

2 Von einer interessanten Fragestellung zur empirischen Untersuchung

2.1 Themensuche – 36

- 2.1.1 Anlegen einer Ideensammlung – 37
- 2.1.2 Replikation von Untersuchungen – 37
- 2.1.3 Mitarbeit an Forschungsprojekten – 38
- 2.1.4 Weitere Anregungen – 38

2.2 Bewertung von Untersuchungsideen – 40

- 2.2.1 Wissenschaftliche Kriterien – 40
- 2.2.2 Ethische Kriterien – 41
- 2.2.3 Informationspflicht – 44

2.3 Untersuchungsplanung – 46

- 2.3.1 Zum Anspruch der geplanten Untersuchung – 46
- 2.3.2 Literaturstudium – 47
- 2.3.3 Wahl der Untersuchungsart – 49
- 2.3.4 Thema der Untersuchung – 59
- 2.3.5 Begriffsdefinitionen und Operationalisierung – 60
- 2.3.6 Messtheoretische Probleme – 65
- 2.3.7 Auswahl der Untersuchungsobjekte – 70
- 2.3.8 Durchführung, Auswertung und Planungsbericht – 75

2.4 Theoretischer Teil der Arbeit – 81

2.5 Durchführung der Untersuchung – 81

- 2.5.1 Versuchsleiterartefakte – 82
- 2.5.2 Praktische Konsequenzen – 83
- 2.5.3 Empfehlungen – 83

2.6 Auswertung der Daten – 85

2.7 Anfertigung des Untersuchungsberichtes – 86

- 2.7.1 Gliederung und Inhaltsverzeichnis – 86
- 2.7.2 Die Hauptbereiche des Textes – 87
- 2.7.3 Gestaltung des Manuskripts – 90
- 2.7.4 Literaturhinweise und Literaturverzeichnis – 90
- 2.7.5 Veröffentlichungen – 93

➤ ➤ Das Wichtigste im Überblick

- Wahl eines geeigneten Themas für eine empirische (Qualifikations-)Arbeit
- Die wichtigsten Varianten empirischer Untersuchungen
- Maßnahmen zur Sicherung interner und externer Validität
- Probleme der Operationalisierung und des Messens
- Auswahl und Anwerbung von Untersuchungsteilnehmern
- Durchführung der Untersuchung und Auswertung der Ergebnisse
- Richtlinien zur Anfertigung des Untersuchungsberichtes über eine empirische Arbeit

Im Folgenden wird in einem ersten Überblick dargestellt, was bei der Planung und Durchführung einer empirischen Untersuchung vorrangig zu beachten und zu entscheiden ist. Die Ausführungen wenden sich in erster Linie an Studierende, die beabsichtigen, eine empirische Qualifikationsarbeit o. Ä. anzufertigen, wobei der Schwerpunkt in diesem Kapitel auf der Planung einer hypothesenprüfenden explanativen Untersuchung liegt. Ausführlichere Informationen über hypothesenerkundende explorative Untersuchungen findet man in ▶ Kap. 5 und 6. (Die wichtigsten Schritte bei der Durchführung empirisch-psychologischer Untersuchungen werden auch bei Hager et al., 2001, dargestellt; zur Anfertigung einer Abschlussarbeit vgl. auch Engel & Slapnicar, 2000, bzw. Beller, 2004.)

Die ▶ Abschn. 2.1 und 2.2 behandeln – gewissermaßen im Vorfeld der eigentlichen Untersuchungsplanung – die Frage, wie sich die Suche nach einem geeigneten Forschungsthema systematisieren lässt und anhand welcher Kriterien entschieden werden kann, welche Themen für eine empirische Untersuchung geeignet sind. In ▶ Abschn. 2.3 (Untersuchungsplanung) stehen Entscheidungshilfen für die Wahl einer dem Untersuchungsthema angemessenen Untersuchungsart, Fragen der Operationalisierung, messtheoretische Probleme, Überlegungen zur Stichprobentechnik sowie die Planung der statistischen Auswertung im Vordergrund. Hinweise zur Anfertigung eines Theorieteils (▶ Ab-

schn. 2.4), zur Durchführung der Untersuchung (▶ Abschn. 2.5), zur Auswertung und Interpretation der Untersuchungsergebnisse (▶ Abschn. 2.6) sowie zur Anfertigung des Untersuchungsberichtes (▶ Abschn. 2.7) beenden dieses Kapitel.

2.1 Themensuche

Die Qualität einer empirischen Untersuchung wird u. a. daran gemessen, ob die Untersuchung dazu beitragen kann, den Bestand an gesichertem Wissen im jeweiligen Untersuchungsbereich zu erweitern. Angesichts einer beinahe explosionsartigen Entwicklung der Anzahl wissenschaftlicher Publikationen befinden sich Studierende, die z. B. die Absicht haben, eine empirische Abschlussarbeit anzufertigen, in einer schwierigen Situation: Wie sollen sie herausfinden, ob eine interessant erscheinende Untersuchungsidee tatsächlich originell ist? Wie können sie sicher sein, dass das gleiche Thema nicht schon bearbeitet wurde? Verspricht die Untersuchung tatsächlich neue Erkenntnisse, oder muss man damit rechnen, dass die erhofften Ergebnisse eigentlich trivial sind?

Eine Beantwortung dieser Fragen bereitet weniger Probleme, wenn im Verlaufe des Studiums – durch Gespräche mit Lehrenden und Mitstudierenden bzw. nach gezielter Seminararbeit und Lektüre – eine eigenständige Forschungsidee heranreife. Manche Studierende vertiefen sich jedoch monatelang in die Fachliteratur in der Hoffnung, irgendwann auf eine brauchbare Untersuchungsidee zu stoßen. Am Ende steht nicht selten ein resignativer Kompromiss, auf dem mehr oder weniger desinteressiert die eigene empirische Untersuchung aufgebaut wird.

McGuire (1967) führt die Schwierigkeit, kreative Untersuchungsideen zu finden, zu einem großen Teil auf die Art der Ausbildung in den Sozialwissenschaften zurück. Er schätzt, dass mindestens 90% des Unterrichts in Forschungsmethodik auf die Vermittlung präziser Techniken zur Überprüfung von Hypothesen entfallen und dass für die Erarbeitung von Strategien, schöpferische Forschungsideen zu finden, überhaupt keine oder nur sehr wenig Zeit aufgewendet wird.

In der Tat fällt es schwer einzusehen, warum der hypothesenüberprüfende Teil empirischer Untersu-

chungen so detailliert und sorgfältig erlernt werden muss, wenn gleichzeitig der hypothesenkreierende Teil sträflich vernachlässigt wird, sodass – was nicht selten der Fall ist – mit einem perfekten Instrumentarium letztlich nur Banalitäten überprüft werden.

Empirische Arbeiten sind meistens zeitaufwendig und arbeitsintensiv. Es ist deshalb von großem Vorteil, wenn es Studierenden gelingt, eine Fragestellung zu entwickeln, deren Bearbeitung sie persönlich interessiert und motiviert. Das eigene Engagement hilft nicht nur, einen frühzeitigen Abbruch der Arbeit zu verhindern, sondern kann auch zu einem guten Gelingen der empirischen Untersuchung beitragen.

Diese Einschätzung rechtfertigt natürlich die Frage, ob die Forderung nach persönlichem Engagement in der Forschung nicht die Gefahr in sich birgt, dass die Wissenschaft Ergebnisse produziert, die durch Vorurteile und Voreingenommenheit der Untersuchenden verzerrt sind. Diese Möglichkeit ist sicherlich nicht auszuschließen.

Shields (1975) behauptet, die Geschichte der Wissenschaften sei voller Belege dafür, wie Wissenschaftler durch bestechende Argumente und phantasiereiche Interpretationen ihre Vorurteile zu bestätigen trachten. Hieraus nun die Forderung nach einer »wertfreien«, von »neutralen« Personen getragenen Wissenschaft ableiten zu wollen, wäre sicherlich illusionär und wohl auch falsch. Kreative und bahnbrechende Forschung kann nur geleistet werden, wenn Forschenden das Recht zugestanden wird, sich engagiert für die empirische Bestätigung ihrer Vorstellungen und Ideen einzusetzen.

Dies bedeutet natürlich nicht, dass empirische Ergebnisse bewusst verfälscht oder widersprüchliche Resultate der wissenschaftlichen Öffentlichkeit vorenthalten werden dürfen. Gerade in der empirischen Forschung ist die präzise Dokumentation der eigenen Vorgehensweise und der Ergebnisse eine unverzichtbare Forderung, die es anderen Forschern ermöglicht, die Untersuchung genau nachzuvollziehen und ggf. zu replizieren. Nur so kann sich Wissenschaft vor vorsätzlicher Täuschung schützen.

Nach diesen Vorbemerkungen geben wir im Folgenden einige Ratschläge, die die Suche nach einem geeigneten Thema erleichtern sollen, denn wie oben angedeutet bereitet die Suche nach dem Thema zuweilen

mehr Schwierigkeiten als dessen Bearbeitung (zur Generierung kreativer Forschungshypothesen vgl. auch McGuire, 1997).

2.1.1 Anlegen einer Ideensammlung

Um spontan interessant erscheinende Einfälle nicht in Vergessenheit geraten zu lassen, ist es empfehlenswert, bereits frühzeitig mit einer breit gefächerten Sammlung von Untersuchungsideen zu beginnen. Diese Untersuchungsideen können durch Lehrveranstaltungen, Literatur, Teilnahme an Untersuchungen als »Versuchsperson«, Gespräche, eigene Beobachtungen o. Ä. angeregt sein. Wird zusätzlich das Datum vermerkt, stellt diese Sammlung ein interessantes Dokument der eigenen »Ideengeschichte« dar, der beispielsweise entnommen werden kann, wie sich die Interessen im Verlaufe des Studiums verlagert haben. Das Notieren der Quelle erleichtert im Falle eines eventuellen späteren Aufgreifens der Idee weiterführende Literaturrecherchen oder Eingrenzungen der vorläufigen Untersuchungsproblematik.

Gewöhnlich werden sich einige dieser vorläufigen, spontanen Untersuchungsideen als uninteressant oder unbrauchbar erweisen, weil sich die eigenen Interessen inzwischen verlagert haben, weil in der Literatur die Thematik bereits erschöpfend behandelt wurde oder weil das Studium Einsichten vermittelte, nach denen bestimmte Themen für eine empirische Untersuchung ungeeignet erscheinen (► Abschn. 2.2). Dennoch stellt diese Gedächtnisstütze für »Interessantes« ein wichtiges Instrument dar, ein Thema zu finden, das mit hoher Motivation bearbeitet werden kann; gleichzeitig trägt es als Abbild der durch die individuelle Sozialisation geprägten Interessen dazu bei, die Vielfalt von Untersuchungsideen und Forschungshypothesen einer Wissenschaft zumindest potenziell zu erweitern.

2.1.2 Replikation von Untersuchungen

Verglichen mit der empirischen Überprüfung eigener Ideen scheint die Rekonstruktion oder Wiederholung einer bereits durchgeführten Untersuchung eine weniger attraktive Alternative darzustellen. Dennoch sind

Replikationen von Untersuchungen unerlässlich, wenn es um die Festigung und Erweiterung des Kenntnisbestandes einer Wissenschaft geht (vgl. Amir & Sharon, 1991).

Replikationen sind vor allem erforderlich, wenn eine Untersuchung zu unerwarteten, mit dem derzeitigen Kenntnisstand nur wenig in Einklang zu bringenden Ergebnissen geführt hat, die jedoch eine stärkere Aussagekraft hätten, wenn sie sich bestätigen ließen.

Völlig exakte Replikationen von Untersuchungen sind schon wegen der veränderten zeitlichen Umstände undenkbar. In der Regel werden Untersuchungen zudem mit anderen Untersuchungsobjekten (z. B. Personen), anderen Untersuchungsleitern oder sonstigen geringfügigen Modifikationen gegenüber der Originaluntersuchung wiederholt (vgl. Neuliep, 1991; Schweizer, 1989). So gesehen können auch Replikationen durchaus originell und spannend sein (zur Problematik von Replikationen vgl. MacKay, 1993). Viele publizierte Studien enthalten zudem im Diskussionsteil Anregungen für Anschlussstudien.

2.1.3 Mitarbeit an Forschungsprojekten

Erheblich erleichtert wird die Themensuche, wenn Studierenden die Gelegenheit geboten wird, an größeren Forschungsprojekten ihres Institutes oder anderer Institutionen mitzuwirken. Hier ergeben sich gelegentlich Teilfragestellungen für eigenständige Qualifikationsarbeiten. Durch diese Mitarbeit erhalten Studierende Einblick in einen komplexeren Forschungsbereich; einschlägige Literatur wurde zumindest teilweise bereits recherchiert, und zu den Vorteilen der Teamarbeit ergeben sich u. U. weitere Vergünstigungen wie finanzielle Unterstützung und Förderung bei der Anfertigung von Publikationen.

An manchen Instituten ist es üblich, dass Untersuchungsthemen aus Forschungsprogrammen der Institutsmitglieder den Studierenden zur Bearbeitung vorgegeben werden. Diese Vergabep Praxis hat den Vorteil, dass die Themensuche erspart bleibt; sie hat jedoch auch den Nachteil, dass eigene Forschungsinteressen zu kurz kommen können.

2.1.4 Weitere Anregungen

Wenn keine der bisher genannten Möglichkeiten, ein vorläufiges Arbeitsthema »en passant« zu finden, genutzt werden konnte, bleibt letztlich nur die Alternative der gezielten Themensuche. Hierfür ist das Durcharbeiten von mehr oder weniger beliebiger Literatur nicht immer erfolgreich und zudem sehr zeitaufwendig. Vorrangig sollte zunächst die Auswahl eines Themenbereiches sein, der gezielt nach offenen Fragestellungen, interessanten Hypothesen oder Widersprüchlichkeiten durchsucht wird. Die folgenden, durch einfache Beispiele veranschaulichten »kreativen Suchstrategien« können hierbei hilfreich sein. (Ausführlichere Hinweise findet man bei Taylor & Barron, 1964, bzw. Golovin, 1964.)

Intensive Fallstudien. Viele berühmte Forschungsarbeiten gehen auf die gründliche Beobachtung einzelner Personen zurück (z. B. Kinderbeobachtungen bei Piaget, der Fall Dora oder der Wolfsmensch bei Freud). Die beobachteten Fälle müssen keineswegs auffällig oder herausragend sein; häufig sind es ganz »normale« Personen, deren Verhalten zu Untersuchungsideen anregen kann (einen Überblick über methodische Varianten zur Untersuchung berühmter Individuen gibt Simonton, 1999).

Introspektion. Eine beinahe unerschöpfliche Quelle für Untersuchungsideen stellt die Selbstbeobachtung (Introspektion) dar. Wenn man bereit ist, sich selbst kritisch zu beobachten, wird man gelegentlich Ungereimtheiten und Widersprüchliches entdecken, was zu interessanten Fragestellungen Anlass geben kann: Warum reagiere ich in bestimmten Bereichen (z. B. in Bezug auf meine Autofahrleistungen) überempfindlich auf Kritik, obwohl es mir im Allgemeinen wenig ausmacht, kritisiert zu werden. Gibt es Belege dafür, dass auch andere Menschen »sensible Bereiche« haben?

Sprichwörter. Im Allgemeinen werden Sprichwörter als inhaltsarme Floskeln abgetan. Dennoch verbergen sich hinter manchen Sprichwörtern die Erfahrungen vieler Generationen; sie können auch für die Gegenwart noch ein »Körnchen Wahrheit« enthalten. – »Besser den Spatz in der Hand, als die Taube auf dem Dach!« In diesem

Sprichwort steckt eine Handlungsregel, bei Wahlentscheidungen eher risikolose Entscheidungen mit geringem Gewinn als risikoreiche Entscheidungen mit hohem Gewinn zu treffen. Wie groß müssen in einer gegebenen Situation die Gewinnunterschiede sein, damit diese Regel nicht mehr befolgt wird? Gibt es Personen, die sich grundsätzlich anders verhalten, als es das Sprichwort rät? (Weitere Hinweise und Anregungen zu dieser Thematik findet man bei Hartmann & Wirrer, 2002, oder auch Preußner, 2003.)

Funktionale Analogien. Interessante Denkanstöße vermitteln gelegentlich die Übertragung bzw. analoge Anwendung bekannter Prinzipien oder Mechanismen (bzw. experimentelle Paradigmen) auf neuartige Probleme. Erschwert wird diese Übertragung durch »funktionale Fixierungen« (Duncker, 1935), nach denen sich Objekte oder Vorgänge nur schwer aus ihrem jeweiligen funktionalen Kontext lösen lassen. Gelingt die Lösung, kann dies zu so interessanten Einfällen wie z. B. die Inokulationstheorie (Impfungstheorie) von McGuire (1964) führen, nach der die Beeinflussbarkeit der Meinungen von Personen in verbalen Kommunikationssituationen (persuasive Kommunikation) z. B. durch Vorwarnungen darüber, dass eine Beeinflussung stattfinden könnte, reduziert wird. Es handelt sich hierbei um eine analoge Anwendung der Impfwirkung: Durch die rechtzeitige Impfung einer schwachen Dosis desjenigen Stoffes, der potenziell eine gefährliche Infektion hervorrufen kann, werden Widerstandskräfte mobilisiert, die den Körper gegenüber einer ernsthaften Infektion immunisieren.

Paradoxe Phänomene. Wer aufmerksam das alltägliche Leben beobachtet, wird gelegentlich Wahrnehmungen machen, die unerklärlich bzw. widersinnig erscheinen. Die probeweise Überprüfung verschiedener Erklärungsmöglichkeiten derartiger paradoxer Phänomene stellt – soweit Antworten noch nicht vorliegen – eine interessante Basis für empirische Untersuchungen dar: Warum verursachen schwere Verwundungen in starken Erregungszuständen keine Schmerzen? Warum kann man sich gelegentlich des Zwanges, trotz tiefer Trauer lachen zu müssen, nicht erwehren? Wie ist es zu erklären, dass manche Menschen bei totaler Ermüdung nicht einschlafen können?

Analyse von Faustregeln. Jahrelange Erfahrungen führen zur Etablierung von Faustregeln, die das Verhalten des Menschen sowie seine Entscheidungen mehr oder weniger nachhaltig beeinflussen. Die Analyse solcher Faustregeln vermittelt gelegentlich Einsichten, die eine bessere Nutzung der in einer Faustregel enthaltenen Erfahrungen ermöglicht: Warum ist eine Ehe in ihrem siebenten Jahr besonders gefährdet? Warum sollte »der Schuster bei seinen Leisten bleiben«? Stimmt es, dass sich »Gleich und Gleich gern gesellt«, obwohl »Gegensätze sich anziehen«?

Veränderungen von Alltagsgewohnheiten. Vieles im alltäglichen Leben unterliegt einer gesellschaftlichen Normierung, der wir uns in der Regel nicht ständig bewusst sind. Erst wenn Veränderungen eintreten, nehmen wir unsere eigene Einbindung wahr. Aus Fragen nach den Ursachen der Veränderung von Alltagsgewohnheiten (Übernahme neuer Moden, veränderte Freizeitgewohnheiten, Veränderungen gesellschaftlicher Umgangsformen etc.) lassen sich eine Fülle interessanter Ideen für sozialpsychologische Untersuchungen ableiten.

Gesellschaftliche Probleme. Wer aufmerksam Politik und Zeitgeschehen verfolgt, wird feststellen, dass in der öffentlichen Diskussion brisanter Ereignisse, wie Naturkatastrophen, Unfälle, Verbrechen, Skandale usw., oftmals ein Mangel an Forschungsergebnissen beklagt wird und man deswegen auf Mutmaßungen angewiesen bleibt. Wird das umstrittene neue Fernsehformat tatsächlich aus »purem Voyeurismus« angeschaut, oder spielen für das Publikum vielleicht Aspekte eine Rolle, mit denen weder die Macher noch die Kritiker der Sendung gerechnet haben? Gerade eine in Eigenregie durchgeführte Qualifikationsarbeit ist ideal geeignet, um aktuelle Fragestellungen rasch aufzugreifen, während größere Forschungsprojekte in der Regel einen 2- bis 3-jährigen Vorlauf für Antragstellung und Bewilligung benötigen (► Anhang E).

Widersprüchliche Theorien. Stößt man auf Theorien, die einander widersprechen (oder einander zu widersprechen scheinen), kann dies zum Anlass genommen werden, eigenständige Prüfmöglichkeiten der widersprüchlichen Theorien bzw. einen allgemeineren theoretischen Ansatz zu entwickeln, der den Widerspruch

aufhebt. Die Brauchbarkeit dieser allgemeineren Theorie muss durch neue empirische Untersuchungen belegt werden. So wurde beispielsweise Anderson (1967) durch die Widersprüchlichkeit des Durchschnittsmodells (Thurstone, 1931: Der Gesamteindruck von einem Menschen entspricht dem Durchschnitt seiner Teilattribute) und des additiven Modells (Fishbein & Hunter, 1964: Der Gesamteindruck ergibt sich aus der Summe der Teilattribute) zu seinem gewichteten Durchschnittsmodell angeregt: Die einzelnen Attribute fließen mit unterschiedlichem Gewicht in eine Durchschnittsbeurteilung ein.

2.2 Bewertung von Untersuchungsideen

Liegt eine Ideensammlung vor, muss entschieden werden, welche Themen für eine empirische Untersuchung in die engere Wahl kommen. Hiermit ist ein Bewertungsproblem angesprochen, das sich nicht nur dem einzelnen Studenten stellt, sondern das auch Gegenstand zentraler, für die gesamte Fachdisziplin bedeutsamer wissenschaftstheoretischer Diskussionen ist (vgl. Ellsworth, 1977; Herrmann, 1976; Holzkamp, 1964; Popper, 1989). Die Argumente dieser Autoren werden hier nur berücksichtigt, wenn sie konkrete Hilfen für die Auswahl eines geeigneten Themas liefern.

Die Einschätzung der Qualität von Untersuchungsideen ist in dieser Phase davon abhängig zu machen, ob die Untersuchungsideen einigen allgemeinen wissenschaftlichen oder untersuchungstechnischen Kriterien genügen und ob sie unter ethischen Gesichtspunkten empirisch umsetzbar sind.

2.2.1 Wissenschaftliche Kriterien

Präzision der Problemformulierung

Vorläufige Untersuchungsideen sind unbrauchbar, wenn unklar bleibt, was der eigentliche Gegenstand der Untersuchung sein soll, bzw. wenn der Gegenstand, auf den sich die Untersuchung bezieht, so vielschichtig ist, dass sich aus ihm viele unterschiedliche Fragestellungen ableiten lassen.

In diesem Sinne wäre beispielsweise das Vorhaben, »über Leistungsmotivation« arbeiten zu wollen, kritik-

würdig. Die Untersuchungsidee ist zu vage, um eine sinnvolle Literaturrecherche nach noch offenen Problemfeldern bzw. nach replikationswürdigen Teilbefunden anleiten zu können. Das Interesse an diesem allgemeinen Thema sollte sich auf eine Teilfrage aus diesem Gebiet wie z. B. die Genese von Leistungsmotivation oder Folgeerscheinungen bei nicht befriedigter Leistungsmotivation (z. B. bei Arbeitslosen) richten.

Unbrauchbar sind vorläufige Untersuchungsideen auch dann, wenn sie unklare, mehrdeutige oder einfach schlecht definierte Begriffe enthalten. Möchte man sich beispielsweise mit der »Bedeutung der Intelligenz für die individuelle Selbstverwirklichung« beschäftigen, wäre von diesem Vorhaben abzuraten, wenn unklar ist, was mit »Selbstverwirklichung« oder »Intelligenz« gemeint ist.

Die Überprüfung der begrifflichen Klarheit und der Präzision der Ideenformulierung kann in dieser Phase durchaus noch auf einem vorläufigen Niveau erfolgen. Die Begriffe gelten vorläufig als genügend klar definiert, wenn sie kommunikationsfähig sind, nach der Regel: Ein Gesprächspartner, der meint, mich verstanden zu haben, muss in der Lage sein, einem Dritten zu erklären, was ich mit meinem Begriff meine. Strengere Maßstäbe an die begriffliche Klarheit werden erst in ► Abschn. 2.3.5 gelegt, wenn es darum geht, das mit den Begriffen Gemeinte empirisch zu erfassen. Genauer beschäftigt sich zudem Westermann (2000, S. 66 ff.) mit diesem Thema.

Empirische Untersuchbarkeit

Es mag selbstverständlich erscheinen, dass eine Themensammlung für empirische Untersuchungen nur solche Themen enthält, die auch empirisch untersuchbar sind. Dennoch wird man feststellen, dass sich die einzelnen Themen in ihrer empirischen Untersuchbarkeit unterscheiden und dass einige ggf. überhaupt nicht oder nur äußerst schwer empirisch zu bearbeiten sind.

In diesem Sinne ungeeignet sind Untersuchungsideen mit religiösen, metaphysischen oder philosophischen Inhalten (z. B. Leben nach dem Tode, Existenz Gottes, Sinn des Lebens) sowie Themen, die sich mit unklaren Begriffen befassen (z. B. Seele, Gemüt, Charakterstärke), sofern keine besondere Strategie zur Präzisierung dieser Ideen verfolgt wird (so lässt sich die Frage nach dem Sinn des Lebens beispielsweise empirisch untersuchen, wenn man sie darauf zuspitzt, welche Vorstellungen über den Sinn des Lebens Personen in unter-

schiedlichen Bevölkerungsgruppen, Lebensaltern oder Kulturen haben). Ferner ist von Untersuchungsideen abzuraten, die bereits in dieser frühen Phase erkennen lassen, dass sie einen unangemessenen Arbeitsaufwand erfordern. Hierzu zählen die Untersuchung ungewöhnlicher Personen (z. B. psychische Probleme bei Kleinkindern) oder ungewöhnlicher Situationen (z. B. Ursachen für Panikreaktionen bei Massenveranstaltungen) bzw. sehr zeitaufwendige Untersuchungen (z. B. eine Längsschnittuntersuchung zur Analyse der Entwicklung des logischen Denkens bei Kindern).

Wissenschaftliche Tragweite

Unbrauchbar sind Themen, die weder eine praktische Bedeutung erkennen lassen noch die Grundlagenforschung bereichern können. Hochschulen und Universitäten sind Einrichtungen, die eine vergleichsweise lange und kostspielige Ausbildung vermitteln. Hieraus leitet sich eine besondere Verantwortung der Hochschulangehörigen ab, sich mit Themen zu beschäftigen, deren Nutzen zumindest prinzipiell erkennbar ist (zum Verhältnis praxisbezogener Forschung und grundlagenorientierter Forschung vgl. Schorr, 1994, oder Wottawa, 1994).

Problematisch, aber notwendig ist die Entscheidung darüber, ob eine Fragestellung bereits so intensiv erforscht wurde, dass die eigene Untersuchung letztlich nur seit langem bekannte Ergebnisse bestätigen würde (z. B. Untersuchungen, mit denen erneut gezeigt werden soll, dass Gruppen unter der Anleitung eines kompetenten Koordinators effizienter arbeiten, dass Bestrafungen weniger lernfördernd sind als Belohnungen, dass sich Reaktionszeiten unter Alkohol verändern oder dass Unterschichtkinder sozial benachteiligt sind). Diese Entscheidung setzt voraus, dass man sich im Verlaufe seines Studiums genügend Wissen angeeignet oder gezielt Literatur aufgearbeitet hat. Zur Frage der Trivialität von Forschungsergebnissen bzw. zu deren Prognostizierbarkeit aufgrund von »Alltagstheorien« findet man bei Holz-Ebeling (1989) sowie Semmer und Tschan (1991) interessante Informationen.

2.2.2 Ethische Kriterien

Empirische Forschung über humanwissenschaftliche Themen setzt in hohem Maße ethische Sensibilität sei-

tens der Untersuchenden voraus. Zahlreiche Untersuchungsgegenstände wie z. B. Gewalt, Aggressivität, Liebe, Leistungsstreben, psychische Störungen, Neigung zu Konformität, ästhetische Präferenzen, Schmerztoleranz oder Angst betreffen die Privatsphäre des Menschen, die durch das Grundgesetz geschützt ist. Neben mangelnder Anonymisierung bzw. möglichem Mißbrauch personenbezogener Daten ist die Beeinflussung bzw. physische oder psychische Beeinträchtigung der Untersuchungsteilnehmer durch den Untersuchungsablauf das wichtigste ethische Problemfeld.

Die Bewertung vorläufiger Untersuchungsideen wäre unvollständig, wenn sie nicht auch ethische Kriterien mit berücksichtigen würde, wenngleich sich die Frage, ob eine Untersuchung ethisch zu verantworten ist oder nicht, häufig erst bei Festlegung der konkreten Untersuchungsdurchführung stellt (► Abschn. 2.3.3, 2.3.5 und 2.5). Dennoch ist es ratsam, sich frühzeitig mit der ethischen Seite eines Untersuchungsvorhabens auseinanderzusetzen.

Für die Psychologie hat praktisch jedes Land seine eigenen berufsethischen Verpflichtungen erlassen (vgl. Schuler, 1980). In Deutschland gelten die vom Berufsverband Deutscher Psychologinnen und Psychologen (BDP) und von der Deutschen Gesellschaft für Psychologie (DGPs) gemeinsam herausgegebenen *Ethischen Richtlinien* (DGPs & BDP, 1999). Diese Richtlinien regeln nicht nur den Umgang mit Menschen und Tieren als Untersuchungsobjekten, sondern beziehen sich unter anderem auch auf die Publikation von Forschungsergebnissen. Schließlich sind nicht nur Versuchspersonen, sondern auch Mitforschende von unethischem Verhalten bedroht, etwa wenn sie trotz nennenswerter Beteiligung an der Arbeit nicht namentlich erwähnt werden. Auszüge aus diesen Richtlinien findet man bei Hussy und Jain (2002, S. 252 f.).

Anlässlich der zunehmenden Internationalisierung (bzw. Amerikanisierung) psychologischer Forschung sei hier auch auf die sehr detaillierten ethischen Richtlinien der American Psychological Association verwiesen (APA, 1992). Generell unterliegt die psychologische Forschung in den USA einer sehr viel strengeren ethischen Kontrolle, als das in Europa bislang der Fall ist. So dürfen in den USA psychologische Fragebögen erst verteilt werden, nachdem sie von der Ethikkommission der jeweiligen Universität genehmigt wurden.

Box 2.1

Gehorsam und Gewalt – Ist diese Untersuchung ethisch vertretbar?

Heftige Kontroversen bezüglich der ethischen Grenzen empirischer Untersuchungen löste eine Studie von Milgram (1963) aus, mit der die Gewissenlosigkeit von Menschen, die sich zum Gehorsam verpflichtet fühlen, demonstriert werden sollte.

40 Personen – es handelte sich um Männer im Alter zwischen 20 und 50 Jahren mit unterschiedlichen Berufen – nahmen freiwillig an dieser Untersuchung teil. Nach einer ausführlichen Instruktion waren sie davon überzeugt, dass sie an einer wissenschaftlichen Untersuchung über den Zusammenhang zwischen Strafe und Lernen teilnehmen würden. Hierfür teilte der Untersuchungsleiter die Untersuchungsteilnehmer scheinbar in zwei Gruppen auf: Die eine Gruppe, so hieß es, würde eine Lernaufgabe erhalten (Paarassoziationsversuch), und die andere Gruppe, die Trainergruppe, erhielt die Aufgabe, den Lernerfolg der »Schüler« durch Bestrafung zu verbessern. Tatsächlich gehörten jedoch alle Untersuchungsteilnehmer der Trainergruppe an; der »Schüler« bzw. das »Opfer« wurde jeweils von einem »Strohmann« des Untersuchungsleiters gespielt.

Der Untersuchungsleiter führte jeden einzelnen Trainer zusammen mit dem »Schüler« in einen Raum, in dem sich ein Gerät befand, das einem elektrischen Stuhl sehr ähnlich sah. Der vermeintliche »Schüler« wurde gebeten, sich auf diesen Stuhl zu setzen. In einem Nebenraum stand ein Gerät, das der Trainer zur Bestrafung des »Schülers« benutzen sollte. Es handelte sich um einen Elektroschockgenerator mit 30 Schaltern für Schockstärken zunehmender Intensität von 15 Volt bis 450 Volt. Einige Schalter waren verbal gekennzeichnet: »leichter Schock«, »mäßiger Schock«, »starker Schock«, »sehr starker Schock«, »intensiver Schock«, »extrem intensiver Schock«, »Gefahr: schwerer Schock!« Zwei weitere Schalter nach dieser letzten Bezeichnung markierten drei Kreuze.



Über eine Anzeige erfuhr der Trainer, ob der »Schüler« die ihm gestellten Aufgaben richtig oder falsch löste. Machte der »Schüler« einen Fehler, erteilte der Trainer ihm einen Schock von 15 Volt. Jeder weitere Fehler musste mit der nächsthöheren Schockstärke bestraft werden. Dem Trainer wurde mitgeteilt, dass die Elektroschocks zwar sehr schmerzhaft, aber ohne bleibende Schäden seien.

Natürlich erhielt der als Schüler getarnte »Strohmann« im Nebenraum keinen Schock. Seine Instruktion lautete, im Verhältnis 3:1 falsche bzw. richtige Antworten zu geben und dies auch nur so lange, bis die Schockstärke 300 erreicht war. Danach signalisierte die Richtig-falsch-Anzeige keine Reaktionen mehr, und stattdessen hörte der Trainer, wie der »Schüler« kräftig gegen die Wand schlug.

In dieser Situation wandten sich die Trainer gewöhnlich an den Untersuchungsleiter mit der Frage, wie sie auf das Schweigen der Richtig-falsch-Anzeige bzw. auf die offenbar heftigen emotionalen Reaktionen des »Schülers« reagieren sollten. Es wurde ihnen bedeutet, dass das Ausbleiben einer Reaktion als Fehler zu werten und damit das Bestrafen mit der nächsthöheren Schockstärke fortzusetzen sei. Nach dem 315-Volt-Schock hörte auch das Pochen an die Wand auf.

Für den Fall, dass ein Trainer darum bat, die Untersuchung abbrechen zu dürfen, waren vier gestaffelte Standardantworten vorgesehen:

1. Bitte fahren Sie fort.
2. Das Experiment erfordert es, dass Sie weitermachen.
3. Es ist absolut erforderlich, dass Sie weitermachen.
4. Sie haben keine andere Wahl, Sie müssen weitermachen.

Erst nachdem auch die vierte Aufforderung den Trainer nicht veranlassen konnte, die Schockstärke weiter zu erhöhen, wurde die Untersuchung abgebrochen. Für jeden Trainer wurde dann als Index für seine »Gehorsamkeit« die Stärke des zuletzt erteilten Schocks registriert.


Ergebnisse: Keiner der 40 Trainer brach die Untersuchung vor dem 300-Volt-Schock ab. (Bei dieser mit der Verbalmarke »Intensiver Schock« versehenen Stärke schlug der »Schüler« gegen die Wand, und der Trainer erhielt keine Rückmeldung mehr bezüglich der gestellten Aufgaben.)

Fünf Trainer kamen der Aufforderung, den nächststärkeren Schock zu geben, nicht mehr nach. Bis hin zur 375-Volt-Marke verweigerten weitere neun Trainer den Gehorsam. Die verbleibenden 26 Trainer erreichten die mit drei Kreuzen

gekennzeichneten maximalen Schockstärken von 450 Volt.

Verhaltensbeobachtungen durch eine Einwegscheibe zeigten Reaktionen der Trainer, die für sozialpsychologische Laborexperimente äußerst ungewöhnlich sind. Es wurden Anzeichen höchster innerer Spannung wie Schwitzen, Zittern, Stottern, Stöhnen etc. registriert.

(Kritische Diskussionen dieser Untersuchung findet man u. a. bei Baumrind, 1964; Kaufman, 1967; Milgram, 1964; Stuwe & Timaeus, 1980.)

Im Folgenden werden einige Aspekte genannt, die bei der Überprüfung der ethischen Unbedenklichkeit empirischer Untersuchungen beachtet werden sollten. Eine ausführliche Behandlung des Themas »Ethik in der psychologischen Forschung« findet man bei Patry (2002).  Box 2.1 führt in die hier zu diskutierende Problematik ein.

Güterabwägung: Wissenschaftlicher Fortschritt oder Menschenwürde

Viele humanwissenschaftliche Studien benötigen Daten, deren Erhebung nur schwer mit der Menschenwürde der beteiligten Personen vereinbar ist. Ob es um die Untersuchung der Schmerztoleranzschwelle, die Erzeugung von Depressionen durch experimentell herbeigeführte Hilflosigkeit oder um Reaktionen auf angstauslösende Reize geht: Es gibt Untersuchungen, die darauf angewiesen sind, dass die untersuchten Personen in eine unangenehme, manchmal auch physisch oder psychisch belastende Situation gebracht werden. Lassen sich derartige Beeinträchtigungen auch nach sorgfältigen Bemühungen, die Untersuchung für die Betroffenen weniger unangenehm zu gestalten, nicht vermeiden, so können sie nur gerechtfertigt werden, wenn die Untersuchung Ergebnisse verspricht, die anderen Personen (z. B. schmerzkranken, depressiven oder phobischen Menschen) zugute kommen.

Hierüber eine adäquate prospektive Einschätzung abzugeben, fällt nicht nur dem Anfänger schwer. Die feste Überzeugung von der Richtigkeit der eigenen Idee erschwert eine umsichtige Einschätzung der Situation. Es ist deshalb zu fordern, dass in allen Fällen, in denen

die eigene Einschätzung auch nur die geringsten Zweifel an der ethischen Unbedenklichkeit der geplanten Untersuchung aufkommen lässt, außenstehende, erfahrene Fachleute und die zu untersuchende Zielgruppe zu Rate gezogen werden.

Persönliche Verantwortung

Bei der Auswahl geeigneter Untersuchungsthemen muss berücksichtigt werden, dass derjenige, der die Untersuchung durchführt, für alle unplanmäßigen Vorkommnisse zumindest moralisch verantwortlich ist. Wann



Der Experimentierfreude sind ethische Grenzen gesetzt. (Zeichnung: R. Löffler, Dinkelsbühl)



immer ethisch bedenklich erscheinende Instruktionen, Befragungen, Tests oder Experimente erforderlich sind, ist der Untersuchungsleiter verpflichtet, die Untersuchungsteilnehmer auf mögliche Gefährdungen und ihr Recht, die Untersuchungsteilnahme zu verweigern, aufmerksam zu machen. Sind physische Beeinträchtigungen nicht auszuschließen, müssen vor Durchführung der Untersuchung medizinisch geschulte Personen um ihre Einschätzung gebeten werden.

2.2.3 Informationspflicht

Die Tauglichkeit einer Untersuchungsidee hängt auch davon ab, ob den zu untersuchenden Personen von vornherein sämtliche Informationen über die Untersuchung mitgeteilt werden können, die ihre Entscheidung, an der Untersuchung teilzunehmen, potenziell beeinflussen. Entschließt sich ein potenzieller Proband nach Kenntnisnahme aller relevanten Informationen zur Teilnahme an der in Frage stehenden Untersuchung, spricht man von »Informed Consent«. Sind Personen an ihren eigenen Untersuchungsergebnissen interessiert, ist es selbstverständlich, dass diese nach Abschluss der Untersuchung schriftlich, fernmündlich oder in einer kleinen Präsentation mitgeteilt werden.

Gelegentlich ist es für das Gelingen einer Untersuchung erforderlich, dass die Untersuchungsteilnehmer den eigentlichen Sinn der Untersuchung nicht erfahren dürfen (Experimente, die durch sozialen Gruppendruck konformes Verhalten evozieren, würden sicherlich nicht gelingen, wenn die Teilnehmer erfahren, dass ihre Konformitätsneigungen geprüft werden sollen; vgl. hierzu die auf ▶ S. 39 erwähnte Inokkulationstheorie). Sind Täuschungen unvermeidlich und verspricht die Untersuchung wichtige, neuartige Erkenntnisse, so besteht die Pflicht, die Teilnehmer nach Abschluss der Untersuchung über die wahren Zusammenhänge aufzuklären (Debriefing). Danach sollten sie auch auf die Möglichkeit aufmerksam gemacht werden, die weitere Auswertung ihrer Daten nicht zu gestatten. In jedem Falle ist bei derartigen Untersuchungen zu prüfen, ob sich Täuschungen oder irreführende Instruktionen nicht durch die Wahl einer anderen Untersuchungstechnik vermeiden lassen.

Freiwillige Untersuchungsteilnahme

Niemand darf zu einer Untersuchung gezwungen werden. Auch während einer Untersuchung hat jeder Teilnehmer das Recht, die Untersuchung abzubrechen.

Diese Forderung bereitet sicherlich Schwierigkeiten, wenn eine Untersuchung auf eine repräsentative Stichprobe angewiesen ist (▶ S. 397f.). Es bestehen aber auch keine Zweifel, dass Personen, die zur Teilnahme an einer Untersuchung genötigt werden, die Ergebnisse erheblich verfälschen können (▶ S. 71 ff.). Hieraus leitet sich die Notwendigkeit ab, die Untersuchung so anzulegen, dass die freiwillige Teilnahme nicht zu einem Problem wird. Hierzu gehört auch, dass die Untersuchungsteilnehmer nicht wie beliebige oder austauschbare »Versuchspersonen« behandelt werden, sondern als Individuen, von deren Bereitschaft, sich allen Aufgaben freiwillig und ehrlich zu stellen, das Gelingen der Untersuchung maßgeblich abhängt.

In manchen Untersuchungen wird die »freiwillige« Untersuchungsteilnahme durch eine gute Bezahlung honoriert. Auch diese Maßnahme ist ethisch nicht unbedenklich, wenn man in Rechnung stellt, dass finanziell schlechter gestellte Personen auf die Entlohnung angewiesen sein könnten, ihre »Freiwilligkeit« also erkauft wird. Im übrigen ist bekannt, dass bezahlte Untersuchungsteilnehmer dazu neigen, sich als »gute Versuchsperson« (Orne, 1962) darzustellen, was – weil die Versuchsperson dem Untersuchungsleiter gefallen möchte – wiederum die Untersuchungsergebnisse verfälscht. Bezahlungen sind deshalb nur zu rechtfertigen, wenn die Untersuchung zeitlich sehr aufwendig ist oder wenn Personen nur gegen Bezahlung für eine Teilnahme an der Untersuchung zu gewinnen sind.

Besonders prekär wird die Frage der Freiwilligkeit der Untersuchungsteilnahme an psychologischen Instituten, deren Prüfungsordnungen die Ableistung einer bestimmten Anzahl von »Versuchspersonenstunden« vorsehen. Hier vertreten wir den Standpunkt, dass angehende Psychologen bereit sein müssen, in psychologischen Untersuchungen ihrer Wahl als »Versuchspersonen« Erfahrungen zu sammeln, die ihnen im Umgang mit Teilnehmern für spätere, eigene Untersuchungen zugute kommen. Ferner gilt, dass Psychologiestudenten dafür Verständnis zeigen sollten, dass eine empirisch orientierte Wissenschaft auf die Bereitschaft von Menschen, sich untersuchen zu lassen, angewiesen ist, sodass ihre Teilnahme an Untersuchungen letztlich auch dem Erkenntnisfortschritt dient. (Auf die Frage, ob Studenten »taugliche« Versuchspersonen sind, wird auf ▶ S. 74 f. eingegangen.)

Vermeidung psychischer oder körperlicher Beeinträchtigungen

Lewin (1979) unterscheidet drei Arten von Beeinträchtigungen:

- vermeidbare Beeinträchtigungen
- unbeabsichtigte Beeinträchtigungen
- beabsichtigte Beeinträchtigungen.

Sie spricht von vermeidbarer Beeinträchtigung, wenn Untersuchungsteilnehmer aus Mangel an Sorgfalt, aus Unachtsamkeit oder wegen überflüssiger, für die Untersuchung nicht unbedingt erforderlicher Maßnahmen zu Schaden kommen. (Wobei mit »Schaden« nicht nur körperliche Verletzungen, sondern auch subtile Beeinträchtigungen wie peinliche Bloßstellungen, unangenehme Überforderungen, Angst, Erschöpfung u. Ä. gemeint sind.) Sie sollten durch eine sorgfältige und schonende Untersuchungsdurchführung vermieden werden.

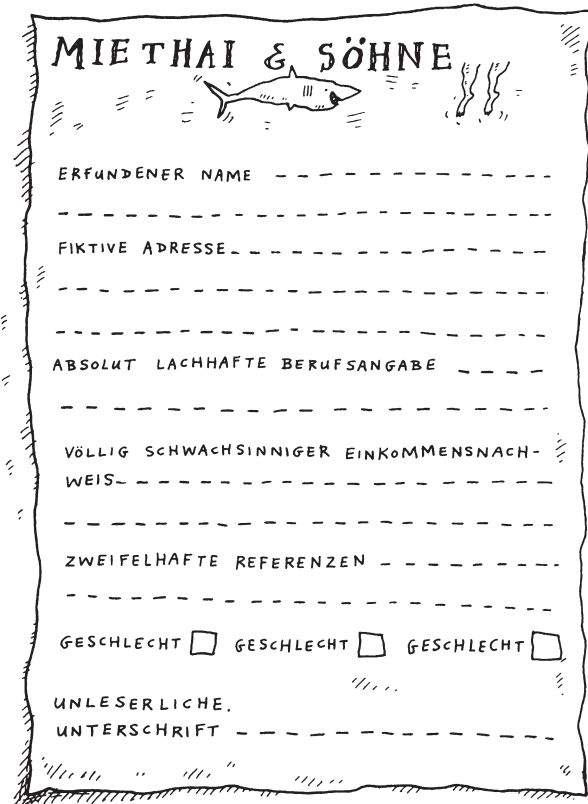
Trotz sorgfältiger Planung und Durchführung einer Untersuchung kann es aufgrund unvorhergesehener Zwischenfälle zu unbeabsichtigten Beeinträchtigungen der Untersuchungsteilnehmer kommen, die den Untersuchungsleiter – soweit er sie bemerkt – zum unverzüglichen Eingreifen veranlassen sollten. Ein einfacher Persönlichkeitstest z. B. oder ein steriles Untersuchungslabor können ängstliche Personen nachhaltig beunruhigen. Die einfache Frage nach dem Beruf des Vaters kann ein Kind zum Schweigen oder gar Weinen bringen, weil der Vater kürzlich einem Unfall erlegen ist. Je nach Anlass können ein persönliches Gespräch oder eine sachliche Aufklärung helfen, über die unbeabsichtigte Beeinträchtigung hinwegzukommen.

Die Untersuchung von Angst, Schuld- und Schamgefühlen, Verlegenheit o. Ä. machen es meistens erforderlich, die Untersuchungsteilnehmer in unangenehme Situationen zu bringen. Diese beabsichtigten Beeinträchtigungen sollten die Untersuchungsteilnehmer so wenig wie möglich belasten. Oftmals reichen bereits geringfügige Beeinträchtigungen für die Überprüfung der zu untersuchenden Fragen aus.

Anonymität der Ergebnisse

Wenn die Anonymität der persönlichen Angaben nicht gewährleistet werden kann, sollte auf eine empirische Untersuchung verzichtet werden. Jedem Untersuchungs-

Du sollst Deine Identität nicht preisgeben



Auch bei psychologischen Untersuchungen sollte die Anonymität gewahrt bleiben. Aus Poskitt, K. & Appleby, S. (1993). Die 99 Lasse-dasse. Kiel: Achterbahn Verlag

teilnehmer muss versichert werden, dass die persönlichen Daten nur zu wissenschaftlichen Zwecken verwendet und dass die Namen nicht registriert werden. Falls erforderlich, kann der Untersuchungsteilnehmer seine Unterlagen mit einem Codewort versehen, dessen Bedeutung nur ihm bekannt ist.

Auskünfte über andere Personen unterliegen dem **Datenschutz**. Vor größeren Erhebungen, in denen auch persönliche Angaben erfragt werden, empfiehlt es sich, die entsprechenden rechtlichen Bestimmungen einzusehen bzw. sich von Datenschutzbeauftragten beraten zu lassen (vgl. Lecher, 1988; Simitis et al., 1981).



2.3 Untersuchungsplanung

Diente die »vorwissenschaftliche Phase« einer ersten Sondierung der eigenen Untersuchungsideen, beginnt jetzt die eigentliche Planung der empirischen Untersuchung. Sie markiert den wichtigsten Abschnitt empirischer Arbeiten. Von ihrer Präzision hängt es ab, ob die Untersuchung zu aussagekräftigen Resultaten führt oder ob die Untersuchungsergebnisse z. B. wegen ihrer mehrdeutigen Interpretierbarkeit, fehlerhafter Daten oder einer unangemessenen statistischen Auswertung unbrauchbar sind. Man sollte sich nicht scheuen, die Aufarbeitung einer Untersuchungsidee abzubrechen, wenn die Planungsphase Hinweise ergibt, die einen positiven Ausgang der Untersuchung zweifelhaft erscheinen lassen. Nachlässig begangene Planungsfehler müssen teuer bezahlt werden und sind häufig während der Untersuchungsdurchführung nicht mehr korrigierbar. (Eine Kurzfassung typischer Planungsfehler und Planungsaufgaben findet man bei Aiken, 1994.)

! Der wichtigste Abschnitt einer empirischen Forschungsarbeit ist die Untersuchungsplanung.

Die Bedeutsamkeit der im Folgenden behandelten Bestandteile eines Untersuchungsplanes hängt davon ab, welche Untersuchungsart für das gewählte Thema am angemessensten erscheint (► Abschn. 2.3.3). Eine Entscheidung hierüber sollte jedoch erst getroffen werden, nachdem der Anspruch der Untersuchung geklärt (► Abschn. 2.3.1) und das Literaturstudium abgeschlossen ist (► Abschn. 2.3.2).

2.3.1 Zum Anspruch der geplanten Untersuchung

Empirische Untersuchungen haben unterschiedliche Funktionen. Eine kleinere empirische Studie, die als Semesterarbeit angefertigt wird, muss natürlich nicht den Ansprüchen genügen, die für eine Dissertation oder für ein mit öffentlichen Mitteln gefördertes Großprojekt gelten. Die Planungsarbeit beginnt deshalb mit einer möglichst realistischen Einschätzung des Anspruchs des eigenen Untersuchungsvorhabens in Abhängigkeit vom Zweck der Untersuchung.

Prüfungsordnungen. Die erste wichtige Informationsquelle hierfür sind Prüfungsordnungen, in deren Rahmen empirische Qualifikationsarbeiten erstellt werden. Auch wenn diese Ordnungen in der Regel nicht sehr konkret über den Anspruch der geforderten Arbeit informieren, lassen sich ihnen dennoch einige interpretationsfähige Hinweise entnehmen. So macht es einen erheblichen Unterschied, ob z. B. »ein selbständiger Beitrag zur wissenschaftlichen Forschung« oder »der Nachweis, selbständig ein wissenschaftliches Thema bearbeiten zu können« gefordert wird. Die zuerst genannte Forderung ist zweifellos anspruchsvoller und wäre für eine Dissertation angemessen; hier geht es um die Erweiterung des Bestandes an wissenschaftlichen Erkenntnissen durch einen eigenen Beitrag. Die zweite, für Studienabschlussarbeiten typische Forderung verlangt »lediglich«, dass die inhaltlichen und methodischen Kenntnisse ausreichen, um ein Thema nach den Regeln der jeweiligen Wissenschaftsdisziplin selbständig untersuchen zu können. Es geht hier also eher um die Befähigung zum selbständigen wissenschaftlichen Arbeiten und weniger um die Originalität des Resultates. Abschlussarbeiten sollen dokumentieren, dass wissenschaftliche Instrumente wie z. B. die Nutzung vorhandener Literatur, die angemessene Operationalisierung von Variablen, der geschickte Aufbau eines Experimentes, der Entwurf eines Fragebogens, die Organisation einer größeren Befragungskampagne, das Ziehen einer Stichprobe, die statistische Auswertung von Daten oder das Dokumentieren von Ergebnissen vom Diplomanden beherrscht werden. Zusätzlich informiert die Prüfungsordnung über den zeitlichen Rahmen, der für die Anfertigung der Arbeit zur Verfügung steht.

Vergleichbare Arbeiten. Die zweite wichtige Informationsquelle, den Anspruch der geplanten Arbeit richtig einzuschätzen, stellen Arbeiten dar, die andere nach derselben Ordnung bereits angefertigt haben. Das Durchsehen verschiedener Qualifikationsarbeiten vermittelt einen guten Eindruck davon, wie anspruchsvoll und wie umfangreich vergleichbare Arbeiten sind. Schließlich sind Studierende gut beraten, sich von erfahrenen Studienkollegen und Mitgliedern des Lehrkörpers bei der Einschätzung der Angemessenheit ihrer Untersuchungsideen helfen zu lassen.

2.3.2 Literaturstudium

Wer ein interessantes Thema gefunden hat, steht vor der Aufgabe, die vorläufige Untersuchungsidee in den bereits vorhandenen Wissensstand einzuordnen. Das hierfür erforderliche Literaturstudium geschieht mit dem Ziel, die eigene Untersuchungsidee nach Maßgabe bereits vorhandener Untersuchungsergebnisse und Theorien einzugrenzen bzw. noch offene Fragen oder widersprüchliche Befunde zu entdecken, die mit der eigenen Untersuchung geklärt werden können. Es empfiehlt sich, das Literaturstudium sorgfältig, planvoll und ökonomisch anzugehen. (Weitere Angaben zur Literatuarbeit finden sich in ► Abschn. 6.2 und ► Anhang C.)

Orientierung

Wenn zur Entwicklung der Untersuchungsidee noch keine Literatur herangezogen wurde, sollten als erstes **Lexika, Wörterbücher** und **Handbücher** eingesehen werden, die über die für das Untersuchungsthema zentralen Begriffe informieren und erste einführende Literatur nennen. Diese einführende Literatur enthält ihrerseits Verweise auf speziellere Monographien oder Zeitschriftenartikel, die zusammen mit den lexikalischen Beiträgen bereits einen ersten Einblick in den Forschungsstand vermitteln.

Von besonderem Vorteil ist es, wenn man bei dieser Suche auf aktuelle **Sammelreferate** (Reviews) stößt, in denen die wichtigste Literatur zu einem Thema für einen begrenzten Zeitraum inhaltlich ausgewertet und zusammengefasst ist. Sehr hilfreich sind in diesem Zusammenhang auch sog. **Metaanalysen**, in denen die empirischen Befunde zu einer Forschungsthematik statistisch aggregiert sind (► Kap. 10). Um die Suche nach derartigen Überblicksreferaten abzukürzen, sollte man sich nicht scheuen, das Bibliothekspersonal zu fragen, in welchen Publikationsorganen derartige Zusammenfassungen üblicherweise erscheinen (für die Psychologie sind dies z. B. die Zeitschriften *Annual Review of Psychology* oder *Psychological Review* bzw. die »Advances«- und »Progress«-Serien für Teilgebiete der Psychologie.) Gute Bibliotheken führen außerdem einen ausführlichen Schlagwortkatalog, der ebenfalls für die Beschaffung eines ersten Überblicks genutzt werden sollte.

Universitätsbibliotheken sind komplizierte wissenschaftliche Organisationen, die zusammen bestrebt sind,

das gesamte Wissen aller wissenschaftlichen Disziplinen zu archivieren. Neben allgemeinen Universitätsbibliotheken helfen auch spezialisierte **Fachinformationsdienste** sowie computergestützte **Datenbanken** und das **Internet** bei der ersten Literaturrecherche (► Anhang C).

In diesem Stadium der Literatuarbeit stellt sich oft heraus, dass das vorläufige Untersuchungsvorhaben zu umfangreich ist, denn allein das Aufarbeiten der im Schlagwortkatalog aufgeführten Literatur würde vermutlich Monate in Anspruch nehmen. Es kann deshalb erforderlich sein, nach einer ersten Durchsicht der einschlägigen Literatur das Thema neu zu strukturieren und anschließend einzugrenzen.

Vertiefung

Die Orientierungsphase ist abgeschlossen, wenn man die in der Literatur zum avisierten Forschungsfeld am ausführlichsten behandelten Themenstränge ebenso kennt wie die zentralen Autorinnen und Autoren und die von ihnen präferierten Methoden und Theorien. Somit ist man dann in der Lage, die eigene Fragestellung einzuordnen und an das bereits Publierte anzuschließen. Interessiert man sich etwa für die Determinanten und Konsequenzen von Schwangerschaften bei Teenagern, so würde man nach der orientierenden Literaturrecherche feststellen, dass die Forschung sich nahezu ausschließlich auf die jungen Mütter konzentriert. Das Profil des eigenen Forschungsvorhabens könnte nun darin bestehen, sich gerade mit den jugendlichen Vätern zu beschäftigen.

In der »zweiten Runde« der Literaturrecherche nutzt man weiterhin Bibliotheken, Buchhandlungen, Fachinformationsdienste, Datenbanken oder das Internet (► Anhang C); allerdings sucht man nun sehr gezielt nach Beiträgen, die das eingegrenzte Themengebiet inhaltlich und methodisch berühren (z. B. Einstellungen männlicher Jugendlicher zur Familienplanung, jugendspezifische Interviewtechniken usw.). Für eine solche Detailsuche sind allgemeine Handbücher oder Einführungswerke wenig geeignet. Stattdessen greift man auf **Bibliographien, Kongressberichte und Abstractbände** (z. B. *Psychological Abstracts*, *Sociological Abstracts*, *Social Science Citation Index*, *Index Medicus*) zurück, die den neuesten Forschungsstand weitgehend lückenlos verzeichnen. Die Nutzung der *Psychological Abstracts* wird in ■ Box 2.2 illustriert. Viele Hochschulbibliotheken

Box 2.2

2

Literatursuche mit Abstracts

Gibt es ein Leben nach dem Tod? Diese Frage scheint Menschen in allen Jahrhunderten zu beschäftigen. Die Esoterikbewegung hat das Interesse an übersinnlichen Phänomenen aufgegriffen und thematisiert das Leben nach dem Tod in unterschiedlicher Weise (Erinnerungen an frühere Leben, Reinkarnation, Kontaktaufnahme mit Verstorbenen etc.). Aber auch die traditionellen Religionen vermitteln Vorstellungen darüber, was uns nach dem Leben erwartet (christliche Vorstellungen von Himmel und Hölle etc.). Lässt sich zu diesem interessanten Themengebiet eine empirische Untersuchung durchführen? Spontan fallen uns diverse psychologische und soziologische Fragestellungen ein: Wie verbreitet ist der Glaube an ein Leben nach dem Tod? Wie unterscheiden sich Menschen, die an ein Leben nach dem Tod glauben, von denjenigen, die diese Vorstellung nicht teilen? Schätzt eine Gesellschaft, in der der Glaube an ein Leben nach dem Tod sehr verbreitet ist, den Wert des irdischen Lebens geringer ein als eine säkularisierte Gesellschaft?

Zu diesen Ideen sollte ein Blick in die Fachliteratur geworfen werden. Dazu könnte man z. B. die Psychological Abstracts des Jahres 1993 heranziehen und im Sachverzeichnis (»Subject Index«) unter den Stichwörtern »Religion«, »Death« und »Death Anxiety« nachschlagen. Unter dem Stichwort »Death Anxiety« befindet sich in der Rubrik »Serials« (Zeitschriften) eine Liste von Zeitschriftenartikeln, wobei die einzelnen Artikel nur mit einigen Stichpunkten skizziert und mit einer Ordnungsnummer versehen sind. Die Ordnungsnummer gibt an, an welcher Stelle in den Abstractbänden des Jahres 1993 der gesuchte Abstract zu finden ist.

Death Anxiety – Serials

afterlife & God beliefs, degree of anxiety perceived in death related pictures, Hindi male 40–60 yr olds with low vs high death site area exposure, India, 1240
attitudes toward & preoccupation with death, college students, 1935 vs 1991, 29434



birth environment & complications & transference & counter-transference, male analysand with fears of death & panic attacks, 30430
correlates of death depression vs anxiety, 16–82 yr olds, 25582
cross cultural & construct validity of Templer's Death Anxiety Scale, nursing students, Philippines, 31844
death anxiety & education, Air Force mortuary officers, conference presentation, 27755
death anxiety & life purpose of future vs past vs present time perspective, 52–94 yr olds, 45033
depression & self esteem & suicide ideation & death anxiety & GPA, 14–19 yr olds of divorced vs nondivorced parents, 25199
development & factor analysis of Revised Death Anxiety Scale, 18–88 yr olds, 15984
didactic vs experiential death & dying & grief workshop. death anxiety, nursing students, 7311
emotional managing function of belief in life after death, death anxiety, Hindu vs Muslim vs Christian 20–70 yr olds, 17278
emotional responses to & fear of child's death from diarrhea, urban vs rural mothers of 0–36 mo olds, Pakistan, 10232
factor analysis of Death Anxiety Scale vs Death Depression Scale, adults, 24103
Der oben fett gesetzte Artikel scheint interessant und wird in den Psychological Abstracts (1993, Band 80/2, Seite 2073) nachgeschlagen. Er sieht folgendermaßen aus:

17278. **Parsuram, Ameeta & Sharma, Maya.** (Jesus & Mary Coll. New Delhi, India) **Functional relevance of belief in life-after-death.** Special Series II: Stress, adjustment and death anxiety. **Journal of Personality & Clinical Studies**, 1992 (Mar–Sep), Vol 8(1–2), 97–100. – Studied the emotion managing function of belief in life after death in dealing with death anxiety. The differences in the concept of afterlife were examined in 20 Ss (aged 60–70 yrs) from each of 3 religions: Hindu, Islam and Christianity. Hindu Ss had the lowest level of death anxiety, followed by Muslim Ss, with the Christian Ss having the highest death anxiety. Hindus had the strongest belief in life after death, Muslims had the weakest belief in afterlife, and Christians fell in the middle. Results are discussed in terms of the theory of functional relevance of beliefs.

Es handelt sich um die Zusammenfassung eines Zeitschriftenartikels mit dem Titel »Functional relevance of belief in life-after-death« (Die funktionale Bedeutung des Glaubens an ein Leben nach dem Tod) aus

dem *Journal of Personality and Clinical Studies* aus dem Jahr 1992 (Band 8, Heft 1–2, Seite 97–100). Die Autoren Ameeta Parsuram und Maya Sharma stammen vom »Jesus & Maria College« in Neu Delhi (Indien). Das Abstract skizziert Fragestellung (1. Satz), Methode (2. Satz), Ergebnisse (3. und 4. Satz) und Schlussfolgerungen (5. Satz) der Untersuchung.

Anhand dieses Abstracts ist nun zu prüfen, ob a) die Studie für die eigene Arbeit relevant ist (in-

haltlicher und methodischer Bezug, Bedeutung der Autoren), b) die genannte Zeitschrift vor Ort zur Verfügung steht oder per Fernleihe beschafft werden muss und c) Aufwand und Nutzen bei der Literaturbeschaffung in angemessenem Verhältnis stehen (Negativpunkte: der Artikel umfasst nur 4 Seiten; es wurden nur 20 Personen – d. h. ca. 7 pro Gruppe! – befragt; die Datenerhebungsmethode – standardisierter Fragebogen, offenes Interview o. Ä. – wird nicht genannt).

ken bieten heute die *Psychological Abstracts* ebenso wie andere Abstractwerke als elektronische Datenbanken an, die mit der Nummer des Bibliotheksausweises auch via Internet zugänglich sind. Findet man auf diesem Wege zwei bis drei aktuelle Zeitschriftenartikel oder Buchbeiträge, so hat man bereits eine Fülle von Quellen erschlossen. Denn die **Literaturverzeichnisse** dieser Artikel werden sich typischerweise als Fundgrube einschlägiger Beiträge erweisen.

Der Fall, dass man vergeblich nach verwertbarer Literatur sucht, tritt relativ selten ein. (Man beachte, dass Arbeiten mit ähnlicher Thematik möglicherweise unter anderen als den geprüften Stichwörtern zusammengefasst sind. Bei der Stichwortsuche unterstützen sog. **Thesauri**, die synonyme und inhaltlich ähnliche Fachbegriffe zu dem jeweiligen Suchbegriff angeben.) Stellt sich dennoch heraus, dass die Literatur für eine hypothesenprüfende Untersuchung keine Anknüpfungspunkte bietet, wird man zunächst eine Erkundungsstudie ins Auge fassen, deren Ziel es ist, plausible Hypothesen zu bilden (► Abschn. 2.3.3).

Dokumentation

Eine Literaturrecherche ist praktisch wertlos, wenn Informationen nachlässig und unvollständig dokumentiert werden. Von den vielen individuellen Varianten, das Gelesene schriftlich festzuhalten, haben sich das traditionelle Karteikartensystem und die elektronische Literaturdatenbank am besten bewährt. Für jede Publikation (Monographie, Zeitschriftenartikel, Lehrbuch usw.) wird eine Karteikarte (auf Papier oder in der Datenbank) angelegt, die zunächst die vollständigen bibliographischen Angaben enthält, die für das Literatur-

verzeichnis (► S. 90 ff.) benötigt werden: Autorenname, Titel der Arbeit sowie Name, Jahrgang und Nummer der Zeitschrift, Anfangs- und Endseitenzahl des Beitrages bzw. bei Büchern zusätzlich Verlag, Ort und Erscheinungsjahr. In Stichworten sollten zudem Angaben über den Theoriebezug, die Fragestellung, die verwendete Methode und die Ergebnisse aufgenommen werden. Wörtliche Zitate (mit Angabe der Seitenzahl!), die für den Untersuchungsbericht geeignet erscheinen, sowie Bibliothekssignaturen, die ein späteres Nachschlagen der Literatur erleichtern, komplettieren die Karteikarte.

2.3.3 Wahl der Untersuchungsart

Im Folgenden wird eine Klassifikation empirischer Untersuchungen vorgestellt, die es Studierenden erleichtern soll, ihr Untersuchungsvorhaben einzuordnen und entsprechende Planungsschwerpunkte zu setzen. Wir befassen uns zunächst mit den Hauptkategorien empirischer Untersuchungen, die in den folgenden Kapiteln ausführlicher dargestellt und ausdifferenziert werden. Für eine gründliche Planung wird empfohlen, die entsprechenden Abschnitte dieser Kapitel ebenfalls vor Durchführung der Untersuchung zu lesen.

Moderne Human- und Sozialwissenschaften müssen einerseits Lösungsansätze für neuartige Fragestellungen entwickeln und andererseits die Angemessenheit ihrer Theorien angesichts einer sich verändernden Realität prüfen. Die Untersuchungsmethoden sind hierbei nicht beliebig, sondern sollten dem Status der wissenschaftlichen Frage Rechnung tragen. Die Wahl der Untersuchungsart richtet sich deshalb zunächst nach dem in der

Literatur dokumentierten Kenntnisstand zu einer Thematik. Dieses erste Kriterium entscheidet darüber, ob mit einer Untersuchung eine oder mehrere Hypothesen überprüft oder ob zunächst Hypothesen erkundet werden sollten. Das zweite Auswahlkriterium betrifft die angestrebte Gültigkeit bzw. die Eindeutigkeit der mit den Untersuchungsergebnissen verbundenen Aussagen.

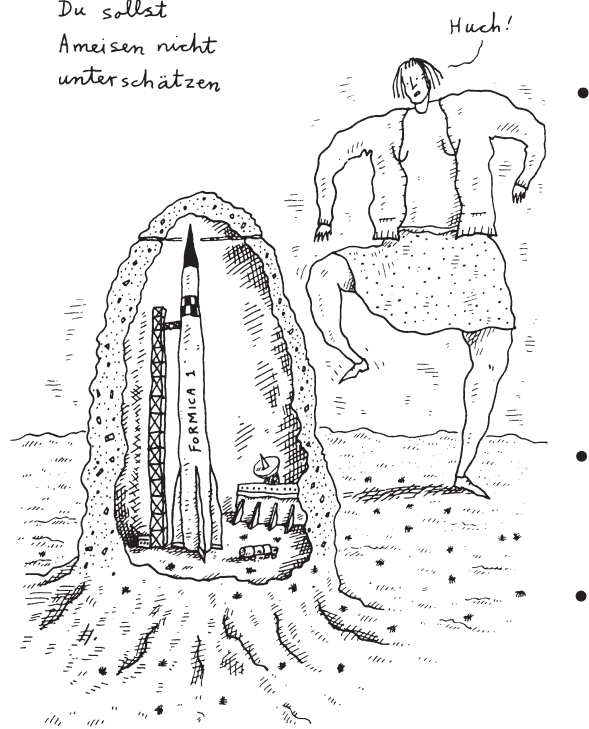
Erstes Kriterium: Stand der Forschung

Nach Abschluss der Literaturlarbeit ist zunächst zu entscheiden, ob der Stand der Forschung die Ableitung und Überprüfung einer gut begründeten Hypothese zulässt (explanative Untersuchung), oder ob mit der Forschungsthematik wissenschaftliches Neuland betreten wird, welches zunächst eine explorative Orientierung bzw. eine gezielte Hypothesensuche erfordert (explorative Untersuchung). Zur Klärung der Frage, ob die Forschungsidee als wissenschaftliche Hypothese formulierbar ist, wird auf ► Abschn. 1.1 verwiesen. Zudem gibt es Fragestellungen, in denen es nicht primär darum geht, Phänomene durch Theorien und Hypothesen zu erklären, sondern darum, Populationen zu beschreiben (deskriptive Untersuchung).

Explorative Untersuchungen. Explorative bzw. erkundende Untersuchungen werden in erster Linie mit dem Ziel durchgeführt, in einem relativ unerforschten Untersuchungsbereich neue Hypothesen zu entwickeln oder theoretische bzw. begriffliche Voraussetzungen zu schaffen, um erste Hypothesen formulieren zu können. Sie sind relativ wenig normiert und lassen der Phantasie und dem Einfallsreichtum des Untersuchenden viel Spielraum. Dementsprechend sind die Richtlinien für die Planung derartiger Untersuchungen und die Anfertigung des Untersuchungsberichtes (► Abschn. 2.7) weniger verbindlich als für hypothesenprüfende Untersuchungen. Charakteristisch für explorative Studien sind beispielsweise die folgenden methodischen Ansätze:

- Durch **offene Befragungen** von Einzelpersonen (z. B. biographische oder narrative Interviews) oder von Gruppen (z. B. Gruppendiskussion) erfährt man, welche Probleme den Betroffenen besonders am Herzen liegen, welche Erklärungen oder Meinungen sie haben und welche besonderen lebensgeschichtlichen Ereignisse ihre aktuelle Situation bestimmen (► Abschn. 5.2.1, 5.4.4).

Du sollst
Ameisen nicht
unterschätzen



Bei der Exploration lohnt es sich zuweilen, auch scheinbar Bekanntes genauer unter die Lupe zu nehmen. Aus Poskitt, K. & Appleby, S. (1993). Die 99 Lassedasse. Kiel: Achterbahn Verlag

- Bei der **Feldbeobachtung** (Feldforschung) nimmt man am sozialen Leben des interessierenden Systems teil und hält dabei nach besonderen Ereignissen und Verhaltensmustern ebenso Ausschau wie nach den unausgesprochenen Gesetzen und Regeln des Zusammenlebens. Auch die Beobachtung von Rollenspielen, bei der Akteure bestimmte Situationen aus ihrem Leben nachspielen, kann die Aufmerksamkeit auf bislang vernachlässigte oder im Alltag nicht öffentlich zutage tretende Details lenken (► Abschn. 5.2.2, 5.4.1).
- Im Verlaufe einer **Aktionsforschung** definieren Wissenschaftler zusammen mit den Betroffenen die Problemstellung, suchen nach Ursachen (Hypothesengenerierung, Theoriebildung) und entwerfen Lösungsvorschläge (Interventionen). Der Erfolg der Intervention wird gemeinsam evaluiert (formative Evaluation; ► Abschn. 3.2.2) und gibt Anlass zur Modifikation von Theorien und Interventionsstrate-

gien. Wesentliche Impulse in diesem Prozess kommen immer von den Betroffenen, denen der Status von gleichberechtigten Experten eingeräumt wird (► Abschn. 5.4.2).

- Die detaillierte **Analyse von Einzelfällen** in Form von Selbstbeobachtung oder Fremdbeobachtung ist oftmals eine sinnvolle Vorbereitung von Stichprobenuntersuchungen, in denen Einzelfälle aggregiert werden (► Abschn. 5.2.2).
- Bei **nonreaktiven Messungen** wird auf der Basis von Verhaltensspuren, Rückständen, Ablagerungen oder Abnutzungen auf vergangenes Verhalten geschlossen. Wichtige Hinweise bei der Untersuchung sozialer Phänomene kann also ergänzend zur Befragung und Beobachtung von Akteuren auch die dingliche Umgebung geben (► Abschn. 5.2.3).
- **Qualitative Inhaltsanalysen** dienen dazu, schrittweise die zentralen Themen und Bedeutungen von Texten oder anderen Objekten (z. B. Kunstwerken, Fotos) herauszuarbeiten. Dabei ist eine minutiöse Wort-für-Wort-Analyse ebenso möglich wie eine orientierende Globalanalyse (► Abschn. 5.3).

Diese Formen der wenig standardisierten Datenerhebung mittels qualitativer Methoden (► Kap. 5) haben nur dann wissenschaftlichen Wert, wenn die gewonnenen Informationen zu neuen Ideen oder Hypothesen verdichtet werden können. Dazu stellt man zweckmäßigerweise Inventare von wichtigen Einflussgrößen auf, bildet durch Zusammenfassung ähnlicher Fälle Typen und Strukturen, schließt auf mögliche Ursachen und Gründe, verfolgt Veränderungen im Zeitverlauf oder konzentriert sich auf das dynamische Zusammenspiel mehrerer Systemelemente (► Abschn. 6.5). Auch die Erfassung quantitativer Daten (► Kap. 4) kann durch entsprechende Aufbereitung (► Abschn. 6.4) die Aufstellung neuer Hypothesen anregen.

Populationsbeschreibende Untersuchungen. Das primäre Ziel dieser Untersuchungsart ist die Beschreibung von Populationen (Grundgesamtheiten) hinsichtlich ausgewählter Merkmale. Diese Untersuchungsart wird vor allem in demoskopischen Forschungen eingesetzt, in denen die Zusammensetzung der Bevölkerung bzw. von Teilen der Bevölkerung in Bezug auf bestimmte Merkmale sowie deren Veränderungen interessieren. Im

Vordergrund stehen Stichprobenerhebungen, die eine möglichst genaue Schätzung der unbekanntem Merkmalsausprägungen in der Population (**Populationsparameter**) gestatten. Wir werden diese Untersuchungsart in ► Kap. 7 ausführlich diskutieren. Diese Diskussion geht auf Techniken ein, welche die Genauigkeit der Parameterschätzungen durch die Nutzung von Vorinformationen aus der Literatur bzw. aufgrund eigener Erfahrungen erhöhen. Wir unterscheiden populationsbeschreibende Untersuchungen

- mit **einfachen Zufallsstichproben** (► Abschn. 7.1),
- mit **geschichteten Stichproben**, bei denen die Stichproben so zusammengesetzt werden, dass die prozentuale Verteilung von Schichtungsmerkmalen (Alter, Geschlecht, Beruf etc.) in der Stichprobe der prozentualen Verteilung dieser Merkmale in der Population entspricht (► Abschn. 7.2.1),
- mit **Klumpenstichproben**, in denen mehrere zufällig ausgewählte »Klumpen« (z. B. Krankenhäuser, Wohnblocks, Schulklassen o. Ä.) vollständig erhoben werden (► Abschn. 7.2.2),
- mit **mehrstufigen Stichproben**, in denen die Auswahl nach mehreren Schichtungs- oder Klumpenmerkmalen erfolgt (► Abschn. 7.2.3),
- Studien nach dem **Bayes'schen Ansatz**, der Stichprobeninformationen und »subjektive« Informationen für eine Parameterschätzung kombiniert (► Abschn. 7.2.5).

Neben populationsbeschreibenden Studien, die mit bevölkerungsrepräsentativen großen Stichproben arbeiten, werden häufig auch kleiner skalierte deskriptive Studien angefertigt. Dies ist beispielsweise bei vielen anwendungsorientierten Studien der Fall. Hier sollen die wissenschaftlichen Befunde eben nicht in erster Linie dazu dienen, Theorien zu prüfen und weiterzuentwickeln, sondern sie sollen praktisches Handeln anleiten. Eine Beschreibung des aktuellen Verhaltens und Erlebens in einem interessierenden Handlungskontext ist dafür ausreichend.

So könnte sich ein Unternehmen beispielsweise fragen, wie die betriebseigene Mitarbeiterzeitung genutzt wird und wie sie verbessert werden kann. Mit Hilfe einer sozialwissenschaftlichen Fragebogenstudie könnten die Nutzungsmuster und Bewertungen aller Unternehmensmitarbeiter erhoben (Vollerhebung) und statistisch be-

schrieben werden, ohne dass es sinnvoll ist, im Vorfeld Hypothesen aufzustellen. Wenn deskriptiv beispielsweise gezeigt werden kann, dass die Mehrzahl der Zielgruppe ein verbessertes Layout mit mehr Grafiken wünscht oder bestimmte Themen stärker berücksichtigt finden möchte, dann lassen sich daraus Handlungsanweisungen für die Praxis ableiten. Eine solche deskriptive bzw. populationsbeschreibende Studie wird strukturiert durch praxisbezogene Fragestellungen (z. B. »Ist die Zielgruppe mit der äußeren Aufmachung der Zeitung zufrieden?«) und eine möglichst umfassende Berücksichtigung aller relevanten Aspekte (z. B. Zufriedenheit mit dem Format, mit dem Schrifttyp, mit der Farbgestaltung, mit den Grafiken etc.).

Sozialwissenschaftlich geschulte Studierende, die im Rahmen von Unternehmenspraktika Untersuchungen durchführen, müssen nicht selten umdenken: Anstelle der in der Wissenschaft präferierten Hypothesenprüfung wird in der Praxis oft eine an anwendungsbezogenen Fragen orientierte Deskription verlangt.

Hypothesenprüfende Untersuchungen. Lassen sich aufgrund des Standes der Theorienentwicklung bzw. aufgrund von Untersuchungen, die zur gewählten Thematik bereits durchgeführt wurden, begründete Hypothesen formulieren, ist die Untersuchung nach den Kriterien einer hypothesenprüfenden bzw. **explanativen Untersuchung** anzulegen. Wir unterscheiden in ► Kap. 8, das diese Untersuchungsart genauer behandelt, zwischen

- Zusammenhangshypothesen (► Abschn. 8.2.3),
- Unterschiedshypothesen (► Abschn. 8.2.4),
- Veränderungshypothesen (► Abschn. 8.2.5) und
- Einzelfallhypothesen (► Abschn. 8.2.6).

Von **unspezifischen Hypothesen** sprechen wir, wenn die Forschung noch nicht genügend entwickelt ist, um genaue Angaben über die Größe des hypothesengemäß erwarteten Zusammenhanges, Unterschiedes oder der Veränderung machen zu können. Hypothesen, die mit dieser Untersuchungsart geprüft werden, behaupten lediglich, dass zwischen zwei oder mehreren Merkmalen ein Zusammenhang besteht, dass sich zwei oder mehrere Populationen in Bezug auf bestimmte Merkmale unterscheiden, dass zwei oder mehrere »Behandlungsarten« (Treatments) unterschiedliche Wirkungen haben oder dass sich ein oder mehrere Merkmale in einer

Population verändern. Beispiele für unspezifische Alternativhypothesen lauten $H_1: \mu_1 \neq \mu_2$ oder $H_1: \mu_1 > \mu_2$. Detailliertere Informationen über die hier angesprochenen Hypothesenarten enthält Box 8.1 (► S. 505 f.).

Unspezifische Hypothesen haben den Nachteil, dass sie bei großen Stichproben eigentlich immer zu einem signifikanten Ergebnis führen. (Zur Begründung dieser Behauptung vgl. ► Abschn. 9.1.) Es ist also letztlich nur eine Frage des Stichprobenaufwandes, ob der statistische Hypothesentest zugunsten der Forschungshypothese entscheidet oder nicht. Dieser Missstand wird behoben, wenn statt einer unspezifischen Alternativhypothese eine spezifische Hypothese mit einer klar definierten Effektgröße geprüft wird.

Spezifische Hypothesen mit Effektgrößen *können* formuliert werden, wenn bereits genügend Erfahrungen mit der Untersuchungsthematik sowie mit den für den Untersuchungsbereich typischen Untersuchungsinstrumenten vorliegen, um die Größe eines erwarteten Zusammenhanges, Unterschiedes oder einer Veränderung (allgemein: einer Effektgröße) angeben zu können. Sie *sollten* formuliert werden, wann immer die Möglichkeit besteht, für einen Zusammenhang, einen Unterschied oder eine Veränderung eine **Mindestgröße** festzulegen, die für praktisch bedeutsam erachtet wird (vgl. hierzu das Good-enough-Prinzip, ► S. 28 f.).

Spezifische Hypothesen mit Effektgrößen ergänzen das Konzept der statistischen Hypothesenprüfung (Signifikanzkriterium) durch Kriterien der praktischen Bedeutsamkeit von Untersuchungsergebnissen.

! Während eine unspezifische Hypothese nur behauptet, dass ein »irgendwie« gearteter Effekt vorliegt und allenfalls noch die Richtung des Effekts angibt, konkretisiert eine spezifische Hypothese auch den Betrag des Effekts bzw. die Effektgröße.

Die Unterscheidung von Hypothesen mit bzw. ohne vorgegebene Effektgröße beeinflusst die Untersuchungsplanung in einem entscheidenden Punkt: Der »optimale« Stichprobenumfang für eine hypothesenprüfende Untersuchung ist nur kalkulierbar, wenn eine spezifische Hypothese mit Effektgröße formuliert wurde. Begründungen hierfür und einfach zu handhabende Anleitungen zur Festlegung einer Effektgröße sowie zur Bestimmung des für eine spezifische Problematik an-

gemessenen Stichprobenumfangs findet man in ► Abschn. 9.1 und 9.2.

Zweites Kriterium: Gültigkeit der Untersuchungsbefunde

Nachdem entschieden ist, welche der genannten Untersuchungsarten dem jeweiligen Forschungsstand und der Fragestellung angemessen ist, muss aus den zahlreichen Varianten für eine bestimmte Untersuchungsart (von denen die wichtigsten in ► Kap. 6–9 behandelt werden) eine konkrete Variante ausgewählt werden. Ein wichtiges Auswahlkriterium hierfür stellt die Gültigkeit bzw. Aussagekraft der erwarteten Untersuchungsergebnisse dar. Wir unterscheiden hierbei die innere Gültigkeit (interne Validität) und die äußere Gültigkeit (externe Validität) von Untersuchungen (vgl. Campbell, 1957; Campbell & Stanley, 1963a,b; eine kritische wissenschaftstheoretische Diskussion dieses Kriteriums findet man bei Gadenne, 1976; Patry, 1991; Moser, 1986). Wie die folgenden Ausführungen belegen, gelingt es nur selten, beide Gültigkeitskriterien in einer Untersuchung perfekt zu erfüllen. Korrekturen einer Untersuchungsplanung zugunsten der internen Validität wirken sich meistens nachteilig auf die externe Validität aus (und umgekehrt), sodass man sich in der Regel mit einer Kompromisslösung begnügen muss.

- Eine Untersuchung ist **intern valide**, wenn ihre Ergebnisse kausal eindeutig interpretierbar sind. Die interne Validität sinkt mit wachsender Anzahl plausibler Alternativerklärungen für die Ergebnisse.
- Eine Untersuchung ist **extern valide**, wenn ihre Ergebnisse über die besonderen Bedingungen der Untersuchungssituation und über die untersuchten Personen hinausgehend generalisierbar sind. Die externe Validität sinkt mit wachsender Unnatürlichkeit der Untersuchungsbedingungen bzw. mit abnehmender Repräsentativität der untersuchten Stichproben.

Cook und Campbell (1979) ergänzen die interne Validität um einen speziellen Aspekt, den sie **statistische Validität** nennen. Zu kleine Stichproben, ungenaue Messinstrumente, Fehler bei der Anwendung statistischer Verfahren etc. sind Gründe, die die statistische Validität einer Untersuchung in Frage stellen. Ein wichtiger Bestandteil der externen Validität ist zudem die »Kon-

struktvalidität«, die durch unzureichende Explikation der verwendeten Konstrukte bzw. durch ungenaue Operationalisierungen der aus den Konstrukten abgeleiteten Variablen gefährdet ist (► S. 18 f. zum Thema »Korrespondenzproblem«). Westermann (2000, S. 297 f.) spricht in diesem Zusammenhang von »Variablenvalidität«.

Weitere Präzisierungen der Validitätskonzepte gehen auf Campbell (1986) zurück. Statt von Internal Validity spricht Campbell von **Local Molar Causal Validity**, wobei mit »local« die Begrenzung der internen Validität auf eine konkrete Untersuchung zum Ausdruck gebracht werden soll. »Molar« steht in diesem Zusammenhang für die Komplexität des mit einem Treatment verbundenen Wirkprozesses, der aus vielen molekularen Teilwirkungen bestehen kann (kausale Mikromediatoren, ► S. 522); »causal« schließlich weist darauf hin, dass Wirkungen tatsächlich eindeutig auf die geprüfte Behandlung zurückführbar sein müssen.

Die External Validity wird nach Campbell (1986) treffender durch die Bezeichnung **Proximal Similarity** gekennzeichnet. »Similarity« soll in diesem Terminus darauf hinweisen, dass spezifische Untersuchungscharakteristika wie die Untersuchungsteilnehmer, die Untersuchungsanlage, der Untersuchungszeitpunkt sowie die Operationalisierung von Treatment und Wirkungen eine hohe Ähnlichkeit zu Populationen und Situationen aufweisen, für die die Untersuchung gültig sein soll. Mit »proximal« wird betont, dass sich die für Generalisierungen erforderliche Ähnlichkeit auf naheliegende bzw. relevante Untersuchungscharakteristika und nicht auf distale, eher nebensächliche Besonderheiten einer Untersuchung bezieht.

Da diese neuen Bezeichnungen bislang kaum Eingang in die Literatur fanden (Cook & Shadish, 1994), verwenden wir zukünftig – mit der hier vorgenommenen Präzisierung – die klassischen Begriffe »interne Validität« und »externe Validität« zur Charakterisierung des Aussagegehaltes empirischer Untersuchungen.

Im Folgenden werden die beiden wichtigsten untersuchungstechnischen Maßnahmen, die die interne bzw. die externe Validität beeinflussen, dargestellt. Weitere Beeinflussungsgrößen der internen und externen Validität nennen wir auf ► S. 502 ff.

! Interne Validität liegt vor, wenn Veränderungen in den abhängigen Variablen eindeutig auf den Einfluss der unabhängigen Variablen zurückzuführen sind bzw. wenn es neben der Untersuchungshypothese keine besseren Alternativerklärungen gibt.

Externe Validität liegt vor, wenn das in einer Stichprobenuntersuchung gefundene Ergebnis auf andere Personen, Situationen oder Zeitpunkte generalisiert werden kann.

Experimentelle vs. quasiexperimentelle Untersuchung.

Der Unterschied zwischen einer experimentellen und einer quasiexperimentellen Vorgehensweise sei zunächst an einem kleinen Beispiel verdeutlicht. Nehmen wir an, es gehe um den Vergleich von zwei Unterrichtsstilen (z. B. »autoritärer« Unterrichtsstil und »demokratischer« Unterrichtsstil) in Bezug auf die Lernleistungen der Schüler. Für beide Untersuchungsarten würde man Lehrer auswählen, deren Unterrichtsstile überwiegend als »autoritär« oder »demokratisch« bezeichnet werden. Eine quasiexperimentelle Untersuchung liefe nun auf einen Vergleich der Schulklassen dieser Lehrer hinaus, d. h., die Schülerstichproben bestehen aus natürlich vorgefundenen Gruppen mit ihren jeweiligen spezifischen Besonderheiten. Bei einer experimentellen Untersuchung hingegen wird über die Schüler, die ein Lehrer zu unterrichten hat und die nachträglich zu vergleichen sind, nach Zufall entschieden.

Eine quasiexperimentelle Untersuchung vergleicht natürliche Gruppen und eine experimentelle Untersuchung vergleicht zufällig zusammengestellte Gruppen.

Unterschiede zwischen den Gruppen, die in quasiexperimentellen Anordnungen mit natürlichen Gruppen (z. B. Schulklassen) nicht nur hinsichtlich der unabhängigen Variablen (z. B. Art des Unterrichtsstils), sondern zusätzlich hinsichtlich vieler anderer Variablen bestehen können (z. B. Intelligenz, Motivation, sozialer Status), werden in experimentellen Untersuchungen durch die zufällige Aufteilung (**Randomisierung**) minimiert. Der Randomisierung liegt das Prinzip des statistischen Fehlerausgleichs zugrunde, das – hier angewandt – besagt, dass sich die Besonderheiten von Personen in der einen Gruppe durch die Besonderheiten von Personen in der anderen Gruppe ausgleichen bzw. dass es zu einer Neutralisierung **personenbezogener Störvariablen** kommt (ausführlicher hierzu Fisher, 1935).

! Bei experimentellen Untersuchungen werden Untersuchungsobjekte per Zufall in Gruppen eingeteilt (Randomisierung), bei quasiexperimentellen Untersuchungen arbeitet man mit natürlichen Gruppen.

Randomisierung bedeutet nicht, dass jedem Individuum der einen Gruppe ein vergleichbares Individuum der anderen Gruppe zugeordnet wird (**Parallelisierung**,

► S. 526 ff. über Kontrolltechniken für quasiexperimenten-

telles Untersuchungen). Die Äquivalenz beider Gruppen wird bei der Randomisierung statistisch erzielt, denn es ist sehr unwahrscheinlich, dass sich beispielsweise nach einer Zufallsaufteilung in der einen Gruppe nur die klügeren und in der anderen Gruppe die weniger klugen Schüler befinden. Im Durchschnitt sind bei genügender Gruppengröße alle für die Untersuchung potenziell relevanten Variablen in beiden Gruppen annähernd gleich ausgeprägt, d. h., mögliche Gruppenunterschiede in Bezug auf die abhängige Variable (d. h. im Beispiel in Bezug auf die Lernleistung) gehen mit hoher Wahrscheinlichkeit auf die unabhängige Variable (Unterrichtsstil) zurück. (Überlegungen zur Kalkulation der Gruppengröße, die erforderlich ist, um Äquivalenz der zu vergleichenden Gruppen herzustellen, findet man bei Mittring & Hussy, 2004.) Ein solches Untersuchungsergebnis wäre (relativ) eindeutig interpretierbar: Die Untersuchung verfügt über eine hohe interne Validität.

! Durch die Randomisierungstechnik werden bei genügender Gruppengröße personenbezogene Störvariablen neutralisiert.

Anders bei quasiexperimentellen Untersuchungen, bei denen die Untersuchungsteilnehmer den Untersuchungsbedingungen (oder Stufen der unabhängigen Variablen) nicht zufällig zugewiesen werden (oder zugewiesen werden können). Hier besteht die Möglichkeit, dass sich die Vergleichsgruppen nicht nur hinsichtlich der unabhängigen Variablen, sondern zusätzlich hinsichtlich weiterer Merkmale (»Confounder«) systematisch unterscheiden. Ergeben sich in einer quasiexperimentellen Untersuchung Gruppenunterschiede in Bezug auf die abhängige Variable, so sind diese nicht eindeutig auf die unabhängige Variable zurückzuführen: Die Untersuchung verfügt im Vergleich zu einer experimentellen Untersuchung über eine geringere interne Validität.

Experimentelle Gruppen werden durch Manipulation der Untersuchungsbedingungen erzeugt, d. h., die Stufen der unabhängigen Variablen werden durch unterschiedliche Behandlungen von Personen hergestellt (z. B. Gruppe 1 erhält einfache Dosis, Gruppe 2 doppelte Dosis); solche unabhängigen Variablen heißen **experimentelle Variablen** oder **Treatmentvariablen**. Quasiexperimentelle Gruppen werden durch Selektion zusammengestellt, d. h., die Stufen der unabhängigen Variablen

Box 2.3

Hatte das Meistertraining einen Effekt?

Die Firma K. beabsichtigt, die Führungsqualitäten ihrer Meister durch ein Trainingsprogramm zu verbessern. Nachdem Herr W., der als Meister die Abteilung »Ersatzteile« leitet, das Trainingsprogramm absolviert hat, überprüft die Firmenleitung das Betriebsklima, die Arbeitszufriedenheit und die Produktivität dieser Abteilung. (Empirische Untersuchungen, die sich mit der Wirksamkeit von Trainingsprogrammen bzw. Interventionen befassen, bezeichnet man als »Evaluationsstudien«; ▶ Kap. 3.) Die Auswertung der Fragebögen führt zu dem Resultat, dass es in dieser Abteilung keine Gründe für Beanstandungen gibt.

Formal lässt sich diese Untersuchung folgendermaßen beschreiben:

$T \rightarrow M$.

Mit T ist die Schulungsmaßnahme der Firmenleitung gemeint. Der Buchstabe kürzt die Bezeichnung »Treatment« ab, die üblicherweise für experimentelle Eingriffe, Manipulationen oder Maßnahmen verwendet wird. M steht für Messung und symbolisiert in diesem Beispiel die Befragung der Mitarbeiter nach dem Treatment.

Diese »**One-Shot Case Study**« (Cook & Campbell, 1976, S. 96) ist kausal nicht interpretierbar, d. h., die Tatsache, dass es in der Abteilung nichts zu beanstanden gibt, kann nicht zwingend auf die Schulung des Meisters zurückgeführt werden, denn vielleicht gab es ja vorher schon nichts zu beklagen. Um Veränderungen in der Abteilung registrieren zu können, hätte die Abteilung nicht nur nach, sondern auch vor der Schulungsmaßnahme befragt werden müssen. Für dieses »**Ein-Gruppen-Pretest-Posttest-Design**« wird die folgende Charakteristik verwendet:

$M_1 \rightarrow T \rightarrow M_2$.

Nach einer Pretestmessung (M_1) erfolgt das Treatment und danach eine erneute Messung, die Posttestmessung (M_2). Ein Vergleich dieser beiden Messungen liefert Hinweise über mögliche, zwischenzeitlich eingetretene Veränderungen.

Aber auch dieser Plan lässt nicht den zwingenden Schluss zu, die Veränderungen seien ursächlich auf das Meistertraining bzw. das Treatment zurückzuführen. Generell muss bei Untersuchungen von diesem Typus damit gerechnet werden, dass eine Veränderung auftritt, weil

- zwischenzeitliche Einflüsse unabhängig vom Treatment wirksam werden (z. B. eine Lohnerhöhung),
- sich die Untersuchungsteilnehmer unabhängig vom Treatment weiter entwickelten (sie werden z. B. mit ihren Aufgaben besser vertraut),
- allein die Pretestmessung das Verhalten veränderte (die Untersuchungsteilnehmer werden z. B. durch die Befragung auf bestimmte Probleme aufmerksam gemacht),
- das gemessene Verhalten ohnehin einer starken Variabilität unterliegt (z. B. könnten die Arbeitsanforderungen saisonalen Schwankungen unterliegen, die den Effekt des Meistertrainings überlagern),
- oder weil sich die Messungen aus formalstatistischen Gründen verändern können (diese »Regressionseffekte« betreffen vorzugsweise Extremwerte, die bei wiederholten Messungen zur Mitte tendieren; Näheres ▶ S. 554ff.).

Auch dieser Plan lässt also keine eindeutige Interpretation zu.

Ein dritter Plan könnte die »behandelte« Gruppe mit einer nichtbehandelten, nichtäquivalenten Kontrollgruppe vergleichen (nichtäquivalent deshalb, weil die Kontrollgruppe, anders als in rein experimentellen Untersuchungen, natürlich angetroffen wird und nicht per Randomisierung zustande kommt). Diese könnte z. B. aus einer anderen Abteilung bestehen, deren Meister keine Schulung erhielt.



$$\frac{T \rightarrow M_1}{M_2}$$

Man bezeichnet diesen Plan als **Ex-post-facto-Plan**, d. h., die vergleichende Messung wird erst nach erfolgtem Treatment vorgenommen. Auch dieser Plan leidet an schlechter Interpretierbarkeit. Unterschiede zwischen den Vergleichsgruppen sind uneindeutig, da man nicht ausschließen kann, dass sie bereits vor Behandlung der Experimentalgruppe bestanden.

Zuverlässigere Interpretationen ließe ein Plan zu, der wiederholte Messungen bei beiden Gruppen vorsieht (**Kontrollgruppenplan mit Pre- und Posttest**):

$$\begin{array}{l} M_{11} \rightarrow T \rightarrow M_{12} \\ M_{21} \rightarrow M_{22} \end{array}$$

Mit M_{11} und M_{12} werden die Pretest- und Posttestmessungen in der Experimentalgruppe (Gruppe 1 mit Meistertraining) verglichen. Besteht hier ein Unterschied, informiert der Vergleich M_{21} und M_{22} in der Kontrollgruppe (Gruppe 2 ohne Meistertraining) darüber, ob die Differenz $M_{11}-M_{12}$ für einen

Treatmenteffekt spricht oder ob andere Ursachen für die Differenz verantwortlich sind, was zuträfe, wenn die gleiche Veränderung auch in der Kontrollgruppe registriert wird.

Zeigen sich nun in der Experimentalgruppe andere Veränderungen als in der Kontrollgruppe, ist dies noch immer kein sicherer Beleg für die kausale Wirksamkeit des Treatments. Es könnte sein, dass der Effekt darauf zurückzuführen ist, dass der trainierte Meister hauptsächlich jüngere Mitarbeiter anleitet, die den neuen Führungsstil positiv aufnehmen. Ältere Mitarbeiter hätten auf den neuen Führungsstil möglicherweise völlig anders reagiert. Das Alter der Mitarbeiter übt damit einen Einfluss auf die abhängige Variable aus; die Wirkung des Treatments richtet sich danach, mit welcher Altersstufe es kombiniert wird.

Eine höhere interne Validität hätte eine experimentelle Studie, bei der per Zufall entschieden wird, welcher Mitarbeiter zur Experimentalgruppe gehört und welcher zur Kontrollgruppe (Randomisierung). Ob jedoch die Geschäftsführung der Firma K. diese Studie genehmigen würde, muss bezweifelt werden. (Mit diesen und ähnlichen Problemen befassen wir uns in ► Kap. 8.)

werden durch die Auswahl bestimmter Probanden realisiert (z. B. Gruppe 1: 20-jährige Probanden, Gruppe 2: 30-jährige Probanden); solche unabhängigen Variablen heißen **Personenvariablen** oder organismische Variablen.

■ Box 2.3 skizziert einige quasiexperimentelle Untersuchungsvarianten mit unterschiedlicher interner Validität.

Die Frage, ob eine Untersuchung experimentell oder quasiexperimentell angelegt werden sollte, erübrigt sich, wenn eine unabhängige Variable natürlich variiert angetroffen wird und damit vom Untersuchungsleiter durch künstliche »Manipulation« nicht variierbar ist (organismische oder Personenvariablen, wie z. B. biologisches Geschlecht, Nationalität, Schichtzugehörigkeit, Art der Erkrankung etc.). Diese Frage wird jedoch bedeutsam, wenn – wie im oben genannten Beispiel – die unabhängige Variable prinzipiell künstlich variiert werden kann,

aber gleichzeitig auch natürlich variiert angetroffen wird. Vorerst bleibt festzuhalten:

! **Experimentelle Untersuchungen haben eine höhere interne Validität als quasiexperimentelle Untersuchungen.**

Detaillierter werden die Vor- und Nachteile experimenteller bzw. quasiexperimenteller Untersuchungen in ► Kap. 8 in Verbindung mit konkreten Untersuchungsplänen diskutiert. (Weitere Ratschläge für die Anlage quasiexperimenteller Untersuchungen findet man bei Bierhoff & Rudinger, 1996; Bungard et al., 1992; Cook & Campbell, 1976, S. 95 ff.; Heinsman & Shadish, 1996; Shadish et al., 2002.)

Bisher wurden experimentelle bzw. quasiexperimentelle Untersuchungsvarianten nur bezüglich des Kriteriums interne Validität diskutiert. Das zweite Gültigkeitskriterium, die externe Validität, ist von diesem Unter-

scheidungsmerkmal praktisch nicht betroffen, wenn man einmal davon absieht, dass externe Validität ein Mindestmaß an interner Validität voraussetzt.

Felduntersuchungen vs. Laboruntersuchungen. Felduntersuchungen und Laboruntersuchungen markieren die Extreme eines Kontinuums unterschiedlich »lebensnaher« bzw. nach Gottschaldt (1942) »biotischer« Untersuchungen. Felduntersuchungen in natürlich belassenen Umgebungen zeichnen sich meistens durch eine hohe externe Validität und streng kontrollierte Laboruntersuchungen durch eine geringe externe Validität aus.

Felduntersuchungen finden »im Feld« statt, d. h. in einer vom Untersucher möglichst unbeeinflussten, natürlichen Umgebung wie beispielsweise einer Fabrik, einer Schule, einem Spielplatz, einem Krankenhaus usw. Der Vorteil dieser Vorgehensweise liegt darin, dass die Bedeutung der Ergebnisse unmittelbar einleuchtet, weil diese ein Stück unverfälschter Realität charakterisieren (hohe externe Validität). Dieser Vorteil geht allerdings zu Lasten der internen Validität, denn die Natürlichkeit des Untersuchungsfeldes bzw. die nur bedingt mögliche Kontrolle störender Einflussgrößen lässt häufig mehrere gleichwertige Erklärungsalternativen der Untersuchungsbefunde zu.

Laboruntersuchungen werden in Umgebungen durchgeführt, die eine weitgehende Ausschaltung oder Kontrolle von Störgrößen ermöglichen, die potenziell auch die abhängige Variable beeinflussen können. Je nach Art der Untersuchung ist dies z. B. in »laborähnlichen«, spartanisch ausgestatteten und schallisolierten Räumen gewährleistet, in denen der Untersuchungsleiter praktisch jede Veränderung des Umfeldes kontrollieren kann.

Anders als die Randomisierung, die wir als Technik zur Kontrolle personenbezogener Störvariablen kennen gelernt haben, liegt der Vorteil von Laboruntersuchungen in der **Kontrolle untersuchungsbedingter Störvariablen**.

! In Laboruntersuchungen legt man besonderen Wert auf die Kontrolle bzw. Ausschaltung untersuchungsbedingter Störvariablen. Felduntersuchungen finden demgegenüber in »natürlichen«, im Zuge des Forschungsprozesses kaum veränderten Umgebungen statt.

Die strikte Kontrolle untersuchungsbedingter Störvariablen macht Laboruntersuchungen zu Untersuchungen mit hoher interner Validität, in denen sich Veränderungen der abhängigen Variablen mit hoher Wahrscheinlichkeit ursächlich auf die unabhängigen Variablen zurückführen lassen. Die Unnatürlichkeit der Untersuchungsumgebung lässt es allerdings häufig fraglich erscheinen, ob die Ergebnisse auch auf andere, »natürlichere« Situationen generalisierbar sind.

Die Entscheidung, eine Untersuchung als Labor- oder als Felduntersuchung zu konzipieren, kann im Einzelfall erhebliche Schwierigkeiten bereiten. Im Zweifelsfall wird man eine Kompromisslösung akzeptieren müssen, die sowohl die an der externen Validität als auch an der internen Validität orientierten Untersuchungsanforderungen berücksichtigt. Liegen zu einem weit fortgeschrittenen Forschungsgebiet vorwiegend Laboruntersuchungen vor, sodass an der internen Validität der Erkenntnisse kaum noch Zweifel bestehen, sollten die Resultate vordringlich mit Felduntersuchungen auf ihre externe Validität hin überprüft werden. Dominieren in einem gut elaborierten Forschungsgebiet hingegen lebensnahe Feldstudien, deren interne Validität nicht genügend dokumentiert erscheint, sollten vorrangig Überlegungen zur Umsetzung der Fragestellung in Laboruntersuchungen angestellt werden.

Kombinationen. Eine zusammenfassende Bewertung der Untersuchungsvarianten »experimentell vs. quasi-experimentell« und »Feld vs. Labor« führt zu dem Ergebnis, dass bezüglich der Kriterien interne und externe Validität die Kombination »experimentelle Felduntersuchung« allen anderen Kombinationen überlegen ist (■ Tab. 2.1). Dies gilt zumindest für die hypothesenprüfende Forschung und für den Fall, dass alle Kombinationen praktisch gleich gut realisierbar sind und dass der Stand der Forschung keine spezielle Kombination dieser Untersuchungsarten erfordert. Als eine empfehlenswerte Darstellung des »State of the Art« zur experimentellen Felduntersuchung sei Shadish (2002) empfohlen.

Die in ■ Tab. 2.1 wiedergegebenen Untersuchungsvarianten, die sich aus der Kombination der Elemente »experimentell–quasiexperimentell« und »Feld–Labor« ergeben, seien im Folgenden anhand von Beispielen ver-

Tab. 2.1. Kombination der Untersuchungsvarianten »experimentell vs. quasiexperimentell« und »Felduntersuchung vs. Laboruntersuchung«

	Experimentell	Quasiexperimentell
Feld	Interne Validität +	Interne Validität –
	Externe Validität +	Externe Validität +
Labor	Interne Validität +	Interne Validität –
	Externe Validität –	Externe Validität –

deutlicht. Man beachte hierbei, dass die Bewertung einer Untersuchung hinsichtlich der Kriterien interne und externe Validität nicht ausschließlich von den Elementen »experimentell–quasiexperimentell« und »Feld–Labor« abhängt, sondern zusätzlich von anderen untersuchungsspezifischen Merkmalen, die ebenfalls zur Eindeutigkeit der Ergebnisinterpretation bzw. zur Generalisierbarkeit der Ergebnisse beitragen können (► Kap. 8). Zudem sei nochmals darauf hingewiesen, dass mit den Bezeichnungen »Feld vs. Labor« die Extreme eines Kontinuums von Untersuchungen mit unterschiedlicher Kontrolle untersuchungsbedingter Störvariablen bezeichnet sind.

Quasiexperimentelle Felduntersuchung: Weber et al. (1971) untersuchten den Einfluss der Zusammenlegung von Schulen mit weißen und schwarzen Schülern auf das akademische Selbstbild der Schüler. Da die Schüler den Stufen der unabhängigen Variablen (schwarze und weiße Schüler) natürlich nicht per Zufall zugewiesen werden können, handelt es sich um eine quasiexperimentelle Untersuchung. Sie fand zudem in einer natürlichen Umgebung (Schule) statt und ist damit gleichzeitig eine Felduntersuchung.

Experimentelle Felduntersuchung: Eine Untersuchung von Bortz und Braune (1980) überprüfte die Veränderungen politischer Einstellungen durch das Lesen zweier überregionaler Tageszeitungen. Den Untersuchungsteilnehmern wurde per Zufall entweder die eine oder die andere Zeitung für einen begrenzten Zeitraum kostenlos ins Haus gesandt. Diese randomisierte Zuteilung qualifiziert die Untersuchung als eine experimentelle Untersuchung. Darüber hinaus wurde das natürliche Umfeld der Untersuchungsteilnehmer nicht beein-

flusst, d. h., die Untersuchung erfüllt die Kriterien einer Felduntersuchung (zur Thematik »Feldexperiment« vgl. auch Frey & Frenz, 1982).

Quasiexperimentelle Laboruntersuchung: In einer sorgfältig angelegten Laboruntersuchung fragte Thanga (1955) nach Unterschieden in der Fingerfertigkeit männlicher und weiblicher Untersuchungsteilnehmer. Auch hier ist es nicht möglich, die Untersuchungsteilnehmer den beiden Stufen der unabhängigen Variablen (männlich und weiblich) zufällig zuzuweisen, d. h., die Laboruntersuchung ist quasiexperimentell.

Experimentelle Laboruntersuchung: Die experimentelle Laboruntersuchung erfordert randomisierte Versuchsgruppen und eine strikte Kontrolle von untersuchungsbedingten Störvariablen. Sie entspricht damit dem »klassischen« psychologischen Experiment, als dessen Urvater Wundt (1898) gilt. Häufig genannte Kriterien des Experimentes sind

- Planmäßigkeit der Untersuchungsdurchführung (Willkürlichkeit),
- Wiederholbarkeit der Untersuchung,
- Variierbarkeit der Untersuchungsbedingungen (vgl. Selg, 1971, Kap. F).

Weitere Definitionen und Anleitungen zur Durchführung von Experimenten findet man z. B. bei Bredenkamp (1996), Hager (1987), Huber (2000), Hussy und Jain (2004), Krauth (2000), Lür (1987), Sarris und Reiß (2005) oder Sarris (1990, 1992).

Als Beispiel für eine experimentelle Laboruntersuchung mag eine Studie von Issing und Ullrich (1969) dienen, die den Einfluss eines Verbalisierungstrainings auf die Denkleistungen von Kindern überprüfte. Dreißig Kinder wurden per Zufall in eine Experimental- und eine Kontrollgruppe aufgeteilt. Über einen Zeitraum von vier Wochen durften die Kinder in einem eigens für diese Untersuchung hergerichteten Raum altersgemäße Spiele spielen. Die unabhängige Variable stellte eine Instruktion dar, die nur die Experimentalgruppe zum Verbalisieren während des Spielens anregte. Die Untersuchung fand unter kontrollierten Bedingungen bei gleichzeitiger Randomisierung statt – wir bezeichnen sie deshalb als eine experimentelle Laboruntersuchung.

■ Box 2.4 beschreibt ein (nicht ganz ernst zu nehmendes) Beispiel für ein »klassisches Experiment«. Eine

Box 2.4**Experimentelle Überprüfung der Sensibilität von Erbsen**

Die folgende Glosse eines »klassischen« Experiments wurde während der Watergate-Anhörungen, die zum Rücktritt des US-Präsidenten Nixon führten, in der New York Times veröffentlicht (mit einigen Modifikationen übernommen aus Lewin, 1979, S. 17).

Abstract

Während des Sommers 1979 nahmen Wissenschaftler aus Petersham, Mass., die seltene Gelegenheit wahr, den Einfluss der amerikanischen Politik auf das Wachstum von Pflanzen zu überprüfen. In einer Serie sorgfältig kontrollierter Experimente konnte der schlüssige Nachweis erbracht werden, dass Pflanzen es vorziehen, uninformiert zu sein.

Untersuchung

Es wurde eine repräsentative Stichprobe von 200.000 Erbsen (*Pisum sativum*) von geschulten Landarbeiterinnen per Zufall in zwei gleich große Stichproben aufgeteilt. Ein Biologe beaufsichtigte dann die Einpflanzung der Erbsen in ein Treibhaus A und ein Treibhaus B. (Den Pflanzern und dem Biologen war nicht bekannt, welches der beiden Treibhäuser das spätere experimentelle Treatment erhalten sollte.) Klima und Lichtbedingungen waren für beide Treibhäuser gleich.

Die experimentelle Erbsengruppe in Treibhaus A wurde danach sämtlichen Rundfunkübertragungen der Watergate-Anhörungen ausgesetzt. Das Radiogerät beschallte das Treibhaus mit einer Durchschnittslautstärke von 50 dB und stellte sich automatisch ein, wenn Anhörungen übertragen wurden. Die gesamte Übertragungszeit betrug 60.000 Sekunden mit durchschnittlich 16,5 schockierenden Enthüllungen pro Tag über eine Gesamtwachstumszeit von 46 Tagen.

Die Erbsen der Kontrollgruppe in Treibhaus B wurden in denselben Zeiträumen in derselben Lautstärke mit monoton gesprochenen, sinnlosen Silben beschallt.

Ergebnisse

Im Vergleich mit den Kontrollerbbsen keimten die Experimentalerbsen langsamer, sie entwickelten verkümmerte Wurzeln, waren erheblich anfälliger für Schädlinge und gingen insgesamt schneller ein.

Interpretation

Die Ergebnisse des Experiments legen den Schluss nahe, dass sich öffentliche Übertragungen von Debatten der Regierungsadministration nachteilig auf den Pflanzenwuchs in den Vereinigten Staaten auswirken.

kritische Auseinandersetzung mit dem Leistungsvermögen psychologischer Experimente findet man bei Dörner und Lantermann (1991).

2.3.4 Thema der Untersuchung

Nachdem man sich anhand der Literatur Kenntnisse über den Stand der Theoriebildung, über wichtige Untersuchungen und über die bisher eingesetzten Methoden verschafft und die Vorstellungen über die Art der Untersuchung präzisiert hat, müsste es möglich sein, einen Arbeitstitel für das Untersuchungsvorhaben zu finden. Die (vorläufige) Festlegung des

Untersuchungsthemas kann folgende Aufgaben akzentuieren:

- Überprüfung spezieller theoretisch begründeter Hypothesen oder Forschungsfragen,
- Replikation wichtiger Untersuchungen,
- Klärung widersprüchlicher Untersuchungen oder Theorien,
- Überprüfung neuer methodischer oder untersuchungstechnischer Varianten,
- Überprüfung des Erklärungswertes bisher nicht beachteter Theorien,
- Erkundung von Hypothesen.

Die Formulierung des Arbeitstitels sollte den Stellenwert der Untersuchung im Kontext des bereits vorhandenen Wissens möglichst genau wiedergeben. Handelt es sich um ein neues Forschungsgebiet, zu dem kaum Untersuchungen vorliegen, verwendet man für den Arbeitstitel allgemeine Formulierungen, die den Inhalt der Untersuchung global charakterisieren. Zusätzlich kann, wie die folgenden Beispiele zeigen, durch einen Untertitel die verwendete Methodik genannt werden:

- Zur Frage des Einflusses verschiedener Baumaterialien von Häusern auf das Wohlbefinden ihrer Bewohner – Eine Erkundungsstudie
- Bürgernahe Sozialpsychiatrie – Aktionsforschung im Berliner Bezirk Wedding
- Die Scheidung von Eltern aus der Sicht eines Kindes – Eine Einzelfallstudie

Untersuchungen, die einen speziellen Beitrag zu einem Forschungsgebiet mit langer Forschungstradition liefern, werden mit einem eindeutig formulierten, scharf abgrenzenden Titel überschrieben:

- Die Bedeutung von Modelllernen und antezedenten Verstärkern für die Entwicklung der Rollenübernahmefähigkeit von 6- bis 9-jährigen Kindern – Eine quasiexperimentelle Längsschnittstudie
- Vergleichende Analyse exosomatischer und endosomatischer Messungen der elektrodermalen Aktivität in einer Vigilanzsituation – Befunde einer Laboruntersuchung

Die endgültige Formulierung des Untersuchungsthemas wird zweckmäßigerweise erst vorgenommen, wenn die Gesamtplanung abgeschlossen ist. Begrenzte Möglichkeiten bei der Operationalisierung der interessierenden Variablen, bei der Auswahl der Untersuchungseinheiten oder auch zeitliche und kapazitive Limitierungen können ggf. eine Neuformulierung oder eine stärker eingrenzende Formulierung des endgültigen Untersuchungsthemas erfordern.

2.3.5 Begriffsdefinitionen und Operationalisierung

Der Arbeitstitel und die Untersuchungsart legen fest, welche Variable(n) erkundet bzw. als unabhängige und

abhängige Variablen in eine hypothesenprüfende Untersuchung aufgenommen werden sollen. Nach einer vorläufigen Kontrolle der begrifflichen Präzision der Variablenbezeichnungen (► S. 40) legt der folgende Planungsabschnitt eindeutig fest, wie die genannten Variablen in die empirische Untersuchung einzuführen sind (vgl. hierzu auch Westermann, 2000, Kap. 5).

Fragen wir beispielsweise nach Ursachen der Schulangst, muss nun festgelegt werden, was unter »Schulangst« genau zu verstehen ist. Interessiert uns der Einfluss der Einstellung zu einer Arbeit auf die Konzentrationsfähigkeit, erfolgen eine genaue Bestimmung der unabhängigen Variablen »Einstellung zu einer Arbeit« und die Festlegung der Messvorschriften für die abhängige Variable »Konzentrationsfähigkeit«.

Steyer und Eid (1993, S. 2) sprechen in diesem Zusammenhang vom **Überbrückungsproblem** und bezeichnen damit die Aufgabenstellung, theoretische Konstrukte wie Aggressivität, Intelligenz oder Ehrgeiz mit konkreten, empirisch messbaren Variablen zu verbinden. Hinter dem Überbrückungsproblem verbirgt sich also die Frage, wie die alltags- oder wissenschaftssprachlich gefassten Begriffe – ggf. unter Verwendung von Hilfstheorien (vgl. Hager, 1987, Abschn. 3.23) – in Beobachtungs- oder Messvorschriften umgesetzt werden können (Operationalisierung der zu untersuchenden Variablen).

Real- und Nominaldefinitionen

Die geschichtliche Entwicklung der Sprache legte für Objekte, Eigenschaften, Vorgänge und Tätigkeiten Namen fest, die im Verlaufe der individuellen Entwicklung eines Menschen gelernt werden. (Dieses Objekt heißt »Messer«; diese Tätigkeit heißt »laufen« etc.). Derartige Realdefinitionen von Begriffen sollten auf geeignete Beispiele für die zu bezeichnenden Objekte, Eigenschaften, Vorgänge oder Tätigkeiten verweisen. Auf der Basis eines Grundstocks an geordneten, realdefinierten Begriffen kann man Sachverhalte auch in der Weise definieren, dass man die nächsthöhere Gattung (Genus proximum) und den artbildenden Unterschied (Differentia specifica) angibt. Beispiel: »Die Psychologie ist eine empirische Sozialwissenschaft (Genus proximum), die sich mit dem Verhalten und Erleben des einzelnen Menschen, mit Dyaden und Gruppenprozessen beschäftigt (Differentia specifica).« Realdefinitionen haben die

Funktion, ein kommunikationsfähiges, ökonomisches Vokabular zu schaffen.

Die gleiche Funktion haben auf Realdefinitionen aufbauende Nominaldefinitionen, in denen der zu definierende Begriff (Definiendum) mit einer bereits bekannten bzw. real definierten Begrifflichkeit (Definiens) gleichgesetzt wird. Zum Beispiel ließe sich »die Gruppe« als »eine Menge von Personen, die häufig miteinander interagieren« definieren, wenn man davon ausgeht, dass alle Begriffe des Definiens bekannt sind. Ist dies nicht der Fall (weil z. B. der Begriff »Interaktion« nicht definiert ist), wird das Definiens zum Definiendum, was letztlich – bei weiteren unklar definierten Begriffen – zu immer neuen Definitionen bzw. zu einem definitorischen Regress führen kann.

Real- und Nominaldefinitionen können weder wahr noch falsch sein. Mit ihnen wird lediglich eine Konvention oder Regel für die Verwendung einer bestimmten Buchstabenfolge oder eines Zeichensatzes eingeführt. Dies verdeutlichen z. B. verschiedene Sprachen, die für dasselbe Definiendum verschiedene Worte verwenden, ohne dass darunter die Verständigung innerhalb einer Sprache beeinträchtigt wäre.

Bedeutung und Umfang der Wörter einer Sprache sind jedoch nicht generell und für alle Zeiten festgelegt, sondern unterliegen einem allmählichen Wandel. Die Sprache wird durch spezielle Dialekte oder Begriffe anderer Sprachen erweitert, regionale Besonderheiten oder gesellschaftliche Subkulturen verleihen Begriffen eine spezielle Bedeutung, die Entdeckung neuartiger Phänomene oder Sinnzusammenhänge macht die Schöpfung neuer Begriffe oder die Neudefinition alter Begriffe erforderlich, oder die Begriffe verlieren für eine Kultur ihre Bedeutung, weil das mit ihnen Bezeichnete der Vergangenheit angehört. Es resultiert eine Sprache, die zwar eine normale Verständigung ausreichend gewährleistet, die aber für wissenschaftliche Zwecke nicht genügend trennscharf ist.

Die Präzisierung eines Alltagssprachlichen Begriffes für wissenschaftliche Zwecke (z. B. durch eine Bedeutungsanalyse oder die Angabe von Operationalisierungen, ► unten) nennt man **Explication**.

! **Eine Realdefinition legt die Bedeutung eines Begriffes durch direkten Verweis auf konkrete reale Sachverhalte (Objekte, Tätigkeiten etc.) fest.**



Eine Nominaldefinition führt einen neuen Begriff unter Verwendung und Verknüpfung bereits definierter Begriffe ein.

Analytische Definitionen

Die wissenschaftliche Verwendung von Begriffen macht deren Bedeutungsanalyse (Hempel, 1954) bzw. analytische Definitionen erforderlich. Hierbei handelt es sich nicht um Konventionen, die von Wissenschaftlern eingeführt werden, sondern um Aussagen, die empirisch überprüfbar sein sollten.

Mollenhauer (1968, zit. nach Eberhard & Kohlmetz, 1973) definiert beispielsweise das Merkmal »Verwahrlosung« als »eine abnorme, charakterliche Ungebundenheit und Bindungsunfähigkeit, die auf eine geringe Tiefe und Nachhaltigkeit der Gemütsbewegungen und Willensstrebungen zurückgeht und zu einer Lockerung der inneren Beziehung zu sittlichen Werten – wie Liebe, Rücksicht, Verzicht, Opfer, Recht, Wahrheit, Pflicht, Verantwortung und Ehrfurcht – führt«. Vermutlich trifft diese Definition das allgemeine Verständnis von Verwahrlosung; aber ist sie damit bereits empirisch überprüfbar?

Mit der analytischen Definition gibt der Forscher zu verstehen, was er mit einem Begriff bezeichnen will. Er legt damit gewissermaßen »seine Karten auf den Tisch« und macht sein Verständnis des Untersuchungsgegenstandes transparent. Es bleibt nun jedermann überlassen, die analytische Definition nachzuvollziehen oder nicht. Ob sich die Definition jedoch bewährt bzw. ob die Definition richtig oder realistisch ist, zeigt letztlich die spätere Forschungspraxis.

Prinzipiell könnte man es bei der analytischen Definition bewenden lassen. Der Forscher, der beispielsweise den Einfluss familiärer Verhältnisse auf die Verwahrlosung von Jugendlichen untersuchen möchte, nennt seine analytische Definition der zentralen Begriffe und berichtet dann über die Ergebnisse seiner Studie.

Dass diese Vorgehensweise noch nicht befriedigend ist, wird deutlich, wenn wir uns erneut der Mollenhauer'schen Definition von Verwahrlosung zuwenden. Er verwendet dort Begriffe wie »charakterliche Ungebundenheit« und »Bindungsunfähigkeit« oder »Tiefe der Gemütsbewegung« und »Nachhaltigkeit der Willensstrebungen«. Wenngleich man ahnt, was mit diesen Begriffen gemeint sein könnte, bleibt der Wunsch

zu erfahren, was diese Begriffe im Kontext dieser Definition genau besagen sollen. Diese Definition von Verwahrlosung verlangt weitere analytische Definitionen der in ihr verwendeten Begriffe.

Aber auch damit wären noch längst nicht alle Uneindeutigkeiten, die das Verständnis einer Untersuchung beeinträchtigen, ausgeräumt. Es bleibt offen, wie die begrifflichen Indikatoren der Verwahrlosung, die die analytische Definition aufzählt, konkret erfasst werden. Es fehlt die Angabe von Operationen, die zur Erfassung der Variablen »Verwahrlosung« führen. Es fehlt die (bzw. eine) operationale Definition des Begriffes Verwahrlosung.

! Eine analytische Definition klärt einen Begriff durch die Analyse seiner Semantik und seiner Gebrauchsweise (Bedeutungsanalyse). Analytische Definitionen müssen empirisch überprüfbar sein.

Ein besonderer Problemfall ist meist auch die negative analytische Definition. So wird »Telepathie« typischerweise darüber definiert, dass eine Kommunikation zwischen zwei Menschen ohne Beteiligung der bekannten Sinneskanäle stattfindet. Weil bislang jeglicher Anhaltspunkt fehlt, wie eine solche telepathische Kommunikation ablaufen könnte, lässt sich auf diese Negativdefinition kaum eine Theorie aufbauen.

Problematisch sind aber nicht nur Konzepte, die das Vorstellungsvermögen übersteigen, sondern gerade auch solche, die allzu vertraut sind. Der Zustand, in dem eine andere Person von uns herausragend positiv bewertet wird, der unsere Gedanken völlig beherrscht und der äußerst angenehme körperliche Reaktionen auslöst, wurde von Tennov (1979) als »Limerence« bezeichnet. Durch diese Wortneuschöpfung versuchte die Autorin, dem analytisch definierten Konstrukt Eindeutigkeit zu sichern, die gefährdet wäre, wenn mit ideologisch aufgeladenen Bezeichnungen wie »Verliebtheit« oder »Liebe« operiert wird.

Tatsächlich ist es ein durchgehendes Problem der Sozial- und Humanwissenschaften, ihre sorgfältig definierten, analytischen Begriffe gegen das eher »schwammige« Alltagsverständnis zu verteidigen. Schließlich will man in der Regel eben doch keine neue Sprache erfinden, sondern die bereits etablierten Bezeichnungen gebrauchen – allerdings mit sehr klar abgesteckten Bedeutungsweisen.

Während in der empirisch-quantitativen Forschung die im Folgenden erläuterten Operationalisierungen von zentraler Bedeutung sind, weil sie die Grundlage von Variablenmessungen bilden, nehmen in der qualitativen Forschung ausgedehnte Bedeutungsanalysen großen Raum ein, die ergründen, welchen Sinn Individuen und Gruppen bestimmten Begriffen geben, wie sie mit diesen Begriffen operieren und welche Konsequenzen diese Praxis für die Akteure hat.

Operationale Definitionen

Der Begriff »operationale Definition« (oder Operationalisierung eines Merkmals) geht auf Bridgman (1927) zurück. Die ursprüngliche, auf die Physik zugeschnittene Fassung lässt sich in folgender Weise zusammenfassen:

1. Die operationale Definition ist synonym mit einem korrespondierenden Satz von Operationen. (Der Begriff »Länge« beinhaltet nicht mehr und nicht weniger als eine Reihe von Operationen, mit denen eine Länge ermittelt wird.)
2. Ein Begriff sollte nicht bezüglich seiner Eigenschaften, sondern bezüglich der mit ihm verbundenen Operationen definiert werden.
3. Die wahre Bedeutung eines Begriffes findet man nicht, indem man beobachtet, was man über ihn sagt, sondern indem man registriert, was man mit ihm macht.
4. Unser gesamtes Wissen ist an den Operationen zu relativieren, die ausgewählt wurden, um unsere wissenschaftlichen Konzepte zu messen. Existieren mehrere Sätze von Operationen, so liegen diesen auch mehrere Konzepte zugrunde.

In seinen Arbeiten von 1945 und 1950 erweiterte Bridgman diese Fassung in einer vor allem für die Sozialwissenschaften bedeutsamen Weise, indem er z. B. nicht nur physikalische, sondern auch geistige und »Paper-and-Pencil-Operationen« zuließ. Jedoch erhielt die operationale Definition zahlreiche verwirrende und einander teilweise widersprechende Auslegungen, die beispielsweise F. Adler zu der in **Box 2.5** wiedergegebenen Karikatur veranlassten.

So sagt beispielsweise die häufig zitierte Behauptung, Intelligenz sei das, was Intelligenztests messen (Boring, 1923) zunächst nichts aus, auch wenn die zur Messung

Box 2.5

Über Sinn und Unsinn operationaler Definitionen

Als zynischen Beitrag zu der häufig zitierten operationalen Definition des Begriffes »Intelligenz« (Intelligenz ist das, was ein Intelligenztest misst; ▶ unten) entwickelte Adler (1947) den folgenden, hier leicht abgewandelten Test zur Messung der Fähigkeit » C_n «:

1. Wieviele Stunden haben Sie in der vergangenen Nacht geschlafen? ...
2. Schätzen Sie die Länge Ihrer Nase in Zentimetern und multiplizieren Sie diesen Wert mit 2. ...
3. Mögen Sie gefrorene Leber (notieren Sie +1 für Ja und -1 für Nein). ...
4. Wieviele Meter hat eine Seemeile? (Falls Sie es nicht wissen, nennen Sie den Wert, der Ihnen am wahrscheinlichsten erscheint.) ...

5. Schätzen Sie die Anzahl der Biergläser, die der Erfinder dieses Tests während seiner Erfindung getrunken hat. ...

Addieren Sie nun die oben notierten Werte. Die Summe stellt Ihren C_n -Wert dar. Sie verfügen über eine hohe C_n -Fähigkeit, wenn Ihre Punktzahl ...

Kommentar

Abgesehen von der Präzision der durchzuführenden mentalen Operationen und der quantitativen Auswertung ist der C_n -Test purer Unsinn. Die Behauptung, C_n -Fähigkeit sei das, was der C_n -Test misst, ist absolut unbefriedigend, solange die C_n -Fähigkeit nicht zuvor analytisch definiert wurde. Operationale Definitionen sind analytischen Definitionen nachgeordnet und damit für sich genommen bedeutungslos.

der Intelligenz vorgeschriebenen Operationen in einem Intelligenztest präzise festgelegt sind. Erst durch eine Bedeutungsanalyse bzw. eine analytische Definition des Begriffes Intelligenz kann nachvollzogen werden, ob die gewählten operationalen Indikatoren sinnvoll sind.

Nach Wechsler et al. (1964, S. 13) ist Intelligenz »die zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen«. Diese zunächst noch recht globale Begriffsbestimmung wird dann im Weiteren Begriff für Begriff näher ausgeführt. Sie endet mit der Aufzählung konkreter Einzelfähigkeiten wie »allgemeines Wissen und allgemeines Verständnis, rechnerisches Denken oder räumliches Vorstellungsvermögen«. Erst auf dieser schon sehr konkreten Ebene der Begriffsbestimmung setzt die Operationalisierung der einzelnen postulierten Teilmerkmale der Intelligenz ein. Dies sind dann präzise formulierte Aufgaben mit vorgegebenen Antwortmöglichkeiten, die zu zehn Untertests zusammengefasst den gesamten Intelligenztest ergeben (in diesem Beispiel den Hamburg-Wechsler-Intelligenztest für Erwachsene oder kurz HAWIE; vgl. hierzu auch Tewes, 1991).

Sind diese analytischen Definitionen bekannt, so macht es durchaus Sinn, Intelligenz als das, was der HAWIE misst, zu definieren. Unabhängig hiervon können andere Wissenschaftlerinnen und Wissenschaftler der Intelligenz anderslautende analytische Definitionen geben (was auch tatsächlich geschieht), die ihrerseits eigene operationale Definitionen erfordern. Welche der konkurrierenden operationalen Intelligenzdefinitionen oder Intelligenztests »richtig« sind, kann gegenwärtig nicht entschieden werden. Ihre Brauchbarkeit hängt letztlich davon ab, wie sich die einzelnen Verfahren in der Praxis bewähren.

! Eine operationale Definition standardisiert einen Begriff durch die Angabe der Operationen, die zur Erfassung des durch den Begriff bezeichneten Sachverhaltes notwendig sind, oder durch Angabe von messbaren Ereignissen, die das Vorliegen dieses Sachverhaltes anzeigen (Indikatoren).

Probleme der Operationalisierung. Eine operationale Definition setzt grundsätzlich eine ausführliche Bedeutungsanalyse des zu definierenden Begriffes voraus. Diese hat eventuell bereits vorliegende wissenschaftliche Auseinandersetzungen mit dem Begriff zu berücksichti-

gen. Den Begriff »Frustration« als »ein unangenehmes Schuldgefühl, das sich bei Misserfolgen einstellt« zu definieren, wäre sicherlich nicht falsch; die Definition übersieht jedoch, dass aus Motivationstheorien bereits einiges über Frustration bekannt ist.

Aber auch präzise analytische Definitionen lassen häufig verschiedenartige Operationalisierungen zu. Frustration ist ein sehr allgemeiner Begriff, der bei Kindern beispielsweise dadurch operationalisiert werden kann, dass man ihnen interessantes Spielzeug zeigt, ohne sie damit spielen zu lassen, dass man ihnen versprochene Belohnungen vorenthält, dass man ihre Freizeit stark reglementiert etc.

Hierbei spielt es keine Rolle, ob die mit einem Begriff gekennzeichnete Variable in einer hypothesenprüfenden Untersuchung als unabhängige Variable oder als abhängige Variable eingesetzt wird. Handelt es sich um eine unabhängige Variable, die vom Untersuchungsleiter manipuliert werden kann, so genügt es häufig, nur eine Ausprägung der unabhängigen Variablen experimentell herzustellen, deren Bedeutung durch Vergleich mit einer Kontrollgruppe eruiert wird (z. B. eine frustrierte Kindergruppe im Vergleich zu einer nichtfrustrierten Gruppe). Bei der Operationalisierung der abhängigen Variablen ist hingegen darauf zu achten, dass diese in möglichst differenzierten Abstufungen gemessen werden kann (► Abschn. 2.3.6 über messtheoretische Probleme).

Die Bedeutungsanalyse eines Begriffes, der eine Variable charakterisiert, schreibt selten zwingend vor, wie der Begriff zu operationalisieren ist. Sixtl (1993, S. 24) verwendet in diesem Zusammenhang das Bild von einem »ausgeleierten Schloß«, das sich von vielen Schlüsseln öffnen lässt. Dieser scheinbare Nachteil kann jedoch für eine weitergehende Bedeutungsanalyse fruchtbar gemacht werden. Führen nämlich verschiedene Operationalisierungen desselben Begriffes zu widersprüchlichen Resultaten, dann ist der Begriff offensichtlich noch nicht präzise genug analysiert (vgl. hierzu auch Schnell et al., 1999, S. 73 ff.). Das Nebeneinander verschiedener, einander widersprechender Operationalisierungen ist daher immer ein sicherer Hinweis darauf, dass sich die Operationalisierungen auf verschiedene Begriffe beziehen und dass damit eine präzisere Bedeutungsanalyse erforderlich ist (hierzu auch Punkt 4 der Position Bridgmans, ► S. 62).

Operationale und analytische Definitionen tragen wechselseitig zu ihrer Präzisierung bei. Wiederum entscheidet der Stand der Forschung über die Genauigkeit der analytischen Definition eines Begriffes und damit auch über die Eindeutigkeit einer Operationalisierung.

Operationalisierungsvarianten. Eine abhängige Variable sollte sensibel und reliabel auf die durch das Treatment bzw. die unabhängige Variable ausgelösten Effekte reagieren. Hierfür sind die im Folgenden genannten fünf Operationalisierungsvarianten (nach Conrad & Maul, 1981, S. 151) besonders geeignet:

- **Häufigkeit:** Wie oft tritt ein bestimmtes Verhalten auf? (Beispiel: Anzahl der Fehler in einem Diktat, Häufigkeit der Blickkontakte, Häufigkeit von Sprechpausen)
- **Reaktionszeit:** Wieviel Zeit vergeht, bis eine Person nach Auftreten eines Stimulus reagiert? (Beispiel: Reaktionslatenz nach Auftreten eines unerwarteten Verkehrshindernisses, Reaktionszeit bis zum Deuten einer Rorschach-Tafel)
- **Reaktionsdauer:** Wie lange reagiert eine Person auf einen Stimulus? (Beispiel: Lösungszeit für eine Mathematikaufgabe, Verweildauer des Auges auf einem bestimmten Bildausschnitt)
- **Reaktionsstärke:** Wie intensiv reagiert eine Person auf einen Stimulus? (Beispiel: Stärke der Muskelanspannung als Indikator für Aggressivität, Rating-skalen, Schreibdruck)
- **Wahlreaktionen:** Welche Wahl trifft eine Person angesichts mehrerer Wahlmöglichkeiten? (Beispiel: Paarvergleichsurteil, Mehrfachwahlaufgaben, Präferenzurteile; ► Abschn. 4.2.2, 4.2.3, 4.3.5).

Die Art der Operationalisierung entscheidet über das Skalierungsniveau der abhängigen Variablen (► Abschn. 2.3.6), das seinerseits bestimmt, wie das Merkmal statistisch auszuwerten ist bzw. welcher Signifikanztest zur Hypothesenprüfung herangezogen werden sollte. Üblicherweise wird man sich um kardinalskalierte abhängige Variablen bemühen bzw. die Operationalisierung so anlegen, dass keine triftigen Gründe gegen die Annahme mindestens einer Intervallskala sprechen (► S. 68). In diesem Sinne unproblematisch dürften die ersten vier Operationalisierungsvarianten sein; sind als abhängige Variablen Wahlreaktionen vorgesehen, so

helfen ggf. die in ► Abschn. 4.2.2 genannten Skalierungsverfahren zur Entwicklung einer Intervallskala bzw. die auf ► S. 215 behandelten Mehrfachwahlaufgaben weiter.

2.3.6 Messtheoretische Probleme

Mit Fragen der Operationalisierung sind messtheoretische Probleme verknüpft. Ist – wie in den meisten Fällen – eine statistische Auswertung der Untersuchungsergebnisse erforderlich (für hypothesenprüfende Untersuchungen stehen hierfür die Methoden der Inferenzstatistik und für erkundende Untersuchungen die Methoden der deskriptiven Statistik zur Verfügung), so sollte in der Planungsphase geklärt werden, wie die zu untersuchenden Merkmale quantifiziert bzw. gemessen werden sollen. ► Kap. 4 (Datenerhebung) fasst die wichtigsten, in den Human- und Sozialwissenschaften gebräuchlichen quantitativen Messmethoden zusammen. Die messtheoretische Einschätzung der dort beschriebenen Verfahren sowie die Auswahl geeigneter statistischer Auswertungsmethoden setzen ein Mindestmaß an messtheoretischen Kenntnissen voraus, die im Folgenden vermittelt werden.

Für diejenigen, die sich mit dieser anspruchsvollen Materie ausführlicher befassen wollen, seien die Arbeiten von Coombs et al. (1975), Gigerenzer (1981), Niederee und Narens (1996), Orth (1974, 1983), Pfanzagl (1971), Roberts (1979), Steyer und Eid (1993) oder Suppes und Zinnes (1963) empfohlen.

Was ist Messen?

Messen wird in der Alltagssprache meistens mit physikalischen Vorstellungen in Verbindung gebracht. Dabei bezeichnet man als **fundamentale Messungen** das Bestimmen einer (Maß-)Zahl als das Vielfache einer Einheit (z. B. Messungen mit einem Zollstock oder einer Balkenwaage). Für derartige Messungen ist der Begriff »Einheit« zentral. Man wählt hierfür eine in der Natur vorgegebene Größe (wie z. B. die Ladung eines Elektrons als Einheit des Merkmals »elektrische Ladung«) oder man legt aus Gründen der Zweckmäßigkeit willkürlich eine Größe als Normeinheit fest (z. B. der in Paris niedergelegte »Archivmeter«). Eine physikalische Messung besteht darin, möglichst genau zu erfassen, wie oft die gewählte Merkmals-einheit in dem zu messenden Objekt enthalten ist.

Eine Übertragung dieser Messvorstellung auf die Sozialwissenschaften scheitert daran, dass »Einheiten« in diesem Sinne in den Sozialwissenschaften bislang fehlen. Dennoch sind auch hier – allerdings mit einer weiter gefassten Messkonzeption – Messoperationen möglich.

Allgemein formuliert besteht eine Messoperation im Zuordnen von Zahlen zu Objekten. Die logisch-mathematische Analyse dieser Zuordnungen und die Spezifizierung von Zuordnungsregeln sind Aufgaben der Messtheorie. Die wichtigsten hierbei zu lösenden Probleme betreffen

- die Repräsentation empirischer Objektrelationen durch Relationen der Zahlen, die den Objekten zugeordnet werden,
- die Eindeutigkeit der Zuordnungsregeln,
- die Bedeutsamkeit der mit Messvorgängen verbundenen numerischen Aussagen.

Diese drei Problembereiche seien kurz erläutert (ausführlicher hierzu Orth, 1983; Steyer & Eid, 1993).

Repräsentationsproblem. Zur Darlegung des Repräsentationsproblems gehen wir von einem **empirischen Relativ** (oder Relationensystem) aus, das aus einer Menge von Objekten sowie einer oder mehreren Relationen besteht, welche die Art der Beziehung der Objekte untereinander charakterisieren. Dieses empirische Relativ wird in ein **numerisches Relativ** abgebildet, deren Zahlen so geartet sein müssen, dass sie die Objektrelationen des empirischen Relativs korrekt repräsentieren. Eine Abbildung mit dieser Eigenschaft bezeichnet man als **homomorph** bzw. strukturerhaltend.

Ein empirisches Relativ, ein numerisches Relativ sowie eine die beiden Relative homomorph verknüpfende Abbildungsfunktion konstituieren eine **Skala**. Als Antwort auf die oben gestellte Frage: »Was ist Messen?« formulieren wir nach Orth (1983, S. 138):

! **Messen ist eine Zuordnung von Zahlen zu Objekten oder Ereignissen, sofern diese Zuordnung eine homomorphe Abbildung eines empirischen Relativs in ein numerisches Relativ ist.**

Beispiel: Ein empirisches Relativ möge aus 10 Tennisspielern sowie der zweistelligen Relation »spielerische Überlegenheit« bestehen. Die spielerische Überlegen-

heit wird in einem Turnier »Jeder gegen Jeden« ermittelt. Ein Spieler i sei einem Spieler j überlegen, wenn er diesen schlägt. Dieser Sachverhalt wird durch $i > j$ (i schlägt j) zum Ausdruck gebracht. Den 10 Spielern sind nun in der Weise Zahlen $\phi(i)$, $\phi(j)$, $\phi(k)$, ... zuzuordnen, dass für jedes Spielerpaar mit $i > j$ die Zahlenrelation $\phi(i) > \phi(j)$ gilt. Die so resultierende Skala heißt Rang- bzw. Ordinalskala (► S. 67 f.).

Wenn man unterstellt, dass das Merkmal »Spielstärke« kontinuierlich ist, die 10 Spieler auf diesem Kontinuum unterschiedliche Positionen einnehmen und diese »wahre« Spielstärke allein über den Ausgang eines jeden Spieles entscheidet, wären die Rangzahlen 1 (schlechtester Spieler) bis 10 (bester Spieler) geeignet, das empirische Relativ homomorph bzw. strukturerhaltend abzubilden.

Man bedenke jedoch, dass aus $i > j$ und $j > k$ nicht unbedingt $i > k$ folgen muss, denn ein dem Spieler j unterlegener Spieler k könnte durchaus Spieler i schlagen ($k > i$), auch wenn Spieler i seinerseits Spieler j besiegt hat. Die Abbildung der Objekte i , j und k mit $\phi(i)=3$, $\phi(j)=2$ und $\phi(k)=1$ wäre in diesem Falle nicht strukturerhaltend, weil die empirische Relation $k > i$ der numerischen Relation $\phi(k) < \phi(i)$ widerspricht.

Die Messbarkeit eines Merkmals ist also an Bedingungen (Axiome) geknüpft, die im empirischen Relativ erfüllt sein müssen. Diese Bedingungen werden in einem **Repräsentationstheorem** zusammengefasst, das die Existenz einer Skala behauptet, wenn diese Bedingungen erfüllt sind. In unserem Beispiel wäre das sog. Transitivitätsaxiom verletzt, wenn für eine beliebige Dreiergruppe von Spielern $i > j$ und $j > k$, aber nicht $i > k$ gilt. (Auf die Möglichkeit äquivalenter Spielstärken gehen wir auf ► S. 67 ein.)

! Unter einer Skala versteht man ein empirisches Relativ, ein numerisches Relativ und eine die beiden Relative verknüpfende, homomorphe Abbildungsfunktion. Die Messbarkeit eines Merkmals bzw. die Konstruierbarkeit einer Skala ist an Bedingungen geknüpft.

Eindeutigkeitsproblem. Mit dem Eindeutigkeitsproblem verbindet sich die Frage, ob sich die Abbildungsfunktion ϕ so in eine andere Abbildungsfunktion ϕ' transformieren lässt, dass die Eigenschaften der Skala erhalten bleiben. Die Lösung des Eindeutigkeitsproblems besteht

dann in der Angabe von Transformationen, gegenüber denen die Skaleneigenschaften invariant sind. Man sagt, eine Messung sei eindeutig bis auf die in diesem Sinne zulässigen Transformationen der ursprünglichen Skala.

Im Beispiel wurden den 10 Tennisspielern die Rangzahlen 1–10 zugeordnet. Sind die Bedingungen für eine Ordinalskala erfüllt, ist davon auszugehen, dass ein Spieler mit einer höheren Zahl einen Spieler mit einer niedrigeren Zahl besiegt. Dieser Informationsgehalt bliebe erhalten, wenn man zu den Rangzahlen 1–10 eine konstante Zahl addiert, wenn man sie mit einer konstanten Zahl c ($c > 0$) multipliziert oder wenn man sie so verändert, dass die Größer-kleiner-Relationen zwischen den ursprünglichen Rangzahlen nicht verändert werden. Transformationen mit dieser Eigenschaft bezeichnet man allgemein als **monotone Transformationen**, sodass wir formulieren können: Messungen auf einer Rang- oder Ordinalskala sind eindeutig bis auf hier zulässige monotone Transformationen.

Bedeutsamkeitsproblem. Unter dem Stichwort Bedeutsamkeit wird gefragt, welche mathematischen Operationen mit den erhobenen Messungen sinnvoll sind. Dass die Beantwortung dieser Frage von der Lösung des Eindeutigkeitsproblems abhängt, lässt sich an unserem Beispiel leicht verdeutlichen: Weder die Aussage: »Spieler i ist doppelt so spielstark wie Spieler j « noch die Aussage: »Spieler i und j unterscheiden sich in ihrer Spielstärke in gleicher Weise wie die Spieler k und l « ist wegen der für Rangskalen zulässigen monotonen Transformation sinnvoll. Addieren wir zu den Rangzahlen 1 und 2 z. B. den Wert 100, bleibt die Größer-Kleiner-Relation zwar erhalten ($101 < 102$); das Verhältnis der Zahlen zueinander hat sich jedoch drastisch verändert. Dass der Vergleich von Spielstärkeunterschieden keinen Sinn macht, verdeutlichen folgenden Zahlen: Die Messungen $\phi(i)=1$, $\phi(j)=3$, $\phi(k)=7$, $\phi(l)=9$ könnten vermuten lassen, dass der Unterschied zwischen i und j genauso groß sei wie der Unterschied zwischen k und l . Da es sich hierbei jedoch um Messungen auf einer Rangskala handelt, sind monotone Transformationen zulässig wie z. B.

$$\begin{array}{ll} \phi'(i) = 1,1 & \text{oder } \phi''(i) = 1,2 \\ \phi'(j) = 2,8 & \text{'' } \phi''(j) = 3,8 \\ \phi'(k) = 6,9 & \text{'' } \phi''(k) = 7,9 \\ \phi'(l) = 9,3 & \text{'' } \phi''(l) = 8,1 \end{array}$$

Bei beiden Transformationen sind die Größer-kleiner-Relationen unverändert; die Spielstärkeunterschiede variieren jedoch beträchtlich: Bei der ersten Transformation wäre der Unterschied zwischen i und j kleiner und bei der zweiten Transformation größer als der Unterschied zwischen k und l . Messwertdifferenzen (oder auch Summen oder Mittelwerte) machen also bei Rangskalen keinen Sinn.

Allgemein sagen wir, dass eine numerische Aussage dann »bedeutsam« ist, wenn sie sich unter den für eine Skala zulässigen Transformationen nicht verändert. Bei Rangzahlen sind nur diejenigen statistischen Verfahren zulässig, die lediglich die Größer-kleiner-Relation der Messungen nutzen.

Eine kritische Analyse der Bedeutsamkeitsproblematik, auch im Hinblick auf die im Folgenden zu behandelnden Skalenarten, findet man bei Niederée und Mausfeld (1996a,b).

Skalenarten

Es werden nun die vier wichtigsten Skalenarten vorgestellt. Dabei werden die für eine Skalenart jeweils gebräuchlichste Messstruktur sowie die Art ihrer Repräsentation im numerischen Relativ kurz beschrieben. Auf eine Behandlung der Axiomatik der Skalen wird unter Verweis auf die bereits erwähnte Spezialliteratur (► S. 65) verzichtet. Ferner werden Eindeutigkeit und Bedeutsamkeit der Skala diskutiert. Die Behandlung der Skalen erfolgt hierarchisch, beginnend mit einfachen, relativ ungenauen Messungen bis hin zu exakten, vor allem in den Naturwissenschaften gebräuchlichen Messungen.

Nominalskala. Eine Nominalskala setzt ein empirisches Relativ mit einer gültigen **Äquivalenzrelation** voraus. Äquivalente Objekte bzw. Objekte mit identischen Merkmalsausprägungen erhalten identische Zahlen, und Objekte mit verschiedenen Merkmalsausprägungen erhalten verschiedene Zahlen.

! Eine Nominalskala ordnet den Objekten eines empirischen Relativs Zahlen zu, die so geartet sind, dass Objekte mit gleicher Merkmalsausprägung gleiche Zahlen und Objekte mit verschiedener Merkmalsausprägung verschiedene Zahlen erhalten.

Ein empirisches Relativ mit einer gültigen Äquivalenzrelation bezeichnet man als eine **klassifikatorische Messstruktur**. Die Auswahl der Zahlen, die den Objektklassen zugeordnet werden, ist für eine Nominalskala unerheblich, solange gewährleistet ist, dass äquivalente Objekte durch identische und nichtäquivalente Objekte durch verschiedene Zahlen abgebildet werden. Vier verschiedenen Parteien könnten also die Zahlen 1, 2, 3 und 4 zugeordnet werden oder auch andere Zahlen wie z. B. 2, 6, 5 und 1. Wir sagen: Die quantitativen Aussagen einer Nominalskala sind gegenüber beliebigen eindeutigen Transformationen invariant.

Unter dem Gesichtspunkt der Bedeutsamkeit ist wegen der für Nominalskalen zulässigen **Eindeutigkeitstransformation** festzustellen, dass nur Aussagen über die Besetzungszahlen bzw. Häufigkeiten für Objektklassen bedeutsam sind. Dementsprechend beschränken sich mathematisch-statistische Operationen für Nominaldaten auf die Analyse von Häufigkeitsverteilungen (vgl. hierzu z. B. Bortz, 2005, Kap. 5.3). Klassifikatorische Begriffe spielen in der qualitativen Forschung eine zentrale Rolle (► Abschn. 5.1.1).

Ordinalskala. Eine Ordinalskala erfordert ein empirisches Relativ, für deren Objektmenge eine sog. **schwache Ordnungsrelation** gilt. Dies bedeutet, dass bei einem beliebigen Objektpaar a und b entscheidbar sein muss, welches Objekt über das andere bezüglich eines untersuchten Kriteriums dominiert, oder ob beide Objekte äquivalent sind. Ferner ist die bereits erwähnte Transitivität gefordert, nach der bei Dominanz von a über b und bei Dominanz von b über c das Objekt a auch über c dominieren muss. Dominiert ein Objekt a über ein Objekt b , so erhält das Objekt a eine Zahl, die größer ist als die dem Objekt b zugeordnete Zahl. Sind Objekte äquivalent, erhalten sie eine identische Zahl.

! Eine Ordinalskala (Rangskala) ordnet den Objekten eines empirischen Relativs Zahlen zu, die so geartet sind, dass von jeweils zwei Objekten das dominierende Objekt die größere Zahl erhält. Bei Äquivalenz sind die Zahlen identisch.

Einer Ordinalskala ist die Rangfolge der untersuchten Objekte bezüglich eines Dominanzkriteriums zu entnehmen (z. B. Beliebtheit von Schülern, gesellschaftliches Prestige von Berufen, Verwerflichkeit von Straf-

delikten). Eine Ordinalskala wird deshalb auch Rangskala genannt, wobei äquivalente Objekte sog. Verbundränge erforderlich machen (► S. 155 f.).

Messungen auf einer Ordinalskala sind eindeutig bis auf hier zulässige **monotone Transformationen**, also Transformationen, durch die die Größer-kleiner-Relationen der Objektmessungen nicht verändert werden (rangerhaltende Transformation). Dementsprechend sind diejenigen quantitativen Aussagen bedeutsam, die gegenüber monotonen bzw. rangerhaltenden Transformationen invariant sind. Die statistische Analyse von Ordinaldaten läuft also auf die Auswertung von Ranginformationen hinaus, über die z. B. bei Bortz et al. (2000, Kap. 6 bzw. Abschnitt 8.2) oder bei Bortz und Lienert (2003, Kap. 3 bzw. Abschnitt 5.2) berichtet wird.

Intervallskala. Eine Intervallskala erfordert ein empirisches Relativ, für das eine schwache Ordnungsstruktur der Dominanzrelationen aller Objektpaare gilt. Anders als bei einer Ordinalskala, bei der die Frage, wie stark ein Objekt über ein anderes dominiert, unerheblich ist, wird hier also gefordert, dass die paarweisen Dominanzrelationen nach ihrer Stärke in eine Rangordnung gebracht werden können. Interpretieren wir eine Dominanzrelation für a und b als Merkmalsunterschied zwischen den Objekten a und b, dann impliziert die Existenz einer schwachen Ordnungsrelation der Objektpaare, dass die Größe des Unterschiedes bei jedem Objektpaar bekannt ist.

Dieses empirische Relativ wird mit dem numerischen Relativ durch folgende Zuordnungsfunktion verknüpft: Wenn der Unterschied zwischen zwei Objekten a und b mindestens so groß ist wie der Unterschied zwischen zwei Objekten c und d, ist die Differenz der den Objekten a und b zugeordneten Zahlen $\phi(a) - \phi(b)$ mindestens so groß wie die Differenz der den Objekten c und d zugeordneten Zahlen $\phi(c) - \phi(d)$.

! **Eine Intervallskala ordnet den Objekten eines empirischen Relativs Zahlen zu, die so geartet sind, dass die Rangordnung der Zahlendifferenzen zwischen je zwei Objekten der Rangordnung der Merkmalsunterschiede zwischen je zwei Objekten entspricht.**

Für eine Intervallskala gilt, dass gleich große Merkmalsunterschiede durch äquidistante Zahlen abgebildet wer-

den, d. h., identische Messwertunterschiede zwischen Objektpaaren entsprechen identischen Merkmalsunterschieden. Hieraus folgt, dass Zahlenintervalle wie z. B. 1 bis 2, 2 bis 3, 3 bis 4 etc. gleich große Merkmalsunterschiede abbilden.

Ein Beispiel für eine Intervallskala ist die Celsius-Skala. Der Temperaturunterschied zwischen 2°C und 4°C ist genauso groß wie z. B. der Temperaturunterschied zwischen 3°C und 5°C, und die Intervalle 1°C bis 2°C, 2°C bis 3°C, 3°C bis 4°C etc. bilden gleich große Temperaturunterschiede ab. Man beachte, dass vergleichbare Aussagen für Ordinalskalen nicht gültig sind.

Eine Intervallskala ist eindeutig bis auf für sie zulässige **lineare Transformationen**: $\phi' = \beta \cdot \phi + \alpha$ ($\beta \neq 0$). Durch β und α werden die Einheit und der Ursprung der Intervallskala im numerischen Relativ festgelegt. Die Celsius-Skala beispielsweise wird durch folgende lineare Transformation in die Fahrenheit-Skala (F) überführt:

$$F = \frac{9}{5}C + 32.$$

Auch die Fahrenheit-Skala bildet identische Temperaturunterschiede durch äquidistante Zahlenintervalle ab.

Bei einer Intervallskala ist die Bedeutung einer numerischen Aussage gegenüber linearen Transformationen invariant. Dies gilt für Differenzen, Summen bzw. auch Mittelwerte von intervallskalierten Messwerten. Die am häufigsten eingesetzten statistischen Verfahren gehen von intervallskalierten Daten aus.

Verhältnisskala. Im empirischen Relativ einer Verhältnisskala sind typischerweise neben einer schwachen Ordnungsrelation der Objekte **Verknüpfungsoperationen** definiert wie z. B. das Aneinanderlegen zweier Bretter oder das Abwiegen von zwei Objekten in einer Waagschale. Dem Verknüpfungsoperator entspricht im numerischen Relativ die Addition.

Bei Merkmalen wie Länge oder Gewicht, auf die der Verknüpfungsoperator sinnvoll angewendet werden kann, sind Aussagen wie: »Durch das Zusammenlegen zweier Bretter a und b resultiert eine Brettlänge, die dem Brett c entspricht« oder: »Zwei Objekte d und e haben gemeinsam das doppelte Gewicht von f« möglich. Man beachte, dass derartige Aussagen bei intervallskalierten Merkmalen nicht zulässig sind, denn weder die Aussage: »An einem Tag mit einer Durchschnittstemperatur von

Tab. 2.2. Die vier wichtigsten Skalenarten

Skalenart	Zulässige Transformationen	Mögliche Aussagen	Beispiele
1. Nominalskala	Eindeutigkeitstransformation	Gleichheit, Verschiedenheit	Telefonnummern, Krankheitsklassifikationen
2. Ordinalskala	Monotone Transformation	Größer-kleiner-Relationen	Militärische Ränge, Windstärken
3. Intervallskala	Lineare Transformation	Gleichheit von Differenzen	Temperatur (z. B. Celsius), Kalenderzeit
4. Verhältnisskala	Ähnlichkeitstransformation	Gleichheit von Verhältnissen	Längenmessung, Gewichtsmessung

10°C ist es doppelt so warm wie an einem Tag mit einer Durchschnittstemperatur von 5°C« noch die Aussage: »Durch das Zusammenfügen der Intelligenz zweier Personen a und b resultiert die Intelligenz einer Person c« macht Sinn.

Ein empirisches Relativ mit den oben genannten Eigenschaften bezeichnet man als **extensive Messstruktur**. Man erhält eine Verhältnisskala, wenn ein empirisches Relativ mit einer extensiven Messstruktur wie folgt in ein numerisches Relativ abgebildet wird: Einem Objekt a, dessen Merkmalsausprägung mindestens so groß ist wie die eines Objektes b, wird eine Zahl $\phi(a)$ zugeordnet, die mindestens so groß ist wie $\phi(b)$. Die Zahl, die der Merkmalsausprägung zugeordnet wird, die sich durch die Verknüpfung von a und b ergibt, entspricht der Summe der Zahlen für a und b. Hieraus folgt (vgl. Helmholtz, 1887, 1959, zitiert nach Steyer & Eid, 1993, Kap. 8.1):

! Eine Verhältnisskala ordnet den Objekten eines empirischen Relativs Zahlen zu, die so geartet sind, dass das Verhältnis zwischen je zwei Zahlen dem Verhältnis der Merkmalsausprägungen der jeweiligen Objekte entspricht.

Messungen auf Verhältnisskalen sind eindeutig bis auf hier zulässige **Ähnlichkeitstransformationen** vom Typus $\phi' = \beta \cdot \phi$ ($\beta > 0$). Beispiele für diese Transformationen sind das Umrechnen von Meter in Zentimeter oder Inches, von Kilogramm in Gramm oder Unzen, von Euro in Dollar, von Minuten in Sekunden. Man beachte, dass die Ähnlichkeitstransformation – anders als die für Intervallskalen zulässige lineare Transformation – den Ursprung der Verhältnisskala, der typischerweise dem Nullpunkt des Merkmals entspricht, nicht verändert.

Die Bedeutung einer numerischen Aussage über verhältnisskalierte Messungen ist gegenüber Ähnlichkeitstransformationen invariant. Für die Aussage: »Ein Objekt a kostet doppelt soviel wie ein Objekt b« ist es

unerheblich, ob die Objektpreise z. B. in Euro oder Dollar angegeben sind.

Verhältnisskalen kommen in der sozialwissenschaftlichen Forschung (mit sozialwissenschaftlichen Merkmalen) nur selten vor. Dementsprechend finden sie in der sozialwissenschaftlichen Statistik kaum Beachtung. Da jedoch Verhältnisskalen genauere Messungen ermöglichen als Intervallskalen, sind alle mathematischen Operationen bzw. statistischen Verfahren für Intervallskalen auch für Verhältnisskalen gültig. Man verzichtet deshalb häufig auf eine Unterscheidung der beiden Skalen und bezeichnet sie zusammengenommen als **Kardinalskalen** oder auch **metrische Skalen**.

Zusammenfassung. Die hier behandelten Skalenarten sowie einige typische Beispiele sind in **Tab. 2.2** noch einmal zusammengefasst. Die genannten »möglichen Aussagen« sind invariant gegenüber den jeweils zulässigen skalenspezifischen Transformationen.

Ein Vergleich der vier Skalen zeigt, dass die Messungen mit zunehmender Ordnungsziffer der Skala genauer werden. Während eine Nominalskala lediglich Äquivalenzklassen von Objekten numerisch beziffert, informieren die Zahlen einer Ordinalskala zusätzlich darüber, bei welchen Objekten das Merkmal stärker bzw. weniger stark ausgeprägt ist. Eine Intervallskala ist der Ordinalskala überlegen, weil hier die Größe eines Merkmalsunterschiedes bei zwei Objekten genau quantifiziert wird. Eine Verhältnisskala schließlich gestattet zusätzlich Aussagen, die die Merkmalsausprägungen verschiedener Objekte zueinander ins Verhältnis setzen.

Praktische Konsequenzen

Nachdem in den letzten Abschnitten messtheoretische Probleme erörtert und die wichtigsten Skalenarten einführend behandelt wurden, stellt sich die Frage, welche praktischen Konsequenzen hieraus für die Anlage einer

empirischen Untersuchung abzuleiten sind. Die Antwort auf diese Frage folgt den bereits bei Bortz (2005, S. 25 f.) genannten Ausführungen.

Empirische Sachverhalte werden durch die vier in ▶ Tab. 2.2 genannten Skalenarten unterschiedlich genau abgebildet. Die hieraus ableitbare Konsequenz für die Planung empirischer Untersuchungen liegt auf der Hand: Bieten sich bei einer Quantifizierung mehrere Skalenarten an, sollte diejenige mit dem höchsten **Skalenniveau** gewählt werden. Erweist sich im Nachhinein, dass die erhobenen Daten dem angestrebten Skalenniveau letztlich nicht genügen, besteht die Möglichkeit, die erhobenen Daten auf ein niedrigeres Skalenniveau zu transformieren. (Beispiel: Zur Operationalisierung des Merkmals »Schulische Reife« sollten Experten intervallskalierte Punkte vergeben. Im Nachhinein stellt sich heraus, dass die Experten mit dieser Aufgabe überfordert waren, sodass man beschließt, für weitere Auswertungen nur die aus den Punktzahlen ableitbare Rangfolge der Kinder zu verwenden.) Eine nachträgliche Transformation auf ein höheres Skalenniveau ist hingegen nicht möglich.

Wie jedoch – so lautet die zentrale Frage – wird in der Forschungspraxis entschieden, auf welchem Skalenniveau ein bestimmtes Merkmal gemessen wird? Ist es erforderlich bzw. üblich, bei jedem Merkmal die gesamte Axiomatik der mit einer Skalenart verbundenen Messstruktur empirisch zu überprüfen? Kann man – um im oben genannten Beispiel zu bleiben – wirklich guten Gewissens behaupten, die Punktzahlen zur »Schulischen Reife« seien, wenn schon nicht intervallskaliert, so doch zumindest ordinalskaliert?

Sucht man in der Literatur nach einer Antwort auf diese Frage, so wird man feststellen, dass hierzu unterschiedliche Auffassungen vertreten werden (z. B. Wolins, 1978). Unproblematisch und im Allgemeinen ungeprüft ist die Annahme, ein Merkmal sei nominalskaliert. Biologisches Geschlecht, Parteizugehörigkeit, Studienfach, Farbpräferenzen, Herkunftsland etc. sind Merkmale, deren Nominalskalenqualität unstrittig ist.

Weniger eindeutig fällt die Antwort jedoch aus, wenn es darum geht zu entscheiden, ob Schulnoten, Testwerte, Einstellungsmessungen, Schätz-(Rating-)Skalen o. Ä. ordinal- oder intervallskaliert sind. Hier eine richtige Antwort zu finden, ist insoweit von Bedeutung, als die Berechnung von Mittelwerten und anderen wichtigen

statistischen Maßen nur bei intervallskalierten Merkmalen zu rechtfertigen ist, d. h., für ordinalskalierte Daten sind andere statistische Verfahren einzusetzen als für intervallskalierte Daten.

Die übliche Forschungspraxis verzichtet auf eine empirische Überprüfung der jeweiligen Skalenaxiomatik. Die meisten Messungen sind **Per-fiat-Messungen** (Messungen »durch Vertrauen«), die auf Erhebungsinstrumenten (Fragebögen, Tests, Ratingskalen etc.) basieren, von denen man annimmt, sie würden das jeweilige Merkmal auf einer Intervallskala messen. Es kann so der gesamte statistische »Apparat« für Intervallskalen eingesetzt werden, der erheblich differenziertere Auswertungen ermöglicht als die Verfahren für Ordinal- oder Nominaldaten (vgl. Davison & Sharma, 1988, oder Lantermann, 1976).

Hinter dieser »liberalen« Auffassung steht die Überzeugung, dass die Bestätigung einer Forschungshypothese durch die Annahme eines falschen Skalenniveaus eher erschwert wird. Anders formuliert: Lässt sich eine inhaltliche Hypothese empirisch bestätigen, ist dies meistens ein Beleg für die Richtigkeit der skalentheoretischen Annahme. Wird eine inhaltliche Hypothese hingegen empirisch widerlegt, sollte dies ein Anlass sein, auch die Art der Operationalisierung des Merkmals und damit das Skalenniveau der Daten zu problematisieren. Wie bereits in ▶ Abschn. 2.3.5 festgestellt, kann die Analyse von Messoperationen erheblich zur Präzisierung der geprüften Theorie beitragen.

2.3.7 Auswahl der Untersuchungsobjekte

Liegen befriedigende Operationalisierungen der interessierenden Variablen einschließlich ihrer messtheoretischen Bewertung vor, stellt sich im Planungsprozess als nächstes die Frage, an welchen bzw. an wie vielen Untersuchungsobjekten die Variablen erhoben werden sollen. Wie bereits in ▶ Abschn. 1.1.1 erwähnt, verwenden wir den Begriff Untersuchungsobjekt sehr allgemein; er umfasst z. B. Kinder, alte Personen, Depressive, Straffällige, Beamte, Arbeiter, Leser einer bestimmten Zeitung, Studierende etc., aber auch – je nach Fragestellung – z. B. Tiere, Häuser, Schulklassen, Wohnsiedlungen, Betriebe, Nationen o. Ä. (Die Problematik vergleichender tierpsychologischer Untersuchungen diskutieren Pritzel

und Markowitsch, 1985.) Wir behandeln im Folgenden Probleme, die sich mit der Auswahl von Personen als Untersuchungsobjekte oder besser: Untersuchungsteilnehmer verbinden, Besonderheiten, die sich bei studentischen Untersuchungsteilnehmern ergeben und das Thema »freiwillige Untersuchungsteilnahme«.

Art und Größe der Stichprobe

Für explorative Studien ist es weitgehend unerheblich, wie die Untersuchungsteilnehmer aus der interessierenden Population ausgewählt werden. Es sind anfallende Kollektive unterschiedlicher Größe oder auch einzelne Untersuchungsteilnehmer, deren Beobachtung oder Beschreibung interessante Hypothesen versprechen.

Untersuchungen zur Überprüfung von Hypothesen oder zur Ermittlung generalisierbarer Stichprobenkennwerte stellen hingegen höhere Anforderungen an die Auswahl der Untersuchungseinheiten. Über Fragen der Repräsentativität von Stichproben, die in derartigen Untersuchungen zu erörtern sind, wird ausführlich in ► Abschn. 7.1.1 berichtet.

Die Festlegung des Stichprobenumfangs sollte ebenfalls in der Planungsphase erfolgen. Verbindliche Angaben lassen sich hierfür jedoch nur machen, wenn eine hypothesenprüfende Untersuchung mit vorgegebener Effektgröße geplant wird (► Abschn. 9.2). Für die Größe von Stichproben, mit denen unspezifische Hypothesen geprüft werden (► Kap. 8), gibt es keine genauen Richtlinien. Wir wollen uns hier mit dem Hinweis begnügen, dass die Wahrscheinlichkeit, eine unspezifische Forschungshypothese zu bestätigen, mit zunehmendem Stichprobenumfang wächst.

Anwerbung von Untersuchungsteilnehmern

Für die Anwerbung der Untersuchungsteilnehmer gelten einige Regeln, deren Beachtung die Anzahl der Verweigerer häufig drastisch reduziert. Zunächst ist es wichtig, potenzielle Untersuchungsteilnehmer individuell und persönlich anzusprechen, unabhängig davon, ob dies in mündlicher oder schriftlicher Form geschieht. Ferner sollte das Untersuchungsvorhaben – soweit die Fragestellung dies zulässt – inhaltlich erläutert werden mit Angaben darüber, wem die Untersuchung potenziell zugute kommt (► Abschn. 2.2.2). Verspricht die Untersuchung Ergebnisse, die auch für den einzelnen Untersuchungsteilnehmer interessant sein könnten, ist dies

besonders hervorzuheben. Hierbei dürfen Angaben darüber, wie und wann der Untersuchungsteilnehmer seine individuellen Ergebnisse erfahren kann, nicht fehlen. Nach Rosenthal und Rosnow (1975) wirkt sich die Anwerbung durch eine Person mit einem möglichst hohen sozialen Status besonders günstig auf die Bereitschaft aus, an der Untersuchung teilzunehmen.

Der in ► Box 2.6 (auszugsweise) wiedergegebene Brief einer männlichen Versuchsperson an einen männlichen Versuchsleiter des humanistischen Psychologen Jourard (1973) illustriert in zugespitzter Weise, welche Einstellungen, Gedanken und Gefühle die Teilnahme an psychologischen Untersuchungen begleiten können. Bereits bei der Anwerbung werden Erwartungshaltungen erzeugt, die die Reaktionen der Untersuchungsteilnehmer auf die spätere Untersuchungssituation nachhaltig beeinflussen. Gerade psychologische Untersuchungen sind darauf angewiesen, dass uns die Teilnehmer persönliches Vertrauen entgegenbringen, denn nur diese Grundeinstellung kann absichtliche Täuschungen und bewusste Fehlreaktionen verhindern (zur Testverfälschung ► Abschn. 4.3.7). Weitere Hinweise zur Anwerbung von Untersuchungsteilnehmern findet man bei Hager et al. (2001, S. 38 ff.).

Determinanten der freiwilligen Untersuchungsteilnahme

Nicht nur Täuschungen und bewusste Fehlreaktionen der Untersuchungsteilnehmer sind Gründe für problematische Untersuchungen, sondern auch eine hohe Verweigerungsrate. Die Verweigerung der Untersuchungsteilnahme wird vor allem dann zum Problem, wenn man davon ausgehen muss, dass sich die Verweigerer systematisch bezüglich untersuchungsrelevanter Merkmale von den Teilnehmern unterscheiden. Die typischen Merkmale freiwilliger Untersuchungsteilnehmer sowie situative Determinanten der Freiwilligkeit sind dank einer gründlichen Literaturdurchsicht von Rosenthal und Rosnow (1975) zumindest für amerikanische Verhältnisse recht gut bekannt. Diese Resultate lassen sich – wie eine Studie von Effler und Böhmeke (1977) zeigt – zumindest teilweise ohne Bedenken auch auf deutsche Verhältnisse übertragen (über Besonderheiten der freiwilligen Untersuchungsteilnahme bei Schülern berichtet Spiel, 1988).

Die folgende Übersicht enthält Merkmale, die zwischen freiwilligen Untersuchungsteilnehmern und Ver-

Box 2.6

Brief einer Vp an einen VI. (Nach Jourard, 1973; Übersetzung: H.E. Lück)

Lieber Herr VI (Versuchsleiter):

Mein Name ist Vp (Versuchsperson). Sie kennen mich nicht. Ich habe einen anderen Namen, mit dem mich meine Freunde anreden, aber den lege ich ab und werde Vp Nr. 27, wenn ich Gegenstand Ihrer Forschung werde. Ich nehme an Ihren Umfragen und Experimenten teil. Ich beantworte Ihre Fragen, fülle Fragebogen aus, lasse mich an Drähte anschließen, um meine physiologischen Reaktionen untersuchen zu lassen. Ich drücke Tasten, bediene Schalter, verfolge Ziele, die sich bewegen, laufe durch Labyrinth, lerne sinnlose Silben und sage Ihnen, was ich in Tintenklecken entdecke – ich mache all den Kram, um den Sie mich bitten. Aber ich frage mich langsam, warum ich das alles für Sie tue. Was bringt mir das ein? Manchmal bezahlen Sie meinen Dienst. Häufiger muß ich aber mitmachen, weil ich Psychologiestudent der Anfangssemester bin und weil man mir gesagt hat, daß ich keinen Schein bekomme, wenn ich nicht an zwei Versuchen teilgenommen habe; wenn ich an mehr Versuchen teilnehme, kriege ich zusätzliche Pluspunkte fürs Diplom. Ich gehöre zum »Vp-Reservoir« des Instituts.

Wenn ich Sie schon mal gefragt habe, inwiefern Ihre Untersuchungen für mich gut sind, haben Sie mir erzählt: »Das ist für die Forschung.« Bei manchen Ihrer Untersuchungen haben Sie mich über den Zweck der Studien belogen. Sie verführen mich. Ich kann Ihnen daher kaum trauen. Sie erscheinen mir langsam als Schwindler, als Manipulator. Das gefällt mir nicht.

Das heißt – ich belüge Sie auch oft, sogar in anonymen Fragebögen. Wenn ich nicht lüge, antworte ich manchmal nur nach Zufall, um irgendwie die Stunde ‚rumzukriegen‘, damit ich wieder meinen Interessen nachgehen kann. Außerdem kann ich oft herausfinden, um was es Ihnen geht, was Sie gern von mir hören oder sehen wollen; dann gehe

ich manchmal auf Ihre Wünsche ein, wenn Sie mir sympathisch sind, oder ich nehme Sie auf den Arm, wenn Sie's nicht sind. Sie sagen ja nicht direkt, welche Hypothesen Sie haben oder was Sie sich wünschen. Aber die Anordnungen in Ihrem Laboratorium, die Alternativen, die Sie mir vorgeben, die Instruktionen, die Sie mir vorlesen, alles das zusammen soll mich dann drängen, irgend etwas Bestimmtes zu sagen oder zu tun. Das ist so, als wenn Sie mir ins Ohr flüstern würden: »Wenn jetzt das Licht angeht, den linken Schalter bedienen!«, und Sie würden vergessen oder bestreiten, daß Sie mir das zugeflüstert haben. Aber ich weiß, was Sie wollen! Und ich bediene den linken oder den rechten Schalter, je nachdem, was ich von Ihnen halte.

Wissen Sie, selbst wenn Sie nicht im Raum sind – wenn Sie nur aus gedruckten Anweisungen auf dem Fragebogen bestehen oder aus der Stimme aus dem Tonbandgerät, die mir sagt, was ich tun soll – ich mache mir Gedanken über Sie. Ich frage mich, wer Sie sind, was Sie wirklich wollen. Ich frage mich, was Sie mit meinem »Verhalten« anfangen. Wem zeigen Sie meine Antworten? Wer kriegt eigentlich meine Kreuzchen auf Ihren Antwortbögen zu sehen? Haben Sie überhaupt ein Interesse daran, was ich denke, fühle und mir vorstelle, wenn ich die Kreuzchen mache, die Sie so emsig auswerten? Es ist sicher, daß Sie mich noch nie danach gefragt haben, was ich überhaupt damit gemeint habe. Wenn Sie fragen würden – ich würde es Ihnen gern erzählen. Ich erzähle nämlich meinem Zimmergenossen im Studentenheim oder meiner Freundin davon, wozu Sie Ihr Experiment gemacht haben und was ich mir dabei gedacht habe, als ich mich so verhielt, wie ich mich verhalten habe. Wenn mein Zimmergenosse Vertrauen zu Ihnen hätte, könnte er Ihnen vielleicht besser sagen, was die Daten (meine Antworten und Reaktionen) bedeuten, als Sie es mit Ihren Vermutungen können. Weiß Gott, wie sehr die gute Psychologie im Ausgüß gelandet ist, wenn mein Zimmergenosse und ich Ihr Experiment und meine Rolle dabei beim Bier diskutieren! ...



Wenn Sie mir vertrauen, vertraue ich Ihnen auch, sofern Sie vertrauenswürdig sind. Ich fände gut, wenn Sie sich die Zeit nehmen und die Mühe machen würden, mit mir als Person vertraut zu werden, bevor wir in den Versuchsablauf einsteigen. Ich möchte Sie und Ihre Interessen gern kennenlernen, um zu sehen, ob ich mich vor Ihnen »ausbreiten« möchte. Manchmal erinnern Sie mich an Ärzte. Die sehen mich als uninteressante Verpackung an, in der die Krankheit steckt, an der sie wirklich interessiert sind. Sie haben mich als uninteressantes Paket angesehen, in dem »Reaktionen« stecken, mehr bedeute ich Ihnen nicht. Ich möchte Ihnen sagen, daß ich mich über Sie ärgere, wenn ich das merke. Ich liefere Ihnen Reaktionen, o.k. – aber Sie werden nie erfahren, was ich damit gemeint habe. Wissen Sie, ich kann sprechen, nicht nur mit Worten, sondern auch mit Taten.

Wenn Sie geglaubt haben, ich hätte nur auf einen »Stimulus« in Ihrem Versuchsraum reagiert, dann habe ich in Wirklichkeit auf Sie reagiert; was ich mir dabei dachte, war folgendes: »Da hast Du's,

Du unangenehmer Soundso!« Erstaunt Sie das? Eigentlich sollte es das nicht ...

Ich möchte mit Ihnen ein Geschäft machen. Sie zeigen mir, daß Sie Ihre Untersuchungen für mich machen – damit ich freier werde, mich selbst besser verstehe, mich selbst besser kontrollieren kann – und ich werde mich Ihnen zur Verfügung stellen wie Sie wollen. Dann werde ich Sie auch nicht mehr verschaukeln und beschummeln. Ich möchte nicht kontrolliert werden, weder von Ihnen noch von sonst jemandem. Ich will auch keine anderen Leute kontrollieren. Ich will nicht, daß Sie anderen Leuten festzustellen helfen, wie »kontrolliert« ich bin, so daß sie mich dann kontrollieren können. Zeigen Sie mir, daß Sie für mich sind, und ich werde mich Ihnen öffnen.

Arbeiten Sie für mich, Herr Vl, und ich arbeite ehrlich für Sie. Wir können dann zusammen eine Psychologie schaffen, die echter und befreiender ist.

Mit freundlichen Grüßen,
Ihre Vp

weigerern differenzieren, sowie Merkmale der Untersuchung, die die Freiwilligkeit beeinflussen (nach Rosenthal & Rosnow, 1975, S. 1955 ff.; ergänzt durch Effler & Böhmeke, 1977).

Merkmale der Person. Die Kontrastierung von freiwilligen Untersuchungsteilnehmern und Verweigerern führte zu folgenden Resultaten:

- Freiwillige Untersuchungsteilnehmer verfügen über eine bessere schulische Ausbildung als Verweigerer (bessere Notendurchschnitte). Dies gilt insbesondere für Untersuchungen, in denen persönliche Kontakte zwischen dem Untersuchungsleiter und den Untersuchungsteilnehmern nicht erforderlich sind. Bei Schülern ist die Schulleistung für die freiwillige Teilnahme irrelevant.
- Freiwillige Untersuchungsteilnehmer schätzen ihren eigenen sozialen Status höher ein als Verweigerer.
- Die meisten Untersuchungsergebnisse sprechen für eine höhere Intelligenz freiwilliger Untersuchungsteilnehmer (z. B. bessere Leistungen in den Unter-

tests »Analogien«, »Gemeinsamkeiten«, »Rechenaufgaben« und »Zahlenreihen« des »Intelligenz-Struktur-Tests« von Amthauer et al., 2001).

- Freiwillige benötigen mehr soziale Anerkennung als Verweigerer.
- Freiwillige Untersuchungsteilnehmer sind geselliger als Verweigerer.
- In Untersuchungen über geschlechtsspezifisches Verhalten geben sich freiwillige Untersuchungsteilnehmer unkonventioneller als Verweigerer.
- Im Allgemeinen sind weibliche Personen eher zur freiwilligen Untersuchungsteilnahme bereit als männliche Personen.
- Freiwillige Untersuchungsteilnehmer sind weniger autoritär als Verweigerer.
- Die Tendenz zu konformem Verhalten ist bei Verweigerern stärker ausgeprägt als bei freiwilligen Untersuchungsteilnehmern.

Aus diesen Befunden folgt, dass einem allgemeinen Aufruf zur Untersuchungsteilnahme generell eher sozial

privilegierte und weibliche Personen nachkommen und dass ggf. Maßnahmen zu ergreifen sind, um den Kreis der Freiwilligen zu erweitern (z. B. gruppenspezifische Werbung für die betreffende Studie). Zudem sollte man sich natürlich auch inhaltlich fragen, welche Motive Personen dazu veranlassen, an einer bestimmten Untersuchung teilzunehmen bzw. auf eine Teilnahme zu verzichten (► S. 249).

Merkmale der Untersuchung. Auch Besonderheiten der Untersuchung können dazu beitragen, die Rate der Verweigerer zu reduzieren bzw. die mit der Verweigerungsproblematik verbundene Stichprobenverzerrung in Grenzen zu halten. Die Durchsicht einer nicht unerheblichen Anzahl diesbezüglicher empirischer Untersuchungen führte zu folgenden Erkenntnissen:

- Personen, die sich für den Untersuchungsgegenstand interessieren, sind zur freiwilligen Teilnahme eher bereit als weniger interessierte Personen.
- Je bedeutender die Untersuchung eingeschätzt wird, desto höher ist die Bereitschaft zur freiwilligen Teilnahme.
- Entlohnungen in Form von Geld fördern die Freiwilligkeit weniger als kleine persönliche Geschenke und Aufmerksamkeiten, die dem potenziellen Untersuchungsteilnehmer vor seiner Entscheidung, an der Untersuchung mitzuwirken, überreicht werden.
- Die Bereitschaft zur freiwilligen Teilnahme steigt, wenn die anwerbende Person persönlich bekannt ist. Erfolgreiche Anwerbungen sind durch einen »persönlichen Anstrich« gekennzeichnet.
- Die Anwerbung ist erfolgreicher, wenn die Untersuchung öffentlich unterstützt wird und die Teilnahme »zum guten Ton« gehört. Empfindet man dagegen eher die Verweigerung als obligatorisch, sinkt die Teilnahmebereitschaft.

Die Ausführungen von Rosenthal und Rosnow (1975) legen es mit Nachdruck nahe, die Anwerbung der zu untersuchenden Personen sorgfältig zu planen. Aber da man keine Person zur Teilnahme an einer Untersuchung zwingen kann und da nicht jede Untersuchung die für die Rekrutierung von Untersuchungsteilnehmern idealen Bedingungen aufweist, wird man mit mehr oder weniger systematisch verzerrten Stichproben rechnen müssen. Es wäre jedoch bereits ein bemerkenswerter

Fortschritt, wenn die Besonderheiten freiwilliger Untersuchungsteilnahme in die Ergebnisdiskussion einfließen würden. Die Ergebnisdiskussion wäre dann gleichzeitig eine Diskussion von Hypothesen darüber, in welcher Weise die Resultate durch Verweigerungen verfälscht sein können.

Empirische Untersuchungen versetzen Personen in soziale Situationen, die sie zuweilen als Einengung ihrer persönlichen Handlungsfreiheit erleben. Der Theorie der **psychologischen Reaktanz** (Brehm, 1966) zufolge muss dann mit Abwehrmechanismen der Untersuchungsteilnehmer gerechnet werden, die vor einer Verletzung der persönlichen Freiheit schützen. Auch nach anfänglicher Teilnahmebereitschaft kann es während einer Untersuchung infolge von Argwohn gegenüber absichtlicher Täuschung durch den Untersuchungsleiter oder wegen erzwungener Verhaltens- und Reaktionsweisen zu den unterschiedlichsten Varianten von »Untersuchungssabotage« kommen. Eine entspannte Anwerbungssituation und Untersuchungsdurchführung, die die persönliche Freiheit und den Handlungsspielraum der Untersuchungsteilnehmer möglichst wenig einengen, helfen derartige Störungen zu vermeiden.

Studierende als Versuchspersonen

Die humanwissenschaftliche Forschung leidet darunter, dass sich viele Untersuchungsleiter die Auswahl ihrer Untersuchungsteilnehmer sehr leicht machen, indem sie einfach anfallende Studentengruppen wie z. B. die Teilnehmer eines Seminars oder zufällig in der Mensa ange-troffene Kommilitonen um ihre Mitwirkung bitten. Hohn (1972) fand unter 700 Originalbeiträgen aus acht deutschsprachigen Zeitschriften der Jahrgänge 1967 bis 1969 475 empirische Untersuchungen, an denen ca. 50.000 Personen mitgewirkt hatten. Von diesen Personen waren 21% Studenten – ein Prozentsatz, der den tatsächlichen Prozentsatz sicher unterschätzt, wenn man bedenkt, dass der Anteil nicht identifizierbarer Probanden mit 23% auffallend hoch war. Diese Vermutung bestätigt eine Kontrollanalyse von Janssen (1979), der in den Jahrgängen 1970 bis 1973 derselben Zeitschriften einen studentischen Anteil von 43% bei 15% nicht identifizierbaren Personen fand.

Noch dramatischer scheinen die Verhältnisse in den USA zu sein. Hier beträgt der studentische Anteil in empirischen Untersuchungen ca. 80%, obwohl diese Gruppe

nur 3% der Gesamtbevölkerung ausmacht. Mit Probanden der Allgemeinbevölkerung wurden nicht einmal 1% aller Untersuchungen durchgeführt (vgl. Janssen, 1979).

Nun kann man zwar nicht generell die Möglichkeit ausschließen, dass sich auch mit Studierenden allgemeine Gesetzmäßigkeiten finden lassen – immerhin wurde das Weber'sche Gesetz (E.H. Weber, 1851) in Untersuchungen mit Studenten und die Vergessenskurve (Ebbinghaus, 1885) sogar in Selbstversuchen entdeckt. Dennoch liegt der Verdacht nahe, dass Untersuchungen über entwicklungsbedingte, sozialisationsbedingte und durch das Altern bedingte Prozesse im kognitiven und intellektuellen Bereich, die vorwiegend mit Studenten durchgeführt wurden, zu falschen Schlüssen führen. Bedauerlicherweise ist jedoch der prozentuale Anteil von Studierenden bzw. jungen Menschen gerade in derartigen Untersuchungen besonders hoch. Leibbrand (1976) ermittelte in 65 Publikationen mit denk- oder lernpsychologischen Fragestellungen, dass 90% der Probanden 25 Jahre oder jünger waren.

Die Fragwürdigkeit humanwissenschaftlicher Forschungsergebnisse, die überwiegend in Untersuchungen mit Studenten ermittelt werden, erhöht sich um ein Weiteres, wenn man in Rechnung stellt, dass an diesen Untersuchungen nur »freiwillige« Studierende teilnehmen. Die Ergebnisse gelten damit nicht einmal für studentische Populationen generell, sondern eingeschränkt nur für solche Studenten, die zur freiwilligen Untersuchungsteilnahme bereit sind (zum Thema Vpn-Stunden als Pflichtleistung im Rahmen psychologischer Prüfungsordnungen ► S. 44). Über die Bedeutsamkeit der individuellen Begründung, an empirischen Untersuchungen teilzunehmen, ist sicherlich noch viel Forschungsarbeit zu leisten.

Empfehlungen

Die Diskussion der Probleme, die mit der Auswahl der Untersuchungsteilnehmer verbunden sind, resultiert in einer Reihe von Empfehlungen, deren Befolgung nicht nur der eigenen Untersuchung, sondern auch der weiteren Erforschung von Artefakten in den Human- und Sozialwissenschaften zugute kommen:

- Die Anwerbung des Untersuchungsteilnehmers und dessen Vorbereitung auf die Untersuchung sollte die Freiwilligkeit nicht zu einem Problem werden lassen. Dies wird umso eher gelingen, je sorgfältiger die von

Rosenthal und Rosnow (1975) erarbeiteten, situativen Determinanten der Freiwilligkeit Beachtung finden.

- Variablen, von denen bekannt ist, dass sie zwischen freiwilligen Untersuchungsteilnehmern und Verweigerern differenzieren, verdienen besondere Beachtung. Überlagern derartige Variablen die unabhängige Variable oder muss man mit ihrem direkten Einfluss auf die abhängige Variable rechnen, sollten sie vorsorglich miterhoben werden, um ihren tatsächlichen Einfluss im Nachhinein kontrollieren zu können.
- Keine empirische Untersuchung sollte auf eine Diskussion möglicher Konsequenzen verzichten, die mit der freiwilligen Untersuchungsteilnahme in gerade dieser Untersuchung verbunden sein könnten.
- In einer die Untersuchung abschließenden Befragung sollte schriftlich festgehalten werden, mit welchen Gefühlen die Untersuchungsteilnehmer an der Untersuchung teilnahmen. Diese Angaben kennzeichnen die Untersuchungsbereitschaft, die später mit dem Untersuchungsergebnis in Beziehung gesetzt werden kann.
- Eine weitere Kontrollfrage bezieht sich darauf, wie häufig die Untersuchungsteilnehmer bisher an empirischen Untersuchungen teilnahmen. Auch die so erfasste Erfahrung mit empirischen Untersuchungen könnte die Ergebnisse beeinträchtigen.
- Die Erforschung der persönlichen Motive, an einer Untersuchung freiwillig teilzunehmen, sollte intensiviert werden. Hierfür könnte die PRS-Skala von Adair (1973; vgl. Timaeus et al., 1977) eingesetzt werden, die die Motivation von Untersuchungsteilnehmern erfassen soll.
- Die externe Validität der Untersuchungsergebnisse ist vor allem bei Untersuchungen mit studentischen Stichproben zu problematisieren.

2.3.8 Durchführung, Auswertung und Planungsbericht

Der Arbeitstitel und die Untersuchungsart liegen fest, die Erhebungsinstrumente sind bekannt und Art und Anzahl der auszuwählenden Untersuchungsteilnehmer sowie deren Rekrutierung sind vorgeplant. Die Untersuchungsplanung sollte nun die Durchführung der Untersuchung vorstrukturieren.

Planung der Untersuchungsdurchführung

Die Verschiedenartigkeit empirischer Untersuchungen bzw. die zeitlichen, finanziellen, räumlichen und personellen Rahmenbedingungen erschweren das Aufstellen genereller Leitlinien für die Untersuchungsdurchführung erheblich. Auch noch so sorgfältige Untersuchungsvorbereitungen können mögliche Pannen in der Untersuchungsdurchführung nicht verhindern. Um die Untersuchungsdurchführung hieran nicht scheitern zu lassen, sollte die Planung der Untersuchungsdurchführung nicht übermäßig rigide sein. Unbeschadet dieser Flexibilität sind jedoch der zeitliche Ablauf sowie Einsatz und Verwendung von Hilfspersonal, Räumen, Apparaten und ggf. auch Finanzen vor der Untersuchungsdurchführung festzulegen.

Wichtig sind einige allgemeine Regeln und Erkenntnisse, die das Verhalten von Untersuchungsleiterinnen und -leitern betreffen. Diese Richtlinien sollten während der Durchführung der Untersuchung im Bewusstsein der Untersuchenden fest verankert sein und sind damit unmittelbar Bestandteil der konkreten Untersuchungsdurchführung. Wir werden hierüber in ► Abschn. 2.5 ausführlich berichten.

Aufbereitung der Daten

Die planerische Vorarbeit setzt zu einem Zeitpunkt wieder ein, nachdem die Untersuchung »gedanklich« durchgeführt ist und die »Daten« erhoben sind. Dies können Beobachtungsprotokolle, Ton- oder Videobänder von Interviews und Diskussionen, ausgefüllte Fragebögen oder Tests, Häufigkeitsauszählungen von Blickbewegungen, Hirnstromverlaufskurven, auf einem elektronischen Datenträger gespeicherte Reaktionszeiten oder ähnliches sein. Der nächste Planungsschritt gilt der Aufbereitung dieser »Rohdaten«.

Die statistische Datenanalyse setzt voraus, dass die Untersuchungsergebnisse in irgendeiner Weise numerisch quantifiziert sind. Liegen noch keine »Zahlen« für die interessierenden Variablen, sondern z. B. qualitative Angaben vor, müssen diese für eine statistische Analyse zu Kategorien zusammengefasst und numerisch kodiert werden (► Abschn. 4.1.4).

In Abhängigkeit vom Umfang des anfallenden Datenmaterials erfolgt die statistische Datenanalyse computergestützt oder manuell, evtl. unterstützt durch einen Taschenrechner.

In deskriptiven Studien ist die Aggregation bzw. Zusammenfassung des erhobenen Datenmaterials vorrangig. Diese kann durch die Ermittlung einfacher statistischer Kennwerte wie z. B. dem arithmetischen Mittel oder einem Streuungsmaß erfolgen, durch die Anfertigung von Graphiken oder aber durch aufwendigere statistische Verfahren wie z. B. eine Clusteranalyse, eine Faktorenanalyse oder Zeitreihenanalyse (► Abschn. 6.4.2).

Planung der statistischen Hypothesenprüfung

Für die Überprüfung von Hypothesen steht ein ganzes Arsenal statistischer Methoden zur Verfügung, das man zumindest überblicksweise beherrschen sollte. Es ist unbedingt zu fordern, dass die Art und Weise, wie die Hypothesen statistisch getestet werden, vor der Datenerhebung feststeht. Auch wenn die Vielseitigkeit und Flexibilität eines modernen statistischen Instrumentariums gelegentlich auch dann eine einigermaßen vernünftige Auswertung ermöglicht, wenn diese nicht vorgeplant wurde, passiert es immer wieder, dass mühsam und kostspielig erhobene Daten wegen begangener Planungsfehler für statistische Hypothesentests unbrauchbar sind. Die Festlegung der Datenerhebung ist deshalb erst zu beenden, wenn bekannt ist, wie die Daten auszuwerten sind.

Stellt sich heraus, dass für die in Aussicht genommenen Daten keine Verfahren existieren, die in verlässlicher Weise etwas über die Tauglichkeit der inhaltlichen Hypothesen aussagen, können vor der Untersuchungsdurchführung meistens ohne große Schwierigkeiten Korrekturen an den Erhebungsinstrumenten, der Erhebungsart oder der Auswahl bzw. der Anzahl der Untersuchungsobjekte vorgenommen werden. Ist die Datenerhebung jedoch bereits abgeschlossen, sind die Chancen für eine verbesserte Datenqualität vertan, und man muss sich bei der eigentlich entscheidenden Hypothesenprüfung mit schlechten Kompromissen begnügen.

Die Planung einer hypothesenprüfenden Untersuchung ist unvollständig, wenn sie den statistischen Test, mit dem die Hypothese zu prüfen ist, nicht nennt. Nachdem die ursprünglich inhaltlich formulierte Hypothese operationalisiert wurde, erfolgt jetzt die Formulierung statistischer Hypothesen. Die Planung der statistischen Auswertung enthält dann im Prinzip Angaben wie z. B.: »Träfe meine Hypothese zu, müsste der Mittelwert \bar{x}_1

größer als der Mittelwert \bar{x}_2 sein« oder: »Nach meiner Vorhersage müsste zwischen den Variablen X und Y eine bedeutsame lineare Korrelation bestehen« oder: »Hypothesengemäß erwarte ich eine erheblich bessere Varianzaufklärung der abhängigen Variablen, wenn zusätzlich Variable Z berücksichtigt wird«.

Die einzusetzenden statistischen Tests sind so auszuwählen, dass sie den mit einer statistischen Hypothese behaupteten Sachverhalt exakt prüfen. Bei derartigen »Indikationsfragen« ist ggf. der Rat von Experten einzuholen (ausführlich wird die Planung der statistischen Hypothesenprüfung bei Hager, 2004, behandelt).

Voraussetzungen. Zum inhaltlichen Kriterium für die Auswahl eines adäquaten statistischen Verfahrens kommt ein formales: es müssen Überlegungen dazu angestellt werden, ob die zu erwartenden Daten diejenigen Eigenschaften aufweisen, die der in Aussicht genommene statistische Test voraussetzt. Ein Test, der z. B. für intervallskalierte Daten gilt, ist für nominalskalierte Daten unbrauchbar. Steht ein Verfahren, das die gleiche Hypothese auf nominalem Niveau prüft, nicht zur Verfügung, wird eine erneute Überprüfung und ggf. Modifikation der Operationalisierung bzw. der Erhebungsinstrumente erforderlich. Verlangt ein Verfahren, dass sich die Messwerte der abhängigen Variablen in einer bestimmten Weise verteilen, muss erwogen werden, ob diese Voraussetzung voraussichtlich erfüllt oder eine andere Verteilungsform wahrscheinlicher ist.

Zuweilen ziehen eingeplante Untersuchungsteilnehmer ihr Einverständnis zur Teilnahme zurück oder fallen aus irgendwelchen Gründen für die Untersuchung aus. Es muss dann in Rechnung gestellt werden, wie sich derartige »Missing Data« auf das ausgewählte statistische Verfahren auswirken (► S. 85).

Die voraussichtliche Genauigkeit der Daten ist ein weiteres Thema, mit dem sich die Planung der statistischen Hypothesenprüfung beschäftigen sollte. Vor allem bei größeren Untersuchungen sind Techniken zur Bestimmung der sog. Reliabilität der Daten einzuplanen, über die wir in ► Abschn. 4.3.3 berichten.

Ferner gehört zur Planung der statistischen Auswertung einer hypothesenprüfenden Untersuchung die Festlegung des **Signifikanzniveaus** (α -Fehler-Niveaus), das als Falsifikationskriterium darüber entscheidet, wann man die eigene Forschungshypothese als durch die

Daten widerlegt ansehen will (► S. 28 f. zum Good-enough-Prinzip; ausführlicher ► S. 635 ff.).

Untersuchungen über eine Thematik mit längerer Forschungstradition lassen nicht nur globale Hypothesen über die Richtung der vermuteten Zusammenhänge oder Unterschiede zu, sondern präzise Angaben über praktisch bedeutsame Mindestgrößen. Zur Planung gehört in diesem Falle auch die Festlegung von **Effektgrößen**. Hierüber wird in ► Abschn. 9.2 ausführlich berichtet. Hier werden wir auch das wichtige Konzept der **Teststärke** kennen lernen, für die es mittlerweile verbindliche Richtlinien gibt. Sind die geplante Teststärke, das Signifikanzniveau und die vermutete Effektgröße fixiert, so hat man damit eine rationale Basis für die Kalkulation des einzusetzenden **Stichprobenumfangs** (»optimaler Stichprobenumfang«, ► S. 604).

Statistische Programmpakete. Statistische Datenanalysen werden heutzutage üblicherweise auf einem Personalcomputer mit entsprechender Statistiksoftware durchgeführt. Bei der Vorbereitung einer computergestützten statistischen Datenanalyse sind folgende Punkte zu klären:

■ **Welche statistischen Verfahren (z. B. t-Test, Varianzanalyse, Faktorenanalyse, Korrelationsanalyse) sollen eingesetzt werden?** Bei der Beantwortung dieser Frage ist auf eigene Statistikkenntnisse zurückzugreifen, ggf. sollte man sich von Experten beraten lassen. Solche Beratungsgespräche ergeben zuweilen, dass die vorgesehenen Verfahren durch bessere ersetzt werden können, die dem Untersuchenden allerdings bislang unbekannt waren. Er steht nun vor der Frage, ob er diese ihm unbekannteren Verfahren übernehmen oder ob er seinen weniger guten, aber für ihn durchschaubaren Vorschlag realisieren soll. In dieser Situation kann man prinzipiell nur empfehlen, sich die Zeit zu nehmen, zumindest die Indikation des besseren Verfahrens und die Interpretation seiner Resultate aufzuarbeiten. Möglicherweise ergibt sich dann zu einem späteren Zeitpunkt die Gelegenheit, die innere Logik und den mathematischen Aufbau des Verfahrens kennen zu lernen. Auf jeden Fall ist davon abzuraten, für die Auswertung ein Verfahren einzuplanen, das einem gänzlich unbekannt ist. Dies führt erfahrungsgemäß zu erheblichen Problemen bei der Ergebnisinterpretation. Die Versuchung,

sich unbekannter oder auch nur leidlich bekannter Verfahren zu bedienen, ist angesichts der immer benutzerfreundlicher gestalteten Statistiksoftware leider recht groß.

— **Welche Statistiksoftware ist für die geplanten Analysen geeignet, zugänglich und in der Benutzung vertraut?** Die gängigen Statistikpakete wie SPSS, SAS, Systat usw. (► Anhang D) unterscheiden sich im Grundangebot ihrer Funktionen kaum, sodass Zugänglichkeit und Nutzungskompetenz die wichtigsten Auswahlkriterien darstellen. Zugänglich ist Statistiksoftware in den universitären PC-Pools. In die Benutzung eines verfügbaren Statistikpakets sollte man sich vor der Untersuchungsdurchführung einarbeiten, was Computervertrauten in der Regel autodidaktisch anhand von Lehrbüchern innerhalb einiger Stunden möglich ist. An vielen Universitäten werden im Rahmen der Statistikausbildung entsprechende Kurse angeboten.

— **Mit welchen Programmbefehlen können die gewünschten Analysen ausgeführt werden, welche Zusatzoptionen sind wichtig?** Auch bei einer computergestützten Datenanalyse sollte man – obwohl es schnell und einfach möglich ist – nicht wahllos verschiedene Prozeduren an den Daten ausprobieren. Dieser »spielerische Umgang« mit den angebotenen Programmen erleichtert zwar das Verständnis der Verfahren und ist deshalb für die Lernphase empfehlenswert; bei der eigentlichen Datenanalyse führt diese Vorgehensweise jedoch sehr schnell zu einem unübersichtlichen Berg von Computerausdrucken und zu Antworten, für die man selbst bislang noch gar keine Fragen formuliert hatte. Gibt das Datenmaterial mehr her als die Beantwortung der eingangs formulierten Fragen – was keineswegs selten ist –, sind diese Zusatzbefunde rein explorativ zu verstehen und auch als solche im Untersuchungsbericht zu kennzeichnen (► S. 498 zum Stichwort HARKING).

Die Unsitte, alle »relevant« erscheinenden Verfahren an den eigenen Daten auszuprobieren, in der Hoffnung, dadurch auf irgend etwas Interessantes zu stoßen, führt nicht selten dazu, dass nach Abschluss der Datenverarbeitungsphase mehr Computerausdrucke vorliegen, als ursprünglich Daten vorhanden waren. Damit ist aber der eigentliche Sinn der Datenverarbeitung, die theoriegeleitete Aggregation und

Reduktion der Rohdaten, ins Gegenteil verkehrt. Die gezielte Kondensierung der Ausgangsdaten in einige wenige hypothesenkritische Indikatoren wird damit aufgegeben zugunsten vieler, mehr oder weniger zufällig zustande gekommener Einzelergebnisse, die übersichtlich und zusammenfassend zu interpretieren nicht nur einen enormen Zeitaufwand bedeutete, sondern auch die Gefahr widersprüchlicher Ergebnisse in sich birgt. Dies ist ein wichtiger Grund dafür, dass sich die Psychologie so schwer damit tut, eine kumulative Wissenschaft zu werden (► S. 601).

— **In welcher Weise sollen die elektronisch erfassten Daten vorbereitet und bereinigt werden?** Nach einer bekannten Redewendung gilt für jedes statistische Verfahren das Prinzip »garbage in, garbage out« (Müll hinein, Müll heraus). Die Ergebnisse einer statistischen Analyse sind immer nur so gut wie die Ausgangsdaten. Neben untersuchungstechnischen Problemen, die zu fehlenden Werten oder Messfehlern führen, birgt die elektronische Datenverarbeitung weitere Fehlerquellen. Ein Engpass ist hierbei die elektronische Datenerfassung (z. B. mittels Textverarbeitungsprogramm, Editor eines Statistikprogramms oder Tabellenkalkulationsprogramm), bei der z. B. die auf einem Paper-Pencil-Fragebogen gegebenen Antworten kodiert und in eine Datendatei übertragen werden müssen. Tippfehler sind hierbei von vornherein unvermeidbar. Eine wichtige Kontrolle wäre es, wenn man einplant, die Dateneingabe stets zu zweit durchzuführen (eine diktiert, einer schreibt), um somit die Wahrscheinlichkeit für grobe Fehler zu reduzieren. Da bei der Dateneingabe in der Regel pro Untersuchungsobjekt bzw. Versuchsperson eine fortlaufende Nummer vergeben wird, sollte diese auch auf dem Originalmaterial verzeichnet werden (z. B. oben auf den Fragebogen), damit man später bei eventuellen Ungereimtheiten die Datendatei noch einmal mit dem Ausgangsmaterial vergleichen kann.

Interpretation möglicher Ergebnisse

Sicherlich werden sich einige Leserinnen und Leser angesichts des hier aufgeführten Planungsschrittes fragen, ob es sinnvoll oder möglich ist, über die Interpretation von Ergebnissen nachzudenken, wenn die Daten noch nicht einmal erhoben, geschweige denn ausgewertet

sind. Dennoch ist dieser Planungsschritt – vor allem für hypothesenprüfende Untersuchungen – wichtig, denn er dient einer letzten Überprüfung der in Aussicht genommenen Operationalisierung und statistischen Auswertung. Er soll klären, ob die Untersuchung tatsächlich eine Antwort auf die formulierten Hypothesen liefern kann, bzw. ob die Resultate der statistischen Analyse potenziell als Beleg für die Richtigkeit der inhaltlichen Hypothesen zu werten sind.

Es könnte z. B. ein signifikanter t-Test (► Anhang B) über die abhängige Variable »Anzahl richtig gelernter Vokabeln« für eine Kontroll- und eine Experimentalgruppe erwartet werden, der eindeutig im Sinne der inhaltlichen Hypothese zu interpretieren wäre, wenn eine Verbesserung der Lernleistungen nach Einführung einer neuen Unterrichtsmethode vorhergesagt wird. Sieht die statistische Planung einer Untersuchung über Ausländerfeindlichkeit jedoch z. B. eine Faktorenanalyse (► Anhang B) über einen Fragebogen zur Ermittlung sozialer Einstellungen vor, ist es sehr fraglich, ob dieser Weg zu einer Entscheidung über die Hypothese führt, dass Ausländerfeindlichkeit politisch bestimmt ist. Das Verfahren ist sicherlich brauchbar, wenn man etwas über die Struktur von Ausländerfeindlichkeit wissen möchte; für die Überprüfung einer Hypothese über Ursachen der Ausländerfeindlichkeit ist es jedoch wenig geeignet.

Es ist deshalb wichtig, sich vor Untersuchungsbeginn alle denkbaren Ausgänge der statistischen Analyse vor Augen zu führen, um zu entscheiden, welche Ergebnisse eindeutig für und welche Ergebnisse eindeutig gegen die inhaltliche Hypothese sprechen (vgl. hierzu auch das Good-enough-Prinzip, ► S. 28 f.). Die Untersuchungsplanung ist unvollständig oder falsch, wenn diese gedankliche Vorarbeit zu dem Resultat führt, dass eigentlich jedes statistische Ergebnis (z. B. weil die entscheidenden Variablen schlecht operationalisiert wurden) oder überhaupt kein Ergebnis (weil z. B. nicht auszuschließen ist, dass andere, nicht kontrollierte Variablen – sog. Confounder – für das Ergebnis verantwortlich sind) eindeutig im Sinne der Hypothese gedeutet werden kann. Eine empirische Untersuchung ist unwissenschaftlich, wenn sie nur die Vorstellungen des Autors, die dieser schon vor Beginn der Untersuchung hatte, verbreiten will und deshalb so angelegt ist, dass die Widerlegung der eigenen Hypothesen von vornherein erschwert oder gar ausgeschlossen ist.

Exposé und Gesamtplanung

Die Planungsarbeit endet mit der Anfertigung eines schriftlichen Berichtes über die einzelnen Planungsschritte bzw. mit einem Exposé. Die hier erörterte Gliederung für die Untersuchungsplanung vermittelt lediglich Hinweise und muss natürlich nicht für jeden Untersuchungsentwurf vollständig übernommen werden. Je nach Art der Fragestellung wird man der Auswahl der Untersuchungsart, der Untersuchungsobjekte oder Fragen der Operationalisierung mehr Raum widmen. Auf jeden Fall aber sollte das Exposé mit der wichtigsten Literatur beginnen und – zumindest bei hypothesenprüfenden Untersuchungen – mit Bemerkungen über die statistische Auswertung und deren Interpretation enden. Im übrigen bildet ein ausführliches und sorgfältiges Exposé nicht nur für die Durchführung der Untersuchung, sondern auch für die spätere Anfertigung des Untersuchungsberichtes eine gute Grundlage.

Nach Abschluss der Planung wird die Untersuchung mit ihrem endgültigen Titel versehen. Dieser kann mit dem ursprünglichen Arbeitstitel übereinstimmen oder aber – wenn sich in der Planung neue Schwerpunkte herausgebildet haben – umformuliert oder präzisiert werden.

Dem Exposé wird ein Anhang beigelegt, der die zeitliche (bei größeren Untersuchungsvorhaben auch die personelle, räumliche und finanzielle) Gesamtplanung enthält. Es müssen Zeiten festgesetzt werden, die für die Entwicklung und Bereitstellung der Untersuchungsinstrumente, die Anwerbung und Auswahl der Untersuchungsteilnehmer, die eigentliche Durchführung der Untersuchung (einschließlich Pufferzeiten für evtl. auftretende Pannen), die Dateneingabe, die Datenanalyse, eine letzte Literaturdurchsicht, die Interpretation der Ergebnisse, die Abfassung des Untersuchungsberichtes sowie die Aufstellung des Literaturverzeichnisses und evtl. notwendiger Anhänge erforderlich sind. ■ Box 2.7 enthält ein Beispiel für die Terminplanung einer Jahresarbeit.

Das Exposé stellt als Zusammenfassung der Planungsarbeiten eine wichtige »Visitenkarte« dar, die einen guten Einblick in das Untersuchungsvorhaben vermitteln sollte.

Für die Beantragung größerer Projekte haben einzelne Förderinstitutionen Antragsrichtlinien festgelegt, die vor Antragstellung angefordert werden sollten. Adressen forschungsfördernder Einrichtungen sind im

Box 2.7

Terminplanung für eine Jahresarbeit (Beispiel)

Befristete Arbeiten geraten zuweilen zum Ende hin in erhebliche Zeitnot, weil der Arbeitsaufwand falsch eingeschätzt und ein ungünstiges Zeitmanagement betrieben wurde. Eine sorgfältige, detaillierte Terminplanung kann solchen Schwierigkeiten vorbeugen. Das folgende Beispiel eines Zeitplans bezieht sich auf eine hypothesenprüfende Jahresarbeit.

- **1. Mai bis 1. Juli**

Literatursammlung und Literaturstudium; Anfertigung einer Problemskizze und eines Exposés; Ableitung der Untersuchungshypothesen und erste Vorüberlegungen zur Operationalisierung der beteiligten Variablen.

- **1. Juli bis 15. Juli**

Schriftliche Ausarbeitung des Theorieteils; Literaturverzeichnis mit den verwendeten Quellen erstellen.

- **15. Juli bis 1. August**

Präzisierung der Operationalisierung der einzelnen Variablen; Entwicklung und Bereitstellung der Untersuchungsinstrumente; weitergehende Überlegungen zur Stichprobe, zum Treatment und zur Untersuchungsdurchführung; Raumfrage klären; Versuchsleiterfrage klären.

- **1. August bis 15. August**

Schriftliche Ausarbeitung des Methodenteils; neben empirischen Fragen der Datenerhebung werden auch die statistischen Auswertungsverfahren festgelegt.

- **15. August bis 15. September**

Urlaub.

- **15. September bis 15. Oktober**

Kleine Voruntersuchungen zur Überprüfung der Untersuchungsinstrumente; Auswertung der Vorversuche und ggf. Revision der Instrumente; Einweisung der Versuchsleiter; Methodenteil vervollständigen und aktualisieren.

- **15. Oktober bis 15. November**

Anwerbung der Versuchspersonen (Aushänge in der Uni etc.); Durchführung der Untersuchungen.

- **15. November bis 1. Dezember**

Datenkodierung; Dateneingabe; Datenbereinigung.

- **1. Dezember bis 20. Dezember**

Kenntnisse über die erforderlichen statistischen Verfahren auffrischen; Testläufe mit der benötigten Statistiksoftware durchführen; Stichprobendeskription; statistische Datenauswertung.

- **20. Dezember bis 10. Januar**

Urlaub.

- **10. Januar bis 1. Februar**

Schriftliche Ausarbeitung des Ergebnisteils; zusammenfassende Darstellung der Ergebnisse zu den einzelnen Hypothesen; Tabellen und Grafiken anfertigen; ggf. Anhänge mit ausführlichem Datenmaterial zusammenstellen.

- **1. Februar bis 1. März**

Interpretation der Ergebnisse in enger Anlehnung an Theorie-, Methoden- und Ergebnisteil, sodass ein »roter Faden« deutlich wird; schriftliche Ausarbeitung der Interpretationen, die teils in den Ergebnisteil, teils in die abschließende Diskussion einfließen.

- **1. März bis 15. März**

Überarbeitung und letzte Ergänzungen des Untersuchungsberichtes; Anfertigung von Inhaltsverzeichnis, Tabellen- und Abbildungsverzeichnis; Einleitung und Zusammenfassung schreiben; Formatierung des Textes und Ausdruck.

- **15. März bis 1. April**

Korrektur lesen und lesen lassen nach Inhalt, Sprach- und Stilfehlern sowie Formatierungsfehlern; Korrekturen eingeben.

- **1. April bis 10. April**

Endausdruck; Arbeit in den Copyshop bringen; mehrere Kopien anfertigen und binden lassen.

- **28. April**

Abgabetermin.

► Anhang E wiedergegeben. Weitere Anschriften sind Oeckl (2000/2001) zu entnehmen. In der Regel werden auch für Qualifikationsarbeiten (Diplom-, Bachelor-, Master-, Doktorarbeiten) Exposés gefordert, deren Modalitäten von den jeweiligen Betreuerinnen und Betreuern bzw. von Prüfungsordnungen festgelegt werden.

2.4 Theoretischer Teil der Arbeit

Nach abgeschlossener Planungsarbeit will man verständlicherweise möglichst schnell zur konkreten Durchführung der Untersuchung kommen. Dennoch ist es ratsam, bereits jetzt den theoretischen Teil der Arbeit (oder zumindest eine vorläufige Version) zu schreiben. Hierfür sprechen zwei wichtige Gründe:

- Der erste Grund betrifft die **Arbeitsökonomie**. Nachdem gerade das Exposé angefertigt wurde, dürfte dessen erster Teil, die theoretische Einführung in das Problem sowie die Literaturskizze, noch gut im Gedächtnis sein. Es sollte deshalb keine besonderen Schwierigkeiten bereiten, den Literaturbericht und – falls der Forschungsgegenstand dies zulässt – die Herleitung der Hypothesen schriftlich niederzulegen.
- Der zweite wissenschaftsimmanente Grund ist schwerwiegender. Solange noch keine eigenen Daten erhoben wurden, kann man sicher sein, dass die Herleitung der Hypothesen oder auch nur Nuancen ihrer Formulierung von den eigenen Untersuchungsbeobachtungen unbeeinflusst sind. Forscherinnen und Forscher dürfen zurecht daran interessiert sein, ihre eigenen Hypothesen zu bestätigen. Legen sie aber die Hypothesen erst nach abgeschlossener Untersuchung schriftlich fest, ist die Versuchung nicht zu leugnen, die Formulierung der Hypothesen so zu akzentuieren, dass deren Bestätigung zur reinen Formsache wird. Den theoretischen Teil einschließlich der Hypothesenherleitung und -formulierung vor Durchführung der Untersuchung abzufassen, ist damit der beste Garant für die **Unabhängigkeit von Hypothesenformulierung und Hypothesenprüfung**.

Der theoretische Teil beginnt mit der Darstellung des inhaltlichen Problems. Es folgt der Literaturbericht, der jedoch die einschlägigen Forschungsbeiträge nicht wahllos aneinanderreihet, sondern kommentierend verbindet

und integriert. Eventuell vorhandene Widersprüche sind zu diskutieren und Informationen, die für die eigene Problematik nur peripher erscheinen, durch inhaltliche Akzentsetzungen auszugrenzen.

Detaillierte Hinweise zur Methodik, den Untersuchungsobjekten oder Erhebungsinstrumenten, die in den zitierten Untersuchungen verwendet wurden, sind erforderlich, wenn die eigene Arbeit hierauf unmittelbar Bezug nimmt. Sie sind auch dann unverzichtbar, wenn die integrierende Diskussion der Forschungsergebnisse andere als vom jeweiligen Autor vorgeschlagene Interpretationen nahe legt.

Die sich anschließende Zusammenfassung des Literaturteils kennzeichnet und bewertet den Stand der Theoriebildung. (Ein spezielles Verfahren zur Integration von Forschungsergebnissen ist die sog. »Metaanalyse«, die in ► Kap. 10 behandelt wird.) Der theoretische Teil endet mit der Ableitung theoretisch begründeter inhaltlicher Hypothesen bzw. der Formulierung statistischer Hypothesen. (Weitere Hinweise zur Literaturarbeit findet man in ► Abschn. 6.2. und auf S. 87 f.).

Auch bei explorativen Studien sollten die theoretischen Überlegungen vor der empirischen Phase abgeschlossen sein. Es wird schriftlich festgelegt, was die Beschäftigung mit dem Untersuchungsgegenstand auslöste, welches Problem die Forschung erforderlich machte, unter welchem Blickwinkel es betrachtet wurde und ggf. in welcher wissenschaftlichen Tradition die Arbeit steht. Dadurch entgeht man der Gefahr, während der Arbeit am Thema die ursprüngliche Fragestellung aus den Augen zu verlieren oder zu modifizieren. Legen die Erfahrungen bei ersten empirischen Schritten eine Veränderung der Forschungsstrategie nahe, so muss dieses dokumentiert werden. Wenn möglich, sollte man schriftliche Ausarbeitungen sachkundigen Korrekturlesern und Kommentatoren vorlegen, um deren Veränderungsvorschläge berücksichtigen zu können.

2.5 Durchführung der Untersuchung

Ist eine Untersuchung sorgfältig und detailliert geplant, dürfte ihre Durchführung keine besonderen Schwierigkeiten bereiten. Was aber durch Planung als potenzielle Störquelle nicht völlig ausgeschlossen werden kann, sind Fehler im eigenen Verhalten bzw. im Verhalten von Drit-

Box 2.8

Der Kluge Hans

Der Kluge Hans war ein anscheinend überaus begabtes Pferd, das die Grundrechenarten beherrschte, lesen und buchstabieren sowie Töne auf der Tonleiter identifizieren konnte. Wilhelm von Osten, ein pensionierter Lehrer, hatte das Pferd gekauft und trainiert und trat mit ihm zwischen 1900 und 1904 bei öffentlichen Veranstaltungen in Berlin auf. Die Aufgaben wurden dem Klugen Hans auf Tafeln präsentiert und das Pferd antwortete beispielsweise auf Rechenaufgaben, indem es die richtige Zahl mit dem Huf auf den Boden klopfte. Der Kluge Hans wurde als überaus intelligentes und kommunikationsfähiges Pferd weltberühmt und Wilhelm von Osten versicherte stets, dass kein Zirkustrick hinter der erstaunlichen Leistung stecke.

Der Psychologe Oskar Pfungst war skeptisch und entwickelte zusammen mit seinem Fachkollegen Carl Stumpf einen systematischen Untersuchungsplan, um auf der Basis empirischer Versuche herauszufinden, wie die besonderen Leistungen des Pferdes zustande kommen. Im Rahmen der Versuche stellte sich heraus, dass der Kluge Hans nur dann richtig antworten konnte, wenn sich die fragende Person für ihn gut sichtbar in der Nähe aufhielt und die richtige Antwort kannte. Offensichtlich konnte das Pferd die Aufgaben doch nicht durch eigenständiges Denken lösen, sondern stützte sich auf externe Hinweise. Auf den ersten Blick

freilich konnte das Publikum nicht erkennen, dass dem Pferd Zeichen gegeben wurden.

Die Versuche von Pfungst belegten jedoch, dass das Pferd sogar kleinste mimische Veränderungen bemerken und darauf entsprechend reagieren konnte (Pfungst, 1907). Der Kommunikationsprozess zwischen Versuchsleiter und Pferd wurde folgendermaßen rekonstruiert: Der Versuchsleiter stellt dem Pferd eine Rechenaufgabe und verfolgt dann gespannt, wie das Pferd durch Klopfen antwortet. So lange die richtige Zahl noch nicht erreicht ist, befindet sich der Versuchsleiter in einer angespannten, neugierigen Verfassung. In dem Moment, in dem das Pferd bei der richtigen Zahl angekommen ist, reagiert der Versuchsleiter unwillkürlich mit innerer Erleichterung und Entspannung. Obwohl der Versuchsleiter dem Pferd keinen bewussten Wink gab, konnte der Kluge Hans die unwillkürlichen mimischen Entspannungssignale aufnehmen und stoppte seine Klopfzeichen somit genau bei der richtigen Zahl.


Der Nachweis derart subtiler, ungeplanter Einflüsse des Versuchsleiters auf ein Tier wurde zum Anlass genommen, auch eine entsprechende Beeinflussung von menschlichen Versuchsteilnehmern ernst zu nehmen. Um solche Versuchsleiterartefakte zu vermeiden, werden »Blindversuche« durchgeführt, bei denen Versuchsleiter nicht wissen, welche Reaktionen der Versuchspersonen im Sinne der Hypothesen wünschenswert sind.

ten, die als Versuchsleiter, Interviewer, Testinstruktoren etc. engagiert werden. Die Literatur spricht in diesem Zusammenhang von Versuchsleiterartefakten.

2.5.1 Versuchsleiterartefakte

Schon die Art und Weise, wie der Untersuchungsleiter die Untersuchungsteilnehmer begrüßt, vermittelt den Teilnehmern einen ersten Eindruck von der für sie in der Regel ungewöhnlichen Situation und kann damit das spätere Untersuchungsverhalten beeinflussen. Eigenarten der dann üblicherweise folgenden Instruktionen sind

ebenfalls ausschlaggebend dafür, wie die Untersuchungsteilnehmer die ihnen gestellten Aufgaben erledigen. Ferner kann es von Bedeutung sein, in welcher emotionalen Atmosphäre die Untersuchung abläuft.

Auf die emotionale Atmosphäre kann der Untersuchungsleiter durch nonverbale Signale massiv Einfluss nehmen. Häufige Blickkontakte und räumliche Nähe gelten als Anzeichen für Sympathie und fördern die Überzeugungskraft der verbalen Äußerungen des Untersuchungsleiters. Dass nonverbale Kommunikation nicht nur das Verhalten eines menschlichen Gegenübers, sondern auch das eines Tieres in unbeabsichtigter Weise beeinflussen kann, zeigt  Box 2.8.

Die Liste möglicher Eigenarten und Verhaltensbesonderheiten des Untersuchungsleiters, die den Ausgang einer Untersuchung beeinflussen, könnte beinahe beliebig fortgesetzt werden. Die Forschung zu den mit dem Namen Rosenthal eng verbundenen Versuchsleiterartefakten oder »Rosenthal-Effekten« füllt inzwischen zahlreiche zusammenfassende Werke, von denen hier lediglich Bungard (1980) und Rosenthal (1976) erwähnt seien.

2.5.2 Praktische Konsequenzen

Die Forschung über Versuchsleiterartefakte belegt zweifelsfrei, dass das Verhalten des Untersuchungsleiters die Ergebnisse seiner Untersuchung beeinflussen kann. Es steht ferner außer Zweifel, dass einige empirisch bestätigte Theorien auf Untersuchungen beruhen, deren Ergebnisse man auch als Versuchsleiterartefakte erklären kann (vgl. Bungard, 1980). Für denjenigen, der mit der konkreten Durchführung seiner Untersuchung befasst ist, gibt diese Forschungsrichtung jedoch nur wenig her. Es ist bisher unmöglich – und wird wohl auch bis auf weiteres unmöglich bleiben –, die Bedeutung der individuellen Eigenarten eines Untersuchungsleiters für eine konkrete Untersuchung vollständig zu erfassen.

Brandt (1971, 1975) sieht in Untersuchungen zur Überprüfung von Versuchsleiterartefakten den Anfang eines unendlichen Regresses, der darin besteht, dass diese Untersuchungen wiederum von Versuchsleitern mit persönlichen Eigenarten durchgeführt werden, die ihrerseits die Untersuchungsergebnisse beeinflussen können und so fort. Sein Vorschlag, die Abhängigkeit der »Messergebnisse« vom Messinstrument »Mensch« (Bridgman, 1959, S. 169) durch die Einbeziehung weiterer Versuchsleiter als neutrale Beobachter des Untersuchungsgeschehens zu reduzieren, kann zumindest für die meisten studentischen Untersuchungen nur als Notlösung bezeichnet werden.

Eine Maßnahme, die die Beeinträchtigung der internen Validität von Untersuchungen durch Versuchsleiterartefakte in Grenzen hält, ist die **Standardisierung der Untersuchungsbedingungen** und vor allem des Versuchsleiterverhaltens (► unten). Wenn – so lässt sich argumentieren – das Verhalten des Versuchsleiters z. B. bei der Instruktion einer Experimental- und einer Kon-

trollgruppe standardisiert ist (z. B. durch Instruktionen per Tonband oder Video), sind Unterschiede zwischen den verglichenen Gruppen nicht als Versuchsleiterartefakte erklärbar.

Dieses Konzept der Standardisierung erfährt durch Kebeck und Lohaus (1985) eine interessante Erweiterung: Sie votieren für ein individuumzentriertes Versuchsleiterverhalten, das sich am Erleben des Untersuchungsteilnehmers orientiert. Aufgabe des Versuchsleiters sei es, die experimentelle Situation so zu gestalten, dass sie von allen Untersuchungsteilnehmern möglichst gleich erlebt wird. Wenn dieses Ziel nur dadurch erreicht werden kann, dass der Versuchsleiter in seinem Verhalten auf individuelle Besonderheiten einzelner Untersuchungsteilnehmer eingeht, so sei dies zu akzeptieren. Das Standardisierungskonzept wird damit also aus der Sicht der Untersuchungsteilnehmer definiert. Auch wenn dieses Standardisierungskonzept, dem vor allem die qualitative Forschung folgt (► Abschn. 5.2.4), theoretisch einleuchtet, muss man befürchten, dass seine praktische Umsetzung nicht unproblematisch ist.

Offensichtlich müssen wir uns mit einer gewissen, letztlich nicht mehr reduzierbaren Ungenauigkeit unserer Untersuchungsergebnisse abfinden. Barber (1972, 1976), der zu den schärfsten Kritikern der durch Rosenthal initiierten Forschungsrichtung zählt, nennt statt eines »Experimentatoreffektes« weitere Effekte, die potenziell Untersuchungsergebnisse beeinflussen oder verfälschen können. Diese Effekte basieren auf der Spannung zwischen einem Projektleiter (Investigator), der für die Untersuchungsplanung und ggf. auch für die Auswertung zuständig ist, und einem für die Untersuchungsdurchführung verantwortlichen Untersuchungsleiter (Experimentator). In der Evaluationsforschung besteht zudem die Gefahr, dass Verpflichtungen gegenüber dem Auftraggeber (bewusst oder unbewusst) die Ergebnisse beeinflussen (► Abschn. 3.1.1).

2.5.3 Empfehlungen

Die folgenden Maßnahmen, deren Realisierbarkeit und Bedeutung natürlich von der Art der Fragestellung und den Untersuchungsumständen abhängen, können dazu beitragen, den Einfluss der eigenen Person oder des Untersuchungsumfeldes auf das Verhalten der Untersu-

chungsteilnehmer (»Demand-Characteristics«, Orne, 1962) gering zu halten. Wichtig ist hierbei der Leitgedanke, dass störende Untersuchungsbedingungen für die Ergebnisse weniger erheblich sind, wenn alle Untersuchungsteilnehmer ihrem Einfluss in gleicher Weise ausgesetzt sind. Konstante störende Bedingungen mindern zwar die Generalisierbarkeit (externe Validität), aber nicht zwangsläufig die Eindeutigkeit der mit der Untersuchung gewonnenen Erkenntnisse (interne Validität).

Die größte Gefährdung einer gleichmäßigen Wirkung störender Untersuchungsbedingungen auf alle Versuchspersonen in allen Untersuchungsgruppen besteht in der Kenntnis der Untersuchungshypothese, die uns unbewusst veranlassen mag, Treatment- und Kontrollgruppe unterschiedlich zu behandeln. Dieses Problem wird ausgeschaltet, wenn die Untersuchungsdurchführung von Helfern übernommen wird, die die Untersuchungshypothese nicht kennen, also der Hypothese gegenüber – ebenso wie die Versuchspersonen – »blind« sind. Man spricht in diesem Zusammenhang auch von **Doppelblindversuchen** (»blinde« Versuchspersonen und »blinde« Versuchsleiter).

Da die Versuchsdurchführung in der Regel ein sehr mühsames Geschäft ist und viele Tage Arbeit bedeutet, wird man aus ökonomischen Gründen – gerade bei Qualifikationsarbeiten – kaum den Luxus externer Versuchsleiter genießen können, sondern stattdessen selbst in Aktion treten müssen. Die folgenden Empfehlungen sollen helfen, Versuchsleitereffekte möglichst gering zu halten.

- Alle Untersuchungsteilnehmer erhalten dieselbe Instruktion, die möglichst standardisiert (z. B. per Tonband- oder Videoaufzeichnung) bzw. schriftlich vorgegeben wird. Sind in quasiexperimentellen oder experimentellen Untersuchungen verschiedene Instruktionen erforderlich (z. B. für die Experimentalgruppe und die Kontrollgruppe), repräsentieren die Instruktionsunterschiede in all ihren Feinheiten die unabhängige Variable.
- Führt eine standardisierte Instruktion bei einzelnen Untersuchungsteilnehmern zu Verständnisproblemen, sind diese individuell auszuräumen (zur Formulierung von Instruktionen vgl. auch Hager et al., 2001, S. 47 ff.).
- Wird eine Untersuchung mit Laborcharakter geplant (► Abschn. 2.3.3), ist auf konstante Untersuchungsbedingungen zu achten. Hierzu zählen Räume, Be-

leuchtung, Geräusche, Arbeitsmaterial, die Temperatur etc., aber auch die äußere Erscheinung (z. B. neutrale Kleidung) des Untersuchungsleiters.

- Zwischenfragen oder andere unerwartete Vorkommnisse während des Untersuchungsablaufes müssen protokolliert werden.
- Besteht die Untersuchung aus mehreren Teilschritten (oder aus mehreren Einzelaufgaben und Fragen), ist deren Abfolge konstant zu halten, es sei denn, man will durch systematische Variation Sequenzeffekte prüfen (► S. 550).
- Erwartet der Untersuchungsleiter bestimmte Ergebnisse, muss er mit eigenen ungewollten nonverbalen Reaktionen rechnen, wenn sich eine Bestätigung seiner Hypothese (oder widersprüchliche Ergebnisse) während des Untersuchungsablaufes abzeichnen. Es sollte deshalb geprüft werden, ob die Untersuchung so angelegt werden kann, dass der Untersuchungsleiter die Ergebnisse der Untersuchungsteilnehmer erst nach Abschluss der Untersuchung erfährt.
- Ursachen für mögliche Pannen, Belastungen der Untersuchungsteilnehmer, störende Reize, ethische Gefährdungen u. Ä. erkennt der Untersuchungsleiter am besten, wenn er den gesamten Untersuchungsablauf zuvor an sich selbst überprüft.
- Eine ähnliche Funktion hat das »Non-Experiment« (Riecken, 1962). Hier werden Personen, die aus derselben Population stammen wie die eigentlichen Untersuchungsteilnehmer, gebeten, den gesamten in Aussicht genommenen Untersuchungsablauf vorzutesten.
- Nach Abschluss des offiziellen Teiles der Untersuchung ist eine Nachbefragung der Untersuchungsteilnehmer zu empfehlen. Sie soll Aufschluss über Empfindungen, Stimmungen, Schwierigkeiten, Aufrichtigkeit, Interesse, Wirkung des Untersuchungsleiters u. Ä. liefern.
- Falls möglich, sollte der gesamte Untersuchungsablauf mit einem Videogerät aufgezeichnet werden. Diese Aufzeichnung kann später auf mögliche Untersuchungsfehler hin analysiert werden.
- Experimente, die computergestützt bzw. im Internet stattfinden und bei denen der Ablauf nicht von einem menschlichen Versuchsleiter, sondern von einem Programm gesteuert wird, sind gegen Versuchsleitereffekte immun, bergen aber wiederum andere spezifische Probleme (Janetzko et al. 2002).

- Sowohl die Untersuchungsumstände als auch sämtliche bewusst in Kauf genommenen oder unerwartet eingetretenen Unregelmäßigkeiten werden in einem abschließenden Untersuchungsprotokoll aufgenommen. Dieses ist – in verkürzter Form – Bestandteil des späteren Untersuchungsberichtes.

2.6 Auswertung der Daten

Die Auswertung des Untersuchungsmaterials erfolgt nach den Vorgaben des Planungsberichtes. Im Mittelpunkt der Auswertung hypothesenprüfender Untersuchungen stehen statistische Signifikanztests, deren Ausgang die Entscheidungsgrundlage dafür ist, ob die forschungsleitende Hypothese als bestätigt gelten oder abgelehnt werden soll. Die inhaltliche Interpretation der Ergebnisse nimmt auf die Theorie Bezug, aus der die Hypothese abgeleitet wurde. Signifikante Ergebnisse bestätigen (vorläufig) die Theorie und nichtsignifikante Ergebnisse schränken (beim Testen von »Minimum-Effekt-Nullhypothesen«; ▶ S. 635 ff.) ihren Geltungsbereich ein. Die Ergebnisse von Auswertungen, die über die eigentliche Hypothesenprüfung hinausgehen, sind explorativ und müssen auch in dieser Weise dargestellt werden.

Vor Beginn der Hypothesenprüfung sollte man mit Hilfe eines Statistikprogramms versuchen, Eingabefehler zu identifizieren bzw. den Datensatz um Fehler zu bereinigen.

Eingabefehler sind oft Werte, die außerhalb des zulässigen Wertebereichs einer Variablen liegen. Hat eine Variable nur wenige Stufen (z. B. Geschlecht: 0, 1), lässt man sich mit einem geeigneten Befehl ausgeben, wie oft die Werte der betrachteten Variablen vorkommen (in SPSS könnte man hierzu den Frequencybefehl nutzen: »Frequency Geschlecht«). Erhielte man nun die Angabe, dass der Wert »0« (für männlich) 456-mal vorkommt, der Wert »1« 435-mal und der Wert »9« 3-mal, hat man damit bereits 3 Eingabefehler identifiziert. Nun lässt man sich die Nummern all derjenigen Fälle ausgeben, bei denen »Geschlecht=9« auftaucht. Bei diesen Personen muss man in den Originalfragebögen nachschauen, welches Geschlecht sie angegeben haben und die entsprechenden Angaben in der Datendatei ändern. Bei Variablen ohne exakt festgelegten Wertebereich (z. B. Alter) ist auf Extremwerte zu achten; so sind Altersangaben größer als 100 z. B. sehr unwahrscheinlich und sollten überprüft werden. Extremwerte springen auch bei graphischen Darstellungen ins Auge (▶ S. 372 ff.).

Hat man die ersten Eingangskontrollen durchlaufen, erstellt man üblicherweise zunächst eine Stichprobendes-kription, bevor man zu den Hypothesentests übergeht. Hierzu berichtet man für die gängigen sozialstatistischen bzw. soziodemographischen Merkmale (Geschlecht, Alter, Familienstand, Bildungsgrad, Tätigkeit, Einkommen, Wohnort etc.) Häufigkeitstabellen und Durchschnittswerte. Unplausible Merkmalsverteilungen können Hinweise auf Eingabe- oder Kodierungsfehler liefern.

Die Datenbereinigung sollte abgeschlossen sein, bevor mit den Hypothesenprüfungen begonnen wird. Stellt sich nämlich erst im nachhinein heraus, dass noch gravierende Kodierungs- oder Eingabefehler in den Daten stecken, müssen die Analysen wiederholt werden. Zudem bestünde die Gefahr, beim Bereinigen der Daten bewusst oder unbewusst im Sinne der eigenen Hypothesen vorzugehen. Dies betrifft auch die Frage, welche Fälle wegen fehlender oder fragwürdiger Angaben ggf. ganz aus den Analysen ausgeschlossen werden sollen.

Damit – und wenn Untersuchungsteilnehmer vollständig ausfielen und die ursprünglich vorgesehenen Stichprobenumfänge nicht realisiert werden konnten – entstehen **Missing-Data-Probleme**. Für die Auswertung derartiger unvollständiger Datensätze stehen spezielle Techniken zur Verfügung (vgl. z. B. Frane, 1976; Lösel & Wüstendörfer, 1974; Madow et al. 1983. Einen ausführlichen Überblick von Missing-Data-Techniken findet man bei Schafer und Graham, 2002, bzw. West, 2001.)

In hypothesenerkundenden Untersuchungen besteht die Auswertung üblicherweise in der Zusammenfassung der erhobenen Daten in statistischen Kennwerten, Tabellen oder Graphiken, die ggf. als Beleg für eine neu zu formulierende Hypothese herangezogen werden (▶ Abschn. 6.4). Am hypothetischen Charakter der Untersuchungsbefunde ändert sich nichts, wenn sich evtl. gefundene Mittelwertunterschiede, Häufigkeitsunterschiede, Korrelationen o. Ä. als statistisch signifikant erweisen sollten (▶ S. 379 f.).

Nicht jede Untersuchung führt zu den erhofften Ergebnissen. Widersprüchliche Ergebnisse, die in Erkundungsstudien keine eindeutige Hypothesenbildung zulassen, und Untersuchungsbefunde, die die Ablehnung zuvor aufgestellter Hypothesen erfordern, sollten uns veranlassen, den Untersuchungsaufbau, die Untersuchungsdurchführung und die statistische Auswertung nochmals kritisch nach möglichen Fehlern zu durchsu-

chen. Sind evtl. entdeckte Fehler nicht mehr korrigierbar, sollten sie offen dargelegt und in ihren Konsequenzen diskutiert werden. Nachträgliche Bemühungen, den Daten unabhängig von den Hypothesen »etwas Brauchbares« zu entnehmen, sind – wenn überhaupt – in einen gesonderten, hypothesenerkundenden Teil aufzunehmen. Hierbei ist die von Dörner (1983) vorgeschlagene »Methode der theoretischen Konsistenz« hilfreich.

2.7 Anfertigung des Untersuchungsberichtes

Der Untersuchungsplan, die bereits vorliegende Aufarbeitung der einschlägigen Literatur (evtl. einschließlich der Herleitung von Hypothesen), die Materialien der Untersuchung, das Protokoll des Untersuchungsablaufes, Tabellen und Computerausdrucke mit den Ergebnissen sowie einzelne Anmerkungen zur Interpretation sind das Gerüst des endgültigen Untersuchungsberichtes. Für die Anfertigung dieses Berichtes gelten – speziell für hypothesenprüfende Untersuchungen – einige Regeln, die möglichst genau eingehalten werden sollten. Noch so gelungene Untersuchungen sind wenig tauglich, wenn es nicht gelingt, diese anschaulich, nachvollziehbar und vollständig zu vermitteln.

Die folgenden Ausführungen orientieren sich an den von der Deutschen Gesellschaft für Psychologie (1997, 2001) herausgegebenen »Richtlinien für die Manuskriptgestaltung« und an den Vorschriften der APA (American Psychological Association, 1994, 2005). Weitere Hinweise zu diesem Thema findet man bei Höge (2002) oder auch in einem vom Deutschen Institut für Normung e.V. (1983) unter DIN 1422 herausgegebenen Informationsblatt.

2.7.1 Gliederung und Inhaltsverzeichnis

Eine empirische Studie gliedert sich in die Hauptteile

- Einleitung,
- Forschungsstand und Theorie,
- Methode,
- Ergebnisse,
- Diskussion,
- Literatur.

Der Einleitung voranzustellen sind Titelseite, Abstract (Kurzzusammenfassung) und Inhaltsverzeichnis sowie ggf. noch ein Tabellen-, ein Abbildungs- und ein Abkürzungsverzeichnis. Am Ende der Arbeit können hinter dem Literaturverzeichnis Anhänge folgen, die z. B. Untersuchungsmaterialien enthalten. Manchmal werden ergänzend auch ein Glossar und ein Personen- und Sachregister (Index) angeboten.

Es zeichnet sich in den Sozialwissenschaften der Trend ab, bereits studentische Qualifikationsarbeiten nach den internationalen Standards für wissenschaftliche Publikationen erstellen zu lassen. Der Idealfall ist dabei die englischsprachige Qualifikationsarbeit, die – mit allenfalls geringen Veränderungen – in einer internationalen Fachzeitschrift publiziert werden kann. Wir halten es für sinnvoll, anstelle persönlicher Dozentenvorlieben, lokaler oder nationaler Konventionen wo immer möglich die internationalen Standards der Scientific Community anzulegen, weil dies zur Qualitätssicherung beiträgt und auf die spätere Wissenschaftspraxis optimal vorbereitet.

Für Gliederung und Inhaltsverzeichnis bedeutet dies, dass man sich an einem relativ standardisierten Raster orientiert. Die gliedernden Überschriften einer Arbeit sind gemäß einem Dezimalsystem zu nummerieren, wobei drei Gliederungsebenen gängig sind. Unterkapitel sollten nur dann mit nummerierten Zwischenüberschriften versehen werden, wenn mindestens zwei Unterkapitel auf derselben Gliederungsebene existieren. Lässt sich ein einzelner Unterbereich nur schwer in einen Hauptbereich integrieren, besteht die Möglichkeit, diesen als Exkurs aus dem normalen Gliederungsschema herauszunehmen.

■ Box 2.9 zeigt den Aufbau eines prototypischen Inhaltsverzeichnisses. In vielen Zeitschriftenartikeln entsprechen die Hauptüberschriften wörtlich dieser Vorlage. Der hohe Standardisierungsgrad mag auf den ersten Blick langweilig erscheinen. Tatsächlich aber hat sich diese standardisierte Gliederung bewährt, da sie der Leserschaft stets eine leichte Orientierung im Text ermöglicht. Kreativität sollte sich also in den Inhalten der Arbeit niederschlagen, nicht in einem gewollt originellen Aufbau.

Box 2.9**Beispiel für Aufbau und Strukturierung einer hypothesenprüfenden empirischen Untersuchung**

- | | |
|---|--|
| <ul style="list-style-type: none"> Titelblatt Inhaltsverzeichnis Abstract (deutsch und englisch) Einleitung 1. Forschungsstand und Theorie <ul style="list-style-type: none"> 1.1 Theoretischer und empirischer Forschungsstand zum Thema 1.2 Theoretisches Modell der Studie 1.3 Fragestellungen und Hypothesen 2. Methode | <ul style="list-style-type: none"> 2.1 Untersuchungsdesign 2.2 Instrumente und Messgeräte 2.3 Stichprobenkonstruktion 2.4 Untersuchungsdurchführung 2.5 Datenanalyse 3. Ergebnisse <ul style="list-style-type: none"> 3.1 Stichprobenbeschreibung 3.2 Ergebnisse zu den einzelnen Fragestellungen und Hypothesen 3.3 Weitere Befunde 4. Diskussion 5. Literatur Anhänge |
|---|--|

2.7.2 Die Hauptbereiche des Textes

Zu den einzelnen Hauptbereichen des empirischen Forschungsberichtes werden im Folgenden jeweils Anregungen vermittelt. Von diesen Hinweisen kann in begründeten Fällen abgewichen werden. Eine ausgesprochen anschauliche und inspirierende Anleitung zum Verfassen wissenschaftlicher Artikel gemäß internationalen Standards liefert Bem (2003). Bem verwendet für den Aufbau eines Forschungsberichtes die Metapher der Sanduhr: Während Anfang (Abstract, Einleitung, Theorie) und Ende der Arbeit (Diskussion) das Thema relativ breit behandeln, muss die Darstellung im Mittelteil (Methode, Ergebnisse) sehr eng auf Detaildarstellungen zugespißt sein.

Abstract

Das Abstract ist eine Kurzzusammenfassung, die in nur 100 bis 120 Worten Thema, Theorie, Methode und Hauptergebnisse zusammenfasst. Für jeden Hauptteil der Arbeit sind also nur ein bis zwei Sätze vorgesehen. Eine derartige Komprimierung einer ganzen Studie erfordert Formulierungskunst. Das Abstract wird erst nach Fertigstellung des gesamten Berichtes geschrieben und in der Regel mehrfach überarbeitet, bis eine bündige Endfassung vorliegt. Das Abstract ist in deutscher und in englischer Sprache zu verfassen, denn Abstracts dienen dazu, lokale und nationale Forschungstätigkeiten international kommunizierbar zu machen. In wachsen-

dem Maße werden die Abstracts studentischer Abschlussarbeiten in Datenbanken aufgenommen. Neben dem hier dargestellten Abstractformat existieren noch sog. Extended Abstracts mit einem Wortumfang von 300 oder auch 800 und mehr Worten. Extended Abstracts werden teilweise bei wissenschaftlichen Konferenzen verlangt, wenn man einen Vortrag anmelden möchte.

Einleitung

Die Einleitung macht deutlich, warum das gewählte Forschungsthema interessant und relevant ist. Sie darf mit einer Anekdote, einem Sprichwort, einem Beispiel, einem Witz oder einem Prominentenzitat beginnen. Denn die Einleitung ist der Türöffner zur Arbeit. Sie soll Interesse wecken und die besonderen inhaltlichen, theoretischen oder methodischen Merkmale der Studie hervorheben.

Forschungsstand und Theorie

Wissenschaftliche Arbeiten setzen immer am bisherigen Forschungsstand an, den es gründlich zu recherchieren gilt. Bevor man behauptet, das gewählte Thema sei bislang von der Forschung vernachlässigt worden, sollten alle Recherchemöglichkeiten ausgeschöpft sein. Ein tatsächlicher Mangel an früheren Untersuchungen ist bei innovativen Themen allerdings durchaus denkbar. Man spricht von »Forschungsdesideraten«, wenn man auf Inhalte hinweist, die bislang ungenügend erforscht wurden und deren Untersuchung wünschenswert erscheint.

Bei der Zusammenfassung des Forschungsstandes orientiert man sich am besten an möglichst aktuellen Übersichtsartikeln. Es gilt herauszuarbeiten, welche Theorien, Methoden und Befunde im Zusammenhang mit dem gewählten Untersuchungsthema in der Literatur auftauchen. Die Zusammenfassung des Forschungsstandes sollte umfassend, aber nicht langatmig ausfallen. Die wichtigsten theoretischen Ansätze sind zu identifizieren. Dabei darf keine seitenlange Nacherzählung der Theorien erfolgen. Stattdessen sollten Kurzcharakterisierungen der Theorien geboten werden. Wichtiger als eine pure Wiedergabe ist die kritische Reflexion, Selektion und Weiterentwicklung der bisherigen Theorien.

Die Arbeit mit und an vorliegenden Theorien mündet in ein eigenes theoretisches Modell, das der Arbeit zugrunde gelegt wird. Dieses eigene theoretische Modell kann beispielsweise Elemente von zwei vorliegenden Theorien miteinander verbinden und um neue Aspekte ergänzen.

Auf der Basis des selbst entwickelten theoretischen Modells, das möglichst auch in einer Grafik veranschaulicht wird, lassen sich dann die Fragestellungen und Hypothesen formulieren. Eine Hypothese über einen Unterschied, einen Zusammenhang oder eine Veränderung sollte nur dann aufgestellt werden, wenn man auf der Basis vorliegender Theorie und Empirie wirklich von dem postulierten Effekt überzeugt ist. Zur Selbstprüfung kann man sich überlegen, wie viel Geld man wetten würde, dass die Hypothese wirklich empirisch gestützt werden kann. Wo immer es nicht möglich oder nicht sinnvoll ist, bestimmte Effekte im Vorfeld mit guter Gewissheit zu postulieren, sollten anstelle von Hypothesen lieber Fragestellungen formuliert werden. Bei Fragestellungen wird kein konkreter Effekt postuliert, sondern nach der Ausprägung von Variablen gefragt.

Methoden

Der Methodenteil muss so exakt sein, dass andere, am gleichen Problem interessierte Forscherinnen und Forscher die Untersuchung nachstellen (replizieren) können. Der Methodenteil beginnt mit einer Charakterisierung des Untersuchungsdesigns (z. B., ob es sich um eine Querschnitt- oder Längsschnittstudie handelt, um eine experimentelle, quasiexperimentelle oder nicht-experimentelle Untersuchung etc.).

Es folgt die Beschreibung der Instrumente (z. B. Fragebögen, Tests, Interviewleitfäden, Beobachtungspläne) sowie – sofern eingesetzt – der Messgeräte und Untersuchungsmaterialien (z. B. psychophysiologische Messgeräte). Die Untersuchungsinstrumente sind jeweils mit ihren Gütekriterien (z. B. Reliabilität, Validität) zu charakterisieren. Handelt es sich um selbst entwickelte Instrumente, so ist der Konstruktionsprozess darzustellen (Originaldokumente wie Fragebogenvorform, Pretestergebnisse und Fragebogenendform sind in den Anhang auszulagern).

Ein weiterer wichtiger Abschnitt des Methodenteils ist die Stichprobenkonstruktion. Hier wird angegeben, wie die Stichprobe zusammengestellt bzw. ausgewählt werden soll (z. B. Gelegenheitsstichprobe, Quotenstichprobe, Zufallsstichprobe), wie groß der Stichprobenumfang sein soll und wie die Anwerbung der Untersuchungsteilnehmer (Rekrutierung) erfolgen soll.

Anschließend wird die Untersuchungsdurchführung beschrieben: Wann und wo erfolgte die Datenerhebung, welche besonderen Vorkommnisse traten auf, wie reagierten die Untersuchungsteilnehmer?

Zuweilen wird im Methodenteil auch auf die Datenanalyse eingegangen und z. B. kurz skizziert, mit welcher Auswertungssoftware und welchen statistischen Verfahren die erhobenen Daten ausgewertet wurden. Eine genaue Beschreibung der Methoden wie z. B. die Wiedergabe von Formeln ist hierbei nicht erforderlich; im Zweifelsfalle genügen Verweise auf einschlägige Statistikkbücher. Handelt es sich jedoch um Eigenentwicklungen oder um relativ neue, wenig bekannte Methoden, so sollten diese nachvollziehbar dargestellt werden. Generell wenden wir uns bei der Abfassung eines Forschungsberichtes an einen methodisch und fachlich vorinformierten Leserkreis. Grundlagen sollten also nicht vermittelt werden, da sonst eher der Charakter eines Lehrbuchtextes entsteht. Andererseits sollte der Text auch nicht nur hochspezialisierten Experten eines engumgrenzten Gebietes verständlich sein.

Ergebnisse

Der Ergebnisteil ist das Herzstück eines Forschungsberichtes. Denn hier werden neue Erkenntnisse dargestellt. Theoretische Vorüberlegungen und methodische Planungen sind lediglich Hilfsmittel für den angestrebten Erkenntnisgewinn. Der Ergebnisteil beginnt mit einer

Stichprobenbeschreibung, die über Merkmale und Zusammensetzung der untersuchten Personengruppen informiert (z. B. Alter, Geschlecht, Bildungsstand, Tätigkeit etc.). Neben allgemeinen soziodemografischen Variablen werden in der Stichprobenbeschreibung auch weitere für das Studienthema relevante Merkmale aufgeführt (z. B. werden in einer Computerspielstudie die Computererfahrungen der Probanden beschrieben).

Nach der Stichprobenbeschreibung sind die Befunde zu den einzelnen Fragestellungen und Hypothesen zu berichten. Im Sinne der Konsistenz bietet es sich an, die Strukturierung aus dem Theorieteil zu übernehmen und die einzelnen Fragestellungen und Theorien in derselben Reihenfolge abzuarbeiten. Dabei werden deskriptivstatistische Ergebnisse für Fragestellungen und inferenzstatistische Ergebnisse für Hypothesen teils in den Fließtext integriert, teils durch Tabellen und Grafiken veranschaulicht. Dieselbe Information sollte allerdings nicht mehrfach wiederholt werden. Grafiken lockern den Fließtext auf und sind besonders aufmerksamkeitssträchtig, deswegen sollte man sie für besonders wichtige Ergebnisse vorsehen.

Im Ergebnisteil muss die richtige Balance gefunden werden zwischen präziser Information durch zahlreiche statistische Befunde einerseits und flüssiger Lesbarkeit andererseits. Im Zweifelsfall lassen sich umfassende Tabellen, die für das Verständnis des Fließtextes nicht zwingend erforderlich sind, in den Anhang auslagern. Die Lesbarkeit von Tabellen wird deutlich erhöht, wenn man Prozentzahlen ganzzahlig rundet (auch wenn dann in der Summe manchmal 99% oder 101% resultiert). Ansonsten werden statistische Kennwerte (z. B. Mittelwerte, Standardabweichungen, Korrelationskoeffizienten etc.) üblicherweise mit einer Genauigkeit von zwei Nachkommastellen angegeben. Sozialwissenschaftliche Daten mit acht Nachkommastellen erwecken den Eindruck von Pseudogenauigkeit. Generell sollte auf die Aufbereitung der Ergebnisse viel Mühe verwendet werden. Ein einfaches Kopieren des Outputs von Statistiksoftware führt zu unbefriedigenden Ergebnissen. Jede Tabelle und jede Grafik im Fließtext muss fortlaufend nummeriert sein und zudem ohne Kenntnis des Fließtextes verständlich sein. Dies wird durch einen aussagekräftigen Tabellen- bzw. Abbildungstitel möglich, dem auch eine Erläuterung aller in der Tabelle bzw. Grafik verwendeten Abkürzungen beizufügen ist.

Weitaus strengere Vorschriften – auch im Hinblick auf Metaanalysen (► Kap. 10) – hat die American Psychological Association für die Publikation empirischer Studien festgelegt (APA, 2001). Ohne dass hier technische Details wiedergegeben werden können – diese sind Gegenstand der ► Kap. 9 und 10 – sei bereits jetzt darauf hingewiesen, dass die folgenden Angaben obligatorisch sind:

- die Größe des in der Untersuchung ermittelten Effekts (► S. 605 ff.)
- das sog. Konfidenzintervall für diesen Effekt (► S. 608 ff.),
- die Teststärke der Untersuchung (► S. 500 f.).

Auf ► S. 640 wird begründet, dass wir auch die Bekanntgabe des erwarteten Effekts, der die Teststärke der Untersuchung maßgeblich mitbestimmt, für erforderlich halten. Schließlich sollte auch angegeben werden, welche Art von Nullhypothese (traditionelle Nullhypothese oder Minimum-Effekt-Nullhypothese im Sinne des Good-enough-Prinzips) geprüft wurde (► S. 635 ff.).

Diskussion

Die Diskussion beginnt mit einer Zusammenfassung der wichtigsten Ergebnisse. Anschließend werden die Einzelbefunde des Ergebnisteils zu einer Gesamtinterpretation und einem Gesamtfazit verarbeitet. Dabei ist ein Rückbezug auf das eigene theoretische Modell wichtig. Es folgt eine kritische Reflexion der Grenzen der eigenen Studie, die sich z. B. aus möglichen Einschränkungen der internen und externen Validität ergeben. Auf die schonungslose Offenlegung der methodischen Schwächen der eigenen Arbeit folgt dann eine Würdigung ihrer Stärken. Es wird diskutiert, wie sich die Befunde der Arbeit für die weitere Forschung (z. B. Ideen für Anschlussstudien) sowie für die Praxis (z. B. Ideen für Interventionsmaßnahmen) fruchtbar machen lassen. Die beiden letztgenannten, zukunftsgerichteten Teile der Diskussion werden manchmal auch in einem sog. **Ausblick** als separatem Gliederungspunkt behandelt. Ideal ist es, wenn die Diskussion mit einer Pointe endet oder im Sinne einer Schließung des Argumentationskreises auf den ersten Satz der Einleitung zurückkommt. Abstract, Einleitung und Diskussion sind die wichtigsten Teile einer Arbeit, da sie häufig als Erstes (und Einziges) gelesen werden.

Literatur

Das Literaturverzeichnis wird in ► Abschn. 2.7.4 ausführlich behandelt.

2.7.3 Gestaltung des Manuskripts

Das Manuskript wird maschinenschriftlich (einseitig, linksbündig beschriebene oder bedruckte DIN-A4-Seiten mit anderthalbfachem Zeilenabstand) fortlaufend geschrieben, d. h., für die einzelnen Hauptbereiche werden keine neuen Blätter angefangen. Für Titel, Vorwort, Zusammenfassung, Inhaltsverzeichnis, Literaturverzeichnis u. Ä. ist jeweils eine neue Seite zu beginnen.

Das Titelblatt enthält

- den vollen Titel der Arbeit,
- Vor- und Familienname der Verfasserin bzw. des Verfassers (ggf. Matrikelnummer),
- Angaben über die Art der Arbeit (Referat, Seminararbeit, Semesterarbeit, Masterarbeit etc.),
- eine Angabe der Institution, bei der sie eingereicht wird, der Lehrveranstaltung, in deren Rahmen sie abgefasst wurde bzw. den Namen des Betreuers,
- Ort und Datum der Fertigstellung der Arbeit.

Fußnoten im laufenden Text sollten nach Möglichkeit vermieden werden, da sie die Lektüre erschweren. Falls diese für technische Hinweise (Danksagungen, Übersetzungshinweise, persönliche Mitteilungen) erforderlich sind, empfiehlt sich eine durchlaufende Numerierung aller Fußnoten. Für Fußnoten ungeeignet sind Literaturhinweise.

Die sprachliche Gestaltung des Textes sollte neutral gehalten sein. Beutelsbacher (1992, S. 70 f.) gibt folgende Empfehlung:

Gehen Sie mit »ich« äußerst vorsichtig um. »Ich« wird nur dann verwendet, wenn der Autor eine persönliche Botschaft zu Papier bringt. Verwenden Sie, wenn immer möglich »wir«. »Wir« kann immer dann benutzt werden, wenn stattdessen auch »der Autor und der Leser« stehen kann. »Wir« ist also kein pluralis majestatis, sondern eine Einladung an den Leser, sich an der Diskussion zu beteiligen und mitzudenken. Wenn es nicht anders geht, benutzen Sie »man«.

Die Auswahl einer gut lesbaren Schrift, eine übersichtliche und ansprechende Formatierung sowie ein Sachre-

gister sollten im Zeitalter der elektronischen Textverarbeitung auch bei Qualifikationsarbeiten zum Standard gehören. Ebenso wie es sich empfiehlt, sich vorbereitend mit der Statistiksoftware zu befassen, sollte man sich rechtzeitig vor Beginn der Arbeit mit den Feinheiten der Textverarbeitung (Gliederungsfunktion, Index, Formatierungsmakros etc.) sowie den Möglichkeiten der computergestützten Grafikerstellung vertraut machen. Erfahrungsgemäß wird der in der Endphase der Arbeit für Formatierung, Einbindung von Grafiken, Erstellen von Verzeichnissen etc. benötigte Zeitaufwand deutlich unterschätzt.

2.7.4 Literaturhinweise und Literaturverzeichnis

Für alle Äußerungen und Gedanken, die man von anderen Publikationen übernimmt, muss deren Herkunft angegeben werden, andernfalls würde man zum Plagiatör. Der wissenschaftliche Quellennachweis verlangt, dass man den Namen des Autors bzw. der Autorin und das Erscheinungsjahr der Publikation im laufenden Text in Klammern nennt:

- Besonders zu beachten ist die Reliabilität des Kriteriums (Abels, 1999).

Ist der Autorenname Bestandteil eines Satzes, wird nur die Jahreszahl, aber nicht der Name in Klammern gesetzt:

- Besonders zu beachten ist nach Abels (1999) die Reliabilität des Kriteriums.

Bei Veröffentlichungen von zwei Autoren werden immer alle beide genannt; sie können im Fließtext durch »und«, als Klammerbeleg durch das Et-Zeichen verbunden werden:

- Besonders zu beachten ist nach Abels und Busch (1998) die Reliabilität des Kriteriums.
- Besonders zu beachten ist die Reliabilität des Kriteriums (Abels & Busch, 1998).

Publikationen, die von mehr als zwei Autoren stammen, können durch den ersten Namen mit dem Zusatz »et al.« (= et alii) gekennzeichnet werden:

- Besonders zu beachten ist nach Abels et al. (1998) die Reliabilität des Kriteriums. (Im Literaturverzeichnis

sollten dagegen stets sämtliche Autoren einer Publikation angeführt werden.)

Verweist eine Arbeit auf Publikationen von Autoren mit gleichem Nachnamen, ist der Anfangsbuchstabe des Vornamens hinzuzufügen:

- Besonders zu beachten ist nach A. Abels (1999) die Reliabilität des Kriteriums.

Mehrere Publikationen eines Autors mit demselben Erscheinungsjahr werden durch Kleinbuchstaben in alphabetischer Reihenfolge, die an die Jahreszahl angehängt werden, unterschieden:

- Besonders zu beachten ist die Reliabilität des Kriteriums (vgl. Abels, 1999a, 1999b). (Diese Kennzeichnung gilt dann auch für das Literaturverzeichnis.)

Ein Aufsatz, der in einem Sammelband oder »Reader« erschienen ist, wird mit dem Namen des Autors und nicht mit dem Namen des Herausgebers zitiert.

Übernommene Gedankengänge sollten wenn möglich durch die Originalliteratur belegt werden. Falls nur Sekundärliteratur verarbeitet wurde, ist dies entsprechend zu vermerken:

- Besonders zu beachten ist die Reliabilität des Kriteriums (Abels, 1998, zit. nach Busch, 1999). (Das Literaturverzeichnis enthält dann *beide* Arbeiten.)

Wörtliche Zitate werden in Anführungszeichen gesetzt und durch zusätzliche Erwähnung der Seitenzahl nachgewiesen:

- Hierzu bemerkt Abels (1999, S. 100): »Besonders hervorzuheben ist die Reliabilität des Kriteriums.«

Erstreckt sich ein Zitat auf die folgende Seite, so steht hinter der Seitenzahl ein »f.« (für »folgende«). Will man auf eine Textpassage Bezug nehmen, die sich nicht nur auf eine, sondern mehrere folgende Seiten erstreckt, so setzt man »ff.« hinter die Seitenzahl.

Ergänzungen eines Zitates stehen in eckigen Klammern und Auslassungen werden durch Punkte gekennzeichnet:

- Hierzu bemerkt Abels (1999, S. 100): »Besonders hervorzuheben ist die Reliabilität des [inhaltlichen] Kriteriums.«

Anführungszeichen in einer wörtlich zitierten Textpassage erscheinen im Zitat als einfache Anführungszeichen (Zitat im Zitat):

- Hierzu bemerkt Abels (1999, S. 100): »Besonders hervorzuheben ist die sog. »Reliabilität« des Kriteriums.«

Hebt der Verfasser im Zitat eine im Original nicht hervorgehobene Stelle, z. B. durch Kursivschrift oder Unterstreichung, hervor, ist dies im laufenden Text zu vermerken: [Hervorhebung durch Verf.]. Befinden sich in einer zitierten Passage kursiv oder fett gedruckte Wörter, so sind diese als Bestandteil des Textes beizubehalten und werden häufig als solche gekennzeichnet: [Hervorhebung im Original]

- Hierzu bemerkt Abels (1999, S. 100): »Besonders hervorzuheben ist die *Reliabilität* [Hervorhebung im Original] des *Kriteriums*« [Hervorhebung durch Verf.].

Alle fremden, im Text erwähnten Quellen müssen im Literaturverzeichnis mit vollständigen bibliographischen Angaben aufgeführt werden. Die Wiedergabe der genauen Literaturnachweise in Fußnoten unmittelbar auf der Seite des Zitates ist nicht mehr üblich. Damit entfallen auch Verweise auf frühere Fußnoten, wie z. B. »a.a.O.« (am angegebenen Ort), »l. c.« (loco citato) oder »op. cit.« (opus citatum).

In Literaturangaben wird stets entweder der Buchtitel (nicht der Titel eines Beitrags aus einem Buch) oder der Name der Zeitschrift kursiv gedruckt. Bei Aufsätzen aus Sammelbänden sowie bei Zeitschriftenaufsätzen sind in jedem Fall Seitenangaben zu machen. Werden englischsprachige Werke zitiert, erscheinen Zusatzangaben wie »Hrsg.« (Herausgeber) und »S.« (Seite) auf Englisch: »Ed.« bzw. »Eds.« (Editor/s) sowie »p.« bzw. »pp.« (pages). Dem Deutschen »S. 3 ff.« entspricht das Englische »pp. 3«; »S. 5–15« wird zu »pp. 5–15«.

Obwohl das Zitieren von Literaturquellen seit jeher zum wissenschaftlichen Handwerk gehört, gibt es bis heute leider keine allgemein verbindlichen Zitierweisen. Überflüssigerweise pflegen unterschiedliche Disziplinen (z. B. deutsche Philologie versus Psychologie) und Publikationsorgane (z. B. Psychologische Rundschau versus Kölner Zeitschrift für Soziologie und Sozialpsychologie) ganz unterschiedliche Zitierstile, sodass Texte

letztlich »zielgruppenspezifisch« formatiert werden müssen: mal wird der Vorname aller Autoren ausgeschrieben, mal abgekürzt; mal werden Ort und Verlag genannt, mal erscheint nur der Ort; mal werden Buchtitel in Anführungsstriche gesetzt, mal kursiv geschrieben, mal »normal« gedruckt.

Die »Zitierwürdigkeit« der »grauen Literatur« (► S. 360, 674) ist strittig. Ebenso sollte man mit Quellen nachweisen für private Mitteilungen (persönliches oder fernmündliches Gespräch, Brief, elektronische Nachricht o. Ä.) sparsam umgehen, denn beide Arten von Quellen sind für Außenstehende schwer nachprüfbar. Allerdings hatten diese Informationsquellen – insbesondere die »graue Literatur« – in der ehemaligen DDR einen besonderen Stellenwert, weil sie frei von Politzensur waren. Beim Zitieren von elektronischen Publikationen aus dem Internet (Rindfuß, 1994) ergibt sich das Problem, dass diese nicht selten verändert, verschoben oder gelöscht werden.

■ Box 2.10 enthält ein kurzes fiktives Literaturverzeichnis mit einigen Beispielen, die zum Teil aus Tröger und Kohl (1977) entnommen wurden. Das Literaturverzeichnis folgt den Richtlinien der Deutschen Gesellschaft für Psychologie (1997, 2001). Für Arbeiten, die in

Zeitschriften oder als Monographien veröffentlicht werden, beachte man zusätzlich die Richtlinien der jeweiligen Verlage. Der folgende Text erläutert, wie auf die im Literaturverzeichnis in Box 2.10 aufgeführten Quellen verwiesen wird und um welche Quellen es sich handelt.

Abavo (1995): Artikel des Autors Abavo aus der Zeitschrift »Die Normalverteilung und ihre Grenzgebiete«. Der Artikel erschien 1995 und steht im 3. Band auf den Seiten 157–158.

American Psychological Association (2000): Von der APA publizierte Webseite zu Zitationsnormen für Onlinequellen. Gemäß diesen Regeln sind erst das Abrufdatum des Dokuments, dann der Internetdienst sowie schließlich die Netzadresse anzugeben. Diese Zitationsweise hat sich bislang jedoch nur bedingt durchgesetzt, stattdessen existieren eine Reihe verwandter Zitierformen (► unten das Beispiel King, 1996).

Bock et al. (1986): Buch der Autoren Bock, Greulich und Pyle mit dem Titel »The Hufnagel-Contributions to Factor Analytic Methods«. Das Buch ist im Jahr 1986 im Verlag Holt, Rinehart & Winston erschienen. Der Verlag hat seinen Hauptsitz in New York.

Frisbie (1975): Hier wird auf ein Buch verwiesen, das von Frisbie herausgegeben wurde. Es heißt »Psycho-

Box 2.10

Ein fiktives Literaturverzeichnis

- Abavo, H.-H. (1995). Bemerkung zur Klumpeneffekt-Stratifikationszerlegung. *Die Normalverteilung und ihre Grenzgebiete*, 3, 157–158.
- American Psychological Association. (2000). *Electronic Reference Formats Recommended by the American Psychological Association*. Retrieved August 20, 2000, from the World Wide Web: <http://www.apa.org/journals/webref.html>.
- Bock, R. D., Greulich, S. & Pyle, D. C. (1986). *The Hufnagel-Contributions to Factor Analysis Methods*. New York: Holt, Rinehart & Winston.
- Frisbie, L. L. (Ed.) (1975). *Psychology and Faking*. Urbana: The University of Wisconsin Press.
- Greulich, S. (1976). *Psychologie der Bescheidenheit* (12. Aufl.). Großhermsdorf: Kaufmann & Trampel.
- Herweg, O. & Peter, G. (1986). *Signifikanz und Transzendenz. Diskussionsbeitrag für das Symposium »Ergodizität infiniter Kausalketten«*. Münster: Katholische Akademie.
- King, S. A. (1996). *Is the Internet Addictive, or Are Addicts Using the Internet?* [Online Document] URL <http://www.concentric.net/~Astorm/iad.html> (20.08.2000).

- Müller, C. & Maier, G. (1913). Intelligenz im Jugendalter. In D. Helfferich (Hrsg.) *Entwicklung und Reife* (S. 5–15). Bad Wimpfen: Uebelhör.
- Picon, J.-J. (1901). Antwort auf Martinis und Pernods Artikel über die »Unbedenklichkeit des Aperitifs«. *Der internationale Wermut-Bruder*, 26, 1041–1043.
- Reydelkorn, H. (1995). Iterative Verfahren zur Zerlegung von Klumpen. *Informationen des Instituts für angewandtes Kopfrechnen in Oldenburg*, 4, 27–58.
- Schlunz, I.I. (1956). *Therapie und Duldung – ein Versuch*. Unveröffentlichte Diplomarbeit. Freiburg: Psychologisches Institut der Universität Freiburg.
- Stiftung VW-Werk. (1993). *Psychologische Forschung im Verkehrswesen*. Wolfsburg: Stiftung VW-Werk.
- Stör, A. von (o. J.). *Anleitung zur Anfertigung von Flugblättern*. Unveröffentlichtes Manuskript. o. O.
- Zielman, P.S. (1991). Questioning Questions. In A. Abel & B. Bebel (Eds.) *More Questions and More Data* (pp. 33–66). New York: Wild Press.

logy and Faking« und wurde 1975 in Urbana von der University of Wisconsin Press gedruckt.

Greulich (1976): Das Buch »Psychologie der Bescheidenheit« ist in der 12. Auflage 1976 im Verlag Kaufmann & Trampel in Großhermannsdorf erschienen.

Herweg und Peter (1986): Hier wird auf einen Diskussionsbeitrag mit dem Titel »Signifikanz und Transzendenz« verwiesen. Der Beitrag wurde auf dem Symposium »Ergodizität unendlicher Kausalketten« gehalten und von der Katholischen Akademie in Münster 1986 publiziert.

King (1996): Dieser 1996 verfasste Aufsatz wird als Onlinedokument auf der persönlichen Homepage des Autors bereitgestellt, wo wir ihn am 20. August 2000 abgerufen haben. Bei späteren Abrufversuchen muss aufgrund der Eigenart des Netzmediums damit gerechnet werden, dass der Beitrag inhaltlich verändert, auf einen anderen Server verschoben oder ganz aus dem Netz genommen wurde.

Müller und Maier (1913): Dieser Beitrag bezieht sich auf einen Aufsatz, den die Autoren Müller und Maier in einem von Helfferich herausgegebenen Sammelband mit dem Titel »Entwicklung und Reife« auf den Seiten 5 bis 15 veröffentlicht haben. Der Sammelband (oder Reader) wurde 1913 im Verlag Uebelhör, Bad Wimpfen, veröffentlicht.

Picon (1901): Hier wird auf einen Aufsatz verwiesen, der die Überschrift »Antwort auf Martinis und Permonds Artikel über die Unbedenklichkeit des Aperitifs« trägt. Der Artikel wurde im Jahre 1901 im 26. Band der Zeitschrift »Der internationale Wermut-Bruder« auf den Seiten 1041–1043 veröffentlicht.

Reydelkorn (1995): Hier wird auf keinen Zeitschriftenartikel verwiesen, sondern auf eine institutsinterne Reihe »Informationen des Instituts für angewandtes Kopfrechnen in Oldenburg«. Der von Reydelkorn in dieser Reihe verfasste Artikel heißt »Iterative Verfahren zur Zerlegung von Klumpen«.

Schlunz (1956): Diese Literaturangabe bezieht sich auf eine unveröffentlichte Diplomarbeit. Der Titel der Arbeit heißt »Therapie und Duldung – ein Versuch«. Die Arbeit wurde 1956 am Psychologischen Institut der Universität Freiburg angefertigt.

Stiftung VW-Werk (1993): In dieser Weise wird auf Literatur verwiesen, die keinen Autorennamen trägt.

Die Stiftung VW-Werk hat 1993 einen Bericht über »Psychologische Forschung im Verkehrswesen« in Wolfsburg herausgegeben.

von Stör (o. J.): Dieser Literaturhinweis bezieht sich auf ein unveröffentlichtes Manuskript, dessen Erscheinungsjahr (o. J. = ohne Jahresangabe) und Erscheinungsort (o. O. = ohne Ortsangabe) unbekannt sind. Das Manuskript trägt den Titel »Anleitung zur Anfertigung von Flugblättern«.

Zielman (1991): Dieser Verweis bezieht sich auf einen Buchbeitrag mit dem Titel »Questioning Questions«, der in dem von Abel und Bebel herausgegebenen Sammelband »More Questions and More Data« auf den Seiten 33 bis 66 abgedruckt ist. Der Sammelband ist in dem in New York ansässigen Verlag »Wild Press« erschienen.

Weitere Hinweise zum Umgang mit Literaturangaben und zur Anfertigung von Manuskripten findet man z. B. bei Hager et al. (2001, Kap. C) oder Höge (2002).

2.7.5 Veröffentlichungen

Gelungene Arbeiten sollte man einer Zeitschrift zur Publikation anbieten. Die wissenschaftlichen Periodika, die für diese Zwecke zur Verfügung stehen, vertreten unterschiedliche inhaltliche Schwerpunkte, die man beim Durchblättern einzelner Bände leicht herausfindet; ggf. lässt man sich bei der Wahl einer geeigneten Zeitschrift von Fachleuten beraten.

In der Regel wird die Version, die zur Veröffentlichung vorgesehen ist, gegenüber dem Original erheblich zu kürzen sein. Lassen umfangreiche Untersuchungen (z. B. Dissertationen) keine erhebliche Kürzung ohne gleichzeitige Sinnentstellung zu, ist die Aufteilung der Gesamtarbeit in zwei oder mehrere Einzelberichte (z. B. Theorieteil, Experiment 1, Experiment 2) zu erwägen. Zu prüfen ist auch, ob sich ein Verlag bereit findet, die gesamte Arbeit als Monographie zu publizieren. (Wichtige Hinweise hierzu bzw. zum Thema »Promotionsmanagement« findet man bei Preißner et al., 1998.)

Besonderheiten der Manuskriptgestaltung und auch des Literaturverzeichnisses entnimmt man am einfachsten den Arbeiten, die in der vorgesehenen Zeitschrift bereits veröffentlicht sind. Im Übrigen sind die »Hinwei-

se für Autoren«, die sich üblicherweise auf der Innenseite des Zeitschrifteneinbandes befinden, zu beachten. (Hier erfährt man auch, an welche Anschrift das Manuskript zu senden ist.)

Für die Anfertigung einer englischsprachigen Publikation sei dem Novizen Huff (1998) empfohlen.

Detaillierte Hinweise für die Anfertigung »professioneller« Publikationen, die den Standards der American Psychological Association genügen (vgl. hierzu vor allem auch die auf ▶ S. 601 genannten Angaben), sind den »Concise Rules of APA Style« zu entnehmen (APA, 2005).

Übungsaufgaben

- 2.1 Was versteht man unter interner und externer Validität?
- 2.2 Wie kann man Menschen für die Teilnahme an einer empirischen Untersuchung motivieren? Was sind günstige Rahmenbedingungen?
- 2.3 Welche der folgenden Aussagen stimmen bzw. stimmen nicht? (Begründung)
 - a) Für einen Mittelwertvergleich zwischen zwei Gruppen muss die abhängige Variable intervallskaliert sein.
 - b) Für experimentelle Untersuchungen ist die Zufallsauswahl der Probanden charakteristisch.
 - c) Externe Validität ist die Voraussetzung für interne Validität.
 - d) In Experimenten wird höchstens eine unabhängige Variable untersucht.
 - e) Experimentelle Laboruntersuchungen haben eine geringere externe, dafür aber eine hohe interne Validität.
 - f) Je höher das Skalenniveau, umso höher die Validität.
- 2.4 Wie ist eine Skala definiert?
- 2.5 Auf welchem Skalenniveau sind folgende Merkmale sinnvollerweise zu messen? Geben Sie Operationalisierungsmöglichkeiten an! Augenfarbe, Haustierhaltung, Blutdruck, Berufserfahrung, Bildungsstand, Intelligenz, Fernsehkonsum.
- 2.6 Worin unterscheiden sich Feld- und Laboruntersuchung?
- 2.7 1993 publizierte H.K. Ma über Altruismus. Im selben Jahr erschien in einer Zeitschrift über Gesundheitsvorsorge der Artikel »Just Cover up: Barriers to Heterosexual and Gay Young Adults' Use of Condoms.« Suchen Sie mit Hilfe der *Psychological Abstracts* nach den kompletten Literaturangaben und zitieren Sie diese korrekt!
- 2.8 Angenommen in einer Telefonbefragung von N=2500 zufällig ausgewählten Berlinerinnen und Berlinern (Zufallsauswahl aus der Liste aller Berliner Telefonnummern) stellte sich heraus, dass 22% »ständig« und 47% »nie« einen Talisman oder Glücksbringer bei sich haben (31% nehmen »manchmal« einen mit). Diejenigen, die ständig einen Talisman bei sich trugen, waren signifikant zufriedener mit ihrem Leben als diejenigen, die nie einen Talisman mitnahmen.
 - a) Um welchen Untersuchungstyp handelt es sich hier?
 - b) Wie lautet die statistische Alternativhypothese zu folgender Forschungshypothese: »Talismanträger sind zufriedener als Nichttalismanträger«. Wie lautet die zugehörige Nullhypothese?
 - c) Beurteilen Sie die interne und die externe Validität dieser Untersuchung (Begründung).
 - d) Welche Rolle spielen Versuchsleitereffekte in dieser Untersuchung?
- 2.9 Wie ist die Aussagekraft von Untersuchungen an Studierenden einzuschätzen?
- 2.10 »Die Behandlung von Höhenangst mit der herkömmlichen Verhaltenstherapie dauert mindestens 6 Monate länger als die Therapie mit einem neuen Hypnoseverfahren.« Kennzeichnen Sie die angesprochenen Variablen (Skalenniveau, Variablenart). Wie lautet das statistische Hypothesenpaar?
- 2.11 Welche Besonderheiten weisen freiwillige Untersuchungsteilnehmer auf?