Balint-Gruppen darstellt (Jugert, 1998; s. Kasten).

Unter der Lupe

Balint-Gruppen

Der ungarische Arzt und Psychoanalytiker Michael Balint (1896–1970) hat in den 1950er-Jahren als Erster damit begonnen, berufliche Selbstreflexion und Instruktion miteinander zu verbinden. Er arbeitete ursprünglich mit Gruppen von Allgemeinmedizinerinnen, die sich vorher gegenseitig noch nicht kannten und die über ihre alltäglichen Erfahrungen mit Klienten berichten sollten. Für die Arbeit in der Gruppe stellte Balint verbindliche Regeln auf, z. B.:

- ▶ Die beruflichen Kompetenzen der Teilnehmer werden grundsätzlich anerkannt.
- ▶ Es wird darauf geachtet, dass keine Abhängigkeit der Gruppenteilnehmer entsteht.
- ▶ Der Fallbericht soll spontan und frei vorgetragen werden.

Balint lädt dazu ein, den freien Fallbericht wie die Erzählung eines Traumes aufzufassen und die Einfälle, Assoziationen und Gedanken der Teilnehmer dazu wie freie Assoziationen zu diesem Traum in der Gruppe zu bearbeiten. Die Sammlung, Aufnahme und das Ver-

stehen der Assoziationen, Einfälle usw. werden vom Leiter teilweise psychoanalytisch gedeutet und dazu genutzt, die Beziehung zwischen dem Supervisanden (Arzt) und seinem Patienten zu reflektieren und zu bearbeiten. Ein zentrales Ziel der Balint-Gruppe ist die »Beziehungsdiagnostik«. Anstelle einer Handlungsanweisung geht es darum, herauszufinden, was in der gegenwärtigen professionellen Beziehung zum Zeitpunkt des Vortrags als problematisch wahrgenommen wird.

Damit war der Schritt in die Supervision in ihrer heutigen Form vollzogen. Supervision in der heutigen Form hat außer der Integration von Selbsterfahrung und Instruktion auch die Analyse der emotionalen und der institutionellen Komponente beruflicher Interaktion zu ihrer Aufgabe gemacht (Rappe-Giesecke, 1994). Die Balint-Gruppe ist inzwischen sehr verbreitet, und zwar weit über ihr ursprüngliches Feld, die Medizin, hinaus in fast allen Tätigkeitsfeldern, in denen überhaupt Supervision durchgeführt wird.

In Deutschland zog die Supervision erst relativ spät in die soziale Arbeit ein. Ein deutlicher Anstieg von Supervision ist erst ab den 1960er-Jahren zu verzeichnen (Jugert, 1998). Seitdem gibt es berufsbegleitende Supervision in der Bundesrepublik in verschiedenen Aufgabenfeldern. Seit den 1970er-Jahren werden auch Supervisionsausbildungen angeboten. Ende der 1980er-Jahre wurde die Deutsche Gesellschaft für Supervision (DGSV) gegründet. Diese hatte im Jahr 2012 über 3.700 Mitglieder und 40 angeschlossene Weiterbildungsstätten.

20.3.3 Formen der Supervision

Folgende Formen der berufsbegleitenden Supervision können unterschieden werden:

- ▶ Einzelsupervision,
- ▶ Gruppensupervision,
- ▶ kollegiale Supervision,
- ▶ Teamsupervision und
- ▶ Coaching (Leitungssupervision).

Einzelsupervision. In der Einzelsupervision wird ein Supervisand von einem Supervisor unterstützt. Dabei

orientiert sich der Supervisor an bestimmten Theorien oder psychotherapeutischen Schulen. Innerhalb der Fort- und Weiterbildungssupervision kommt die Einzelsupervision v. a. dann zum Einsatz, wenn es um die Kontrolle und Verbesserung der beruflichen Kompetenz geht (z. B. in der Ausbildung zum Therapeuten, Erziehungsberater). Der Nachteil der Einzelsupervision ist, dass sie sehr zeit- und kostenaufwendig ist.

Gruppensupervision. Die Gruppensupervision findet mit Personen aus dem gleichen Berufsfeld, aber nicht zwingend aus der gleichen Einrichtung statt. Sie wird nach Belardi (1994) häufig in Anlehnung an das Konzept der Balint-Gruppenarbeit (s. Abschn. 20.3.2) in Fort- und Weiterbildungsveranstaltungen oder bei der Betreuung von Praktikanten universitärer Studiengänge eingesetzt. Im Gegensatz zur Einzelsupervision erfordern die in der Gruppensupervision auftretenden gruppendynamischen Prozesse aufseiten des Supervisors zusätzliche Qualifikationen. Insgesamt kann in der Gruppensupervision auf ein reiches Repertoire an theoretischen und methodischen Möglichkeiten zurückgegriffen werden. Die Kosten-Nutzen-Relation ist bei der Gruppensupervision vergleichsweise günstig.

Unter der Lupe

Methoden der Gruppensupervision

Zur Verdeutlichung des konkreten Vorgehens in einer Gruppensupervision kann exemplarisch auf zwei Methoden hingewiesen werden, die bei verschiedenen Problemen eingesetzt werden (vgl. Ehinger & Hennig, 1994).

Rollenspiel. Das Rollenspiel ist eine geeignete Methode zur nachträglichen Analyse einer problematischen Situation (z. B. schwierige Interaktion mit Kollegen oder Klienten; Kritik von außen) und zur gemeinsamen Erarbeitung alternativer Handlungs- und Reaktionsmöglichkeiten. Der Vorteil des Rollenspiels liegt darin, dass es einer Realsituation so nahe kommt wie möglich. Außerdem können damit neue Verhaltensweisen und Reaktionen ausprobiert, eingeübt und zu einem späteren Zeitpunkt auf eine reale Situation übertragen werden. Rollenspiele werden entweder einer Zuschauergruppe dargeboten und dann mit der gesamten Gruppe besprochen, oder mit allen Teilnehmern einer Kleingruppe durchgeführt und nachbearbeitet. In beiden Fällen kommt es vor allen Dingen darauf an, dass vorher geklärt wird, welche Frage beantwortet werden soll, ob alle Teilnehmer selber im Spiel konkrete Erfahrungen machen sollen oder ob die Diskussion des Problems in der Großgruppe das zentrale Ziel ist.

Gruppendiskussion. Diese Methode ist besonders geeignet, wenn ein Teilnehmer eine selbst erlebte schwierige Situation schildert und Lösungsmöglichkeiten für dieses Problem erarbeitet werden sollen, die auch für alle anderen Teilnehmer relevant sind. Eine Gruppendiskussion besteht in der Regel aus den folgenden drei Schritten:

- (1) Der Supervisand trägt sein Problem in die Gruppe. Der Supervisor achtet darauf, dass die zu klärende Frage kurz und möglichst präzise formuliert wird.
- (2) Die Gruppenmitglieder stellen Informations- und Verständnisfragen, die zur Erfassung der äußeren und inneren Wirklichkeit des Supervisanden wichtig sind.
- (3) Die Gruppenmitglieder äußern reihum ihre Gedanken zu dem Problem oder der Frage, ihre Wahrnehmungen, Hypothesen, Gefühle und spontane Handlungsimpulse. Der Supervisand hört aufmerksam zu und nimmt die Ideen der Gruppenmitglieder auf, ohne diese zu werten, und prüft, welche Gedanken ihm zur Problemlösung weiterhelfen.

Kollegiale Supervision. Diese besondere Form der Gruppensupervision wird ohne formelle Anleitung durch einen außenstehenden Supervisor realisiert. Sie wird auch als »Intervision« bezeichnet und stammt ursprünglich aus der Psychotherapie. Mittlerweile findet sie auch in der Schule immer mehr Verbreitung, denn es hat sich gezeigt, dass die meisten alltäglichen Probleme in einer Gruppe erfahrener Kollegen, die die erforderlichen Kommunikationsformen beherrschen, in zufriedenstellender Weise bearbeitet werden können. So betrachtet, ist die kollegiale Supervision die effektivste und kostengünstigste Form der Supervision. Allerdings bestehen auch gewisse Gefahren, z. B. dass Probleme nicht gründlich genug bearbeitet werden und deshalb nicht das Qualitätsniveau einer »expertengeleiteten« Gruppensupervision erreicht wird (Fengler, 1986). Mittlerweile gibt es allerdings Trainingsprogramme, die eine gute Anleitung für Gruppen ohne externen Leiter bieten (z. B. Macha, Lödermann & Bauhofer, 2010).

Teamsupervision. Bei der Teamsupervision wird eine Gruppe von Kollegen supervidiert, die in derselben Institution arbeiten und ähnliche Aufgaben zu bewältigen haben (z. B. die Erzieher in einer Kindertagesstätte). Die Teamsupervision sollte im Idealfall eine hohe Kontinuität aufweisen. Teamsupervision ist inzwischen die häufigste Form der Supervision (Belardi, 1994). Sie markiert den Weg der Supervision von der Einzelsituation über die arbeitsplatzferne Gruppensupervision bis hin zur Organisations- und Teamsupervision. Bei der Teamsupervision kommen gemeinsame Arbeitsprobleme zur Sprache, ebenso Hierarchie- und Leitungsfragen oder Beziehungsprobleme zwischen den Teammitgliedern (Jugert, 1998). Grundsätzlich spielt sich die Supervisionsarbeit innerhalb des Teams auf drei Ebenen ab: individuelles Verhalten des jeweils behandelten Einzelfalls, Teaminteraktion und Institution.

Coaching (Leitungsberatung). In pädagogisch-psychologischen Kontexten bezeichnet Coaching eine besondere Form der Supervision von Personen mit Leitungs-

funktionen. Der Begriff stammt ursprünglich aus dem sportlichen Bereich und bezieht sich auf die professionelle Unterstützung einzelner Personen bei der Bewältigung schwieriger Probleme oder der Förderung beruflicher Selbstgestaltungspotenziale. Ein häufiger Anlass für Coaching ist die selbstkritische Einschätzung von Führungskräften, den eigenen Erwartungen an die Leitungsrolle nicht gerecht werden zu können. Die Leitungssupervision soll dem Supervisanden zum einen dabei helfen, die Organisation im Sinne der Erwartungen und Anforderungen des Trägers zu führen, gleichzeitig aber auch den Erwartungen der Mitarbeiter gerecht zu werden, ihre Interessen gegenüber dem Träger

zu vertreten und auch den Erwartungen des Klientels, für die die Organisation ihre Dienstleistungen anbietet, entgegenzukommen. Leitungssupervision kann daher als eine gewinnbringende Form der Einzelberatung für Personen in leitender Funktion angesehen werden. Ein wichtiger Aspekt von Coaching für Personen in Leitungsfunktionen ist zudem, dass diese eine besondere Stellung im sozialen Netzwerk einer Institution einnehmen, die ihnen oft die Offenlegung oder Diskussion persönlicher Probleme in der Gruppe der anderen Führungskräfte oder der Mitarbeiter erschwert. Unter diesen Umständen liegt es für eine Führungskraft nahe, Unterstützung bei einem Coach zu suchen, der entwe-



Abbildung 20.8 Handlungsmodell der pädagogischen Supervision (nach Petermann, 1995)

der gar nicht in die Organisation eingebunden ist oder bei dem aus anderen Gründen damit gerechnet werden kann, dass die thematisierten Sachverhalte streng vertraulich behandelt werden.

Weitere Supervisionsformen. Neben den genannten berufsbegleitenden Supervisionsformen beschreibt Rapp-Giesecke (2009) weitere Formen der Supervision, wie die »Ausbildungssupervision«, bei der Experten ihres Fachs den Erwerb berufsspezifischer Techniken überwachen (z. B. Chirurgie-Ausbildung) oder die »administrative Supervision« in Organisationen, bei der Fachvorgesetzte die Funktion der fachlichen Begleitung sowie der Personalführung und -entwicklung übernehmen.

Phasen der Supervision

Für alle Formen der Supervision gelten im Prinzip die gleichen Ablaufschritte, wenngleich ihre jeweilige Gewichtung von Fall zu Fall erheblich variieren kann. Ein von Petermann (1995) vorgeschlagenes allgemeines Modell der Supervision unterteilt den Supervisionsprozess in sechs Phasen:

- Orientierungsphase
- Informationsphase
- Gewichtungphase
- Rückmeldungsphase
- Umsetzungsphase
- Kontrollphase

In Abbildung 20.8 sind die Ziele und Aufgabenstellungen der einzelnen Phasen näher beschrieben.

20.3.4 Anwendungsfelder im pädagogischen Bereich

Die Anwendungsfelder der Supervision im pädagogischen Bereich sind sehr vielfältig. Sie reichen von der Einzelsupervision für Lehrkräfte, Trainer, Erziehungsberater über Gruppen- und Teamsupervision in erzieherischen und pädagogischen Einrichtungen, Leitungssupervision für Führungskräfte in Bildungseinrichtungen bis hin zur Supervision für ganze Organisationen. Zur pädagogischen Supervision im schulischen Bereich zählt zudem die Supervision für den unterrichtlichen Kontext (Staub, 2004), um z. B. die didaktischen und methodischen Fertigkeiten von Lehrkräften zu unterstützen. Bei allen Anlässen und Themen ist es wichtig, dass Supervision in einem vertraulichen Rahmen stattfindet und die Schweigepflicht eingehalten wird. Nur so kann eine vertrauensvolle Atmosphäre geschaffen werden, in der auch schwierige und persönlich belastende Themen besprochen werden können.

Anlässe und Themen für Supervision

Die Themen, mit denen sich Supervision inhaltlich beschäftigen kann, sind ebenso vielfältig wie die Adres-

Tabelle 20.4 Anlässe und Themen der Supervision in pädagogischen Handlungsfeldern

Probleme und Schwierigkeiten	Supervisionsform
Persönliche Fragestellungen: z. B. Überforderung, Burn-out, private Probleme, berufliche Veränderungswünsche, Verbesserung der beruflichen Kompetenz in speziellen Aspekten (etwa der fachdidaktischen Kompetenz)	überwiegend Einzelsupervision, aber auch Teamsupervision oder Balint-Gruppen
Berufliche Belastungen: z. B. Konflikte mit Vorgesetzten, Kollegen, Klienten, Angehörigen von Klienten; mangelnde Kooperation, Differenzen über pädagogisches Konzept; Sympathie/Antipathie, Einfluss/Macht	überwiegend Einzel- oder Gruppensupervision
Fallarbeit im Rahmen regelmäßiger Fallbesprechungen von Klienten oder zu schwierigen Anlässen mit einzelnen Klienten: z. B. soziale Probleme, Verhaltensauffälligkeiten, Sucht, Lernschwierigkeiten	überwiegend Einzel- oder Gruppensupervision
Konflikte mit Mitarbeitern, Verwaltungssystemen, Finanzgebern, Rollenkonflikte: z. B. Einführung neuer Regelungen, Kürzungen von Geldern, Schwierigkeiten unter Mitarbeitern	überwiegend Leitungssupervision
Veränderungen in der Organisation: z. B. Schulentwicklung, Qualitätsmanagement	überwiegend Supervision zur Organisationsentwicklung

saten für Supervision. Tabelle 20.4 gibt eine Übersicht über einige Anlässe von Supervision in pädagogischen Handlungsfeldern.

Anwendungsbeispiele für Supervision in der Schule

Als Beispiele für angewandte Supervision werden zwei Konzepte aus dem schulischen Bereich vorgestellt. Das erste Konzept zielt auf die Verbesserung der Unterrichtskompetenz von Lehrkräften, das zweite auf die Prävention von Burn-out. Beide Konzepte wurden durch wissenschaftliche Begleituntersuchungen evaluiert und machen auf exemplarische Weise deutlich, dass Supervision im schulischen Kontext positive Effekte erzielen kann.

Fachspezifisch-Pädagogisches Coaching zur Verbesserung der Unterrichtskompetenz. Das Fachspezifisch-Pädagogische Coaching (*Content-Focused Coaching*; vgl. Staub, 2004) ist ein Ansatz zur Fort- und Weiterbildung von Lehrpersonen zur Unterstützung ihres Unterrichts. Den Lehrkräften wird dabei im entsprechenden Fach ein kompetenter Coach/Mentor zur Seite gestellt. Dieser übernimmt Mitverantwortung für die Gestaltung der Unterrichtseinheit. Das unmittelbare Ziel besteht in der gemeinsamen Gestaltung eines für die Schüler möglichst optimalen Unterrichts. Das langfristige Ziel ist die nachhaltige Entwicklung von Unterrichtskompetenz. Es entsteht dabei eine Interaktion zwischen Wissenschaft und Praxis als Kooperation von theoriebasierten Settings und Werkzeugen zur Entwicklung innovativer Praxis. Coaching verweist hier auf die individualisierte und situationsbezogene Unterstützung eines Lehrers bei der Bearbeitung einer komplexen Aufgabenstellung durch eine Person, die in der Bewältigung solcher Aufgaben selbst über eine hohe Expertise verfügt. Das Coaching findet innerhalb des aktuellen beruflichen Settings als Teil der normalen Arbeit statt.

Das Konzept des Fachspezifisch-Pädagogischen Coachings zielt darauf ab, die Lehrperson in ihrer Unterrichtsplanung, -durchführung und -reflexion zu unterstützen. Dies kann sowohl in dyadischen Settings als auch in verschiedenen Gruppen von Lehrpersonen umgesetzt werden. Der Ablauf lässt sich in drei Phasen einteilen:

- (1) Vor der Unterrichtseinheit besprechen der Coach und die Lehrkraft (Dyade) die Gestaltung des Unterrichts; diese Unterrichtsvorbereitung dient der Verständigung über Unterrichtsziele und den Lehrplan. Der Coach beteiligt sich hierbei als mit-

verantwortlicher Partner. Ein guter Unterricht setzt voraus, dass die Lehrkraft gut auf den Unterricht vorbereitet ist, dass sie sich in der Materie auskennt und die Bearbeitung der Inhalte auch aus der Sicht der Lernenden durchdacht hat. Nur so wird ein an die Voraussetzungen der Schüler angepasster Unterricht möglich.

- (2) Der Coach ist während der Umsetzung ein aktives Mitglied im Unterrichtsgeschehen; häufig unterstützt er mit der Lehrkraft gemeinsam oder beteiligt sich anderweitig am Unterricht (z.B. in Form von Rückmeldungen oder nützlichen Tipps).
- (3) Im Anschluss wird der Unterricht gemeinsam reflektiert. Bei der Nachbesprechung ist der Fokus wiederum auf den inhaltspezifischen Lernprozess der Schüler gerichtet.

Für einen Coach ist es nicht ganz einfach, zwischen aktivem Zuhören und dem Einbringen von Gestaltungsvorschlägen ein angemessenes Gleichgewicht zu finden. Außerdem muss er gleichzeitig zwei Hauptziele im Auge behalten: Einerseits geht es um die optimale Förderung der aktuell unterrichteten Schüler, andererseits soll die Unterrichtskompetenz der Lehrperson weiterentwickelt werden. Zur Ausbildung der Coaches werden beim Fachspezifisch-Pädagogischen Coaching in der Regel auch Videoaufzeichnungen von Unterricht sowie dazugehörige Vor- und Nachbesprechungen verwendet (Staub, 2004).

Coaching für Lehrkräfte als Burn-out-Prävention. Das Projekt von Unterbrink et al. (2010) wurde durch das Bundesministerium für Arbeitssicherheit und Gesundheit in Auftrag gegeben und hatte das Ziel, die Gesundheitsvorsorge von Lehrkräften zu verbessern, indem ihre sozialen und emotionalen Kompetenzen gestärkt werden. Das Lehrer-Coaching nach dem sog. »Freiburger Modell« umfasst fünf Module:

- (1) Informationen zu Auswirkungen von Beziehungserfahrungen auf die Gesundheit; Entspannungsübungen
- (2) Persönliche Einstellungen: Identität und Identifikation
- (3) Beziehungsgestaltung zu Schülern
- (4) Beziehungsgestaltung mit Eltern
- (5) Spaltungstendenzen vs. kollegialer Zusammenhalt

An der Studie nahmen Lehrkräfte aus zwei verschiedenen Schulformen (Hauptschule und Gymnasium) teil. Die Supervisionen fanden über den Zeitraum von einem Jahr statt und beinhalteten zehn Sitzungen à 90 Minu-

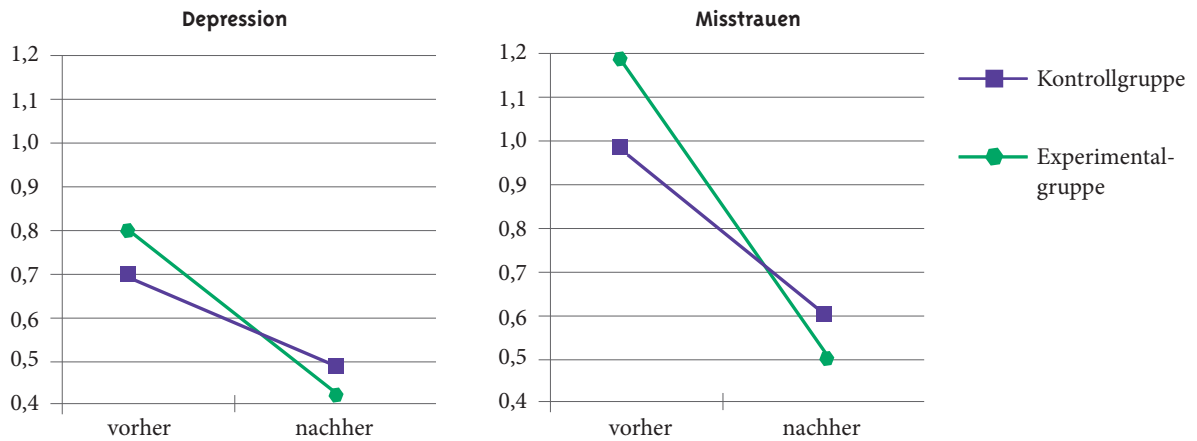


Abbildung 20.9 Veränderung der Gesundheitseinschätzung bezüglich Depression und Misstrauen in der Coaching-Studie von Unterbrink et al. (2010)

ten. Es zeigten sich bei den Teilnehmern der Coaching-Gruppe (Experimentalgruppe) im Vergleich zu den Teilnehmern, die kein Coaching erhalten hatten (Kontrollgruppe) positive Veränderungen in der subjektiven

Gesundheitseinschätzung bezüglich der beiden Kriterien Depressivität und Misstrauen. Abbildung 20.9 zeigt die Ergebnisse für die Gesundheitseinschätzung für die beiden Gruppen vor und nach dem Coaching.

Zusammenfassung

- ▶ Die gemeinsamen Ziele von Beratung, Intervention und Supervision in pädagogischen Handlungsfeldern bestehen in der Unterstützung bei persönlichen, beruflichen und erzieherischen Problemen sowie der Optimierung von Entwicklungs- und Bildungsprozessen.
- ▶ In der pädagogischen Beratung werden unterschiedliche Beratungsansätze eingesetzt, z. B. klientenzentrierte, verhaltensbezogene, systemische oder lösungsorientierte Beratung. Zu den pädagogischen Anwendungsfeldern gehören die Erziehungs- und Familienberatung, die Beratung in Schule und Hochschule sowie in der Aus- und Weiterbildung.
- ▶ Mit dem Ziel der Erarbeitung von Lösungen für spezifische Problemfälle wird professionelle Beratung stets planvoll und von einem fachkundigen und methodisch geschulten Berater durchgeführt, der über eine Vielzahl an Kompetenzen (z. B. soziale, diagnostische, Prozess- und Gesprächsführungskompetenz sowie Fachwissen) verfügen sollte.
- ▶ Als pädagogische Interventionen werden Maßnahmen bezeichnet, die durch ein relativ stark von außen gesteuertes zielgerichtetes und theorie-

geleitetes Eingreifen in Entwicklungs- und Bildungsprozesse auf Individual- oder Systemebene gekennzeichnet sind und die eine Optimierung dieser Prozesse anstreben.

- ▶ Die Untersuchung der Wirksamkeit von Interventionen ist Gegenstand der (quasi-)experimentellen Interventionsforschung. Sie dient u. a. der Überprüfung kausaler Hypothesen bezüglich der im Rahmen der Intervention eingesetzten Maßnahmen.
- ▶ Supervision stellt eine personenbezogene Form der Beratung dar, bei der berufliche Zusammenhänge besprochen und bearbeitet werden, mit dem Ziel einer Reflexion und Optimierung des professionellen Handelns.
- ▶ Pädagogische Supervision richtet sich vorwiegend an Personengruppen, die in Bildungseinrichtungen beratend, unterrichtend, menschenführend oder helfend tätig sind.
- ▶ Die verschiedenen Formen der Supervision reichen von Einzelsupervision und Coaching über Gruppensupervision und kollegiale Beratung bis hin zu Teamsupervision und Balint-Gruppen.