

**Einführung in die
psychologische Methodenlehre**

Skript zur Vorlesung

WS 2002 / 2003

M. Schreier

Vorbemerkung

Das Skript enthält eine Zusammenfassung der zentralen Inhalte der Vorlesung *Einführung in die psychologische Methodenlehre*, im Wesentlichen ohne Abbildungen u.ä. Es ist daher auf jeden Fall durch die Folien zur Vorlesung zu ergänzen; in den Folien finden sich erläuternde Abbildungen, Definitionen wesentlicher Begriffe usw.

Am Ende jedes Abschnitts ist die Prüfungsliteratur angegeben, durch die das Skript vertieft werden sollte. Dabei ist das Skript auf jeden Fall Grundlage. Wenn also als Prüfungsliteratur beispielsweise ein Kapitel aus Bortz & Döring 2002 angegeben ist, dann stellen das Skript und das entsprechende Kapitel gemeinsam die relevante Prüfungsliteratur bezogen auf die Vorlesung dar.

Natürlich enthält die Vorlesung nur einen ersten Teil des Stoffes für die Vordiplomprüfung in Methodenlehre. Ergänzt werden diese Grundlagen insbesondere durch die Vorlesung zur Versuchsplanung im zweiten sowie die Veranstaltung zur Datenerhebung und –auswertung im dritten Semester (und natürlich im Rahmen der beiden Praktika).

Dies ist eine erste, in unruhigen Zeiten entstandene Fassung des Skripts – die daher sicher auch einige Unklarheiten, vielleicht auch noch den einen oder anderen Fehler enthält. Um das Skript zu verbessern, bin ich auf Ihre Rückmeldung angewiesen. Es wäre daher schön, wenn Sie sich mit mir in Verbindung setzen könnten, wenn Sie einen Fehler entdecken oder wenn Ihnen etwas unklar erscheint (und natürlich auch, wenn Ihnen etwas besonders gut gefallen sollte...), und zwar am besten per e-mail:

- bis 15.3.03 unter: m.schreier@uni-koeln.de
- danach unter: m.schreier@iu-bremen.de

Ich würde mich freuen, von Ihnen zu hören!

Köln, im Februar 2003

Margrit Schreier

Inhaltsverzeichnis

I. Einführung: Psychologie als Wissenschaft

I.1.	Psychologie und ihre Gegenstände	S.	4
I.2.	Wissenschaftliche Methoden	S.	5
I.3.	Wissenschaftliche Erkenntnis ist relativ	S.	7
I.4.	Grundbegriffe psychologischer Methodenlehre	S.	8
I.4.1.	Messung	S.	8
I.4.2.	Gütekriterien	S.	9
I.4.3.	Operationalisierung	S.	10
I.5.	Aufgaben und Ziele der Psychologie als Wissenschaft	S.	10
I.5.1.	Beschreibung	S.	11
I.5.2.	Erklärung	S.	13
I.5.3.	Prognose	S.	15
I.5.4.	Veränderung (Technologie)	S.	16
I.5.5.	Bewertung	S.	17

II. Quantitative Forschung

II.1.	Grundbegriffe	S.	19
II.1.1.	Untersuchungsarten quantitativer Forschung	S.	19
II.1.2.	Stichprobenarten	S.	20
II.1.3.	Populationsbeschreibende Untersuchungen	S.	22
II.1.4.	Hypothesentestende Untersuchungen	S.	22
II.2.	Methoden der Datenerhebung	S.	23
II.2.1.	Befragung	S.	23
II.2.2.	Beurteilung mittels Ratingskalen	S.	26
II.2.3.	Tests	S.	28
II.2.4.	Beobachten, Zählen, Messen	S.	32
II.2.5.	Gütekriterien	S.	35
II.3.	Forschungsmethoden	S.	37
II.3.1.	Das (Labor-)Experiment	S.	38
II.3.2.	Arten des Experiments	S.	43
II.3.3.	Die Zusammenhangsstudie	S.	44
II.3.4.	Ethische Probleme quantitativer Forschung	S.	45

III. Qualitative Forschung

III.1.	Qualitative Forschung im Überblick	S.	48
III.1.1.	Forschungsbeispiel: Die Marienthalstudie	S.	48
III.1.2.	Merkmale qualitativer Forschung	S.	49
III.1.3.	Qualitative und quantitative Forschung im Vergleich	S.	53
III.1.4.	Arten qualitativer Forschung	S.	54
III.2.	Qualitative Verfahren der Datenerhebung	S.	55
III.2.1.	Qualitative und quantitative Befragung im Vergleich	S.	56
III.2.2.	Das halbstandardisierte Interview (Leitfadeninterview)	S.	58
III.2.3.	Das nonstandardisierte Interview	S.	59
III.2.4.	Bewertung des Interviews als Forschungsmethode	S.	61
III.2.5.	Gruppendiskussion	S.	62
III.2.6.	Verfahren der Transkription	S.	64

III.2.7. Verfahren qualitativer Beobachtung	S.	66
III.3. Qualitative Verfahren der Datenanalyse	S.	67
III.3.1. Hermeneutik	S.	68
III.3.2. Objektive Hermeneutik	S.	69
III.3.3. Inhaltsanalytische Verfahren	S.	71
III.3.4. Datenaufbereitende Verfahren: Typenbildung	S.	77
III.4. Qualitative Forschungsansätze (Komplexe Verfahren)	S.	79
III.4.1. Stichprobenziehung in der qualitativen Forschung	S.	80
III.4.2. Gegenstandsbezogene Theorienbildung (Grounded Theory)	S.	81
III.4.3. Deskriptive Feldforschung	S.	82
III.4.4. Handlungsforschung (Aktionsforschung)	S.	85
III.4.5. Forschungsprogramm Subjektive Theorien (FST)	S.	86
III.5. Gütekriterien qualitativer Forschung (einschließlich Ethik)	S.	90
III.5.1. Übertragbarkeit quantitativer Kriterien auf qualitative Forschung	S.	90
III.5.2. Eigenständige Gütekriterien qualitativer Forschung	S.	91
III.5.3. Ethische Probleme qualitativer Forschung	S.	92
III.6. Zur Vereinbarkeit von qualitativer und quantitativer Forschung	S.	94
IV Im Skript verwendete Prüfungsliteratur	S.	95

I. Einführung: Psychologie als Wissenschaft

Gegenstand der Vorlesung ‚Einführung in die psychologische Methodenlehre‘ und dieses Skripts sind Methoden der Psychologie.

Definition: Methode

Eine Methode lässt sich verstehen als eine Vorgehensweise, um ein bestimmtes Ziel zu erreichen. Psychologische Methoden sind entsprechend definiert als Vorgehensweisen, um etwas über den Gegenstand der Psychologie zu erfahren.

Im Folgenden wird zunächst der Gegenstand psychologischer Forschung genauer bestimmt. Im Anschluss wird genauer auf die Anforderungen an wissenschaftliche Methoden und auf die Aufgaben und Ziele psychologischer Forschung eingegangen.

I.1. Psychologie und ihre Gegenstände

Als Gegenstand der Psychologie gilt heute ganz allgemein das *menschliche Erleben und Verhalten (Handeln)*.

Diese Gegenstandsbestimmung ist historisch gesehen jedoch ausgesprochen kontrovers. Dabei lassen sich zwei verschiedene Auffassungen der Psychologie und ihres Gegenstandes unterscheiden:

- *Naturwissenschaftliche Auffassung:* Diese Auffassung von Psychologie ist, wie der Name schon sagt, in den Naturwissenschaften verankert. Im Mittelpunkt steht das Verhalten des Menschen, das unter Anwendung sog. quantitativer Methoden (insbesondere des Experiments) erklärt werden soll. Dahinter steht die Auffassung, dass Wissenschaft unabhängig vom konkreten Gegenstand immer nach denselben Prinzipien vorgehen sollte (*‚monistische‘* Auffassung von Wissenschaft). Historisch gesehen sind beispielsweise die Elementpsychologie Wundts oder der Behaviorismus dieser Auffassung von Psychologie zuzuordnen.
- *Geisteswissenschaftliche Auffassung:* Die geisteswissenschaftliche Auffassung von Psychologie wurde vor allem von Dilthey propagiert und entwickelt. Programmatisch ist insbesondere der Satz „Die Natur erklären wir, das Seelenleben verstehen wir.“ (1884). Dilthey vertritt hier eine sog. *dualistische* Wissenschaftskonzeption. Danach erfordert die Erforschung verschiedener Gegenstände auch je verschiedene Methoden. Dilthey bestimmt den Gegenstand der Psychologie als das Seelenleben, legt also den Schwerpunkt auf das menschliche Erleben. Dieses lässt sich nach dualistischer Auffassung nicht nach dem Modell der Naturwissenschaften erklären; Zielsetzung ist vielmehr das Verstehen. Dazu dienen sog. qualitative Methoden.

In der Geschichte der Psychologie hat meist die naturwissenschaftliche Auffassung des Faches und seines Gegenstandes dominiert. Allerdings hat es immer auch relevante ‚Gegenbewegungen‘ aus geisteswissenschaftlicher Sicht gegeben. So hat beispielsweise Wundt selbst die Auffassung vertreten, dass die experimentelle Psychologie sich nur zur Erforschung ‚einfacher‘ Vorgänge eignet; der Untersuchung ‚komplexer‘ Vorgänge im Sinne von Kultur, Religion, Sprache usw. hat er dagegen in der ‚Völkerpsychologie‘ festgehalten (einem 10bändigen Werk). Auch die Würzburger Schule oder die Gestaltpsychologie haben wichtige Beiträge zur Entwicklung des Faches aus einer geisteswissenschaftlichen Auffassung heraus geleistet.

Aus den Beschreibungen der beiden Auffassungen von Psychologie geht bereits hervor, dass je nach Gegenstandsverständnis unterschiedliche Aspekte von Psychologie im Mittelpunkt des Forschungsinteresses stehen. Je nach Gegenstandsverständnis erscheinen auch jeweils andere Methoden angemessen: quantitativ-naturwissenschaftlich oder qualitativ-geisteswissenschaftlich orientierte. Die Komplexität des Gegenstandes der Psychologie ist also einer der Gründe dafür, dass in diesem Fach so viele unterschiedliche Methoden zur Anwendung kommen.

Prüfungsliteratur: nur Skript

I.2. Wissenschaftliche Methoden

Was ist überhaupt ‚Wissenschaft‘?

Der Begriff ‚Wissenschaft‘ wird auf unterschiedliche Art und Weise verwendet. Es lassen sich mindestens drei Verwendungsweisen unterscheiden:

- *Wissenschaft als Institution:* In diesem Sinn bezeichnet ‚Wissenschaft‘ die Gesamtheit der Forschenden und Lehrenden, meist an Hochschulen.
- *Wissenschaft als Tätigkeit (Methode):* ‚Wissenschaft‘ in diesem Sinn bezieht sich auf den Prozess der Generierung von Wissen. Wissenschaft besteht darin, dass Wissenschaftler/innen versuchen, auf eine bestimmte Art und Weise Probleme zu lösen.
- *Wissenschaft als Produkt:* Schließlich kann sich der Begriff ‚Wissenschaft‘ auch auf die Ergebnisse wissenschaftlicher Tätigkeit beziehen, auf wissenschaftliche Erkenntnisse also.

Im Folgenden steht Wissenschaft in der Bedeutung einer Tätigkeit im Mittelpunkt.

Psychologie als Erfahrungswissenschaft

Es existieren verschiedene Arten von Wissenschaften, in denen auch je verschiedene Methoden zur Anwendung kommen: logisch-mathematische Wissenschaften, Erfahrungswissenschaften und philosophische Wissenschaften. Die Psychologie ist eine *Erfahrungswissenschaft*. Das bedeutet, dass die Aussagen, die sie hervorbringt, nicht nur stringent und logisch widerspruchsfrei sein sollten, sondern sich außerdem durch Erfahrungen stützen lassen: Die Aussagen der Psychologie (z.B. Beschreibungen oder Erklärungen) sind durch Beobachtungen überprüfbar. Dieser Wissenschaftsauffassung liegt die Annahme zugrunde, dass es Ordnung (Uniformität) im jeweils interessierenden Gegenstandsbereich gibt. Ohne eine solche Geordnetheit ließen sich keine Theorien, Gesetze oder Regeln entwickeln.

Methoden im Alltag

Um psychologische Aussagen hervorzubringen oder zu überprüfen, ist die Anwendung von Methoden erforderlich. Methoden existieren keineswegs nur in der Wissenschaft, sondern auch im Alltag: Wenn jemand ein Fahrrad repariert, geht er oder sie nach einer bestimmten Methode vor – oder jemand, der/die sich durch sämtliche indischen Restaurants Kölns isst, um das beste zu identifizieren. Eine Methode im Alltag ist also dadurch charakterisiert, dass in einer regelhaften Weise vorgegangen wird, um ein bestimmtes Ziel zu erreichen.

Anforderungen an wissenschaftliche Methoden

In der Wissenschaft will man durch die Anwendung von Methoden zu gültigen, möglichst zutreffenden, objektiven Aussagen über die Welt gelangen, die unabhängig sind von der Person der einzelnen Wissenschaftlerin oder des Wissenschaftlers. Um dieses Ziel zumindest potenziell zu erreichen, müssen Methoden in der Wissenschaft weiteren Anforderungen genügen (die an Methoden im Alltag so nicht gestellt werden):

- *Systematik*: Eine wissenschaftliche Methode muss systematisch und regelhaft sein. In der naturwissenschaftlich-quantitativen Auffassung von Psychologie wird mit dem Kriterium der Systematik außerdem die Forderung nach einer Kontrolle von Einflussvariablen mit dem Ziel einer Ursachenerklärung verbunden; diese Spezifikation findet sich in einer geisteswissenschaftlich-qualitativ orientierten Psychologie nicht.
- *Operationale Definition*: Die Anwendung einer wissenschaftlichen Methode beinhaltet, dass man genau angibt, was unter einem bestimmten Begriff zu verstehen ist. Der Begriff wird definiert und die Methoden zu seiner Erfassung werden angegeben. Auch diese Anforderung wird in der quantitativen Forschung noch weiter spezifiziert: Operationale Definition bedeutet hier meist auch Quantifizierung: Ausprägungen des Begriffs werden gemessen (z.B. Hunger als Zeitdauer des Nahrungsentzugs).
- *Replizierbarkeit*: Diese Anforderung findet sich nur innerhalb des quantitativen Ansatzes. Wenn die Untersuchung unter denselben Bedingungen unter Anwendung derselben Methode wiederholt wird, müssen nach dem Kriterium der Replizierbarkeit auch dieselben Ergebnisse resultieren.

Systematik psychologischer Methoden

Psychologische Methoden lassen sich somit zusammenfassend bestimmen als Mittel, um zu möglichst objektiven Erkenntnissen über menschliches Erleben, Verhalten und Handeln zu gelangen.

Psychologische Methoden lassen sich erstens weiter danach unterteilen, ob sie in der naturwissenschaftlich-quantitativen oder in der geisteswissenschaftlich-qualitativen Tradition verortet sind.

Ein zweites Kriterium zur Unterscheidung zwischen Arten psychologischer Methoden betrifft die Forschungsphase, in der die Methoden zur Anwendung kommen. Man unterscheidet zwischen Methoden der Datenerhebung, der Datenanalyse sowie Untersuchungsanlagen bzw. Forschungsstrategien. Forschungsstrategien legen eine Sequenz bestimmter Schritte bei der Untersuchungsdurchführung fest.

Es ergibt sich die folgende Gliederung psychologischer Methoden:

Qualitative Ausrichtung	Quantitative Ausrichtung
Forschungsmethoden	
Deskriptive Feldforschung Handlungsforschung Qualitatives Experiment	Experiment Quasi-Experiment Zusammenhangsstudie
Forschungsprogramm Subjektive Theorien	
Erhebungsmethoden	
Teilnehmende Beobachtung Interview Gruppendiskussion	Befragen Testen Beobachten
Analysemethoden	
Hermeneutische Verfahren Kodiervverfahren Typenbildung	Deskriptive Statistik Inferenzstatistik Multivariate Datenanalyse
Inhaltsanalyse	

Methoden im Kontext

Methoden sind Gegenstand der *Methodenlehre*; Methodenlehre wiederum stellt einen Teilbereich der *Wissenschaftstheorie* dar. An der Wissenschaftstheorie (und folglich auch an der Methodenlehre) lassen sich zwei Aspekte abheben:

- *deskriptiv*: Wissenschaftstheorie in einem deskriptiven Sinn befasst sich mit der Beschreibung und Rekonstruktion wissenschaftlichen Handelns.
- *präskriptiv*: Unter präskriptiver Perspektive wird das Handeln von Wissenschaftlern/innen kritisch hinterfragt und es wird versucht, Regeln für eine verbesserte wissenschaftliche Praxis zu entwickeln.

Prüfungsliteratur: nur Skript

I.3. Wissenschaftliche Erkenntnis ist relativ

Im Alltag findet sich häufig die Vorstellung, dass eine wissenschaftliche Untersuchung zu sicherer Erkenntnis führt. Wenn man beispielsweise etwas darüber erfahren will, wie Entscheidungsprozesse in Gruppen verlaufen, dann variiert man relevante Einflussvariablen, beobachtet das Verhalten – und schon hat man die gewünschten Erkenntnisse gewonnen.

Diese Auffassung ist teilweise sicher richtig – ohne empirische Untersuchungen wären keine Erkenntnisse über den Menschen. Allerdings reicht es meist nicht aus, ‚einfach nur hinzuschauen‘; um sich wissenschaftlichen Erkenntnissen anzunähern, sind vielmehr mehrere Untersuchungen durchzuführen, die sich hinsichtlich ihrer Aussagekraft wechselseitig ergänzen. Jede einzelne Untersuchung ist fehleranfällig; insbesondere liefert eine Methode kein ‚objektives‘, vollständiges Bild des interessierenden Gegenstandes. Im Folgenden sind einige Gründe dafür aufgeführt, weshalb Methoden einen Gegenstand nicht einfach abbilden.

Gegenstands-Methodik-Interaktion

Keine Methode kann einen Gegenstand ganz, in all seinen Aspekten erfassen. Eine Methode hebt an einem Gegenstand vielmehr immer nur bestimmte Merkmale ab. Je nachdem, welche Methode man wählt werden, also je unterschiedliche Aspekte eines Gegenstandes sichtbar (vgl. z.B. Methoden zur Erhebung der Innensicht vs. physiologische Messungen bei der Erfassung von Emotionen).

Methoden sollten entsprechend so gewählt werden, dass relevante Gegenstandsaspekte erfassbar sind. Ein Hammer ist z.B. in der Geologie ein hervorragendes Werkzeug; in der Psychologie ist die Methode des Klopfens auf den Gegenstand dagegen meist wenig weiterführend. Welche Merkmale eines Gegenstandes die relevanten sind, ist allerdings ebenfalls Ansichtssache (s.u. Vorverständnis).

Aus der Annahme einer wechselseitigen Abhängigkeit von Gegenstand und Methode folgt weiterhin, dass Methoden nicht für sich genommen ‚gut‘ oder ‚schlecht‘ sind. Eine Methode ist immer nur für einen bestimmten Gegenstand mehr oder weniger gut geeignet.

Gegenstandsvorverständnis

Wissenschaftler/innen tragen an ihren Gegenstand immer schon bestimmte Vorannahmen heran. Forschung beginnt also nie ‚bei null‘. Vielmehr beeinflussen Vorannahmen, was man sieht und wie man es sieht (vgl. z.B. Auffassung vom Menschen als vernunftbegabtem Wesen vs. Auffassung vom Menschen als auf Umweltreize reagierend); Vorannahmen fungieren quasi als Perspektiven, aus denen ein Gegenstand betrachtet wird. Entsprechend beeinflussen Vorannahmen auch die Auswahl einer Methode. Wissenschaftler/innen können ihre Vorannahmen zwar nicht ‚ausschalten‘; sie können sie aber offen legen und damit auch für andere sichtbar (und kritisierbar) machen.

Der Forschungsgegenstand Mensch

Die Forschungssituation in der Psychologie ist eine grundsätzlich andere als in anderen Wissenschaften. In der Geologie oder der Chemie beispielsweise können Forscher/innen auf den Gegenstand einwirken und dann die Auswirkungen ihrer Intervention beobachten. Forschungsgegenstand der Psychologie dagegen ist ein denkender und fühlender Mensch, eine Forschungssituation ist immer auch eine soziale Situation, und Untersuchungsteilnehmer/innen machen sich ihre eigenen Gedanken darüber, was in einer Untersuchungssituation mit ihnen geschieht. Daraus wiederum folgt, dass es in der Psychologie in der Regel nicht möglich ist, die Auswirkungen einer Veränderung unmittelbar zu erfassen. Erfasst werden kann meist nur das Ergebnis einer von den Untersuchungsteilnehmer/innen *verarbeiteten* Einwirkung.

Wissenschaftspsychologie und -soziologie

Wissenschaft wird nicht in einem Vakuum betrieben, sondern stellt eine gesellschaftliche Institution dar. In einer solchen Institution gelten bestimmte Regeln, die nicht unbedingt mit dem Prinzip der ‚Wahrheitsfindung‘ vereinbar sind. So werden beispielsweise Untersuchungen, in denen es gelingt, die Testhypothese zu bestätigen, deutlich häufiger in wissenschaftlichen Zeitschriften veröffentlicht als Untersuchungen, die eine Falsifikation der Testhypothese ergeben. Weiterhin besteht die Institution aus Menschen, die nicht zuletzt auch ihre eigenen Interessen verfolgen, selber bestimmte Vorannahmen haben usw. So hielten z.B. die wissenschaftlichen Zeitgenossen Edisons Erfindung der Glühlampe 1879 für Scharlatanerie – selbst nachdem Edison die Straßen rund um sein Labor mit Glühlampen beleuchtet hatte. Wenn eine Untersuchung zu relevanten Erkenntnissen führt, ist damit also noch nicht sichergestellt, dass diese Erkenntnis auch verbreitet wird.

Konsequenz?

Daraus, dass auch Wissenschaft nicht fehlerfrei ist, folgt aber nicht notwendig, dass Wissenschaft keinerlei Erkenntnisse erbringt. Zum einen gehen zumindest einige Wissenschaftstheoretiker/innen von einem kumulativen, allmählichen Erkenntnisfortschritt aus. Zum anderen lässt sich die Rationalität wissenschaftlichen Vorgehens steigern, indem Wissenschaftler/innen sich der Fehlerquellen in ihrer Arbeit bewusst sind und vor diesem Hintergrund die eigenen Auffassungen und die eigene Vorgehensweise offen legen und somit für andere nachvollziehbar und auch kritisierbar machen.

Prüfungsliteratur: nur Skript

I.4. Grundbegriffe psychologischer Methodenlehre

1.4.1. Messung

Messbegriff in der Psychologie

Wenn im Alltag von ‚Messung‘ die Rede ist, dann ist damit meist die Erhebung von Daten in Form von Zahlen gemeint (z.B. Größe; Temperatur des Badewassers usw.).

In der Psychologie wird der Begriff ‚Messen‘ dagegen meist in einem weiteren (und zugleich genauer definierten) Sinn verwendet.

Definition: Messung

Messung bedeutet hier, dass beobachteten Verhaltensweisen, Äußerungen usw. Zahlen so zugeordnet werden, dass die Beziehungen zwischen den Beobachtungen durch die Beziehungen zwischen den Zahlen abgebildet werden.

(in Fachbegriffen: Abbildung eines empirischen Relativs, d.h. der Beobachtungen, durch ein numerisches Relativ, d.h. die Zahlen, und zwar in homomorpher Weise, d.h. so, dass die Beziehungen erhalten bleiben).

Wenn ein Mensch z.B. hilfsbereiter ist als ein anderer, dann sollte dem hilfsbereiteren Menschen bei der Erfassung dieser Eigenschaft auch eine höhere Zahl zugeordnet werden als dem weniger hilfsbereiten. Um dies zu gewährleisten, erfolgt die Zuordnung der Zahlen zu den Beobachtungen in regelgeleiteter Weise.

Skalenniveaus psychologischer Messung

Eine Regel für die Zuordnung von Zahlen zu Beobachtungen wird als *Skala* bezeichnet. Es lassen sich folgende Skalen unterscheiden:

- *Nominalskala*: Die beobachteten Ereignisse sind voneinander verschieden. Verschiedene Zahlen drücken hier also lediglich Non-Identität aus (z.B. Geschlecht mit den Ausprägungen männlich = 1 und weiblich = 2). In der Psychologie ist umstritten, ob die Erhebung von Daten auf diesem Niveau als ‚Messen‘ zu bezeichnen ist.
- *Ordinalskala*: Die beobachteten Ereignisse lassen sich in eine Rangreihe bringen; die Abstände zwischen den Rängen können jedoch je unterschiedlich sein (z.B. Schulnoten). Die Zahlen sagen also lediglich aus, dass eine Beobachtung mehr oder weniger von einer bestimmten Eigenschaft aufweist.
- *Intervallskala*: Die den beobachteten Ereignissen zugeordneten Skalen repräsentieren nicht nur die Tatsache, sondern auch das Ausmaß eines Größenunterschieds (z.B. Punktzahl bei einem Intelligenztest).
- *Verhältnisskala*: Die den beobachteten Ereignissen zugeordneten Zahlen repräsentieren nicht nur Differenzen, sondern auch Proportionen. Die Skala hat einen absoluten Nullpunkt (z.B. Temperatur). Solche Skalen sind in der Psychologie sehr selten.

Je höher das Skalenniveau einer Messung, desto genauer sind die Aussagen, die mit der Messung möglich sind (ein Tisch hat eine andere Länge als ein zweiter Tisch; ein Tisch ist länger als ein zweiter Tisch; ein Tisch ist 20 cm länger als ein anderer Tisch; ein Tisch ist doppelt so lang wie ein zweiter Tisch). Je genauer die Aussagen über die Relationen der beobachteten Fälle, desto mehr Rechenoperationen sind in der Regel mit den Messungen möglich. Aus diesen beiden Gründen (Genauigkeit und Auswertungsmöglichkeiten) werden in der Psychologie meist Messungen auf Intervallskalenniveau angestrebt. Je nach Gegenstandsbereich ist dies jedoch nicht immer möglich, so dass die Psychologie es häufig auch mit Messungen auf Nominalskalenniveau zu tun hat.

1.4.2. GütekriterienDefinition

Psychologische Methoden müssen bestimmten Anforderungen genügen, die in Form sog. Gütekriterien spezifiziert sind (s. ausführlich II.4.2.). Dies sind:

- *Objektivität*: Unabhängigkeit der Untersuchungsergebnisse von den Personen, die Untersuchung durchführen (also die Methoden anwenden); eine Annäherung an Objektivität er-

folgt durch Intersubjektivität, d.h. die Ähnlichkeit der Ergebnisse, zu denen verschiedene Personen unter Anwendung derselben Methoden gelangen

- *Reliabilität*, d.h. Zuverlässigkeit, Messgenauigkeit: Grad der Genauigkeit, mit der ein Merkmal gemessen wird
- *Validität*, d.h. Gültigkeit: Ausmaß, in dem ein Instrument das erfasst, was erfasst werden soll

Beispiel

- Wenn in einer Prüfung über psychologische Methodenlehre dieselbe Studierende bei dem einen Prüfer eine 1.0 erzielt, bei dem anderen Prüfer dagegen eine 3.7 (unter ansonsten vergleichbaren Bedingungen), dann wäre die Prüfung nicht objektiv.
- Wenn dieselbe Studierende bei ein- und demselben Prüfer in einer Prüfung über psychologische Methodenlehre heute eine 1.0 erzielen würde und morgen (unter ansonsten gleichbleibenden Bedingungen) eine 3.7, dann wäre die Prüfung nicht reliabel.
- Wenn eine Prüfung nicht den Wissensstand einer Studierenden erfasst, sondern den Grad ihrer Nervosität, dann wäre die Prüfung nicht valide.

1.4.3. Operationalisierung

Im Zusammenhang mit den Anforderungen an wissenschaftliche Methoden wurde bereits erläutert, dass in der Wissenschaft vor Untersuchungsbeginn genau definiert werden muss, was unter einem bestimmten Begriff zu verstehen ist. Im Hinblick auf die empirische Untersuchung des Gegenstandes bedeutet das, dass der Begriff einer empirischen Erfassung zugänglich gemacht werden muss: Der Begriff muss *operationalisiert* werden. Das bedeutet, dass man beobachtbare Kriterien dafür anzugeben sind, wann der Sachverhalt vorliegt (und ggf. auch, in welcher Ausprägung).

Wann ist jemand beispielsweise aggressiv? Wenn er oder sie jemand anderen körperlichen Schaden zufügt? Wenn er oder sie jemand anderen beleidigt? Wenn er oder sie jemand anderen demonstrativ ignoriert?

Ein Begriff ist mittels beobachtbarer Kriterien jedoch nie vollständig operationalisierbar; es bleibt immer ein *Bedeutungsüberschuss*. Weiterhin erfasst jedes Kriterium immer nur einen bestimmten Bedeutungsaspekt. Vor diesem Hintergrund wird verständlich, weshalb die Ergebnisse psychologischer Untersuchungen zu ein und demselben Gegenstand manchmal so unterschiedlich ausfallen: Wenn ein Begriff wie Aggression in verschiedenen Untersuchungen auch je verschieden operationalisiert wird, dann wundert es auch nicht, wenn je verschiedene Ergebnisse resultieren.

Prüfungsliteratur: Bortz & Döring 2002, Kap. 2.3.5., 2.3.6.

I.5. Aufgaben und Ziele der Psychologie als Wissenschaft

Als Aufgaben und Ziele der Psychologie gelten: Beschreibung, Erklärung, Vorhersage und Veränderung (Technologie). Der Gegenstand der Psychologie lässt sich damit wie folgt spezifizieren: menschliches Erleben und Verhalten (Handeln) zu beschreiben, zu erklären, vorherzusagen und ggf. auch zu verändern. Inwieweit auch die Bewertung zu den Aufgaben und Zielen der Psychologie zählt, wird kontrovers diskutiert.

1.5.1. Beschreibung

Überblick

Die Beschreibung stellt insofern die grundlegendste Aufgabe der Psychologie dar, als alle anderen Aufgaben und Zielsetzungen auf einer Beschreibung aufbauen. Bevor man beispielsweise versucht, einen Sachverhalt zu erklären, sollte man zunächst ermitteln, um was für einen Sachverhalt es sich genau handelt. In der Psychologie sind Methoden der Beschreibung weitgehend gleichbedeutend mit Methoden der Datenerhebung (s. unten II.2.).

Klassifikation von Methoden der Beschreibung

Methoden der Beschreibung lassen sich nach drei Gesichtspunkten weiter unterteilen:

- *konkret vs. abstrakt:*
 - Konkrete Beschreibungen liegen vor, wenn etwa einzelne Beobachtungen wiedergegeben werden (z.B.: Die Studierende X hat sich in der ersten Sitzung der Vorlesung ‚Einführung in die psychologische Methodenlehre‘ dreimal gemeldet.). Methoden der konkreten Beschreibung sind das Beobachten und das Zählen.
 - Abstrakte Beschreibungen beinhalten dagegen subjektive *Einschätzungen* (z.B.: Die Studierende X hat sich für die genannte Vorlesung interessiert.). Abstrakte Beschreibungsverfahren sind das Beurteilen, das Befragen und das Testen.

- *qualitativ, quantitativ, klassifizierend:*
 - Beschreibungen können *qualitativ* sein, d.h. die genaue Beschaffenheit eines Sachverhalts wiedergeben (z.B. Irene spielt mit dem Bleistift).
 - Sie können aber auch *quantitative* Aussagen umfassen, z.B. darüber, wie häufig eine bestimmte Handlung ausgeführt wurde oder wie stark eine Eigenschaft bei einer Person ausgeprägt ist. Quantitative Formen der Beschreibung werden auch als *Messung* bezeichnet (s. unten I.5.).
 - Schließlich kann eine Beschreibung auch dadurch erfolgen, dass ein Sachverhalt als ein Sachverhalt eines bestimmten Typs *klassifiziert* wird (z.B. Unterteilung von Persönlichkeitseigenschaften in fünf Gruppen: Neurotizismus, Extraversion, Liebenswürdigkeit, Gewissenhaftigkeit, Offenheit für Erfahrung).

- *selbst vs. fremd:* Beschreibungen lassen sich weiterhin danach unterteilen, ob sie sich auf die eigene Person beziehen (*Selbstbeschreibung*) oder auf andere Personen (*Fremdbeschreibung*).

Die drei Klassifikationsdimensionen sind voneinander unabhängig. Eine quantitative Beschreibung kann also beispielsweise sowohl abstrakt als auch konkret erfolgen, kann sich auf die eigene Person beziehen oder auf andere usw.

Beobachtung

Wissenschaftliche Beobachtung unterscheidet sich von der Beobachtung im Alltag vor allem darin, dass Zufälligkeiten entgegengewirkt wird.

Im Alltag kann es z.B. sein, dass man gar nicht alle relevanten Sachverhalte auch bewusst wahrnimmt, sondern nur diejenigen, die der eigenen Vermutung entsprechen. Bei einem Streit in der Partnerschaft darüber, dass der andere morgens nie das Waschbecken saubermacht, nimmt man vermutlich vor allem die Tage wahr, in denen der andere das Waschbecken schmutzig hinterlässt – und nicht die, an denen er daran gedacht hat, es sauberzuwischen. Und wenn man die relevanten Sachverhalte wahrnimmt, dann kann es im Alltag sein, dass sie sich

zufällig zum Zeitpunkt der Beobachtung häufen (oder im Gegenteil ungewöhnlich selten auftreten).

Diesen Zufälligkeiten versucht man in der Wissenschaft entgegenzuwirken, indem man systematisch beobachtet. *Systematisch* beobachten heißt, dass man von vornherein festlegt, was man beobachten will, zu welchen Zeitpunkten, in welchen Zeiträumen und in welchen Situationen. Dabei ist die Auswahl so zu treffen, dass für die Fragestellung repräsentative Ergebnisse zu erwarten sind.

Einschätzung

Subjektive Einschätzungen sind (möglichst wertneutrale) zusammenfassende Kurzbeschreibungen, in denen einem Gegenstand eine bestimmte Eigenschaft zugeschrieben wird (z.B.: Irene ist nervös.). Subjektive Einschätzungen unterscheiden sich von Beobachtungen vor allem darin, dass Einschätzungen über das Beobachtete hinausgehen. Beobachtungen basieren im Wesentlichen auf der menschlichen Wahrnehmung; Einschätzungen greifen dagegen auf das menschliche Urteilsvermögen zurück.

Das Problem bei der Einschätzung besteht darin, dass etwas beschrieben werden soll, was der Beobachtung eigentlich nicht zugänglich ist (z.B. die Kreativität oder Intelligenz eines Menschen). Häufig kommen bei der Erfassung sog. Einschätzinstrumente zur Anwendung. Diese beinhalten eine Reihe von Aussagen, die beurteilt werden sollen; aus dem Ergebnis über alle Beurteilungen hinweg ergibt sich dann die Ausprägung einer Person auf dem interessierenden Merkmal. Die Aussagen stellen dabei die Kriterien im Rahmen der Operationalisierung des Begriffs dar (s.o. 1.4.3.).

Beispiel für ein Einschätzinstrument						
1. Ich fühle mich unabhängig.	1	2	3	4	5	Ich fühle mich abhängig.
2. Ich fälle leicht Entscheidungen.	1	2	3	4	5	Ich fälle schwer Entscheidungen.
3. Ich gebe leicht auf.	1	2	3	4	5	Ich gebe fast nie auf.
4. Ich bin nicht selbstsicher.	1	2	3	4	5	Ich bin sehr selbstsicher.
5. Ich fühle mich überlegen.	1	2	3	4	5	Ich fühle mich unterlegen.
6. Ich kann Druck schlecht standhalten.	1	2	3	4	5	Ich kann Druck gut standhalten.
7. Ich bin sehr aktiv.	1	2	3	4	5	Ich bin nicht aktiv.
<i>M+-Skala aus dem German Extended Personality Attributes Questionnaire (GEPAQ; Runge et al. 1981)#</i>						

1.5.2. Erklärung

Überblick

In der Psychologie gibt man sich häufig nicht mit der Beschreibung von Sachverhalten zufrieden (z.B. dass die Kinder an einer bestimmten Schule hochgradig aggressiv sind), sondern man versucht, diese Sachverhalte darüber hinaus auch zu erklären (z.B. herauszufinden, warum die Kinder so aggressiv sind). Erklärungen sind Antworten auf Warum-Fragen. Sie enthalten Angaben über Bedingungen von bzw. Abhängigkeiten zwischen Sachverhalten.

Subsumptionsmodell der Erklärung (Hempel-Opppenheim-Schema)

Die häufigste Form der Erklärung in der Psychologie ist die Erklärung nach dem sog. Subsumptionsmodell: Der zu erklärende Sachverhalt wird unter ein allgemeines Gesetz subsumiert.

Die Erklärung nach dem Subsumptionsmodell besteht im Wesentlichen aus zwei Teilen:

- dem *Explanandum*: dem Sachverhalt, der erklärt werden soll; z.B.: Tanja ist aggressiv gegenüber Maja (indem Tanja Maja schubst)
- dem *Explanans*: der Erklärung selbst, die sich weiter untergliedert in:
 - ein allgemeines Gesetz, z.B.: Frustration erzeugt Aggression;
 - eine Antezedens- bzw. Randbedingung, die dem zu erklärenden Sachverhalt vorausgeht, z.B.: Maja hat Tanja frustriert (indem Maja zuvor Tanja deren Lieblingsente weggenommen hat).

Genau genommen ist der Erklärungsprozess noch etwas komplizierter. Vor der Erklärung muss zunächst sichergestellt werden, dass die Sachverhalte im Explanandum und in der Antezedensbedingung sich in der Tat den Beschreibungsklassen zuordnen lassen, die im Gesetz spezifiziert sind. Es wäre also beispielsweise zu sichern, dass das Wegnehmen der Lieblingsente in der Tat eine Form der Frustration darstellt und das Schubsen in der Tat eine Form der Aggression.

Erklärungen setzen somit immer Beschreibungen voraus. Man sagt auch: Ein Sachverhalt ist nie als solcher eine Erklärung, sondern immer nur unter einer bestimmten Beschreibung. In einem anderen Zusammenhang könnte die Antezedensbedingung der einen Erklärung selbst Explanandum einer anderen Erklärung sein (Warum z.B. hat Maja Tanja die Lieblingsente weggenommen?).

Erklärung nach dem Subsumptionsmodell

Gesetz: Frustration erzeugt Aggression.

Antezedensbedingung: Maja frustriert Tanja (indem Maja Tanja deren Lieblingsente wegnimmt).

Explanandum: Tanja ist aggressiv gegenüber Maja (indem Tanja Maja schubst).

Erklärungen erster und zweiter Ordnung

Erklärungen wie die beschriebene, bei denen ein Bedingungsverhältnis zwischen zwei Sachverhalten angegeben wird, werden auch als *Erklärung erster Ordnung* bezeichnet. Der bedingende Sachverhalt (also der Sachverhalt im Explanans) wird auch als unabhängige Variable

bezeichnet, der bedingte bzw. zu erklärende Sachverhalt (im Explanandum) als abhängige Variable.

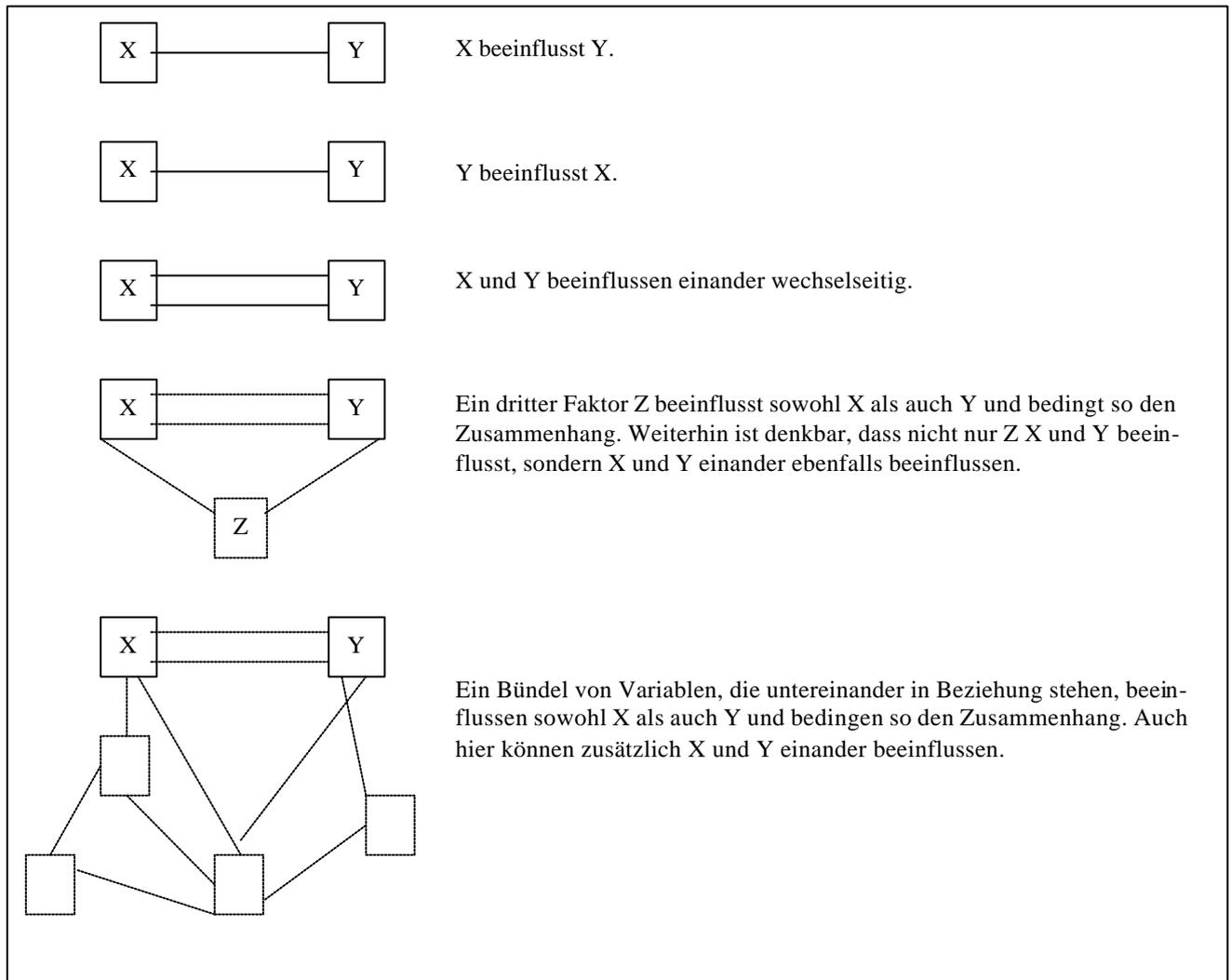
Gerade in der Psychologie besteht aber häufig ein Interesse an ‚Erklärungen der Erklärungen‘. Warum sollte Frustration zu Aggression führen? Warum ist Lernen mit Pausen in der Regel effektiver als Lernen ohne Pausen? Solche ‚Erklärungen von Erklärungen‘ werden als *Erklärungen zweiter Ordnung* bezeichnet. Ihr Schwerpunkt liegt in der Regel auf den vermittelnden Prozessen zwischen den relevanten Sachverhalten im Individuum (z.B. auf Prozessen der Gedächtnisorganisation und –struktur bezogen auf das Beispiel Lernen mit vs. ohne Pausen). Solche vermittelnden Konstrukte, die der Beobachtung nicht direkt zugänglich sind, werden auch *hypothetische Konstrukte* genannt.

Vorsicht: Zusammenhang ist nicht gleich Ursache

Es wurde schon mehrfach darauf hingewiesen, dass Vorannahmen einen Einfluss darauf haben, wie Wissenschaftler/innen ihren Gegenstand sehen. Solche Vorannahmen können leicht zu falschen Schlussfolgerungen führen, wenn man empirisch einen Zusammenhang zwischen zwei Sachverhalten ermittelt hat (z.B. zwischen der Aggressivität von Jugendlichen und der Höhe ihres Konsums von Gewaltfilmen und –videos). Wenn man beispielsweise der Überzeugung ist, dass ein hoher Konsum von Medienprodukten mit Gewaltinhalten die Rezipienten/innen zu Gewalttaten anregt, dann wird man den Zusammenhang vermutlich im Sinne eines Ursachengefüges interpretieren: Hoher medialer Gewaltkonsum führt zu erhöhter Aggressivität. Der Medienkonsum wird als Erklärung für die gewalttätigen Handlungen angesetzt.

Zusammenhänge können aber auch auf andere Art und Weise zustande kommen:

- So kann dem Zusammenhang durchaus ein Ursachengefüge zugrunde liegen – aber was ist Ursache, was ist Wirkung? So ist durchaus denkbar, dass Jugendliche, die zu Gewalt neigen, bevorzugt Filme mit Gewaltinhalten auswählen.
- Oder es ist denkbar, dass die beiden Sachverhalte einander wechselseitig beeinflussen: Wer aggressiv ist, bevorzugt Filme mit Gewaltinhalten. Diese erhöhen die Aggressivität, was wiederum zu einer vermehrten Bevorzugung von Filmen mit Gewaltinhalten führt – usw.
- Möglich wäre auch, dass die Aggressivität und die Vorliebe für gewalttätige Filme eigentlich gar nichts miteinander zu tun haben – dass der empirische Zusammenhang nur durch eine Drittvariable zustande kommt: Beide Sachverhalte stehen in Zusammenhang mit der Drittvariablen, aber nicht miteinander. Vielleicht besteht ein Zusammenhang zwischen Kontaktarmut und Aggressivität einerseits sowie zwischen Kontaktarmut und Vorliebe für gewalthaltige Filme andererseits. In Erweiterung dieses Modells wäre auch denkbar, dass alle drei Variablen untereinander in Zusammenhang stehen und sich gegenseitig beeinflussen.
- Schließlich könnte der Zusammenhang zwischen zwei Variablen wie Aggressivität und Präferenz für Gewaltfilme auch Teil eines ganzen Geflechts von Variablen sein, die in komplexer Weise miteinander interagieren: Kontaktarmut, passives Familienklima, soziale Schicht, Bildung usw.



1.5.3. Prognose

Prognose als vorwärts gerichtete Erklärung

Ein drittes Ziel der Psychologie als Wissenschaft besteht darin, Vorhersagen über künftige Sachverhalte zu treffen (z.B. in Form von Hypothesen, die dann einer empirischen Prüfung unterzogen werden). Prognosen lassen sich als vorwärts gerichtete Erklärungen auffassen. Es liegt ihnen derselbe Bedingungs-zusammenhang zwischen zwei Sachverhalten zugrunde wie bei der Erklärung. Nur geht es hier nicht darum, die abhängige Variable auf der Grundlage der unabhängigen zu erklären, sondern die abhängige Variable vorherzusagen. Was in der Erklärung die abhängige Variable darstellt, wird bei der Prognose als *Kriterium* bezeichnet (das, was vorhergesagt werden soll); was bei der Erklärung die unabhängige Variable ist, gilt bei der Prognose als *Prädiktor* (dasjenige, worauf sich die Prognose stützt).

Strukturparallelität von Erklärung und Prognose

Wenn man Prognosen als vorwärts gerichtete Erklärungen betrachtet, so liegt dieser Auffassung die Annahme zugrunde, dass Erklärung und Prognose eine vergleichbare Struktur aufweisen – lediglich die Ausgangssituation ist eine andere.¹

¹ Die Annahme der Strukturparallelität von Prognose und Erklärung wird in der heutigen Wissenschaftstheorie kritisch gesehen – mehr dazu in der Vorlesung zur Wissenschaftstheorie im 3. Semester.

- Bei der Erklärung geht man vom Explanandum aus (Tanja schubst Maja) und fragt nach den Gesetzen und Randbedingungen, die diesen Sachverhalt erklären können.
- Bei der Prognose geht man dagegen von der Randbedingung aus (Maja hat Tanja gerade die Lieblingsente weggenommen) und sucht nun nach einem geeigneten Gesetz, das es erlaubt, zusammen mit der Randbedingung einer Vorhersage über das Folgeereignis abzugeben.

Bedingungen der Vorhersage-Genauigkeit

Die Genauigkeit einer Prognose hängt u.a. von den folgenden Faktoren ab:

- Auswahl relevanter Prädiktoren;
- Präzise Beschreibung von sowohl den Prädiktoren als auch dem Kriterium, d.h. eine Beschreibung, die den Gütekriterien von Objektivität, Reliabilität und Validität genügt;
- Gewichtung der Prädiktoren nach ihrer empirischen Bedeutung;
- Zeitraum der Prognose: Je länger der Zeitraum, über den sich eine Prognose erstreckt, desto ungenauer wird sie.

Statistische vs. individuelle Prognose

Beschreibungen, Erklärungen und Prognosen in der Psychologie beziehen sich in der Regel nicht auf den Einzelfall, sondern auf Menschen im Allgemeinen oder auf eine bestimmte Personengruppe. Gute Prädiktoren sind in diesem Zusammenhang solche Prädiktoren, die sich empirisch bei der Überprüfung an einer größeren Personengruppe möglichst gut bewähren, die also das Auftreten eines Kriteriums auch tatsächlich beeinflussen. Die Stärke des Einflusses eines Prädiktors auf ein Kriterium wird auch als der statistische Prognosewert des Prädiktors bezeichnet.

Dieser Prognosewert bezieht sich allerdings immer nur auf den Grad der Bewährung des Prädiktors an einer größeren Personengruppe, nicht auf den Einzelfall. Wenn man es in der Praxis mit einem Einzelfall zu tun hat, also eine individuelle Prognose erstellen möchte, muss man notwendig auf solche Prädiktoren zurückgreifen, die sich statistisch am besten bewährt haben. Im Einzelfall erlauben diese Prädiktoren jedoch nie eine genaue Prognose. Gute Prädiktoren sind solche, die häufig zutreffen – ob sie auch in dem speziellen Einzelfall zutreffen, hängt von einer Vielzahl weiterer Faktoren ab.

1.5.4. Veränderung (Technologie)

Typen von Veränderungsaufgaben

Eine vierte Zielsetzung der Psychologie besteht darin, gezielt auf den Gegenstand einzuwirken, ihn in einer bestimmten Weise zu verändern. Solche Veränderungs- und Interventionsaufgaben sind vor allem in der praktisch-psychologischen Tätigkeit von Bedeutung.

Es lassen sich drei Arten von Veränderungsaufgaben unterscheiden:

- *Korrektur*: Ein problematischer Ist-Zustand soll aufgehoben werden (z.B. eine Depression im Rahmen einer Psychotherapie).
- *Förderung* (Optimierung): Ein künftiger positiver Zustand wird angestrebt (z.B. Erziehung, Ausbildung).
- *Prävention*: Es soll verhindert werden, dass ein problematischer Ausgangszustand überhaupt erst eintritt (z.B. Erlernen von Techniken zur Stressbewältigung).

Zwischen diesen drei Veränderungsaufgaben bestehen fließende Übergänge; so kann z.B. eine Beratung im Rahmen einer schulpsychologischen Tätigkeit der Korrektur, Förderung und Prävention gleichermaßen dienen.

Probleme bei Veränderungsaufgaben

Veränderungen gestalten sich in der Praxis jedoch meist schwierig. Insbesondere können die folgenden Probleme auftreten:

- *Zielprobleme:* So stellt sich beispielsweise die Frage, worauf genau die Intervention abzielt, welcher Zielzustand z.B. den problematischen Ausgangszustand ersetzen soll (Beschreibungsproblem). Wenn ein Mann z.B. mit dem Tod der Lebenspartnerin nicht zurechtkommt – besteht das Ziel dann darin, dass er wieder gesellschaftlich ‚funktioniert‘ und seiner Arbeit nachgehen kann? Oder darin, die Trauer bei sich selbst zuzulassen? Oder darin, eine neue Lebenspartnerin zu finden? Zielsetzungen enthalten immer auch Bewertungen. In diesem Zusammenhang stellt sich außerdem die Frage, ob ein bestimmtes Ziel überhaupt erstrebenswert ist (wenn besagter Mann beispielsweise in die Therapie kommt, um wieder zu ‚funktionieren‘).
- *Mittelprobleme:* Wenn ein Ziel einmal festgelegt ist, schließt sich die Frage an, mit welchen Mitteln das Ziel erreicht werden soll. Die Mittel einer psychologischen Intervention werfen ebenfalls Probleme auf. Zunächst stellt sich die Frage nach der Ursache des Problems; das Problem ist also zu erklären (Verfügt der Mann beispielsweise nicht über ein soziales Netz? Stellt er vielleicht zu hohe Ansprüche an sich selbst?). Geeignete Mittel der Veränderung können immer nur vor dem Hintergrund von Erklärungen festgelegt werden. Auch bei gegebenen Mitteln ergibt sich wiederum ein Bewertungsproblem: Sind die Mittel akzeptabel? So wurde beispielsweise zur Therapie von Phobien die Technik des ‚flooding‘ entwickelt, bei der Klienten/innen massiv und ohne Ausweichmöglichkeiten mit den Angst auslösenden Stimuli konfrontiert werden. Diese Technik mag effektiv sein; es stellt sich aber die Frage, inwieweit es ethisch gerechtfertigt ist, Menschen dem resultierenden Stress auszusetzen.

Nicht zuletzt ergeben sich im Zusammenhang mit Veränderungsaufgaben auch Probleme hinsichtlich der Durchführbarkeit. Interventionen finden immer in einem gesellschaftlichen Kontext statt, der die Auswahl an möglichen Zielen und Mitteln einschränkt. Ein trauernder Mensch bräuchte vielleicht im Optimalfall eine ‚Auszeit‘ von mehreren Monaten – eine Situation, in der aber sein Lebensunterhalt nicht gesichert wäre.

1.5.5. Bewertung

Beschreibung und Bewertung

Die wissenschaftstheoretische Diskussion darüber, ob die Bewertung zu den Aufgaben und Zielen der Wissenschaft zählt, basiert auf der grundlegenden Unterscheidung zwischen Beschreibung und Bewertung, zwischen Deskription und Präskription. Deskriptive Sätze sind Ist-Sätze, die Sachverhalte beschreiben (Die Wand ist weiß; Emma erzielt auf der Skala ‚need for cognition‘ einen Punktwert von 20). Präskriptive Sätze sind z.B. Soll-Sätze, in denen Bewertungen vorgenommen oder Handlungsempfehlungen ausgesprochen werden (z.B. Horrorfilme sind schädlich; Horrorfilme sollten nicht für Jugendliche unter 18 Jahren freigegeben sein).

Warum ist die Bewertung so kontrovers?

Bewertungen sind in der Wissenschaft vor allem aus zwei Gründen umstritten:

- Bewertungen sind nicht vollständig begründbar. Bei einer Begründungskette stösst man irgendwann auf einen ‚Grundwert‘, der eine Setzung darstellt (Warum soll man anderen

beispielsweise keinen Schaden zufügen?). Es wurde argumentiert, dass Bewertungen in diesem Sinne subjektiv sind und den Anforderungen an wissenschaftliche Aussagen nicht genügen.

- Gegenstand der Wissenschaft sind deskriptive Sätze. Aus deskriptiven sind jedoch keine präskriptiven Sätze herleitbar.

Vor diesem Hintergrund formulierte Max Weber zu Beginn des 20. Jahrhunderts das sog. Werturteilsfreiheitspostulat. Darin wird die Forderung aufgestellt, dass Wissenschaftler/innen in dem Gegenstandsbereich ihrer Forschung keine Bewertungen vornehmen sollten; Bewertung wird hier als Gegenstandsbereich von Wissenschaft ausgeschlossen.

Mit dem Werturteilsfreiheitspostulat ist allerdings zweierlei *nicht* gemeint:

- Wissenschaft beinhaltet insofern notwendig Wertungen, als beispielsweise Hypothesen auf ihre Richtigkeit hin geprüft werden. Diese Form von Wertung ist auch nach Weber durchaus zulässiger bzw. sogar notwendiger Bestandteil von Wissenschaft.
- Wissenschaft kann sich damit beschäftigen, auf welcher Grundlage Menschen Bewertungen vornehmen (z.B. Entstehung von Vorurteilen). Auch wissenschaftliche Forschung über den Gegenstand ‚Bewertung‘ ist nach Weber durchaus zulässig.

Nicht zulässig ist dagegen im Rahmen des Werturteilsfreiheitspostulats die Ableitung von Bewertungen aus den eigenen Forschungsergebnissen, z.B.: Weil der Konsum von Horrorfilmen mit einer erhöhten Aggressivität einhergeht, sollten Horrorfilme verboten werden.

Lässt sich das Werturteilsfreiheitspostulat aufrecht erhalten?

In der wissenschaftstheoretischen Diskussion hat sich allerdings gezeigt, dass Wertung sich aus der Wissenschaft nur schwer ausklammern lässt:

- Viele Begriffe in der Wissenschaft sind kryptonormativ: Es scheint sich um reine Deskriptionen zu handeln, die aber bei näherem Hinsehen durchaus Wertungen enthalten: intelligent, aggressiv, kreativ usw.
- Ziel- und Mittelprobleme, wie sie sich im Rahmen von Veränderungsaufgaben stellen, beinhalten wesentlich auch Bewertungsfragen.

Die Position in Bezug auf diese Frage ist letztlich kontrovers: Manche Wissenschaftler/innen halten das Postulat aufrecht; andere vertreten dagegen die Position, dass es nicht möglich ist, Bewertungen zu vermeiden, und dass diese daher möglichst offen gelegt und somit auch für andere kritisierbar gemacht werden sollten.

Begründete Kritik von Bewertungen

Wertungen mögen zwar nicht bis ins Letzte begründbar sein; sie sind aber auch nicht gänzlich subjektiv und willkürlich.

Bewertungen bauen immer auf zwei Prämissen auf: einer präskriptiven und einer deskriptiven.

Begründung einer Bewertung: Beispiel

Bewertung: Beim Lernen Musik zu hören ist schlecht.

Begründung – Prämissen:

- deskriptiv: Wenn man beim Lernen Musik hört, kann man sich nicht gut konzentrieren.
- präskriptiv: Geringe Konzentration beim Lernen ist schlecht.

Die Kritik, dass Bewertungen nicht schlüssig begründbar sind, bezieht sich auf die deskriptive Prämisse. Hier könnte man weiterfragen: Warum ist es schlecht, wenn man sich beim Lernen nicht konzentrieren kann usw. In Bezug auf die deskriptive Prämisse sind Bewertungen aber durchaus kritisierbar: Wenn sich beispielsweise empirisch herausstellen sollte, dass es durch-

aus Menschen gibt, die sich mit Musik beim Lernen ebenso gut oder sogar besser konzentrieren können als ohne, dann ist die Bewertung in dem obigen Beispiel offensichtlich falsch – ganz unabhängig davon, ob jemand die präskriptive Prämisse teilt oder nicht.

Prüfungsliteratur: Nolting & Paulus 1999, Kap. 7

II. Quantitative Verfahren

Im Folgenden werden zunächst einige Grundbegriffe quantitativer Forschung dargestellt. Im Anschluss wird (in dieser Reihenfolge) genauer auf Methoden der Datenerhebung sowie auf Forschungsansätze eingegangen. Die Datenanalyse wird in diesem Skript nicht behandelt; s. dazu Selg, Klapprott & Kamentz 1992, Teil II; Hussy & Jain 2002, Kap. 4.1.

II.1. Grundbegriffe

II.1.1. Untersuchungsarten quantitativer Forschung

Überblick

In der quantitativen Forschung wird zwischen drei Untersuchungsarten unterschieden: Erkundungsstudien, populationsbeschreibenden Untersuchungen sowie hypothesentestenden (explanativen) Untersuchungen.

Erkundungsstudien

Erkundungsstudien (explorative Studien) dienen dem Erkunden, dem Beschreiben eines Gegenstandes. In der quantitativen Forschung haben Erkundungsstudien häufig den Zweck, zu einer Hypothese zu gelangen, um im Anschluss eine hypothesentestende Untersuchung durchführen zu können. Den Erkundungsstudien kommt daher im quantitativen Bereich oft nur eine Hilfsfunktion zu. Erkundungsstudien sind für quantitative Forschung wenig charakteristisch. Eine eigenständige bedeutsame Rolle kommt solchen erkundenden Untersuchungen jedoch im qualitativen Bereich zu.

Ziel einer Erkundungsstudie könnte es beispielsweise sein, die Geschlechterrollenvorstellungen von Jugendlichen zu erheben.

Populationsbeschreibende Untersuchungen

Populationsbeschreibende Untersuchungen (auch: Survey-Studien genannt) dienen dazu, eine Personengruppe (die Grundgesamtheit: s. unten II.1.2.) hinsichtlich relevanter Merkmale zu beschreiben. Zu diesem Zweck werden die Merkmalsausprägungen an einer kleineren Teilgruppe (der Stichprobe) erhoben; von der Stichprobe wird auf die Grundgesamtheit zurückgeschlossen. Populationsbeschreibende Untersuchungen werden häufig in der demoskopischen Forschung sowie in der Medienwissenschaft durchgeführt.

Eine populationsbeschreibende Untersuchung läge beispielsweise vor, wenn es darum geht, die Häufigkeit psychogener Erkrankungen in der Bevölkerung zu ermitteln.

Hypothesenprüfende Untersuchungen

Hypothesenprüfende (explanative) Untersuchungen stellen in der quantitativen Psychologie die häufigste Untersuchungsart dar. Diese Untersuchungsart ist auf Erklärungen ausgerichtet; Ziel ist es, eine Kausalhypothese zu überprüfen (wiederum, indem von einer Stichprobe auf eine Grundgesamtheit zurückgeschlossen wird).

Eine Überprüfung, ob und ggf. wie sich die bildliche Vorstellbarkeit von Begriffen auf das Behalten dieser Begriffe auswirkt, wäre ein Beispiel für eine hypothesenprüfende Untersuchung.

Variable: Definition

In praktisch allen Untersuchungsarten wird der interessierende Gegenstand in Form von Variablen erhoben. Unter Variablen versteht man veränderliche Beobachtungsgrößen. Jede Variable hat mindestens zwei (in der Regel mehrere) Ausprägungen. Geschlecht stellt z.B. eine Variable mit den beiden Ausprägungen ‚männlich‘ und ‚weiblich‘ dar.

Beobachtungsgrößen, die nur eine Ausprägung annehmen können, werden als ‚Konstante‘ bezeichnet (z.B. die Lichtgeschwindigkeit).

In empirischen Untersuchungen müssen relevante Variablen zunächst operationalisiert werden, d.h. es sind Kriterien für die Variablenausprägungen festzulegen, die der Beobachtung zugänglich sind (vgl. oben I.4.3.).

Prüfungsliteratur: nur Skript

II.1.2. Stichprobenarten

Überblick

In empirischen Untersuchungen werden meistens Aussagen über größere Personengruppen angestrebt (z.B.: alle Menschen; Frauen; Jugendliche usw.). Aufgrund der Größe dieser Gruppen ist es aber in der Regel nicht möglich, die Untersuchung auch tatsächlich an allen Gruppenmitgliedern durchzuführen. Untersucht wird daher nur eine sog. Stichprobe, von der (mittels inferenzstatistischer Verfahren) auf die Eigenschaften der gesamten Gruppe zurückgeschlossen wird.

Population - Definition

Die gesamte Gruppe von Elementen, über die eine Aussage gemacht werden soll, bezeichnet man als *Grundgesamtheit* bzw. als *Population*. Populationen können *endlich* sein oder *unendlich*. Bei endlichen Populationen lässt sich die Anzahl der Elemente genau angeben (z.B. die Anzahl der Studierenden im Diplomstudiengang Psychologie im ersten Semester an der Universität zu Köln). Bei unendlichen Populationen ist dies nicht möglich (z.B. alle Menschen – die derzeit leben, jemals gelebt haben und künftig leben werden); unendliche Populationen können notwendiger Weise nicht in ihrer Gesamtheit untersucht werden. Die Variablenwerte einer Population bezeichnet man als *Parameter* (z.B. das mittlere Einkommen aller Jugendlichen in Deutschland unter 21 Jahren).

Stichprobe - Definition

Eine *Stichprobe* ist als eine Teilmenge aus einer Population definiert, die mittels eines bestimmten Auswahlverfahrens gewonnen wurde. Die Variablenwerte einer Stichprobe bezeichnet man als *Kennwerte*. In einer empirischen Untersuchung will man also von Stichprobenkennwerten auf Populationsparameter schließen. Damit man bei diesem Schluss auch zu gültigen Aussagen kommt, muss die Stichprobe für die Grundgesamtheit *repräsentativ* sein, d.h. die Stichprobe muss der Grundgesamtheit in ihrer Zusammensetzung möglichst stark ähneln.

Stichprobenarten

Verschiedene Verfahren der Stichprobenziehung stellen die Repräsentativität der Stichprobe in unterschiedlichem Maß sicher. Auf höchstem Abstraktionsniveau lassen sich zwei Arten von

Stichproben unterscheiden: *probabilistische* und *nicht-probabilistische* Stichproben. Grundsätzlich sind probabilistische Stichproben vorzuziehen, da hier die Repräsentativität höher ist.

Probabilistische Stichproben	Nicht-probabilistische Stichproben
Einfache Zufallsstichprobe	Quotenstichprobe
Geschichtete Stichprobe	Ad hoc-Stichprobe
Klumpenstichprobe	Qualitative Verfahren der Stichprobenziehung

Probabilistische Stichproben

Probabilistische Stichproben sind dadurch definiert, dass die Auswahlwahrscheinlichkeit der einzelnen Elemente bekannt (und im besten Fall für alle Elemente gleich) ist. Probabilistische Stichproben basieren auf einer möglichst vollständigen Auflistung der Grundgesamtheit. Aus dieser Liste werden die einzelnen Elemente der Stichprobe nach einem Zufallsverfahren ausgewählt (z.B. unter Nutzung einer Tabelle von Zufallszahlen).

- *Einfache Zufallsstichprobe*: Bei einer Zufallsstichprobe hat jedes Element der Grundgesamtheit die gleiche und unabhängige Chance, in die Stichprobe aufgenommen zu werden. Die Auswahl eines Elements hat keinerlei Auswirkungen auf die Auswahl anderer Elemente.
- *Geschichtete Stichprobe*: Die Population wird anhand bestimmter Merkmale (meist soziodemographischer Merkmale, wie Geschlecht, Alter, soziale Schicht usw.) in Teilpopulationen zerlegt. Aus jeder Teilpopulation wird eine Zufallsstichprobe gezogen.
- *Klumpenstichprobe*: Aus einer Population, die in natürliche Gruppen („Klumpen“) gegliedert ist (z.B. Schulklassen), werden per Zufall einige Gruppen ausgewählt und in ihrer Gesamtheit untersucht.

Nicht-probabilistische Stichproben

Bei nicht-probabilistischen Stichproben ist die Wahrscheinlichkeit der Auswahl der einzelnen Elemente nicht bekannt. Es werden u.a. die folgenden Arten unterschieden:

- *Quotenstichprobe*: Es wird die Zusammensetzung der Grundgesamtheit hinsichtlich relevanter Merkmale ermittelt (wiederum meist hinsichtlich soziodemographischer Merkmale wie Geschlecht usw.). Die Stichprobe wird so zusammengestellt, dass sie in ihrer Zusammensetzung in Bezug auf diese Merkmale der Grundgesamtheit entspricht. Wenn beispielsweise bekannt ist, dass die Population Deutschlands sich zu 52% aus Frauen und zu 48% aus Männern zusammensetzt, dann würde eine entsprechende Quotenstichprobe ebenfalls zu 52% aus Frauen und zu 48% aus Männern bestehen.
- *Ad hoc-Stichprobe*: Eine ad hoc-Stichprobe (auch: anfallende Stichprobe, Gelegenheitsstichprobe) setzt sich aus denjenigen Elementen der Population zusammen, die gerade für eine Untersuchung verfügbar sind. Ad hoc-Stichproben sind nicht repräsentativ und können zu verzerrten Schlussfolgerungen über die Population führen.
- *Qualitative Verfahren der Stichprobenziehung*: Auch bei den Verfahren der Stichprobenziehung, wie sie in der qualitativen Forschung entwickelt wurden (z.B. Theoretical sampling) handelt es sich um nicht-probabilistische Stichproben (vgl. III.4.1.).

Prüfungsliteratur: Bortz & Döring 2002, Kap. 7.1.; Hussy & Jain 2002, Kap. 3.5.1.

II.1.3. Populationsbeschreibende Untersuchungen

Bei populationsbeschreibenden Untersuchungen geht man wie folgt vor:

- Finden einer Fragestellung
- Operationalisierung der Variablen
- Untersuchungsplanung
- Stichprobenziehung
- Datenerhebung
- Auswertung
- Diskussion

Prüfungsliteratur: nur Skript

II.1.4. Hypothesentestende Untersuchungen: Die Prüfung von Kausalhypothesen

Überblick

Hypothesentestende unterscheiden sich von populationsbeschreibenden Untersuchungen vor allem darin, dass eine Kausalhypothese geprüft wird.

Hypothesen

Hypothesen stellen vermutete Antworten auf Fragen dar (z.B. die Vermutung, dass Menschen aggressiv reagieren, wenn sie frustriert werden; oder dass sich die Behaltensleistung steigern lässt, wenn man beim Lernen Pausen einlegt). Hypothesen werden meist als *Kausalhypothesen* formuliert, d.h. als wenn-dann oder je-desto-Sätze, z.B.:

- Wenn Personen frustriert sind, dann reagieren sie aggressiv.
- Je stärker Personen frustriert sind, desto aggressiver reagieren sie.

Diese Varianten von Kausalhypothesen sind besonders günstig, weil die beiden Hypothesenteile (Explanandum und Explanans) bereits in der Formulierung getrennt sind (vgl. oben I.5.2.).

Anforderungen an wissenschaftliche Hypothesen

An wissenschaftliche Hypothesen werden vier Anforderungen gestellt:

- Die Hypothese muss *widerspruchsfrei* sein. Diese Forderung ist bei Einzelhypothesen vergleichsweise leicht zu realisieren; bei komplexeren Theorien kann es allerdings durchaus vorkommen, dass sich Widersprüche einschleichen.
- Die Hypothese muss *prinzipiell widerlegbar* sein. Es muss also mögliche Sachverhalte geben, die zu der Hypothese im Widerspruch stehen; solche Sachverhalte werden als *Falsifikatoren* bezeichnet. Eine Person, die frustriert wird, jedoch nicht aggressiv reagiert, würde z.B. einen Falsifikator für die obige Frustrations-Aggressions-Hypothese darstellen. Die Hypothese „Wenn der Hahn kräht auf dem Mist, ändert sich das Wetter, oder es bleibt, wie es ist“ ist dagegen nicht falsifizierbar; sie wäre daher als wissenschaftliche Hypothese nicht akzeptabel. Je mehr Falsifikatoren es für eine Hypothese gibt, desto höher ist ihr *empirischer Gehalt*.
- Die in der Hypothese enthaltenen Begriffe müssen *operationalisierbar* sein; es müssen ihnen also Beobachtungen zuordenbar sein.
- Schließlich sollte die Hypothese *begründet* sein, d.h. die Herleitung der Hypothese sollte nachvollziehbar sein.

Vorgehen bei hypothesenprüfenden Untersuchungen

Aus dem Aufstellen einer Kausalhypothese ergeben sich für das Vorgehen bei hypothesentestenden Untersuchungen zwei Veränderungen gegenüber populationsbeschreibenden Untersuchungen.

- Erstens ist es erforderlich, einen Versuchsplan aufzustellen (s. unten II.3.1.).
- Zweitens erfolgt die Testung der statistischen Hypothese mittels eines inferenzstatistischen Signifikanztests. Aus den Ergebnissen wird auf die Geltung der theoretischen Hypothese in der Population zurückgeschlossen.

Einschränkungen bei der Hypothesenprüfung

- Hypothesen wie die oben aufgeführten Kausalhypothesen, die für alle Mitglieder einer Population gelten sollen, können anhand einer Stichprobe niemals verifiziert, sondern nur *falsifiziert* werden. Es kann nicht mit Sicherheit ausgeschlossen werden, dass in der Population ein Element existiert, das im Widerspruch zu der Hypothese steht.
- Eine theoretische Hypothese wie die Frustrations-Aggressions-Hypothese kann in einer empirischen Untersuchung nie in ihrem vollen Umfang geprüft werden. Geprüft wird die Hypothese immer nur für eine ganz bestimmte Operationalisierung auf der Ebene bestimmter statistischer Kennwerte (s. genauer unten II.3.1.).

Prüfungsliteratur: Bortz & Döring 2002, Kap. 1; Hussy & Jain 2002, Kap. 1.4.

II.2. Methoden der Datenerhebung

Methoden der Datenerhebung in der quantitativen Psychologie lassen sich unterteilen in Einschätzverfahren und Beobachtungsverfahren. Zu den Einschätzverfahren zählen Methoden des Befragens, Urteilens und Testens; unter die Beobachtungsverfahren fallen die Beobachtung selbst, das Zählen sowie das Messen.

II.2.1. Befragung

Arten der Befragung

Befragungsverfahren lassen sich nach verschiedenen Kriterien weiter unterteilen:

- *Schriftliche versus mündliche Befragung:* Die schriftliche Befragung wird mittels *Fragebogen* durchgeführt; sie findet sich häufiger in der quantitativen als in der qualitativen Forschung. Die mündliche Form der Befragung (*Interview*) ist dagegen für die qualitative Forschung typischer.
- *Standardisierte versus nonstandardisierte Befragung:* Der Grad der Standardisierung einer Befragung bezieht sich auf die *Freiheitsgrade der befragten Person*. Bei einer standardisierten Befragung werden (in Form *geschlossener Fragen*) Antwortmöglichkeiten vorgegeben, unter denen die befragte Person auswählt. Bei der unstandardisierten Befragung (in Form *offener Fragen*) kann die befragte Person die Antwort in ihren eigenen Worten formulieren. Quantitative Befragungsverfahren sind standardisierte Verfahren.
- *Strukturierte versus unstrukturierte Befragung:* Der Grad der Strukturiertheit einer Befragung bezieht sich auf die *Freiheitsgrade der Forscher/innen*. Bei einer strukturierten Befragung sind Wortlaut und Reihenfolge der Fragen genau vorgegeben. Bei einer *halbstrukturierten* Befragung ist als Orientierung ein Leitfaden mit vorformulierten Fragen

vorgegeben; der/die Forscher/in kann jedoch die genaue Formulierung und die Reihenfolge der Fragen variieren. In der *nonstrukturierten* Befragung schließlich muss der/dieForscher/in hinsichtlich Frageformulierung und Reihenfolge der Fragen keine Vorgaben einhalten. In der quantitativen Forschung dominiert die strukturierte Befragung.

- *Anzahl der befragten Personen*: An einer *Einzelbefragung* nimmt nur eine befragte Person teil. Bei einer *Gruppenbefragung* oder –diskussion werden die Meinungen und Einstellungen einer kleineren Gruppe (zwischen 5 und 15 Personen) erhoben. Bei einem *Survey* schließlich kann die Anzahl der Befragten sehr hoch sein.
- Weitere Kriterien der Unterteilung von Formen der Befragung sind beispielsweise die *Art des Kontaktes* (face-to-face, telefonisch, schriftlich usw.) *Anzahl der Forscher/innen* (Einzelperson, Tandem, Panel) sowie der *Autoritätsanspruch* (weich, neutral, hart).

Schriftliche Befragung: Überblick

Eine schriftliche Befragung ist immer dann gegeben, wenn Untersuchungsteilnehmer/innen schriftlich vorgelegte Fragen selbst wieder in schriftlicher Form beantworten.

Die Fragen in Fragebögen werden auch als *Items* bezeichnet. Ein Item ist definiert als eine als Frage oder als Urteil formulierte Aussage, zu der die befragte Person ihre Zustimmung oder Ablehnung – ggf. in unterschiedlicher Intensität – äußern kann.

Es lassen sich zwei Arten von Fragebögen unterscheiden: solche zur Erfassung von Persönlichkeitsmerkmalen oder -einstellungen sowie solche zur Erfassung konkreter Verhaltensweisen.

Fragebögen enthalten meist geschlossene Fragen mit vorgegebenen Antworten. Um solche Fragen zu formulieren, ist Vorwissen über den Gegenstand erforderlich.

Vorteile und Nachteile der schriftlichen Befragung

- *Vorteile*: Die schriftliche Befragung ist *kostengünstig* und gut geeignet, wenn *größere Personengruppen* untersucht werden sollen.
- *Nachteile*: Wenn die Befragung nicht als Gruppenuntersuchung durchgeführt wird, sondern die Fragebögen den Untersuchungsteilnehmern/innen nach Hause mitgegeben oder zugeschickt werden, entzieht sich die Untersuchungssituation der Kenntnis und der Kontrolle der Forscher/innen. In diesem Fall kann nicht ausgeschlossen werden, dass unbekannte Faktoren das Untersuchungsergebnis beeinflussen (z.B. dass Fragebögen nicht in der vorgegebenen Reihenfolge beantwortet werden; dass andere als die intendierten Personen die Fragen beantworten usw.).

Frageformulierung

Die folgende Zusammenstellung von Gesichtspunkten, die bei der Frageformulierung zu beachten sind, gelten für die Formulierung von Fragen im allgemeinen, sowohl bei der mündlichen als auch bei der schriftlichen Befragung:

- Die Formulierung der Fragen sollte *an den sprachlichen Gewohnheiten der Zielgruppe ausgerichtet* sein.
- Die Fragen sollten *eindeutig interpretierbar* sein. Eindeutige Interpretierbarkeit ist insbesondere dann nicht gegeben, wenn eine Frage verschiedene Teilfragen enthält. In einem solchen Fall ist nicht eindeutig feststellbar, auf welchen Teil der Frage die Antwort sich bezieht.
- Die Frage soll gut *verständlich* sein.
- Eine Frage sollte möglichst *keine doppelten Verneinungen* enthalten.

- Ein Fragebogen sollte möglichst *keine Items* enthalten, die von praktisch allen Befragten *in gleicher Weise beantwortet* werden. Solche Items sind nicht geeignet, um zwischen den Untersuchungsteilnehmern/innen zu differenzieren.
- Auch die *Anordnung der Fragen* ist von Bedeutung. So sollte ein Fragebogen z.B. nicht mit allzu schwierigen Fragen beginnen; auch Fragen, die sehr persönlich sind, sollten nicht gleich zu Beginn des Fragebogens gestellt werden.
- Gerade bei schriftlichen Fragebögen ist außerdem eine *Instruktion* erforderlich, wo den Befragten erläutert wird, welche Arten von Fragen gestellt werden und welche Möglichkeiten der Beantwortung zur Verfügung stehen.

Varianten der schriftlichen Befragung

Im Folgenden sind einige wichtige Formen der schriftlichen Befragung aufgeführt. Die verschiedenen Arten der Befragung schließen einander nicht aus, können also auch in Kombination realisiert werden.

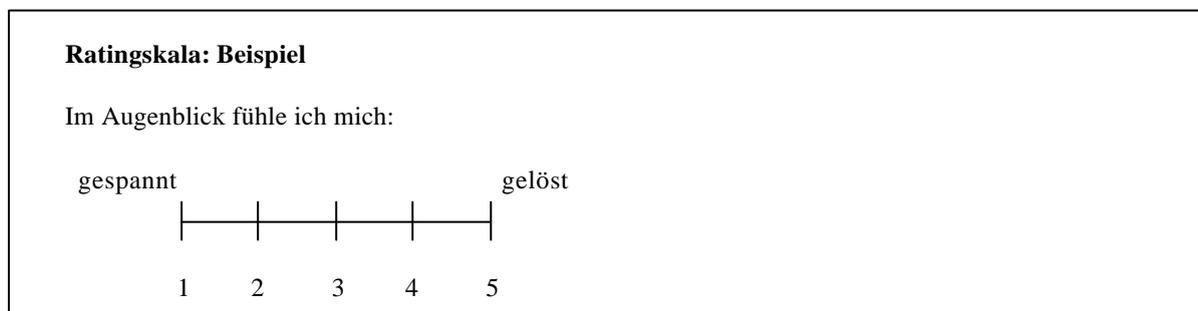
- *Befragung vor Ort*: Unter einer Befragung vor Ort versteht man eine Befragung, die in Universitätsräumen durchgeführt wird. Befragungen vor Ort ermöglichen ein hohes Ausmaß von Kontrolle über die Untersuchungssituation, sind aber aufgrund von Problemen bei der Terminabsprache oft schwer zu realisieren.
- *Postalische Befragung*: Häufig werden Befragungen in der Form durchgeführt, dass die Fragebögen den Teilnehmern/innen – nach vorheriger Absprache – nach Hause zugeschickt werden. Bei postalischen Befragungen haben die Forscher/innen allerdings keine Kontrolle über die Untersuchungssituation (s.o.). Außerdem ergibt sich bei postalischen Befragungen häufig das Problem, dass die *Rücklaufquote* niedrig ausfällt, dass also nur ein Teil der Personen in der Stichprobe die Bögen auch tatsächlich ausgefüllt zurückschickt. Eine geringe Rücklaufquote ist vor allem dann problematisch, wenn die Nicht-Rücksendung systematisch mit bestimmten Merkmalen der Stichprobe zusammenhängt. Wenn das der Fall ist, muss die Stichprobe der zurückgesandten Fragebögen als verzerrt gelten; die Ergebnisse können dann nicht mehr ohne Weiteres verallgemeinert werden.
- *Befragung in einem Panel*: Bei Panel-Befragungen handelt es sich um längsschnittlich angelegte Untersuchungen, bei denen die Teilnehmer/innen in bestimmten Zeitabständen wiederholt befragt werden. Das zentrale Problem solcher Untersuchungen besteht in dem ‚Schwund‘ von Teilnehmern/innen: Je mehr Befragungszeitpunkte vorgesehen sind, desto größer wird der Anteil der Personen in der Stichprobe, die für eine Befragung nicht mehr erreichbar sind. Längsschnittliche Untersuchungen sind vor allem in der Entwicklungs- und in der Gesundheitspsychologie von Bedeutung.
- *Computervermittelte Befragung*: Mit der Verbreitung von Computer und Internet haben computervermittelte Formen der Befragung zunehmend an Bedeutung gewonnen. So werden beispielsweise Fragebögen im WWW online zur Verfügung gestellt oder den Probanden/innen per e-mail zugesandt. Der Vorteil dieser Befragungsverfahren besteht darin, dass eine große Personenanzahl mit wenig Aufwand erreicht werden kann. Diesem Vorteil steht allerdings das Problem gegenüber, dass die so erreichte Stichprobe eine sehr selektive ist, so dass die Ergebnisse nur eingeschränkt verallgemeinert werden können.
- *Delphi-Methode*: Bei der Delphi-Methode handelt es sich um ein mehrstufiges Verfahren der Befragung von Experten/innen .

Prüfungsliteratur: Bortz & Döring 2002, Kap. 4.4.

II.2.2. Beurteilung mittels Ratingskalen

Überblick

Es existieren verschiedene Formen von Beurteilungsverfahren (z.B. Paarvergleiche, Ratingskalen usw.). Unter diesen Verfahren wird hier die Beurteilung mittels Ratingskalen genauer dargestellt. Es handelt sich dabei um eine besondere Variante von Items in Fragebögen. Ratingskalen dienen dazu, einen Gegenstand (sich selbst, andere Personen, Objekte) hinsichtlich eines bestimmten Merkmals zu beurteilen (z.B. Schüchternheit, Gefallen usw.). Ratingskalen geben gleich große markierte Abschnitte eines Merkmalskontinuums vor. Die Urteiler/innen werden gebeten, diejenige Stufe der Ratingskala anzukreuzen, die ihrem persönlichen Empfinden hinsichtlich der Merkmalsausprägung bei dem zu beurteilenden Gegenstand entspricht (also z.B. anzugeben, für wie schüchtern sie sich selbst halten; wie gut ihnen ein Witz gefällt usw.).



Ratingskalen sind als Instrument in der Psychologie sehr beliebt

- weil sie, zumindest auf den ersten Blick, einfach zu erstellen sind;
- weil für Ratingskalen –aufgrund der gleich großen Abstände zwischen den verschiedenen Ausprägungen des Merkmalskontinuums – Intervallskalenniveau angenommen wird.

Varianten von Ratingskalen

- Die Items zu Ratingskalen können als *Frage* oder als *Aussage* formuliert sein (,Wie schüchtern sind Sie?' vs. ,Ich bin sehr schüchtern'). Welche Formulierungsvariante günstiger ist, hängt von der Fragestellung der Untersuchung ab.
- Ratingskalen können hinsichtlich der *Endpunkte* der Skalen *unipolar* (z.B. von: ,kein Lärm' bis ,viel Lärm') oder *bipolar* (von ,leise' bis ,laut') sein. Bipolare Skalen haben den Vorteil, dass die Begriffe an den Endpunkten der Skala einander wechselseitig definieren. Unipolare Skalen sind vor allem dann sinnvoll, wenn der Gegensatz eines Begriffs unklar ist (Was ist z.B. genau der Gegensatz von ,schüchtern?'); sie eignen sich auch zur Beurteilung auf Dimensionen mit einem natürlichen Nullpunkt (z.B. Höhe des Lärmpegels).
- Die *Anzahl der Stufen* einer Ratingskala kann *gerade* oder *ungerade* sein. Bei einer geraden Anzahl von Stufen (für eine bipolare Skala) wird ein Urteil in Richtung des einen oder anderen Pols der Skala quasi erzwungen. Bei einer ungeraden Anzahl von Stufen steht dagegen auch eine neutrale Urteilmöglichkeit zur Verfügung. Ein solches neutrales Urteil ist allerdings problematisch, wenn diese Urteilmöglichkeit zu häufig gewählt wird. Problematisch sind neutrale Urteile auch insofern, als sie nicht eindeutig interpretierbar sind: Eine neutrale Beurteilung kann ebenso Ausdruck einer gleichgültigen wie auch einer

zwiespältigen Einstellung im Hinblick auf den fraglichen Gegenstand sein (*Ambivalenz-Indifferenz-Problem*).

- Ratingskalen können nur wenige (drei oder vier) oder viele (20 oder mehr) *Stufen* aufweisen. Bei allzu wenigen Stufen sind relevante Unterschiede zwischen verschiedenen Beurteilungen nicht abbildbar; andererseits sollte die Anzahl der Stufen auch die Differenziertheit des Urteils nicht übersteigen. In der Praxis haben sich *fünfstufige Ratingskalen* am besten bewährt.
- Die Abstufungen der Skalen können durch Zahlen (*numerische Marken*) oder durch verbale Bezeichnungen (*verbale Marken*) gekennzeichnet sein. Numerische Marken haben den Vorteil, dass sie eindeutig sind und dass die Abstände zwischen den Stufen der Ratingskala gleich sind. Verbale Marken sind dagegen leichter verständlich (vor allem für Personen, die mit Forschung wenig vertraut sind); auf der anderen Seite ergibt sich hier das Problem, wie man den gleichen Abstand zwischen den Stufen der Skala sicherstellt. Auch andere Arten von Marken sind möglich, z.B. *symbolische* (z.B. Ankreuzen trauriger bis lächelnder Gesichter).

Probleme bei Ratingskalen

- Bei der Beurteilung von Objekten auf Ratingskalen kann es zunächst zu *Urteilsfehlern* kommen (s. dazu ausführlich unten).
- Weiterhin können sich *messtheoretische* Probleme ergeben.
 - Ein erstes messtheoretisches Problem betrifft die Frage, ob es gerechtfertigt ist, Urteile auf Ratingskalen als intervallskaliert zu betrachten – ob man also davon ausgehen kann, dass die Abstände zwischen den Marken gleich groß sind. Zu dieser Frage existieren unterschiedliche Auffassungen: ‚Puristen‘ gehen davon aus, dass Intervallskaliertheit nicht einfach unterstellt, sondern jeweils konkret gesichert werden sollte. ‚Pragmatiker‘ nehmen dagegen an, dass Analyseverfahren wie z.B. die Varianzanalyse auch dann noch aussagekräftige Ergebnisse liefern, wenn leichte Verstöße gegen die Intervallskaliertheit vorliegen; ‚Pragmatiker‘ gehen daher davon aus, dass Daten, die mittels Ratingskalen erhoben wurden, wie intervallskalierte Daten betrachtet und ausgewertet werden können.
 - Ein zweites messtheoretisches Problem tritt auf, wenn Beurteilungen mittels derselben Ratingskala je verschiedene Einheiten oder einen je anderen Ursprung voraussetzen. Dies ist beispielsweise dann der Fall, wenn eine Person im Verlauf ihrer Beurteilungen den Bezugspunkt wechselt (intraindividuelle Inkonsistenz) oder wenn verschiedene Personen den Ursprung der Skala je unterschiedlich definieren. Das Problem der intraindividuellen Inkonsistenz lässt sich zumindest mildern, wenn die Untersuchungsteilnehmer/innen die Möglichkeit haben, vor der eigentlichen Beurteilung alle Objekte kennenzulernen. Außerdem sollten die zu beurteilenden Objekte sich in etwa symmetrisch über die Hälften der Skala verteilen.

Arten von Urteilsfehlern bei Ratingskalen

- *Halo-Effekt*: Dieser Fehler bezieht sich auf die Tendenz, die Beurteilung einer Person (oder eines Gegenstandes) hinsichtlich verschiedener Merkmale von der Beurteilung auf einem dieser Merkmale abhängig zu machen. Dieses eine Merkmal bildet sozusagen einen Hof (einen ‚halo‘, ähnlich dem Hof des Mondes) für die Beurteilung auf den anderen Merkmalen. Der Fehler tritt vor allem bei der Personenbeurteilung auf: Wer beispielsweise für gutmütig gehalten wird, wird auch als zuverlässig, freundlich usw. beurteilt. Zu einem Halo-Effekt kommt es besonders dann, wenn das ‚Schlüsselmerkmal‘ ungewöhnlich oder unklar definiert ist. Verringern lässt sich der Effekt entsprechend, indem klare Infor-

mationen über die zu beurteilenden Merkmale und auch über den Fehler selbst gegeben werden.

- *Milde-Härte-Fehler*: Dieser Fehler liegt vor, wenn eine Person systematisch entweder zu negativ oder zu positiv beurteilt wird. Vom Ergebnis her ähnelt der Milde-Härte-Fehler dem Halo-Effekt; allerdings hängt die Beurteilung hier nicht von einem ‚Schlüsselmerkmal‘ ab. Das Problem lässt sich verringern, indem die Untersuchungsteilnehmer/innen vor der Beurteilung über den Fehler informiert werden.
- *Effekt der zentralen Tendenz*: Darunter wird die Tendenz verstanden, Extremurteile zu vermeiden. Dieser Fehler tritt besonders dann auf, wenn die Urteilsobjekte wenig bekannt sind oder wenn die Skalen an den Extrempunkten nicht verankert sind (wenn die Extreme also unklar bleiben). Der Fehler lässt sich vermindern, indem die Untersuchungsteilnehmer/innen hinreichend über die zu beurteilenden Objekte informiert werden. Unter Umständen ist auch eine Neukonstruktion der Skala erforderlich.
- *Primacy-Recency-Effekt*: Der Primacy-Recency-Effekt stellt eine Urteilsverzerrung dar, die mit der Reihenfolge der zu beurteilenden Objekte zusammenhängt. Die Verzerrung tritt auf, wenn als erstes Objekte mit extremer Merkmalsausprägung beurteilt werden. Die Beurteilung der folgenden Objekte kann dann (im Sinne eines Kontrastes) von der Beurteilung des ersten Objekts abhängen.
- *Rater-Ratee-Interaktion*: Hierbei handelt es sich um eine Urteilsverzerrung, die in Abhängigkeit von der Person bzw. der Einstellung der Urteiler/innen abhängt. Die Verzerrung tritt vor allem bei Personen mit extremer Merkmalsausprägung auf. Man unterscheidet weiter zwischen einem Ähnlichkeits- und einem Kontrastfehler:
 - *Ähnlichkeitsfehler*: Die Merkmalsausprägungen anderer Personen werden der eigenen Merkmalsausprägung angepasst. Ein Mitglied der SPD mag beispielsweise der Meinung sein, dass andere SPD-Mitglieder eine Aufhebung des Kündigungsschutzes in mittleren Unternehmen ebenso ablehnen wie er oder sie selbst – was aber in der Tat gar nicht der Fall ist.
 - *Kontrastfehler*: Die Merkmalsausprägungen anderer Personen werden in Richtung auf das andere Extrem verschätzt. Ein Mitglied der CDU mag beispielsweise einem Mitglied der SPD eine viel gewerkschaftsfreundlichere Haltung unterstellen, als diese/r sie tatsächlich vertritt.

Prüfungsliteratur: Bortz & Döring 2002, Kap. 4.2.4.

II.2.3. Tests

Test: Definition

Ein Test ist definiert als ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung. Ziel der Testung ist also eine Individualdiagnose.

Arten von Tests

Es wird zwischen zwei Arten von Tests unterschieden: Leistungs- und Persönlichkeitstests.

- Bei *Leistungstests* existiert ein objektiver Maßstab zur Beurteilung der Güte der Antworten; die Antworten können also ‚richtig‘ oder ‚falsch‘ sein. Intelligenztests, Eignungstests

usw. zählen beispielsweise zu den Leistungstests. Um zwischen mehr oder weniger guten Leistungen differenzieren zu können, müssen in dem Test verschiedene Schwierigkeitsgrade realisiert sein. Dies kann auf zwei verschiedene Arten erfolgen:

- Bei *Speed-Tests* ist die Bearbeitungszeit zu knapp angesetzt, so dass in der Regel nicht alle Aufgaben bearbeitet werden können.
 - Bei *Power-Tests* wird das Niveau der Aufgaben sukzessive gesteigert.
- Bei *Persönlichkeitstests* (z.B. MMPI, Freiburger Persönlichkeitsinventar) spielen objektive Beurteilungsmaßstäbe dagegen keine Rolle. Es wird zwischen objektiven und subjektiven Persönlichkeits- tests unterschieden. Bei subjektiven Persönlichkeits- tests ist der Zweck des Tests für die getesteten Personen leicht durchschaubar; bei objektiven Persönlichkeits- tests wird versucht, den Zweck zu verschleiern, um so Beurteilungsfehler zu minimieren (s. auch unten: Testverfälschung – Gegenmaßnahmen).

Unterschiede zwischen Testverfahren und Fragebögen mit Ratingskalen

Auch bei Tests können Ratingskalen zum Einsatz kommen. Von Fragebögen mit Ratingskalen unterscheiden sich Tests jedoch in den folgenden Aspekten:

- Fragebögen werden meist zu Forschungszwecken eingesetzt (*Forschungsinstrument*), Tests meist zu Zwecken der Individualdiagnose (*Diagnoseinstrument*).
- In der Forschung unter Verwendung von Fragebögen interessiert meistens die Hypothesenprüfung über *Aggregatwerte* (statistische Kennwerte für Personengruppen), während bei Tests gerade die *individuellen Werte* einzelner Personen von Interesse sind.
- Fragebögen können sich auf beliebige Inhalte beziehen (häufig sind insbesondere *Objektbeurteilungen*); Tests beziehen sich speziell auf den *Persönlichkeits- und Leistungsbe- reich*.

Testskala: Definition und Voraussetzungen

Diejenigen Items, die im Rahmen eines Tests der Erhebung eines bestimmten Merkmals dienen, werden als *Itemsatz* bzw. als *Testskala* bezeichnet. Eine Testskala muss den folgenden Bedingungen genügen:

- Die Items einer Skala müssen *homogen* sein; d.h., die Items dienen sämtlich der Erhebung eines einzelnen Merkmals. Wenn ein Merkmal mehrere Dimensionen beinhaltet (wie z.B. Intelligenz), dann ist es erforderlich, mehrere Testskalen zu erstellen.
- Die Items der Testskala müssen es erlauben, möglichst *viele verschiedene Ausprägungs- grade* des interessierenden Merkmals zu erheben.
- Die Testskala muss eine *eindeutige Unterscheidung* zwischen Personen mit hoher und mit geringer Merkmalsausprägung ermöglichen.
- Die Items müssen *objektiv* sein; die Testergebnisse müssen also unabhängig von der Person sein, die den Test durchführt, auswertet und interpretiert.
- Die Messung der Merkmalsausprägung muss verlässlich bzw. *reliabel* sein.
- Die Items müssen *valide* sein, d.h. sie müssen das erfassen, was erfasst werden soll.

Die ersten drei Bedingungen ergeben sich aus der generellen Anforderung der Individualdiag- nose. Um Items zu erstellen, die diese Bedingungen erfüllen, muss die Testskala bestimmten messtheoretischen Voraussetzungen genügen (z.B. Likert-Skala; Guttman-Skala); diese sind im Rahmen des Grundstudiums jedoch nicht weiter von Bedeutung.

Die gebräuchlichsten Skalen und Tests basieren meist auf der *Klassischen Testtheorie*. In der Klassischen Testtheorie wird angenommen, dass jeder Wert einer Person auf einem konkreten Item aus zwei Komponenten zusammengesetzt ist: aus dem wahren Testwert der Person und aus einem Fehleranteil. Dieser Fehleranteil ist zugleich einer der Gründe dafür, weshalb zur

Erfassung eines Merkmals mehrere Items formuliert werden: Mehrere Items ermöglichen eine bessere Annäherung an den wahren Testwert einer Person. Außerdem erlauben erst mehrere Items eine differenzierte Erfassung unterschiedlicher Merkmalsausprägungen.

Arten von Items

In einem Test können drei verschiedene Arten von Items zur Anwendung kommen:

- *Items mit offener Beantwortung*: Bei Tests dieser Art wird ein Reiz vorgegeben; die untersuchte Person wird nach ihren Assoziationen zu diesem Reiz gefragt (z.B. Tintenklecks im Rorschach-Test) oder sie wird z.B. gebeten, die Situation genauer zu schildern, die in dem Reiz dargestellt wird (z.B. TAT zur Erfassung verschiedener Motive). Items mit offener Beantwortung schneiden in Bezug auf die Testgütekriterien der Objektivität und Reliabilität in der Regel schlechter ab als Items mit Antwortvorgaben.
- *Items mit halboffener Beantwortung*: Bei Items mit halboffener Beantwortung werden die untersuchten Personen beispielsweise gebeten, einen Teil-Satz in ihren eigenen Worten zu vervollständigen. Auch bei diesen Items kann es zu Problemen bezüglich der Objektivität und der Reliabilität kommen.
- *Items mit Antwortvorgaben*: Hier werden verschiedene Antwortmöglichkeiten vorgegeben (z.B. im multiple choice-Verfahren), zwischen denen die untersuchte Person auswählen soll. Für solche Items sind Objektivität und Reliabilität deutlich leichter sicherzustellen. Items mit Antwortvorgaben sind unter den verschiedenen Arten von Testitems am weitesten verbreitet.

Ursachen von Testverfälschung

Testergebnisse sind nicht notwendiger Weise valide, sondern können auch verfälscht sein. Verfälschungen haben bei Leistungs- und bei Persönlichkeitstests je verschiedene Ursachen.

- Bei *Leistungstests* kommt es vor allem dann zu Verfälschungen, wenn es den Probanden/innen gelingt, die richtige Antwort zu erraten. Bei Multiple Choice-Aufgaben lässt sich solchen Verfälschungen entgegenwirken, indem neben der korrekten Antwort möglichst gute *Distraktoren* konstruiert werden. Distraktoren sind Antwortmöglichkeiten, die auch durchaus plausibel erscheinen; sie werden nur von solchen Personen als falsch erkannt, die über das relevante Wissen verfügen. Eine weitere Möglichkeit, Verfälschungen durch Erraten richtiger Antworten entgegenzuwirken, besteht darin, bei der Auswertung des Tests eine *Ratekorrektur* vorzunehmen. Die Korrektur besteht darin, dass von der Punktezahl, die eine Person in dem Test erreicht hat, so viele Punkte abgezogen werden, wie die Person allein durch Erraten hätte erzielen können.
- Bei *Persönlichkeitstests* werden vor allem Verfälschungen wirksam, die sich aus dem Bemühen ergeben, die eigene Person möglichst günstig darzustellen. Diese Fehlerquellen werden in den folgenden Abschnitten ausführlich dargestellt.

Testverfälschung in Persönlichkeitstests: Ursachen

Zu einer Testverfälschung in einem Persönlichkeitstest kann es im Wesentlichen aus drei Gründen kommen: Selbstdarstellung, soziale Erwünschtheit sowie Antworttendenzen.

- *Selbstdarstellung*: Die getestete Person versucht, einen ganz bestimmten Eindruck von der eigenen Person zu vermitteln, sich in einer bestimmten Weise darzustellen.

- *Soziale Erwünschtheit:* Bei der sozialen Erwünschtheit handelt es sich um eine spezielle Form der Selbstdarstellung. Die untersuchte Person ist bemüht, sich selbst so darzustellen, wie es den sozialen Erwartungen und Normen der Gruppe entspricht, von der der Test durchgeführt wird. Da die Erwartungen an andere in verschiedenen sozialen Gruppen durchaus unterschiedlich sind (und ggf. zusätzlich in Abhängigkeit vom Interaktionszweck variiert), kann eine Selbstdarstellung im Sinne der sozialen Erwünschtheit in verschiedenen Kontexten auch ganz unterschiedlich ausfallen. Wenn jemand sich bei einem alteingesessenen Bankhaus bewirbt, wird er oder sie vermutlich versuchen, sich anders darzustellen als im Fall einer Bewerbung bei einer jungen Computerfirma.
- *Antworttendenzen:* Zu Testverfälschungen kann es schließlich auch aufgrund von Antworttendenzen kommen. Manche Menschen tendieren z.B. dazu, immer möglichst neutral zu antworten; andere wiederum stimmen Testitems in der Regel zu (Ja-Sage-Tendenz), andere antworten meist ablehnend (Nein-Sage-Tendenz). Solche Antworttendenzen sind den betreffenden Personen meistens gar nicht bewusst.

Testverfälschung in Persönlichkeitstests: Gegenmaßnahmen

Die Mehrzahl von Maßnahmen zur Kontrolle der Testverfälschung wirken Tendenzen zur Selbstdarstellung und der sozialen Erwünschtheit entgegen.

- *Ausbalancierte Antwortvorgaben:* Bei dieser Kontrolltechnik formuliert man erstens unterschiedlich ‚gepolte‘ Items, d.h. Items, bei denen eine Zustimmung mit einer je unterschiedlichen inhaltlichen Antwortrichtung einhergeht. Zweitens versucht man, die Formulierungen so zu wählen, dass beide Antwortrichtungen, beide ‚Pole‘ des Items, gängigen gesellschaftlichen Werten und Normen entsprechen. So könnte man z.B. das Interesse am Studienfach mit den Items erheben: „Mit meinem Studienfach beschäftige ich mich täglich mehrere Stunden“ und „Ich finde es wichtig, neben meinem Studium auch andere Interessen nicht zu vernachlässigen“. Diese Technik eignet sich zur Kontrolle aller drei Formen der Selbstdarstellung gleichermaßen (während die folgenden Maßnahmen ausschließlich der Kontrolle von Tendenzen zur Selbstdarstellung und zur sozialen Erwünschtheit dienen).
- *Kontrollskalen:* Kontrollskalen sollen im Sinne eines Tests die je individuelle Tendenz zur sozialen Erwünschtheit erfassen (z.B. die Social Desirability Scale von Crowne & Marlowe 1964). In den Items sind Verhaltensweisen beschrieben, die gesellschaftlich eindeutig negativ bewertet werden – die zugleich aber so häufig sind, dass es ausgesprochen unwahrscheinlich ist, dass jemand die entsprechenden Verhaltensweisen selbst tatsächlich noch nie gezeigt hat. So haben die meisten Menschen schon einmal zu einer Notlüge gegriffen, obwohl Lügen gesellschaftlich negativ sanktioniert sind. Je häufiger jemand angibt, sich ‚nie‘ in der beschriebenen Weise zu verhalten, desto höher ist seine oder ihre Tendenz zur sozialen Erwünschtheit.
- *Objektive Tests:* Bei objektiven Tests wird versucht, den Zweck des Tests möglichst zu verschleiern.
- *Aufforderung zu korrektem Testverhalten:* Die Probanden/innen werden in der Instruktion gebeten, die Items ehrlich zu beantworten. Manchmal wird hinzugefügt, dass die Forscher/innen in der Lage sind, Lügen als solche zu erkennen. Das ist allerdings erstens selbst eine Lüge; zweitens wird dadurch psychischer Druck auf die Probanden/innen ausgeübt, so dass ein solcher Zusatz ethisch bedenklich ist (die bloße Aufforderung zu korrektem Testverhalten ist dagegen in keiner Weise ethisch problematisch).

- *Random-Response-Technik*: Bei der Random Response-Technik werden die Probanden/innen aufgefordert, vor dem Beantworten jeder Frage zunächst z.B. zu würfeln. Jedesmal, wenn sie eine bestimmte Zahl würfeln (etwa die Vier), sollen sie die entsprechende Frage nicht wahrheitsgemäß beantworten. Da die Person, die den Test durchführt, das Ergebnis des Würfeln nicht kennt, ist bei der Testauswertung für die Forscher/innen nicht rekonstruierbar, welche Fragen wahrheitsgemäß beantwortet wurden und welche nicht. Forscher/innen versprechen sich von der Verwendung dieser Technik, dass die Probanden/innen sich unter den genannten Bedingungen freier fühlen, auch weniger sozial erwünschte Antworten zu geben. Die unzutreffenden Antworten werden in der Hoffnung in Kauf genommen, dass die übrigen Fragen wahrheitsgemäß beantwortet werden – so dass die Antworten insgesamt bei der Random Response-Technik eine höhere Validität aufweisen, als dies der Fall wäre, wenn die Probanden/innen auf alle Fragen in sozial erwünschter Weise antworten würden.

Prüfungsliteratur: Bortz & Döring 2002, Kap. 4.3.5. u. 4.3.7.

II.2.4. Beobachten, Zählen, Messen

Merkmale und ihre Erhebung

In den Sozialwissenschaften wird zwischen zwei Arten von Merkmalen unterschieden: quantitativen und qualitativen. *Quantitative Merkmale* (z.B. Körpergröße, Intelligenzquotient) lassen sich unmittelbar in Zahlen ausdrücken; sie sind mindestens ordinalskaliert und im engeren Sinne des Begriffs *messbar*. *Qualitative Merkmale* (z.B. Geschlecht, Art einer aggressiven Handlung) sind dagegen nominalskaliert; feststell- im Sinne von *beobachtbar* sind nur die Ausprägungen des Merkmals (z.B. für Geschlecht: männlich / weiblich). Die Ausprägungen quantitativer Merkmale werden also gemessen, die Ausprägungen qualitativer Merkmale beobachtet. Sowohl an eine Messung als auch an eine Beobachtung kann sich (im Übergang zur Auswertung) eine Häufigkeitsauszählung anschließen.

Quantitative Merkmale: Unterscheidungen

Bei quantitativen Merkmalen wird weiter zwischen stetigen und diskreten Merkmalen unterschieden.

- *Stetige Merkmale* (wie z.B. Körpergröße) können viele verschiedene Ausprägungen annehmen. Stetige Merkmale sind *messbar*.
- *Diskrete Merkmale* (wie z.B. Anzahl der Geschwister) können nur als Häufigkeiten in ganzen Zahlen erfasst werden; ihre Ausprägungen sind *zählbar*.

Qualitative Merkmale: Unterscheidungen

Bei qualitativen Merkmalen sind nach zwei Gesichtspunkten weitere Unterscheidungen möglich:

- *dichotom vs. mehrkategorial*: Qualitative Merkmale mit zwei Ausprägungen (z.B. biologisches Geschlecht) werden dichotom, solche mit mehr als zwei Ausprägungen (z.B. soziales Geschlecht) mehrkategorial genannt.
- *natürlich vs. künstlich*: Natürliche qualitative Merkmale sind sozusagen im Gegenstandsbereich vorgegeben (z.B. Augenfarbe); bei künstlichen Merkmalen handelt es sich dagegen um Konstrukte, die von den Forschern/innen selbst erzeugt/postuliert werden (z.B. soziale Schicht).

Formen der Beobachtung

- *Standardisierte vs. nonstandardisierte Beobachtung:* Eine standardisierte Beobachtung basiert auf einem Beobachtungsschema, in dem Klassen möglicher Ereignisse vorgegeben sind. Die Anforderungen an Beobachtungsschemata entsprechen denen an inhaltsanalytische Kategoriensysteme (s. dort: III.3.3.). Die Beobachter/innen kreuzen an, welche Ereignisklassen im Beobachtungszeitraum vorkommen; um die Objektivität der Beobachtungen zu erfassen, wird der Zusammenhang zwischen den Ereignisklassifikationen mehrerer Beobachter/innen ermittelt. Bei der nonstandardisierten Beobachtung ist dagegen lediglich festgelegt, auf welche Arten von Ereignissen sich die Beobachtung richten soll. Beobachtungsdimensionen und Ereignisklassen werden erst im Verlauf des Beobachtungsprozesses ermittelt. In der quantitativen Forschung dominiert die standardisierte Form der Beobachtung.
- *Offene vs. verdeckte Beobachtung:* Im Fall einer offenen Beobachtung wissen die Untersuchungsteilnehmer/innen, dass sie beobachtet werden; eine verdeckte Beobachtung erfolgt dagegen ohne das Wissen der beobachteten Personen. Eine offene Beobachtung, die nur für einen kürzeren Zeitraum durchgeführt wird, kann zur Folge haben, dass die Teilnehmer/innen sich anders verhalten, als sie das normalerweise tun (*Reaktivität* des Verfahrens); in diesem Fall wäre die Validität der Untersuchung betroffen. Eine verdeckte Beobachtung ist dagegen unter ethischen Gesichtspunkten problematisch. Außerdem verfügt jeder Mensch über ein Recht auf informationelle Selbstbestimmung; spätestens zum Zeitpunkt der Veröffentlichung ist daher in jedem Fall die Zustimmung der beobachteten Personen einzuholen.
- *Nonreaktive Beobachtung:* Vor dem Hintergrund der Problematik, dass Menschen sich z.T. anders verhalten, wenn sie sich beobachtet wissen, wurden nonreaktive Formen der Beobachtung entwickelt. Hier kommen die Beobachter/innen und die beobachteten Personen gar nicht miteinander in Kontakt; die Privatsphäre der beobachteten Personen wird nicht verletzt, und die beobachteten Personen reagieren nicht (z.B. durch Veränderung ihrer Handlungsweise) auf die Beobachtung. Eine nonreaktive Form der Beobachtung liegt z.B. vor, wenn Abnutzungserscheinungen des Teppichbodens in einem Museum verwendet werden, um daraus auf die häufigsten ‚Pfade‘ der Besucher/innen durch eine Ausstellung zu schließen. Nonreaktive Formen der Beobachtung sind allerdings nicht immer möglich.
- *Apparative Beobachtung:* Bei der apparativen Beobachtung erfolgt die Beobachtung nicht unvermittelt, sondern durch technische Geräte vermittelt (z.B. durch eine Videokamera). Das hat den Vorteil, dass die Beobachter/innen (deren Beobachtungen unvollständig, verzerrt usw. sein können) als Fehlerquelle ausgeschlossen sind. Der Nachteil dieser Form der Beobachtung besteht darin, dass die Untersuchungsteilnehmer/innen sich durch das Wissen um eine Kamera stark gehemmt fühlen können; je länger die Beobachtung dauert, desto schwächer wird jedoch in der Regel diese ‚Kamerascheu‘. Es lassen sich weiter verschiedene Formen der apparativen Beobachtung unterscheiden: Ein erstes Differenzierungskriterium bezieht sich darauf, ob die Apparatur von den Beobachtern/innen bedient wird oder von den Untersuchungsteilnehmern/innen selbst. Auch lassen sich Formen der apparativen Beobachtung, die das Offensichtliche aufzeichnen (das auch mit dem menschlichen Auge sichtbar wäre), gegenüber solchen Formen differenzieren, bei denen das nicht unmittelbar Sichtbare Gegenstand der Beobachtung ist (z.B. Blickbewegungskameras).
- *Automatische Beobachtung:* Bei der automatischen Beobachtung wird das Verhalten der Untersuchungsteilnehmer/innen eben automatisch aufgezeichnet, häufig in verdeckter

Form (und verbunden mit den entsprechenden Problemen). Die automatische Beobachtung findet sich insbesondere bei der Beobachtung im Internet, z.B. bei der Aufzeichnung von Chat-Protokollen oder von Bewegungen durch das WWW in Form von sog. Logfiles.

- *Selbst- vs. Fremdbeobachtung:* Weiterhin lässt sich, je nach Beobachtungsgegenstand, zwischen Selbst- und Fremdbeobachtung unterscheiden. Die Selbstbeobachtung (also die Beobachtung der eigenen Person) war vor allem in den Anfangszeiten der Psychologie in Form der Introspektion stark verbreitet; heute ist die Fremdbeobachtung (also die Beobachtung von Untersuchungsteilnehmern/innen durch Forscher/innen) die üblichere Form. In Medienwissenschaft und -psychologie kommen jedoch weiterhin auch Varianten der Selbstbeobachtung zur Anwendung, z.B. in Form von Medientagebüchern.
- *Beobachtung durch eine Person oder durch mehrere Personen:* Die Beobachtung durch mehrere Personen ist einerseits für die Untersuchungsteilnehmer/innen störender, hat jedoch gegenüber der Beobachtung durch nur eine Person den Vorteil, dass eventuelle Beobachtungsfehler besser ausgleichbar sind.

Systematische Beobachtung

Beobachtung in der Wissenschaft unterscheidet sich von der Beobachtung im Alltag insbesondere darin, dass wissenschaftliche Beobachtung immer systematische Beobachtung ist. Systematisch zu beobachten heißt, dass man von vornherein festlegt, was man beobachten will, zu welchen Zeitpunkten, in welchen Zeiträumen und Situationen, wobei die Auswahl so zu treffen ist, dass für die Fragestellung repräsentative Ergebnisse zu erwarten sind.

Stichprobenziehung: Auswahl der Beobachtungseinheiten

Die Repräsentativität der Ergebnisse einer wissenschaftlichen Untersuchung hängt wesentlich von der Stichprobenziehung ab. Bei einer Beobachtungsstudie betrifft die Stichprobenziehung nicht nur die Auswahl der untersuchten Personen, sondern weiterhin die Auswahl der Beobachtungseinheiten. Es lassen sich zwei Formen der Auswahl von Beobachtungseinheiten unterscheiden:

- *Ereignisstichprobe:* Bei der Ereignisstichprobe wird ermittelt, wie häufig ein bestimmtes Ereignis im gesamten Beobachtungszeitraum auftritt (z.B. wie häufig eine Schülerin sich im Unterricht meldet). Diese Form der Stichprobenziehung erlaubt es, auch solche Ereignisse zu untersuchen, die nur selten auftreten. Insgesamt eignet sich die Ereignisstichprobe am besten zur Dokumentation bestimmter Verhaltensweisen oder Ereignisse.
- *Zeitstichprobe:* Bei der Zeitstichprobe wird das gesamte Geschehen (bzw. die daraus interessierenden Aspekte) in vorher festgelegten Zeitabständen festgehalten (z.B. Notierung in 5-Minuten-Abständen, was eine Schülerin gerade macht). Die Zeitstichprobe eignet sich am besten zur Beobachtung des gesamten Geschehens. Sie ist allerdings nicht gut geeignet, wenn der Verlauf der interessierenden Ereignisse zeitlich unregelmäßig erfolgt. Sie stellt außerdem hohe Anforderungen an die Konzentrationsfähigkeit der Beobachter/innen.

Durchführung einer Beobachtungsstudie

Eine Beobachtungsstudie erfordert in der Regel die Aufstellung eines Beobachtungsschemas oder -plans, in dem die interessierenden Ausprägungen der zu beobachtenden Ereignisse so festgehalten sind, dass die Beobachter/innen diese nur noch ankreuzen müssen. Vor Beginn der Beobachtung findet meist ein Beobachtertraining statt. Hier werden die Forscher/innen mit dem Beobachtungsplan und den interessierenden Ereignissen vertraut gemacht. Im Anschluss an eine Probephase erfolgt zur Sicherung der Objektivität und Reliabilität der Beobachtung eine Berechnung der Übereinstimmung zwischen den Beobachtern/innen.

Beispiel für ein Beobachtungsschema

Vom Lehrer ausgehendes Verbalverhalten	1 Gibt Informationen oder Meinungen 2 Gibt Anweisung 3 Stellt enge Frage 4 Stellt weite Frage
Lehrer-Erwidernng	5 Akzeptanz von (a) Gedanken, (b) Verhalten, (c) Gefühlen 6 Ablehnung von (a) Gedanken, (b) Verhalten, (c) Gefühlen
Schüler-Erwidernng	7a Voraussagbare Antwort an Lehrer 7b Nicht-voraussagbare Antwort an Lehrer 8 Antwort an Mitschüler
Vom Schüler ausgehendes Verbalverhalten	9 Beginnt Gespräch mit Lehrer 10 Beginnt Gespräch mit Mitschüler
Anderes	11 Stille 12 Durcheinander

Verbal Interaction Category System von Amidon & Hunter, entnommen aus Grell 1980, S. 52

Prüfungsliteratur: Bortz & Döring 2002, Kap. 4.5.

II.2.5. Gütekriterien

Überblick

Alle quantitativen Verfahren der Datenerhebung müssen drei Gütekriterien erfüllen: Objektivität, Reliabilität, Validität. Dies gilt insbesondere für Testverfahren; deswegen werden die Gütekriterien hier auch am Beispiel von Testverfahren erläutert.

Objektivität

Ein Verfahren ist objektiv, wenn es Ergebnisse liefert, die unabhängig von der Person des Forschers bzw. der Forscherin sind. Es lassen sich drei Arten von Objektivität unterscheiden:

- *Durchführungsobjektivität:* Durchführungsobjektivität wäre etwa dann nicht gegeben, wenn die Probanden/innen bei einem Test je verschieden antworten – je nachdem, wer den Test durchführt. Die Durchführungsobjektivität lässt sich erhöhen, indem die Vorgehensweise bei der Anwendung eines Verfahrens standardisiert wird, d.h. indem die Interaktionen von Forscher/in und erforschter Person möglichst genau festgelegt werden.
- *Auswertungsobjektivität:* Auswertungsobjektivität wäre dann nicht gegeben, wenn zwei Forscher/innen die Antworten einer untersuchten Person auswerten und dabei zu je verschiedenen Ergebnissen kommen. Die Auswertungsobjektivität lässt sich erhöhen, indem das Vorgehen bei der Auswertung genau festgelegt wird – indem z.B. bei einem Test genau spezifiziert ist, welche Antworten mit wie vielen Punkten in die Ergebnisse eingehen und wie sich die Punkte zu einem Gesamtergebnis aufsummieren.
- *Interpretationsobjektivität:* Die Interpretationsobjektivität wäre gering, wenn zwei verschiedene Forscher/innen dieselben Ergebnisse einer Person unterschiedlich interpretieren bzw. bewerten – wenn z.B. die eine Forscherin zu dem Schluss käme, ein Punktwert von

50 stelle ein vergleichsweise gutes Ergebnis dar, während der andere Forscher die Meinung vertreten würde, dass 50 Punkte im Verhältnis zum Bevölkerungsdurchschnitt einen eher unterdurchschnittlichen Wert ergeben. Die Interpretationsobjektivität speziell von Tests lässt sich steigern, indem der Test geeicht wird: Für die Population, für die der Test intendiert ist, werden Normwerte ermittelt; der Vergleich eines individuellen Ergebnisses mit den Normwerten erlaubt eine objektive Interpretation des Ergebnisses in Relation zum Bevölkerungsdurchschnitt (vgl. etwa den durchschnittlichen IQ von 100).

Die Objektivität wird bestimmt, indem der durchschnittliche Zusammenhang (die durchschnittliche Korrelation) zwischen der Durchführung (bzw. der Auswertung oder der Interpretation) eines Verfahrens durch verschiedene Forscher/innen ermittelt wird. Je näher bei 1 dieser Zusammenhang liegt, desto objektiver ist das Verfahren.

Reliabilität

Unter der Reliabilität versteht man die Zuverlässigkeit bzw. die Genauigkeit der Erfassung eines Merkmals. Es existieren verschiedene Verfahren zur Bestimmung der Reliabilität:

- *Retest-Reliabilität*: Ein Test wird denselben Personen zweimal nacheinander in einem gewissen Abstand vorgegeben; die Reliabilität (im Sinne der Stabilität) ergibt sich als Maß des Zusammenhangs zwischen den beiden Messwertreihen. Dieses Verfahren führt allerdings aufgrund von Erinnerungseffekten leicht zu einer Überschätzung der Reliabilität; bei Leistungstests ist das Verfahren eben wegen dieser Erinnerungseffekte gar nicht erst anwendbar.
- *Paralleltest-Reliabilität*: Es werden zwei äquivalente Testformen erstellt, und dieselbe Stichprobe von Personen bearbeitet beide Versionen kurz nacheinander. Die Reliabilität des Verfahrens entspricht wiederum dem Zusammenhang zwischen diesen beiden Messwertreihen. Dieses Verfahren ist sehr aufwändig, weil ja nicht nur ein Test konstruiert werden muss, sondern gleich zwei Tests. Der Aufwand lohnt sich nur dann, wenn von vornherein zwei Testversionen erforderlich sind.
- *Testhalbierungs-Reliabilität*: Ein Test wird in zwei gleichwertige Hälften unterteilt; die Reliabilität ergibt sich als der Zusammenhang zwischen den beiden Werten einer Person (über alle Teilnehmer/innen hinweg). Dieses Verfahren ist nicht weiter aufwändig; die Güte der Reliabilitätsbestimmung hängt jedoch davon ab, wie gut es gelingt, auch tatsächlich zwei äquivalente Testhälften herzustellen.
- *Interne Konsistenz*: Hier wird jedes einzelne Item als eine Art Paralleltest betrachtet; aus den Zusammenhängen der Items untereinander ergibt sich deren interne Konsistenz bzw. Homogenität, die hier als Maß für die Reliabilität angesetzt wird. Die interne Konsistenz erlaubt stabilere Schätzungen der Reliabilität als die Testhalbierung. Gebräuchlichstes Maß ist Cronbach's Alpha.

Die Reliabilität eines Tests sollte mindestens 0.8 betragen (bei einem Maximalwert = 1.0).

Validität

Ein Verfahren ist in dem Maß valide bzw. gültig, in dem es auch tatsächlich das erfasst, was es erfassen soll. Es lassen sich folgende Formen der Validität unterscheiden:

- *Inhaltsvalidität* (Augenscheinvalidität, face validity): Inhaltsvalidität ist dann gegeben, wenn ein Test das interessierende Merkmal direkt und in seinen wesentlichen Aspekten

erfasst (z.B. eine Maschinenschreibe-Probe, um zu erfassen, ob jemand auch tatsächlich Schreibmaschine schreiben kann). Die Feststellung dieser Form der Validität erfolgt ‚per Augenschein‘; diese Form der Validität lässt sich entsprechend auch nicht in Zahlen ausdrücken.

- *Kriteriumsvalidität*: Kriteriumsvalidität liegt in dem Maß vor, in dem die Messwerte bei dem interessierenden Test mit den Messwerten für ein Kriterium übereinstimmen. So könnte man beispielsweise davon ausgehen, dass Planungsfähigkeit auch etwas mit Intelligenz zu tun hat und entsprechend Intelligenz als Kriterium für Planungsfähigkeit ansetzen. Ein Test zur Erfassung von Planungsfähigkeit wäre dann valide, wenn die Werte der Probanden/innen in dem Test zur Erfassung der Planungsfähigkeit in einem positiven Zusammenhang mit den Intelligenzwerten derselben Probanden/innen stehen. Je nachdem, zu welchem Zeitpunkt die Kriteriumswerte erfasst werden, sind zwei Formen der Kriteriumsvalidität zu unterscheiden:
 - *Übereinstimmungsvalidität*: Die Messwerte auf dem zu validierenden Instrument und für das Kriterium werden zum selben Zeitpunkt erfasst (s. Beispiel Planungsfähigkeit und Intelligenz).
 - *Prognostische Validität*: Der zu validierende Test dient als Prädiktor für ein Kriterium, das in der Zukunft liegt. Die Testwerte werden in der Gegenwart erhoben, die Kriteriumswerte zu einem späteren Zeitpunkt. In diesem Sinne wurde beispielsweise die Validität von Schulnoten als Prädiktor für das Kriterium ‚Berufserfolg‘ verwendet.
- *Konstruktvalidität*: Bei der Konstruktvalidität handelt es sich um die ‚anspruchsvollste‘, aber zugleich auch aussagekräftigste Form der Validität. Das interessierende Merkmal wird vor einem theoretischen Hintergrund zu mehreren anderen Variablen in Beziehung gesetzt (z.B. Einsamkeit mit einem geringen Selbstwertgefühl, einer hohen sozialen Ängstlichkeit, mit dem Familienstand (bei Geschiedenen in der Regel höher als bei Verheirateten) usw.). Die resultierenden Hypothesen werden empirisch geprüft; das Instrument zur Erfassung des interessierenden Merkmals kann als valide gelten, wenn die Hypothesen sich bestätigen lassen. Wenn die Hypothesen dagegen widerlegt werden, lässt dies nicht unbedingt einen Schluss auf die mangelnde Validität des Verfahrens zu; es ist auch denkbar, dass die zugrunde gelegten Hypothesen fehlerhaft sind. Der Unterschied zwischen der Kriteriumsvalidität und der Konstruktvalidität besteht darin, dass die Konzepte, die bei der Kriteriumsvalidität herangezogen werden, untereinander nicht in einem kohärenten theoretischen Zusammenhang stehen.

Prüfungsliteratur: nur Skript

II.3. Forschungsmethoden

Überblick

Unter Forschungsmethoden versteht man die generelle Vorgehensweise beim Erarbeiten einer Fragestellung, bei der Planung, Durchführung und Auswertung einer Untersuchung. Im quantitativen Bereich wird zwischen zwei Arten von Forschungsmethoden unterschieden: experimentellen und non-experimentellen.

II.3.1. Das (Labor-)Experiment

Merkmale des Experiments

Das Experiment stellt eine Forschungsmethode zur Überprüfung von Kausalhypothesen dar (z.B.: Wenn Menschen beim Lernen Pausen machen, dann ist ihre Behaltensleistung höher (als beim Lernen ohne Pausen). Ziel des Experiments ist also die Erklärung von Sachverhalten.

Das Experiment ist durch zwei Merkmale bestimmt:

- systematische Variation mindestens einer Variablen und Registrieren des Effekts dieser Variation:
- gleichzeitiges Ausschalten der Wirkung von potenziellen Störvariablen.

Diese beiden Merkmale gemeinsam sind es, die die Überprüfung speziell von Kausalhypothesen erlauben. Durch die systematische Variation (mindestens) einer Variablen lässt sich prüfen, ob diese Variable den interessierenden Gegenstandsbereich verändert. Durch das gleichzeitige Ausschalten von Störvariablen lässt sich ausschließen, dass eine eventuelle Veränderung im Gegenstandsbereich auf andere Ursachen als die systematische Variation des potenziellen Ursachenfaktors zurückzuführen ist.

Arten von Variablen im Experiment

- *Unabhängige Variable (UV)*: diejenige Variable, die variiert wird; zugleich die Variable im Wenn-Teil der Hypothese (z.B. Lernen mit Pausen / am Stück); die in der Untersuchung realisierten Ausprägungen der unabhängigen Variablen werden auch *Stufen* der UV genannt;
- *Abhängige Variable (AV)*: diejenige Variable, an der die Wirkung der Variation der unabhängigen Variable beobachtet wird; zugleich die Variable im Dann-Teil der Hypothese (z.B. Behaltensleistung);
- (potenzielle) *Störvariablen*: andere Variablen als die UV, die die Ausprägung der AV ebenfalls beeinflussen könnten und die daher zu kontrollieren sind (z.B. Interesse am Thema, Vorwissen, Intelligenz, Geräuschpegel usw.).

Störvariablen: Definition

Eine Variable wird dann als Störvariable bezeichnet, wenn sie erstens für die Ausprägung der AV von Bedeutung und wenn sie zweitens mit der UV *konfundiert* ist. Eine Konfundierung einer unabhängigen mit einer Stör-Variablen liegt dann vor, wenn die Stufen der Störvariable in systematischer Weise mit den Stufen der unabhängigen Variable variieren. In diesem Fall lässt sich eine Veränderung in den Ausprägungen der AV nicht mehr eindeutig auf die Variation der UV zurückführen (sondern sie könnte auch durch den Einfluss der Störvariable zustande gekommen sein).

Wenn man (s.o.) beispielsweise den Einfluss der Lernform (mit Pausen / am Stück) auf die Behaltensleistung überprüfen würde, dann wäre Vorwissen eine potenzielle Störvariable, denn auch das Vorwissen kann die Behaltensleistung beeinflussen. Das Vorwissen wird dann tatsächlich zu einer Störvariable, wenn z.B. die Untersuchungsteilnehmer/innen in der Untersuchungsbedingung ‚Lernen mit Pausen‘ mehrheitlich auch über ein hohes Vorwissen verfügen, während die Untersuchungsteilnehmer/innen in der Bedingung ‚Lernen ohne Pausen‘ meist ein eher geringes Vorwissen haben. Wenn die Teilnehmer/innen in der Bedingung ‚Lernen mit Pausen‘ nun tatsächlich eine höhere Behaltensleistung aufweisen, dann kann dies ebenso auf das höhere Vorwissen dieser Personen wie auf ihren Lernstil (mit Pausen) zurückzuführen sein. Die Ergebnisse sind also nicht mehr eindeutig interpretierbar.

Es wird zwischen allgemeinen und speziellen Störvariablen unterschieden. Allgemeine Störvariablen sind solche, die in jeder experimentellen Untersuchung zu kontrollieren sind. Spezielle Störvariablen treten dagegen nur in Untersuchungen mit Messwiederholung auf. Im Folgenden wird auf allgemeine Störvariablen genauer eingegangen.

Arten von allgemeinen Störvariablen

- *Vpn-Merkmale*: Alle beliebigen Merkmale der Untersuchungsteilnehmer/innen (Versuchspersonen: Vpn) können potenziell eine Störvariable darstellen (oben z.B. Vorwissen).
- *Vl-Merkmale*: Auf Seiten des Versuchsleiters bzw. der Versuchsleiterin wird zwischen zwei Arten potenzieller Störmerkmale unterschieden:
 - *biosoziale Merkmale*: Das sind erstens verschiedenste biosoziale Faktoren, z.B. das Geschlecht oder der soziale Status der Versuchsleitung, unterschiedliche Grade von Freundlichkeit, mit denen er oder sie den Untersuchungsteilnehmern/innen begegnet usw.
 - *Erwartungen*: Zweitens können sich die Erwartungen der Versuchsleitung im Sinne einer Störvariable auswirken. Dies wurde erstmals von Rosenthal in den sechziger Jahren nachgewiesen: Er führte eine Untersuchung durch, in der die Teilnehmer/innen Ratten beim Erlernen eines Labyrinths trainieren sollten. Der einen Gruppe von Teilnehmern/innen wurde gesagt, ihre Ratten seien ganz besonders intelligente Tiere; der anderen Gruppe wurde mitgeteilt, ihre Ratten seien eher unterdurchschnittlich intelligent. In der Tat bestanden zu Untersuchungsbeginn zwischen den Ratten keinerlei Unterschiede. Zum Untersuchungsende zeigte sich allerdings, dass die angeblich intelligenteren Ratten in der Tat deutlich besser abgeschnitten hatten. Die Untersuchungsteilnehmer/innen in den beiden Gruppen hatten den Ratten also ihre je unterschiedlichen Erwartungen vermittelt, sie auch je anders behandelt und so tatsächlich unterschiedliche Ergebnisse erzielt. In der Folge konnte gezeigt werden, dass sich auch Menschen in Abhängigkeit von den Erwartungen, die an sie herangetragen werden, je unterschiedlich verhalten.
- *Situationsmerkmale*: Auch verschiedenste Merkmale der Untersuchungssituation können zu Störvariablen werden, wenn sie systematisch mit den Stufen der UV konfundiert sind (Helligkeit, Geräuschpegel usw.).

Kontrolle von Vp-seitigen Störvariablen

Zur Kontrolle vp-seitiger Störvariablen stehen vor allem zwei Techniken zur Verfügung: Randomisieren und Parallelisieren.

- *Randomisieren*: Randomisieren bedeutet, dass die Untersuchungsteilnehmer/innen den Untersuchungsbedingungen bzw. den Stufen der UV (z.B. Lernen mit vs. ohne Pausen) per Zufall zugewiesen werden. Dahinter steht die Annahme, dass verschiedenste Personenmerkmale in der Bevölkerung normalverteilt sind. Bei einer hinreichend großen Stichprobe ist daher davon auszugehen, dass bei einer Randomisierung die Verteilungen potenzieller Störvariablen über die verschiedenen Untersuchungsbedingungen statistisch äquivalent sind. Der große Vorteil dieser Kontrolltechnik besteht entsprechend darin, dass die Störvariablen gar nicht bekannt sein müssen, um sie zu kontrollieren. Allerdings setzt die Randomisierung hinreichend große Stichproben voraus (ca. 80 Personen oder mehr).
- *Parallelisieren*: Bei einer kleineren Stichprobe ist die Kontrolltechnik des Parallelisierens besser geeignet; sie setzt allerdings voraus, dass die fragliche Störvariable bekannt ist (z.B. Vorwissen im obigen Beispiel). Parallelisieren bedeutet, dass im ersten Schritt die Werte der Untersuchungsteilnehmer/innen auf der Störvariable erhoben werden. Zweitens werden die Teilnehmer/innen entsprechend ihren Werten auf der Störvariablen in eine Rangreihe gebracht. Drittens werden, auf der Grundlage dieser Rangreihe, Paare (bei

zweistufigen Uven) von Personen gebildet, die auf der Störvariablen vergleichbare Werte aufweisen. Diese beiden Personen werden viertens den Untersuchungsbedingungen per Zufall zugewiesen. Auf diese Weise wird sichergestellt, dass die Untersuchungsgruppen hinsichtlich der relevanten Störvariable ebenfalls statistisch äquivalent sind. Andere Störvariablen sind damit jedoch nicht kontrolliert. Außerdem eignet sich das Verfahren lediglich zur Kontrolle von einer, maximal von zwei Störvariablen; bei mehr als zwei bekannten Störvariablen wird das Verfahren zu kompliziert.

Vp-seitige Störvariablen können auch durch systematische und durch Zufalls-Variation kontrolliert werden (s.u.); Randomisieren und Parallelisieren sind jedoch die üblicheren Techniken.

Achtung – Nicht verwechseln...

Manchmal wird die Kontrolltechnik des Randomisierens mit dem Ziehen einer Zufallsstichprobe verwechselt:

- Das Ziehen einer Zufallsstichprobe dient in erster Linie der Sicherung der externen Validität einer Untersuchung (durch Sicherung der Repräsentativität der Stichprobe für die Population). Zufallsstichproben können verschiedensten Untersuchungsdesigns zugrunde liegen, können also auch im Rahmen non-experimenteller Untersuchungen zur Anwendung kommen.
- Das Randomisieren stellt eine Technik zur Kontrolle vp-seitiger Störvariablen dar. Sie dient der Sicherung der internen Validität und kommt nur im Rahmen experimenteller Untersuchungsanordnungen zur Anwendung. Die Technik des Randomisierens kann auf verschiedenste Arten von Stichproben angewandt werden; prinzipiell ist es durchaus möglich, die Personen in einer Gelegenheitsstichprobe den Untersuchungsbedingungen per Zufall zuzuteilen. Allerdings zeigt die Kontrolltechnik bei der Anwendung auf eine Zufallsstichprobe die größte Wirkung.

Kontrolle von Versuchsleiter-Erwartungseffekten und biosozialen VI-Merkmalen

Ein VI-Erwartungseffekt (Rosenthal-Effekt) kann bei bestimmten Arten von Untersuchungen auftreten: nämlich erstens vorzugsweise in Einzeluntersuchungen sowie zweitens nur dann, wenn der/die Untersuchungsleiter/in weiß, welcher Untersuchungsbedingung die gerade untersuchte Person zugeteilt ist. Zur Kontrolle solcher Erwartungseffekte und anderer VI-Merkmale eignen sich die folgenden Maßnahmen:

- *Standardisieren des Versuchsablaufs:* Die Interaktion zwischen Versuchsleitung und Untersuchungsteilnehmern/innen ist genau festgelegt.
- *Elimination der Versuchsleitung:* Untersuchungsteilnehmer/innen interagieren nicht mit einer Person, sondern die Instruktionen werden beispielsweise von einem Tonband abgespielt oder der Versuch wird per Computer gesteuert. Dabei ist allerdings zu bedenken, dass die Untersuchungssituation auch dann noch eine soziale Situation bleibt, wenn kein/e Versuchsleiter/in anwesend ist. Der soziale Charakter der Situation lässt sich nicht ausschalten.
- *Training der Versuchsleitung:* Versuchsleiter/innen werden vor Untersuchungsbeginn darin geschult, in möglichst gleichförmiger Weise mit verschiedenen Personen zu interagieren. Dies beinhaltet insbesondere auch die Kontrolle über die eigene Körpersprache.
- *Blindversuche:* Eine der wichtigsten Techniken zur Kontrolle von Erwartungseffekten ist der Blindversuch: Hier wird die Untersuchung beispielsweise von anderen Personen durchgeführt als denen, die die Untersuchung geplant haben. Die Versuchsleitung weiß nicht, welcher Bedingung die untersuchte Person zugeteilt ist und kann daher erst gar keine Erwartungen ausbilden, die dann dem Gegenüber vermittelt werden könnten.

Kontrolle von Situationsmerkmalen

Zur Kontrolle von potenziellen Störvariablen auf Seiten der Untersuchungssituation stehen vier Techniken zur Verfügung:

- *Elimination*: Potenziell störende Situationsmerkmale werden ausgeschaltet, z.B. Geräusche in einem schalldichten Labor (s. auch oben V-Merkmale).
- *Konstanthalten*: Relevante Situationsmerkmale werden für alle Teilnehmer/innen gleich gehalten, z.B. sind alle demselben Geräuschpegel ausgesetzt.
- *Zufallsvariation*: Bei dieser Technik werden alle möglichen (oder möglichst viele) Ausprägungen der potenziellen Störvariable realisiert; die Untersuchungsteilnehmer/innen werden diesen Ausprägungen per Zufall zugewiesen.
- *Systematische Variation*: Bei diesem Kontrollverfahren wird die potenzielle Störvariable als zusätzliche UV in die Untersuchung einbezogen. Auf diese Weise kann die Wirkung der Variable auf die AV genau ermittelt werden; auch das Zusammenwirken der Störvariable mit den unabhängigen Variablen ist erfassbar.

Interne und externe Validität

Zentrale Gütekriterien der Forschungsmethode des Experiments sind die interne und die externe Validität.

- *Interne Validität*: Ein Experiment ist in dem Maß intern valide, indem es gelungen ist, potenzielle Störvariablen zu kontrollieren.
- *Externe Validität*: Der Begriff der externen Validität bezieht sich auf die Verallgemeinerbarkeit der Ergebnisse des Experiments. Es werden drei Arten externer Validität unterschieden:
 - *Populationsvalidität*: Verallgemeinerbarkeit von der Stichprobe auf die Population (Von einer Untersuchung mit Studierenden kann beispielsweise nicht ohne weiteres auf die Bevölkerung insgesamt verallgemeinert werden.)
 - *Konstruktvalidität* (auch: Variablenvalidität): Verallgemeinerbarkeit auf andere Operationalisierungen der UV und der AV (Eine Erfassung der Lernleistung mittels multiple choice-Fragen kann beispielsweise andere Ergebnisse erbringen als die Erfassung der Lernleistung durch freies Reproduzieren.)
 - *Situationsvalidität*: Verallgemeinerbarkeit auf andere Situationen (Von einer der Lernleistung in einem schalldichten Raum lässt sich nicht ohne weiteres auf die Lernleistung am eigenen Schreibtisch schließen.)

Interne und externe Validität gelten als gegenläufig. Wenn man die interne Validität erhöht, senkt man damit häufig zugleich die externe; erhöht man dagegen die externe Validität, beeinträchtigt dies meistens die externe Validität (s. ausführlicher II.3.2.).

Dieses Problem ist allerdings weniger gravierend, wenn man berücksichtigt, dass die Erforschung eines Gegenstandsbereichs meist mehrere aufeinander folgende Untersuchungen beinhaltet. Dabei ist es sinnvoll, zunächst Untersuchungen mit hoher interner Validität durchzuführen, um den fraglichen Kausalzusammenhang auch möglichst gut abzusichern. Im Anschluss können dann weitere Untersuchungen durchgeführt werden, in denen mittels gezielter Replikationen der ursprünglichen Untersuchung der Geltungsbereich des Kausalzusammenhangs sukzessive erweitert wird. So kann man z.B. zunächst unter Laborbedingungen prüfen, ob Pausen die Behaltensleistung tatsächlich steigern. Daran können sich dann weitere Untersuchungen anschließen, in denen geprüft wird, inwieweit dies auch für verschiedenartige Lernstoffe (Vokabellernen, Auswendiglernen, Lernen auf Zusammenhänge hin usw.) und für verschiedene Personengruppen gilt.

Achtung – Nicht verwechseln...

Interne und externe Validität werden häufig mit Kriteriumsvalidität, Konstruktvalidität usw. verwechselt. Es besteht jedoch ein wesentlicher Unterschied zwischen diesen Validitätskonzepten:

- Inhaltsvalidität, Kriteriums- und Konstruktvalidität beziehen sich auf Methoden der Datenerhebung und der Auswertung.
- Interne und externe Validität beziehen sich dagegen auf ganze Forschungsansätze wie z.B. das Experiment.

Forschungsprozess

Im Folgenden ist noch einmal der gesamte Forschungsprozess für eine experimentelle Untersuchung aufgeführt:

Forschungsprozess im Experiment

- Finden einer Fragestellung
- Aufstellen einer Sachhypothese
- Operationalisierung der Variablen
- Aufstellen eines Versuchsplans
- Kontrolle der Störvariablen
- Stichprobenziehung
- Ableitung der empirischen und der statistischen Hypothesen
- Durchführung (Datenerhebung)
- Auswertung
- Rückschluss von den statistischen Ergebnissen auf die Sachhypothese
- Diskussion

Dabei ergeben sich gegenüber den bisherigen Darstellungen der Vorgehensweise zwei Spezifikationen. Diese betreffen erstens die Hypothesen und zweitens das Aufstellen eines Versuchsplans.

Hypothesen im Experiment

Einer experimentellen Studie liegt in der Regel eine theoretisch-inhaltliche Hypothese zugrunde (z.B. Wenn man beim Lernen Pausen macht, dann verbessert das die Behaltensleistung.). Diese kann jedoch nicht in ihrem vollen Umfang getestet werden. Vielmehr wird die ursprüngliche Hypothese im Verlauf der experimentellen Testung immer weiter konkretisiert.

- *Empirisch-inhaltliche Hypothese:* Die empirisch-inhaltliche Hypothese leitet sich durch die Operationalisierung der Variablen aus der theoretisch-inhaltlichen Hypothese ab, z.B.: Wenn Personen beim Lernen von wahrnehmungspsychologischen Inhalten nach vier Stunden eine Pause von dreißig Minuten machen, dann beantworten sie im Anschluss an die Lernphase mehr multiple choice-Fragen über den Stoff richtig als Personen, die acht Stunden am Stück gelernt haben.
- *Statistische Hypothese:* Die empirisch-inhaltliche Hypothese wird wiederum in Form einer statistischen Hypothese weiter konkretisiert, d.h. als Hypothese über statistische Kennwerte der Stichprobe, z.B.: Der Mittelwert richtig beantworteter Fragen ist in der Gruppe mit Pause höher als in der Gruppe ohne Pause.
- *Testhypothese:* Die statistische Hypothese ist häufig mit der Testhypothese identisch, d.h. mit der Hypothese, die auch statistisch geprüft wird. Die Testhypothese wird auch als *Alternativhypothese* bezeichnet und postuliert meistens (wie auch in dem obigen Beispiel) einen Unterschied zwischen den untersuchten Gruppen. Exakt komplementär zur Alternativhypothese ist die *Nullhypothese* formuliert (z.B.: Der Mittelwert richtig beantworteter Fragen ist in der Gruppe mit Pause kleiner oder gleich dem in der Gruppe ohne Pause). Da

Alternativ- und Nullhypothese einander wechselseitig ausschließen, kann von der Bewährung oder Zurückweisung der Testhypothese auf die Gültigkeit der Nullhypothese zurück geschlossen werden. Wenn sich die Testhypothese z.B. bestätigen lässt und sie entsprechend vorläufig angenommen wird, dann bedeutet das zugleich, dass die Nullhypothese vorläufig zurückgewiesen wird.

Versuchsplan

Experimentelle Untersuchungen können ganz unterschiedlich angelegt sein. So kann eine experimentelle Untersuchung eine UV und eine AV beinhalten, es können aber auch mehrere Uven gleichzeitig angesetzt werden. Bei manchen Untersuchungen durchlaufen die Teilnehmer/innen eine experimentelle Variation nicht nur einmal, sondern mehrfach (Untersuchung mit vs. ohne Messwiederholung). Der logische Aufbau der Untersuchung im Hinblick auf die Hypothesentestung wird als *Versuchsplan* bezeichnet.

Beispiel für einen Versuchsplan

UV B (z.B. Lernart)	B1 (mit Pausen)	B2 (ohne Pausen)
UV A (z.B. Vorwissen)		
hoch	a1b1	a1b2
niedrig	a2b1	a2b2

Prüfungsliteratur: Huber 1999, Kap. 3, 4, 7

II.3.2. Arten des Experiments

Überblick

Neben dem klassischen Laborexperiment, das bisher im Vordergrund stand, lassen sich noch weitere Arten des Experiments unterscheiden: das Feldexperiment, das Quasi-Experiment sowie die Feldstudie.

Arten des Experiments		
	Labor	Feld
Zufallszuteilung d. Vpn möglich	Labor- experiment	Feld- experiment
Zufallszuteilung d. Vpn nicht möglich	Quasi- experiment	Feld- studie

Das Feldexperiment

Beim Feldexperiment handelt es sich um eine Form des Experiments, die – im Gegensatz zum Laborexperiment – in der natürlichen Umgebung der Untersuchungsteilnehmer/innen durchgeführt wird. Dadurch erhöht sich die Situationsvalidität (als Teilbereich der externen Validität). Zugleich sind potenzielle Störvariablen auf Seiten der Situation jedoch kaum kontrollierbar, so dass die interne Validität des Feldexperiments geringer anzusetzen ist als die des Laborexperiments.

Das Quasi-Experiment

In experimentellen Untersuchungen können auch unabhängige Variablen von Interesse sein, die fest mit den Personen der Untersuchungsteilnehmer/innen verbunden sind (*organismische Variablen*: z.B. Geschlecht, Intelligenz, Vorwissen usw.). Das bedeutet zugleich, dass eine Kontrolle versuchspersonenseitiger Störvariablen, etwa durch Randomisierung, in solchen Fällen nicht möglich ist (Teilnehmer/innen können nicht per Zufall den Faktorstufen ‚männlich‘ und ‚weiblich‘ zugewiesen werden). Experimentelle Anordnungen, in denen mindestens eine UV aktiv variiert wird, bei denen jedoch eine zufällige Zuteilung der Teilnehmer/innen zu den Untersuchungsbedingungen nicht möglich ist, werden als Quasi-Experiment bezeichnet. Wegen der unzureichenden Kontrolle von Störvariablen auf Seiten der Versuchspersonen können die Ergebnisse quasi-experimenteller Untersuchungen nur eingeschränkt kausal interpretiert werden. Eine Kontrolle der übrigen Störvariablen ist jedoch möglich.

Quasi-experimentelle Untersuchungen liegen (wie im obigen Beispiel vor), wenn eine der unabhängigen Variablen eine organismische Variable ist. Auch Untersuchungen mit Messwiederholung (bei denen die Teilnehmer/innen die Untersuchungsbedingungen mehrfach durchlaufen) stellen meist Quasi-Experimente dar. Schließlich kann es bei angewandten Untersuchungen der Fall sein, dass nur eine quasi-experimentelle Anordnung möglich ist. So kann z.B. eine Schulklasse nicht einem beliebigen Stadtbezirk zugeteilt werden.

Die interne Validität des Quasi-Experiments ist im Vergleich zum Laborexperiment eingeschränkt; die externe Validität ist ebenfalls eher niedrig anzusetzen.

Die Feldstudie

Die Feldstudie stellt ein Quasi-Experiment dar, das in der natürlichen Umgebung der Untersuchungsteilnehmer/innen durchgeführt wird. Entsprechend sind sowohl Störvariablen auf Seiten der Untersuchungsteilnehmer/innen (wie beim Quasi-Experiment) und auf Seiten der Situation (wie beim Feldexperiment) nur schwer zu kontrollieren; die interne Validität dieser experimentellen Anordnung ist also in Relation zu der der anderen Anordnungen vergleichsweise am niedrigsten. Die Situationsvalidität als Teilaspekt der externen Validität ist dagegen eher hoch anzusetzen (wiederum wie beim Feldexperiment).

Prüfungsliteratur: Huber 1999, Kap. 3; Hussy & Jain 2002, Kap. 5.1. – 5.5.

II.3.3. Die Zusammenhangsstudie

Überblick

Bei den bisher aufgeführten Forschungsmethoden handelt es sich sämtlich um experimentelle Anordnungen, die (trotz aller Einschränkungen der internen Validität z.B. beim Quasi-Experiment oder der Feldstudie) auf die Erklärung von Sachverhalten ausgerichtet sind. Neben erklärenden sind aber auch in der quantitativen Forschung rein beschreibende Untersuchungsanordnungen möglich. Von zentraler Bedeutung ist dabei die Zusammenhangsstudie.

Deren Ziel ist es, den Zusammenhang zwischen zwei Variablen zu beschreiben (z.B. zwischen Intelligenz und Kreativität).

Zusammenhangsstudie vs. experimentelle Anordnungen

Die Zusammenhangsstudie unterscheidet sich von den experimentellen Untersuchungsanordnungen vor allem in den folgenden Punkten:

- Es wird nicht zwischen abhängiger und unabhängiger Variable unterschieden.
- Die interne Validität stellt kein Kriterium bei der Beurteilung der Zusammenhangsstudie dar.
- Die Versuchsleitung greift nicht in das Geschehen ein.
- Überprüft wird eine Zusammenhangshypothese, keine Kausalhypothese.
- Zielsetzung ist die Beschreibung, nicht die Erklärung.

Prüfungsliteratur: nur Skript

II.3.4. Ethische Probleme quantitativer Forschung

Ethische Probleme psychologisch-quantitativer Forschung

Im Kontext psychologischer Forschung können vor allem drei Arten ethischer Probleme auftreten:

- *Schädigungen* oder Beeinträchtigungen der Teilnehmer/innen: z.B. wenn die Teilnehmer/innen eine negative Rückmeldung über eigene Fähigkeiten erhalten;
- *Eindringen in die Privatsphäre* der Teilnehmer/innen: z.B. bei verdeckter Beobachtung;
- *Missinformation / Täuschung*: wenn die Teilnehmer/innen über den Zweck der Studie getäuscht werden und / oder darüber hinaus im Untersuchungsverlauf unzutreffende Informationen erhalten.

Beispiel: Die Milgram-Studie

Die Milgram-Studie stellt eine Untersuchung dar, bei der die Probleme der Täuschung und der potenziellen Schädigung von Teilnehmern/innen besonders deutlich werden.

Milgram ging es in seiner Studie um die Frage, inwieweit Menschen aus reinem Autoritätsgehorsam zentralen moralischen Prinzipien zuwiderhandeln. Den Teilnehmern/innen wurde gesagt, es handle sich um ein Lernexperiment: Sie selbst sollten als ‚Lehrer‘ fungieren. In einem anderen Raum saß ein ‚Schüler‘ (in Wirklichkeit ein/e Vertraute/r Milgrams), dessen Aufgabe darin bestand, Worte zu lernen. Immer, wenn der ‚Schüler‘ bei der Wiedergabe einen Fehler machte, sollte der ‚Lehrer‘ ihn mit einem Stromschlag bestrafen. Mit jedem Fehler steigerte sich die Schockintensität, die von 30 bis 450 Volt reichte (in der Tat erhielten die Personen im anderen Raum natürlich keine Stromschläge). Die Stromstärken waren verbal markiert, von ‚leichte Intensität‘ bis ‚Lebensgefahr‘. Die ‚Lehrer‘ erhielten zu Untersuchungsbeginn selbst einen Stromschlag von 75 Volt. Als Autoritätsfigur fungierte der Untersuchungsleiter, der die gesamte Zeit anwesend war.

Im Untersuchungsverlauf begann der ‚Schüler‘, ab einer Stromstärke von 75 Volt zu jammern, bei höherer Stromstärke zu schreien, er hätte eine Herzschwäche, und den ‚Lehrer‘ anzuflehen, er solle aufhören. Die Untersuchungsteilnehmer/innen beschwerten sich daraufhin, baten den Untersuchungsleiter, aufhören zu dürfen. Dieser befahl den Teilnehmer/innen jedoch, weiterzumachen; die Untersuchung erfordere es, dass sie weitermachten. Obwohl sie protestierten, gingen zwei Drittel der Teilnehmer/innen bis zur höchsten Stromstärke.

Die Untersuchung wurde im Folgenden in verschiedenen Ländern unter Heranziehung verschiedener unabhängiger Variablen wiederholt. So zeigte sich z.B., dass der Autoritätsgehorsam mit der Entfernung zur Autoritätsperson und mit größerer körperlicher Nähe zum ‚Opfer‘ geringer ausfällt.

Die Teilnehmer/innen wurden nach der Untersuchung über deren Zweck aufgeklärt. Milgram weist darauf hin, dass die Teilnehmer/innen die Untersuchung und ihr eigenes Handeln als eine Form der Selbsterkenntnis bewerteten. Trotzdem wurde die Untersuchung in der Folge unter ethischer Perspektive intensiv diskutiert: zum einen wegen der Beeinträchtigung, die den Teilnehmern/innen dadurch entsteht, dass sie zu moralwidrigem Handeln aufgefordert werden, zum anderen wegen der Rolle, die die Täuschung der Teilnehmer/innen in der Untersuchung spielt.

Weshalb wird in psychologischen Untersuchungen häufig getäuscht?

Unter den drei ethischen Problemen psychologischer Forschung ist Täuschung vermutlich dasjenige, das am häufigsten auftritt und am intensivsten diskutiert wird. Dass Täuschung in psychologischen Experimenten eine so große Rolle spielt, hat seine Ursache in einem grundlegenden Konflikt zwischen methodischen Anforderungen einerseits und ethischen Kriterien andererseits. Da der Mensch ein denkendes und fühlendes Wesen ist, macht er sich grundsätzlich seine Gedanken über den Zweck eines psychologischen Experiments und handelt ggf. auch diesen Gedanken entsprechend. Damit entspricht seine Handlungsweise allerdings nicht mehr der, wie er sie spontan zeigen würde, und die Validität der Ergebnisse ist beeinträchtigt. Damit die Validität der Untersuchung nicht leidet, werden die Teilnehmer/innen über den Zweck der Untersuchung zunächst fehlinformiert – was jedoch wiederum einen Verstoß gegen moralische Prinzipien darstellt.

Dieser Konflikt zwischen methodischen Anforderungen und ethischen Kriterien muss jedoch nicht dazu führen, dass in jeder psychologischen Untersuchung zwangsläufig getäuscht wird. Zum einen beeinträchtigt die wahrheitsgemäße Information der Teilnehmer/innen über den Zweck der Untersuchung deren Validität meist nur dann, wenn die Untersuchung einen Gegenstandsbereich betrifft, der sozial stark normiert ist (z.B. Stereotype, Vorurteile); nur in solchen Fällen ist aus methodologischen Überlegungen heraus eine Täuschung erforderlich. Zum anderen existieren ja auch viele Fragestellungen, bei denen das Denken und Fühlen des Menschen gerade Teil des interessierenden Gegenstandsbereichs ist.

Potenzielle Folgen von Täuschung

Wenn in einer Untersuchung getäuscht wird, kann das negative Konsequenzen nach sich ziehen:

- Es kann zu einem Vertrauensverlust in Bezug auf diese/n eine/n Forscher/in kommen
- Es kann auch ein Vertrauensverlust in Bezug auf psychologische Forschung im allgemeinen resultieren. In einem solchen Fall werden Personen grundsätzlich skeptisch sein, ob der genannte Zweck einer Untersuchung auch mit dem tatsächlichen Zweck übereinstimmt, ob die Anonymität ihrer Daten wirklich gewährleistet ist usw. Gegebenenfalls stellen sich Teilnehmer/innen, die einmal getäuscht wurden, für eine weitere Untersuchung nicht mehr zur Verfügung.
- Untersuchungsteilnehmer/innen können in ihrem Selbstwertgefühl beeinträchtigt sein, Schuldgefühle entwickeln usw. – je nachdem, worin die Täuschung im Einzelnen bestanden hat.

Begründungsrichtungen in der Diskussion um die Ethik psychologischer Forschung

- *Utilitaristisch*: Unter utilitaristischer Perspektive werden Handlungen im Hinblick auf ihre Konsequenzen beurteilt; dies beinhaltet in der Regel eine *Kosten-Nutzen-Abwägung*. Diese Begründungsrichtung wird meist von Befürwortern einer Täuschung vertreten. Unter dieser Perspektive ließe sich z.B. für die Milgram-Studie anführen, dass sie Einsichten geführt hat, die anders nicht zu gewinnen waren.
- *Deontologisch*: Unter der deontologischen Perspektive ist der moralische Wert von Handlungen an sich unter Rückgriff auf allgemeingültige ethische Prinzipien thematisch. Diese Begründungsrichtung findet sich häufiger bei Gegnern der Täuschung. Diese würden z.B. das Prinzip der Selbstanwendung gegen die Täuschung anführen: Wenn man selbst nicht getäuscht werden möchte, sollte man auch andere nicht täuschen.

Maßnahmen gegen negative Konsequenzen psychologischer Forschung

Es lassen sich zwei Arten von Maßnahmen gegen negative Konsequenzen psychologischer Forschung unterscheiden: solche auf der *institutionellen* und solche auf der *individuellen* Ebene:

Institutionelle Maßnahmen

- *Berufsethische Richtlinien:* Psychologische Gesellschaften und Verbände (wie z.B. die Deutsche Gesellschaft für Psychologie, die American Psychological Association usw.; für ein Beispiel siehe: <http://www.dgps.de/dgps/kommissionen/ethik/003.php4>) haben in der Regel berufsethische Richtlinien verabschiedet. In diesen Richtlinien werden die Forscher/innen auf ihre Verantwortlichkeit gegenüber Untersuchungsteilnehmern/innen hingewiesen, es werden Maßnahmen zur Vermeidung von Täuschung und Irreführung genannt usw. Die Einhaltung dieser Maßnahmen wird allerdings nicht streng kontrolliert.
- *Begutachtung durch Ethikkommissionen:* Insbesondere in den USA müssen Forscher/innen, die Untersuchungen am und mit Menschen durchführen, ihre Pläne vor Untersuchungsbeginn einer Ethikkommission vorlegen und genehmigen lassen. In Deutschland sind solche Ethikkommissionen bisher weniger verbreitet. Aber auch hier spielen ethische Gesichtspunkte bei der Genehmigung von Forschungsanträgen z.B. durch die Deutsche Forschungsgemeinschaft durchaus eine Rolle.
- *Einbringen berufsethischer Prinzipien in die Ausbildung:* Studierende der Psychologie werden schon frühzeitig während ihrer Ausbildung für die ethischen Probleme psychologischer Forschung sensibilisiert.
- *Rechtsvorschriften:* In einigen Bereichen existieren schließlich auch klare Rechtsvorschriften, so z.B. das schon mehrfach erwähnte Recht auf informationelle Selbstbestimmung; auch müssen Forscher/innen die Daten von Untersuchungsteilnehmern/innen vertraulich behandeln.

Individuelle Maßnahmen

Letztlich liegt die Entscheidung über die genaue Untersuchungsplanung und das genaue Vorgehen jedoch meist bei der einzelnen Forscherin bzw. dem einzelnen Forscher. Wenn diese/r entscheidet, dass Täuschungen oder auch Beeinträchtigungen der Untersuchungsteilnehmer/innen unvermeidbar sind, bieten sich – sozusagen als Ausgleich - die folgenden Maßnahmen an:

- *Postexperimentelle Aufklärung:* Die Teilnehmer/innen werden (wie bei Milgram) zumindest im Anschluss an die Untersuchung über deren Zweck aufgeklärt und haben die Gelegenheit, Fragen zu stellen. Eventuelle Missinformationen im Untersuchungsverlauf werden richtig gestellt. Allerdings kann eine solche nachträgliche Aufklärung eventuelle Beeinträchtigungen der Untersuchungsteilnehmer/innen nicht verhindern.
- *Informierte Einwilligung (informed consent):* Die Teilnehmer/innen werden vor Untersuchungsbeginn soweit als möglich über den Zweck der Untersuchung informiert; wenn dies nicht möglich sein sollte, wird den Teilnehmer/innen mitgeteilt, dass eine vollständige Information über den Untersuchungszweck im voraus nicht möglich ist. Die Teilnehmer/innen unterzeichnen daraufhin einen Vertrag, dass sie unter den genannten Bedingungen zu einer Untersuchungsteilnahme bereit sind. Es bleibt ihnen freigestellt, die Untersuchung jederzeit abzubrechen, ohne dass ihnen daraus negative Konsequenzen entstehen.
- *Kompensation der Untersuchungsteilnehmer/innen:* Die Teilnehmer/innen sollen für ‚Kosten‘ entschädigt werden, die ihnen u.U. entstanden sind, z.B. in Form einer Vergütung oder durch die Vergabe von ‚Versuchspersonenstunden‘.

Fazit

Genauere Richtlinien dafür, wann eine Täuschung der Untersuchungsteilnehmer/innen gerechtfertigt ist, existieren nicht. Hier muss der/die Forscher/in selbst entscheiden. Dabei sollte man bedenken, dass Menschen Situationen sehr unterschiedlich beurteilen – was einem selbst ganz unproblematisch erscheint (z.B. verdeckte Beobachtung der Interaktionen mit anderen Personen beim Warten vor Untersuchungsbeginn), kann für andere durchaus problematisch sein.

Die Entscheidung sollte daher möglichst unter Rücksprache mit anderen Forschern/innen getroffen werden.

Prüfungsliteratur: Huber 1999, Kap. 9; Hussy & Jain 2002, Kap. 6

III. Qualitative Forschung

Wie fühlt es sich eigentlich an, Tag für Tag mit einer körperlichen Behinderung zu leben? Was geht in jemandem vor sich, der gerade einen nahe stehenden Menschen verloren hat? Wie ist die Kultur der Fans von Horrorfilmen organisiert? Solche Fragen lassen sich mit den quantitativen Forschungsmethoden, wie sie bisher erläutert wurden, nicht beantworten. Es geht hier nicht um die Testung einer Hypothese; im Vordergrund steht vielmehr die Exploration eines ausgewählten Gegenstandsbereichs. Und auch Formen der standardisierten Beobachtung oder Befragung sind nicht anwendbar: Um beispielsweise einen Fragebogen zum Alltagserleben körperlicher Behinderung zu erstellen, müsste man über die entsprechenden Beschreibungskategorien schon verfügen – diese sollen aber mittels der Untersuchung ja überhaupt erst ermittelt werden. Für solche explorativen, erkundenden Fragestellungen sind in der Regel die sog. *qualitativen Methoden* am besten geeignet.

Im Folgenden soll qualitative Forschung zunächst genauer beschrieben werden (III.1.). Daran schließt sich eine Darstellung ausgewählter Methoden der Datenerhebung (III.2.), der Analyse (III.3.) sowie einzelner komplexer Forschungsansätze an (III.4.).

III.1. Qualitative Forschung im Überblick

III.1.1. Forschungsbeispiel: Die Marienthalstudie

Die Marienthalstudie stellt eine der ‚klassischen‘ Untersuchungen innerhalb des qualitativen Ansatzes dar. Die Untersuchung wurde zu Beginn der 30er Jahre von Jahoda, Lazarsfeld und Zeisel (1933) durchgeführt, um unter natürlichen Bedingungen die Auswirkungen von Arbeitslosigkeit zu erforschen. Der Name der Studie leitet sich von dem Ort her, auf den sich die Untersuchung bezieht: Marienthal war ein österreichisches Dorf, dessen Bewohner/innen praktisch alle in einer Textilfabrik arbeiteten. Als diese Anfang der 30er Jahre Massenentlassungen vornahm, war praktisch der gesamte Ort von Arbeitslosigkeit betroffen.

Die Untersuchung erstreckte sich über mehrere Monate: Eine Projektmitarbeiterin wohnte fast zwei Monate lang selbst in Marienthal; die gesamte Arbeitsgruppe traf sich ein- bis zweimal wöchentlich dort.² Der Kontakt zu den Dorfbewohnern/innen wurde vorsichtig und allmählich hergestellt, vor allem dadurch, dass das Untersuchungsteam Aktivitäten zur Unterstützung der arbeitslosen Menschen anbot. So wurden beispielsweise diverse Kurse abgehalten (ein Schnitzzeichenkurs sowie ein Turnkurs für Mädchen), eine Frauen- und Kinderärztin hielt einmal pro Woche eine kostenfreie Sprechstunde ab; es wurde die Möglichkeit für Beratungsgespräche über Probleme der Erziehung und des häuslichen Lebens geschaffen. Die Forscher/innen verteilten in einer Kleideraktion Kleidungsstücke aus Wien an die Marienthaler

² Dass die Forscher/innen selbst am Ort der Untersuchung (im ‚Feld‘) leben, stellt ein Merkmal der sog. Deskriptiven Feldforschung dar (s. genauer unter II.4.). Die Marienthalstudie insgesamt kann jedoch nicht als typisches Beispiel für Deskriptive Feldforschung gelten – der Versuch, ganz gezielt die Lebensbedingungen der Bevölkerung zu verbessern, ist den Grundsätzen der Feldforschung gerade entgegengesetzt.

Bevölkerung; und außerdem engagierten sich die Forscher/innen (die selbst in einer sozialdemokratischen Tradition standen) politisch in den örtlichen Verbänden. Diese Aktivitäten boten vielfältige Möglichkeiten für eine allmähliche Kontaktaufnahme zwischen den Forschern/innen und der Bevölkerung; außerdem dienten die Aktivitäten auch selbst bereits der Datenerhebung. Sie waren allerdings keineswegs nur Mittel zum Zweck; vielmehr gehörte die Verbesserung der Lebensbedingungen der betroffenen Menschen von vornherein zu den Zielen der Forscher/innen.³

Das im Verlauf der Untersuchung gesammelte Datenmaterial war ausgesprochen umfangreich und divers. Es beinhaltete unter anderem:

- ausführliche Lebensgeschichten von 32 Männern und 30 Frauen
- Zeitverwendungsbögen der Tagesabläufe von 80 Personen
- Inventare von Mahlzeiten in 40 Familien über eine Woche hinweg
- Beschreibung der Weihnachtsgeschenke von 80 Kleinkindern
- Gesprächsthemen und Beschäftigungen in öffentlichen Lokalen
- Entleihzahlen in der öffentlichen Bibliothek.

Diese Materialvielfalt zeigt zugleich, dass bei der Erhebung unterschiedlichste Methoden zum Einsatz kamen: Interviews (zur Erhebung von Lebensgeschichten), freie Beobachtung (Beschäftigungen in öffentlichen Lokalen), standardisierte Beobachtung (Zeitverwendungsbögen), Zählen (Entleihzahlen) usw.

Ein wesentlicher Aspekt der Auswertung (die zu komplex war, als dass sie hier im Einzelnen dargestellt werden könnte) war die Erstellung eines induktiven Kategoriensystems (auf der Grundlage der genannten Materialien) zur Charakterisierung der Grundhaltung der Menschen in Marienthal. Es wurden vier Grundhaltungen unterschieden:

- eine ungebrochene, die durch aufrecht erhaltene Lebenslust, Aktivitäten und unverminderte Versuche der Arbeitsbeschaffung gekennzeichnet war (16%);
- eine resignierte, für die das Gefühl charakteristisch war, doch nichts gegen die Arbeitslosigkeit tun zu können (48%);
- eine verzweifelte, gekennzeichnet durch Verzweiflung, Depression und Hoffnungslosigkeit (11%);
- sowie eine apathische Grundhaltung im Sinne von Tatenlosigkeit, unbeteiligtem Zusehen, wie die Umgebung verfällt (25%).

Die Prozentzahlen zeigen, dass es nur einer Minderheit der Menschen in Marienthal gelungen war, sich eine ungebrochene Hoffnung auf die Zukunft zu erhalten; die meisten Menschen reagierten dagegen auf Dauer mit Resignation, Verzweiflung oder Apathie. Entsprechend wurde Marienthal von den Forschern/innen auch zusammenfassend als „müde Gemeinschaft“ bezeichnet. Die Untersuchung, die den Verlust von Hoffnung aufgrund von anhaltender Arbeitslosigkeit dokumentiert, besitzt auch heute durchaus noch Gültigkeit.

Prüfungsliteratur: nur Skript

II.1.2. Merkmale qualitativer Forschung

Eine genauere Beschreibung des qualitativen Ansatzes wird dadurch erschwert, dass die qualitative Forschung – anders als die quantitative – nicht durch eine einheitliche Ausrichtung

³ In dieser Hinsicht steht die Marienthalstudie auch in der Tradition der sog. Aktionsforschung. Auch für die Aktionsforschung ist sie allerdings kein typisches Beispiel, denn dort werden die angezielten Veränderungen der Lebenssituation der erforschten Personen von Forschern/innen und erforschten Personen gemeinsam bestimmt.

gekennzeichnet ist. ‚Qualitative Forschung‘ stellt vielmehr einen Sammelbegriff für eine kaum mehr überschaubare Methodenvielfalt dar. Eine Beschreibung qualitativer Forschung an Hand charakteristischer Merkmale kann daher immer nur eine idealtypische sein: Qualitative Forschung in der Praxis wird in der Regel nur einige dieser Merkmale in mehr oder minder hoher Ausprägung aufweisen, aber kaum jemals sämtliche Merkmale in voller Ausprägung.

Solche Merkmale qualitativer Forschung hat beispielsweise Mayring (2002) in Form von fünf Postulaten zusammengestellt, die er in Form von 13 Säulen qualitativen Denkens weiter spezifiziert.

Postulate qualitativer Forschung (nach Mayring, 2002)

1. Orientierung am Subjekt
Ganzheitlichkeit, Historizität, Problemorientierung
2. Deskription als Ausgangspunkt
Einzelfallbezug, Offenheit, Kontrolle
3. Interpretation
Vorverständnis, Introspektion, Forscher-Gegenstands-Interaktion
4. Alltagsnähe
5. Schrittweise Verallgemeinerung
Argumentative Verallgemeinerung, Induktion, Quantifizierung, Regelbegriff

- Ein erstes Postulat stellt die *Orientierung am Subjekt* dar. Damit ist gemeint, dass qualitative Forschung immer am Subjekt, am Menschen selbst ansetzt.
 - Dazu gehört vor allem, dass der Mensch in seiner *Ganzheit* betrachtet wird; es interessieren also nicht einzelne Merkmale (Variablen) der Person, sondern der gesamte Mensch. In der Marienthalstudie wird dieses Prinzip daran deutlich, dass die Gesamtheit der Reaktionen der betroffenen Menschen auf die Arbeitslosigkeit untersucht wurden (Denken, Fühlen, Handeln), nicht nur einzelne Aspekte davon (z.B. Anzahl der unternommenen Versuche zur Arbeitsbeschaffung).
 - Dazu gehört auch seine Lebensgeschichte (*Historizität*), also die Art und Weise, wie der Mensch zu dem geworden ist, was er ist. In der Marienthalstudie manifestiert sich die Historizität des Vorgehens darin, dass die Auswirkungen der Arbeitslosigkeit im zeitlichen Prozess untersucht wurden, nämlich über mehrere Monate hinweg. So konnten beispielsweise auch Veränderungen der Reaktionen in dieser Zeit erfasst werden.
 - Als dritte Spezifikation dieses ersten Postulats nennt Mayring die *Problemorientierung*: dass qualitative Forschung an den praktischen Problemen des Menschen ansetzt, also stets praxis- und lösungsorientiert ist. In der Marienthalstudie zählt die Problemorientierung im Sinne einer Verbesserung der Lebensbedingungen zu den ausdrücklichen Zielen der Forscher/innen und wurde im Rahmen der verschiedenen Aktionen (Kleideraktion, Kurse usw.) auch konkret im forschungspraktischen Handeln umgesetzt. Zur Problemorientierung finden sich in der qualitativen Forschung insgesamt jedoch durchaus unterschiedliche Standpunkte: So gibt es qualitative Forschungsansätze, die gerade durch ihre Praxis- und Handlungsorientierung gekennzeichnet sind (z.B. die sog. Aktionsforschung); es existiert aber ebenso qualitative Grundlagenforschung, für die ein Praxisbezug nicht charakteristisch ist (z.B. Deskriptive Feldforschung; Ethnomethodologie).

- Mayrings zweites Postulat ist das der *Deskription*: „Am Anfang der Analyse muss eine genaue und umfassende Beschreibung (Deskription) des Gegenstandsbereiches stehen.“ (2002, S. 21).
 - Das bedeutet zunächst, dass qualitative Forschung in der Regel am Einzelfall ansetzt (*Einzelfallbezogenheit*). Dieser Einzelfall kann, je nach Fragestellung, eine Person, eine Gruppe, eine Gemeinschaft o.ä. sein (in Marienthal z.B. die Dorfgemeinschaft, auch die einzelnen Familien). Die Orientierung am Einzelfall bedeutet nicht, dass ein Vergleich von Einzelfällen ausgeschlossen ist. Ein solcher Vergleich stellt aber einen späteren (zudem fakultativen) Schritt im qualitativen Forschungsprozess dar, der auf der Einzelfallbeschreibung aufbaut.
 - Das Postulat der Deskription beinhaltet weiterhin eine grundsätzliche *Offenheit* gegenüber dem Forschungsgegenstand. Damit ist gemeint, dass qualitative Forschung nicht primär theoriegeleitet-deduktiv vorgeht, sondern induktiv – Kategorien zur Materialbeschreibung werden also beispielsweise nicht in erster Linie aus theoretischen Vorannahmen abgeleitet, sondern ergeben sich aus dem Material selbst. So stellt das Kategoriensystem zur Beschreibung der Grundhaltungen der Marienthaler Bevölkerung das Ergebnis einer induktiven Analyse des erhobenen Materials dar.
 - Diese Offenheit darf Mayring zu Folge allerdings nicht dazu führen, dass qualitative Forschung in unsystematischer Weise durchgeführt wird. Mayring fordert vielmehr auch bei qualitativer Forschung im Rahmen der Deskription ein gewisses Maß an *Methodenkontrolle*, beispielsweise in Form einer genauen Dokumentation der Verfahrensschritte. Dabei handelt es sich allerdings um ein (in der Tat durchaus wünschenswertes) Zielkriterium, das jedoch nicht als charakteristisch für qualitative Forschung im Allgemeinen gelten kann. Speziell die Marienthalstudie ist allerdings in ausführlicher Form dokumentiert.
- Als drittes Postulat nennt Mayring die *Interpretation*. Damit ist gemeint, dass der Gegenstand qualitativer Forschung nicht unmittelbar zugänglich ist, sondern erst durch Interpretation erschlossen werden muss.⁴
 - Diese Interpretationsbedürftigkeit des Forschungsgegenstandes manifestiert sich erstens darin, dass Forscher/innen einem Gegenstand nie gänzlich unvoreingenommen gegenüber treten, sondern immer schon mit bestimmten Vorannahmen, mit einem sog. *Vorverständnis*. Dieses Vorverständnis (das sozusagen das methodologische ‚Gegenstück‘ zur Offenheit darstellt) beeinflusst notwendiger Weise die Sichtweise und Interpretation des Gegenstandes durch die Forscher/innen, ist aber zugleich unvermeidbar (In der heutigen Wissenschaftstheorie herrscht Einigkeit darüber, dass eine theoriefreie Forschung gänzlich ohne alle Vorannahmen nicht möglich ist). Um die je eigene Interpretation des Gegenstandes auch für andere nachvollziehbar zu machen, ist es daher erforderlich, dass qualitative Forscher/innen ihr jeweiliges Vorverständnis des Forschungsgegenstandes explizit machen.
 - Interpretation im Rahmen qualitativer Forschung vollzieht sich nach Mayring immer auch in Form der Beobachtung eigener Gedanken und Reaktionen auf den Gegenstand, d.h. in Form von *Introspektion*. Wie das Vorverständnis sind auch Prozesse der Introspektion möglichst offen zu legen.
 - Qualitative Forschung vollzieht sich in Interaktion zwischen den Forschern/innen und dem Forschungsgegenstand, meist anderen Menschen. Insofern entstehen Interpretationen des Gegenstandes in der Regel im Rahmen solcher *Forscher-Gegenstands-Interaktionen*. Qualitative Forschung wird somit wesentlich als sozialer Prozess konzipiert, in dessen Verlauf sich sowohl der/die Forscher/in als auch die erforschte Per-

⁴ Aufgrund der Kürze der Darstellung der Marienthalstudie können die Spezifikationen dieses Postulats nicht anhand der Studie veranschaulicht werden.

son verändern können (dies hat Konsequenzen für die qualitative Sichtweise des quantitativen Gütekriteriums der Reliabilität; s. Gütekriterien). Die erforschte Person kann im Forschungsprozess beispielsweise zum Nachdenken über ihre gegenwärtige Situation angeregt werden. Forscher/innen reagieren als Menschen auf ihr Gegenüber, entwickeln Sympathien und Antipathien, die notwendigerweise in die Interaktionen mit der erforschten Person einfließen und so auch wiederum den Gegenstand und das Gegenstandsverständnis bzw. die Interpretation seitens der Forscher/innen beeinflussen. Die Reaktionen von Forscher/innen auf ihren Gegenstand stellen daher eine wichtige Erkenntnisquelle im Rahmen qualitativer Forschung dar.

- Ein viertes Postulat qualitativer Forschung stellt die *Alltagsnähe* dar. Damit verbindet sich die Forderung, Gegenstände qualitativer Forschung möglichst in ihrem natürlichen Kontext und Umfeld zu untersuchen, damit der Gegenstand durch den Forschungsprozess möglichst wenig verzerrt wird. In der Marienthalstudie ist diese Alltagsnähe dadurch gegeben, dass die Untersuchung vor Ort, in der betroffenen Gemeinde selbst stattfindet. Dieses Postulat, das Mayring in unmittelbarer Nähe zum ersten Postulat der Orientierung am Subjekt sieht, wird nicht weiter in Unterpunkte spezifiziert.
- Das fünfte und letzte Postulat bezieht sich schließlich auf die *Verallgemeinerbarkeit* qualitativer Forschung.³ Im Rahmen des Postulats der Deskription wurde bereits auf den charakteristischen Einzelfallbezug qualitativer Forschung hingewiesen; außerdem basiert qualitative Forschung in der Regel auf kleinen Stichproben. Die Verallgemeinerbarkeit qualitativer Ergebnisse ist daher ggf. schrittweise zu begründen. Auch hier handelt es sich wieder um ein methodologisches Zielkriterium (ähnlich wie z.B. die Methodenkontrolle), das jedoch nicht als charakteristisch für qualitative Forschung generell gelten kann. Als Verfahren zur Begründung der Verallgemeinerbarkeit von Ergebnissen qualitativer Forschung nennt Mayring:
 - *argumentative Verallgemeinerung*,
 - *Induktion*
 - sowie *Quantifizierung*. Eine Verallgemeinerung qualitativer Ergebnisse durch Quantifizierung stellt zugleich eine Kombination qualitativer und quantitativer Vorgehensweisen dar. In der Marienthalstudie werden beispielsweise sowohl qualitative als auch quantitative Verfahren der Datenerhebung verwendet, und mittels des Kategoriensystems zur Unterscheidung verschiedener Grundhaltungen wird eine Quantifizierung der Verbreitung dieser verschiedenen Haltungen vorgenommen. Es gibt aber durchaus auch Vertreter/innen eines qualitativen Ansatzes, die einen Rückgriff auf Quantifizierung im Rahmen qualitativer Forschung strikt ablehnen.
 - Bei der Verallgemeinerung geht es in der qualitativen Forschung nicht um die Aufstellung von allgemeingültigen Gesetzen; Gleichförmigkeiten (z.B. über verschiedene Fälle hinweg) lassen sich Mayring zu Folge besser durch den Begriff der *Regel* abbilden.

Prüfungsliteratur: Mayring 2002, Kap. 2

III.1.3. Qualitative und quantitative Forschung im Vergleich

Qualitativ	Quantitativ
geisteswissenschaftlich	naturwissenschaftlich
Verstehen	Erklären
Erleben	Verhalten
idiographisch	nomothetisch
Feld	Labor
induktiv	deduktiv
zyklisch	linear
fallorientiert	variablenorientiert

Die Spezifik qualitativer Forschung wird auch noch einmal deutlicher, wenn man die Merkmale der Forschung im qualitativen und im quantitativen Ansatz einander gegenüberstellt. Dabei handelt es sich allerdings wiederum um einen idealtypischen Vergleich; in der Praxis besteht (wie auch die Marienthalstudie zeigt) nicht notwendig ein solcher Gegensatz zwischen qualitativem und quantitativem Vorgehen.

- Wie eingangs bereits erläutert, ist die qualitative Forschung in der Tradition von Dilthey eher einem *geisteswissenschaftlich*-dualistischen, die quantitative Forschung dagegen eher einem *naturwissenschaftlich*-monistischen Ansatz verpflichtet.
- Mit diesen beiden Ansätzen verbinden sich zugleich unterschiedliche Zielsetzungen: Geisteswissenschaftliche Ansätze sind eher darauf ausgerichtet, den Forschungsgegenstand zu *verstehen*. Für den naturwissenschaftlichen Ansatz ist dagegen die Zielsetzung des *Erklärens* (in Form einer Überprüfung von Kausalhypothesen) charakteristisch.
- Entsprechend diesen unterschiedlichen Zielsetzungen stehen in den beiden Ansätzen auch je unterschiedliche Gegenstandsbereiche im Vordergrund: Gegenstand der qualitativen Forschung ist das menschliche *Erleben*, während die quantitative Forschung eher auf das *Verhalten* ausgerichtet ist.
- Die Vorgehensweise innerhalb des qualitativen Ansatzes lässt sich als *idiographisch* bzw. einzelfallorientiert charakterisieren. Die Vorgehensweise in der quantitativen Forschung ist dagegen meist *nomothetisch* orientiert; im Vordergrund stehen nicht Einzelfälle, sondern über die Einzelfälle hinweg zusammengefasste („aggregierte“) Daten (z.B. Gruppenvergleiche).
- Während qualitative Forschung meist in der natürlichen Umgebung des interessierenden Gegenstandsbereichs stattfindet, also im *Feld*, ist der bevorzugte Ort quantitativer Forschung das *Labor*. Diese unterschiedlichen Präferenzen sind das Ergebnis je verschiedener Gegenstandsauffassungen: In der quantitativen Forschung wird versucht, variable Merkmale des Umfelds im Sinne einer Kontrolle von Störvariablen auszuschalten; Kontextmerkmale werden also nicht als Teil des Gegenstandes aufgefasst. In der qualitativen Forschung wird dagegen davon ausgegangen, dass der Gegenstand ohne Berücksichtigung seines Kontextes nicht adäquat erfasst werden kann.
- Qualitative Forschung ist *induktiv* orientiert; das Vorgehen verläuft von den Daten, vom Einzelnen, hin zum Allgemeinen. Quantitative Forschung ist dagegen eher *deduktiv* ausgerichtet; die Argumentationsrichtung verläuft vom Allgemeinen zum Besonderen (z.B. bei der Ableitung von Hypothesen aus Theorien).
- Der quantitative Forschungsprozess ist streng *linear* angelegt und verläuft von der Hypothesenableitung über die Datenerhebung und –analyse hin zum Schluss hinsichtlich der (vorläufigen) Bewährung oder Falsifikation der Hypothese; insbesondere sind Datenerhebung und –auswertung klar getrennt. Der qualitative Forschungsprozess kann dagegen auch *zyklisch* verlaufen (genauer: spiralenförmig): Erste Daten werden sofort ausgewertet;

es resultiert eine Hypothese, aufgrund derer erneut Daten erhoben und ausgewertet werden, was wiederum zu einer Modifikation der Hypothese führt; dieser Kreislauf wird so lange fortgesetzt, bis eine erneute Datenerhebung keine weitere Modifikation der Hypothese mehr nach sich zieht. Das vertiefte Gegenstandsverständnis, wie es aus der wiederholten Datenerhebung resultiert, kann zur Folge haben, dass anfängliche Auswertungen der zuerst erhobenen Daten in der Folge modifiziert werden. Datenerhebung und –auswertung greifen bei der qualitativen Forschung also häufig ineinander.

- Während quantitative Forschung *variablenorientiert* ist (Fälle werden hinsichtlich ihrer Ausprägung auf ausgewählten Variablen miteinander verglichen), ist qualitative Forschung als *fallorientiert* zu charakterisieren (s. oben II.2. das Postulat des Ansetzens am Subjekt).
- In der qualitativen Forschung wird der Mensch tendenziell eher als selbstbestimmt gesehen (z.B. über innere Gründe zum Handeln motiviert), in der quantitativen Forschung eher als fremdbestimmt (Zielsetzung der Erklärung menschlichen Verhaltens).

Prüfungsliteratur: Bortz & Döring 2002, Kap. 5.1.

III.1.4. Arten qualitativer Forschung

Wie eingangs bereits erwähnt, existiert eine kaum überschaubare Vielzahl qualitativer Methoden. Dies ist nicht zuletzt darauf zurückzuführen, dass qualitative Forscher/innen danach streben, Methoden zu finden, die ihrem jeweiligen Gegenstand möglichst optimal gerecht werden; entsprechend existieren fast so viele qualitative Methoden wie Gegenstände. Allerdings wird eine Systematisierung des Methodenspektrums dadurch natürlich auch erschwert. Im Folgenden werden drei Systematisierungsvarianten dargestellt. Diese sind allerdings nicht erschöpfend; es sind also durchaus noch andere Kriterien denkbar, nach denen zwischen verschiedenen Arten qualitativer Methoden unterschieden werden kann.

Unterscheidung nach Untersuchungsphasen

Eine erste Unterscheidung zwischen verschiedenen Arten qualitativer Methoden greift auf denselben Gesichtspunkt zurück, der bereits bei der Differenzierung verschiedener Arten von quantitativen Methoden zu Grunde gelegt wurde, nämlich die *Forschungsphase*, in der die Methode zur Anwendung kommt. Entsprechend lassen sich auch im qualitativen Bereich unterscheiden:

- *Methoden der Datenerhebung*, z.B. Interview, Gruppendiskussion, freie Beobachtung usw.
- *Methoden der Datenanalyse*, z.B. Hermeneutik, Kodierung, Inhaltsanalyse usw.
- *Komplexe Forschungsansätze* (Untersuchungsdesigns), z.B. Deskriptive Feldforschung, Gegenstandsbezogene Theorienbildung usw.

Unterscheidung zwischen qualitativen Methoden und qualitativer Forschung

Unter II.1.3. wurde eine zyklische Vorgehensweise, bei der Datenerhebung und –auswertung ineinandergreifen, als charakteristisch für qualitative Forschung bezeichnet. Genau genommen wird eine solche zyklische Vorgehensweise allerdings nur im Rahmen der o.g. komplexen Forschungsansätze (etwa bei der Gegenstandsbezogenen Theorienbildung) realisiert. Qualitative Methoden der Datenerhebung und –analyse sind demgegenüber nicht an einen bestimmten Forschungsprozess gebunden: Interviews oder Inhaltsanalysen können beispielsweise auch im Rahmen einer hypothesenprüfenden, linear angelegten Untersuchung zur Anwendung kommen. Entsprechend ergibt sich als zweites Kriterium der Unterscheidung zwischen verschiedenen Verfahren im qualitativen Bereich die Differenzierung zwischen *qualita-*

tiven Methoden (die sowohl in zyklischen als auch in linearen Forschungsprozessen eingesetzt werden können) und *qualitativer Forschung* im engeren Sinne, die vom Forschungsansatz her zyklisch konzipiert ist.

Unterscheidung nach Gegenstandsbereich und Zielsetzung

Weiterhin ist in der Literatur zur qualitativen Forschung vielfach versucht worden, qualitative Verfahren nach inhaltlichen Gesichtspunkten zu unterteilen. Eine verbindliche, allgemein akzeptierte Unterscheidung existiert hier nicht. Es soll daher hier lediglich beispielhaft eine solche Differenzierung vorgestellt werden (nach Lamnek, 1995, Kap. 2.3.):

- *Nachvollzug/Rekonstruktion subjektiv gemeinten Sinns*: Ziel dieser qualitativen Verfahren ist es, die Sichtweise des Gegenüber zu verstehen. Diese Zielsetzung liegt beispielsweise vielen Interviewverfahren zugrunde.
- *Deskription sozialen Handelns*: Bei Verfahren dieses Typs steht die Frage im Vordergrund, wie Menschen in Interaktion mit anderen soziale Realität konstruieren. Die historischen Wurzeln dieser Verfahren liegen in der sog. Chicagoer Schule. Hier sind beispielsweise die ethnografische Forschung, bestimmte Formen der Gruppendiskussion oder auch die Durchführung narrativer Interviews mit dem Ziel der Erschließung allgemeintypischer „Prozessstrukturen des Lebenslaufs“ einzuordnen.
- *Rekonstruktion von Strukturen*: Bei einer dritten Gruppe von Verfahren geht es darum, die impliziten Regeln zu erschließen, die menschlichen Handlungen zugrunde liegen (Tiefenstrukturen). Die erforschten Personen fungieren hier quasi nur als ‚Regelträger‘. Während solchen Verfahren, die der Rekonstruktion subjektiv gemeinten Sinns dienen, die Annahme zugrunde liegt, dass Menschen in der Lage sind, ihre eigenen Gedanken und Handlungsgründe zu erkennen und im Interview wiederzugeben, sind Verfahren dieses Typs durch die Annahme gekennzeichnet, dass Menschen gerade nicht über Einsicht in handlungsleitende Strukturen verfügen; diese müssen vielmehr erst von den Forschern/innen erschlossen werden. Dieser dritten Gruppe von Verfahren sind beispielsweise die Objektive Hermeneutik oder die Konversationsanalyse zuzuordnen.

Prüfungsliteratur: nur Skript

III.2. Qualitative Verfahren der Datenerhebung

Die wichtigsten Verfahren der Datenerhebung im qualitativen Bereich sind Formen der *Befragung* und der *Beobachtung*. Während in der quantitativ orientierten Forschung meist mit standardisierten Instrumenten gearbeitet wird (z.B. Fragebogen mit geschlossenen Antworten, Beobachtungsschema), dominieren im qualitativen Bereich freie, weniger bis gar nicht standardisierte Verfahren. Während Befragungen in der quantitativen Forschung häufiger in schriftlicher Form stattfinden (also per Fragebogen), ist für die qualitative Forschung eher die mündliche Befragung charakteristisch. Im Folgenden wird in erster Linie auf Verfahren der qualitativen mündlichen Befragung eingegangen (Formen des Interviews und Gruppendiskussion). Die qualitative Beobachtung wird dagegen vor allem im Zusammenhang mit dem Forschungsansatz der Deskriptiven Feldforschung thematisiert (II.4.), wo sie als zentrale, geradezu definierende Methode gelten kann.

II.2.1. Qualitative und quantitative Befragung im Vergleich

Qualitative Befragung	Quantitative Befragung
teil- oder nonstandardisiert	standardisiert
eher mündlich	eher schriftlich
offene (freie Antworten)	geschlossen
Antworten in der Begrifflichkeit der befragten Person	Antworten in der Begrifflichkeit des Forschers bzw. der Forscherin
Forscher/in geht individuell auf befragte Person ein	gleichförmiges Verhalten von Forscher/in
Berücksichtigung der Eindrücke des Forschers / der Forscherin als Erkenntnisquelle	

In der obigen Tabelle sind noch einmal die zentralen Unterschiede zwischen der Befragung in der quantitativen und der qualitativen Forschung zusammenfassend dargestellt. Die dort aufgeführten Merkmale der qualitativen Befragung gelten tendenziell für verschiedene Formen der Befragung, insbesondere jedoch für das Interview.

- Grad der *Standardisierung / Strukturiertheit*: In der quantitativen Forschung sind sowohl der Wortlaut als auch die Reihenfolge der Fragen fest vorgegeben. Bei der qualitativen Forschung kann der Wortlaut der Fragen an das Gegenüber angepasst werden; die Reihenfolge ist flexibel bzw. richtet sich nach der befragten Person. Im nonstandardisierten Interview sind allenfalls Themenkomplexe, aber keine konkreten Fragen vorformuliert. Die qualitative Befragung ist hier also flexibler als die quantitative; zugleich sind Objektivität, Reliabilität sowie Vergleichbarkeit verschiedener Interviews im qualitativen Bereich jedoch deutlich eingeschränkt.
- Wie bereits erwähnt, findet die qualitative Befragung eher *mündlich*, die quantitative eher schriftlich statt. Es existieren aber durchaus auch standardisierte Formen des Interviews, die in der qualitativen Forschung Anwendung finden, wie auch nonstandardisierte, offene Formen der schriftlichen Befragung (wenn befragte Personen z.B. gebeten werden, ihre persönliche ‚Lektürebiographie‘ aufzuschreiben: wie sie zum Lesen gekommen sind, welche Bücher sie in der Kindheit vielleicht besonders beeindruckt haben usw.). Auf diese untypischeren Befragungsformen wird im Folgenden jedoch nicht eingegangen.
- In der qualitativen Befragung dominieren *offene Fragen* ohne vorgegebene Antwortmöglichkeiten, auf die die befragte Person entsprechend auch in ihrer eigenen Begrifflichkeit antworten kann. Für die quantitative Forschung sind dagegen eher geschlossene Fragen mit vorgegebenen Antwortmöglichkeiten die Regel.
- Bei der Befragung innerhalb des quantitativen Ansatzes wird versucht, die Befragungssituation für alle Untersuchungsteilnehmer/innen möglichst vergleichbar zu gestalten; entsprechend soll auch das Verhalten des Forschers bzw. der Forscherin verschiedenen Personen gegenüber möglichst gleichförmig sein. Im qualitativen Ansatz wird dagegen davon ausgegangen, dass ein/e Untersuchungsteilnehmer/in nur dann bereit ist, sich im Interview zu öffnen und (gerade auch persönliche) Fragen wahrheitsgemäß zu beantworten, wenn er oder sie sich im Gespräch wohlfühlt und Vertrauen zum/r Forscher/in fasst. Die Bedingungen für ein solches vertrauensvolles Gespräch sind jedoch von Person zu Person verschieden. Entsprechend besteht die Zielvorstellung in der qualitativen Forschung darin, dass Forscher/innen sich von außen gesehen verschiedenen befragten Person gegenüber

ggf. je anders verhalten, um auf diese Weise eine *innere Vergleichbarkeit* der verschiedenen Interviewsituationen zu erzielen.

- Bei der Darstellung von qualitativer und quantitativer Forschung im Vergleich wurde bereits darauf hingewiesen, dass die Reaktionen des Forschers oder der Forscherin selbst gegenüber der befragten Person innerhalb des qualitativen Ansatzes ebenfalls als Erkenntnisquelle genutzt werden. So kann es beispielsweise die Validität des Interviews beeinträchtigen, wenn ein/e Forscher/in sich von den Äußerungen der befragten Person unangenehm berührt fühlt. Innerhalb des quantitativen Ansatzes werden solche Eindrücke des Forschers oder der Forscherin bei der Auswertung der Interviews nicht weiter berücksichtigt.

Ungeachtet dieser Unterschiede zwischen qualitativer und quantitativer Befragung gelten die Richtlinien für die Frageformulierung, wie sie im Zusammenhang mit der quantitativen Befragung bereits aufgeführt wurden, auch für das qualitative Interview. In der Tat kommt der Verständlichkeit der Fragen bei der mündlichen Befragung sogar eine erhöhte Bedeutung zu: Eine schriftlich vorgelegte Frage können Untersuchungsteilnehmer/innen ggf. noch einmal lesen, wenn sie sie nicht sofort verstanden haben. In der Interviewsituation werden die Untersuchungsteilnehmer/innen dagegen vielleicht zwei- oder dreimal nachfragen, wenn sie etwas nicht verstanden haben (wenn sie überhaupt nachfragen); danach geben sie bei Nicht-Verstehen vermutlich eher vage, ausweichende Antworten. Entsprechend ist im Interview ganz besonders darauf zu achten, kurze, eindeutige Fragen zu stellen, die keine doppelten Verneinungen enthalten und sprachlich der befragten Person angepasst sind.

Ein besonderes Problem beim Interview stellen Suggestivfragen dar, in denen der/die Forscherin mehr oder weniger deutlich eigene Bewertungen und Präferenzen zum Ausdruck bringt und auf diese Weise der befragten Person eine bestimmte Antwortrichtung nahe legt (z.B.: „Aber heute haben Sie dieses Alkoholproblem doch sicher überwunden?“). Interviewfragen sollten möglichst neutral formuliert sein, so dass die Untersuchungsteilnehmer/innen sich in ihren Antworten frei fühlen.

Weiterhin ist darauf zu achten, dass Untersuchungsteilnehmer/innen im Interview nicht überfordert werden. Nach einer Stunde lässt die Konzentration in der Regel schon etwas, nach anderthalb Stunden deutlich nach. Das gilt im Übrigen nicht nur für die Konzentration der befragten Personen, sondern ebenfalls für die der Forscher/innen.

Interviews werden in der Regel auf Kassette aufgezeichnet. Dazu ist natürlich das Einverständnis der befragten Personen einzuholen. Außerdem sollte der/die Forscherin sich rechtzeitig mit der Technik vertraut machen.

Was für qualitative Verfahren insgesamt gilt, trifft ebenfalls auf das Interview zu: Es existieren viele verschiedene Formen des Interviews: Tiefeninterview, Leitfadeninterview, narratives Interview, offenes Interview, episodisches Interview, lebensgeschichtliches Interview usw. Im Folgenden soll auf zwei Formen des Interviews genauer eingegangen werden: das halbstandardisierte Leitfadeninterview und das nonstandardisierte offene Interview.

Prüfungsliteratur: nur Skript

II.2.2. Das halbstandardisierte Interview (Leitfadeninterview)

Überblick

Eine erste Form des qualitativen Interviews ist das sog. Leitfadeninterview. Zentrales Merkmal des Leitfadeninterviews ist es, dass vor Beginn der Untersuchung zunächst auf theoretischer Grundlage relevante Aspekte des Gegenstands identifiziert werden. In einer Untersuchung zur Wertorientierung Jugendlicher wären beispielsweise die Bedeutung von Erfolg/Karriere, Familie/Freunden, Genussorientierung usw. solche theoretisch vorab identifizierbaren Aspekte. Zur Erfassung dieser Aspekte werden Fragen vorformuliert und in einem Interviewleitfaden zusammengestellt. Dadurch ist zum einen sichergestellt, dass alle relevanten Aspekte im Verlauf des Interviews auch tatsächlich angesprochen werden. Zum anderen ist durch den Leitfaden eine grundsätzliche Vergleichbarkeit der Interviews gegeben; es ist also im Rahmen der Auswertung möglich, zu vergleichen, was verschiedene Personen oder auch Personengruppen auf eine bestimmte Frage geantwortet haben. Ein solcher Leitfaden ist vor Beginn der eigentlichen Untersuchung auf jeden Fall zu erproben. Wenn sich im Verlauf dieser Probephase herausstellen sollte, dass wichtige Aspekte des Gegenstandsbereichs im Leitfaden fehlen, können diese vor Beginn der Untersuchung noch ergänzt werden. Daran wird zugleich deutlich, dass das Leitfadeninterview häufig im Rahmen eines eher linearen Forschungsprozesses eingesetzt wird.

Merkmale des Leitfadeninterviews

Beim Leitfadeninterview handelt es sich um ein halbstandardisiertes Interview. Der Wortlaut der Fragen ist entsprechend variabel, und die Reihenfolge der Fragen kann dem Gesprächsverlauf angepasst werden. Wenn ein Untersuchungsteilnehmer beispielsweise auf die Frage, wie wichtig die Karriere für ihn ist, ausführt, dass die Karriere für ihn im Vergleich zur Familie nur wenig Relevanz besitzt – und dann spontan ausführlich auf die Bedeutung der Familie eingeht, dann haben sich damit einige Fragen zur Bedeutung von Familie für diese Person bereits erübrigt, brauchen also nicht mehr gestellt zu werden. Andere vertiefende Fragen in diesem Zusammenhang, die erst für einen späteren Zeitpunkt des Interviews vorgesehen waren, wären dagegen vorzuziehen. Aus dieser Flexibilität des Leitfadeninterviews ergeben sich hohe Anforderungen an die Konzentration des/der Forscher/in. Eine Interviewerschulung ist daher unbedingt erforderlich. Sie findet meist, gemeinsam mit der Testung des Leitfadens selbst, in der Probephase statt.

Ablauf

Damit ergibt sich zusammenfassend der folgende Ablauf für eine Untersuchung unter Anwendung des Leitfadeninterviews: Auf theoretischer Grundlage werden relevante Merkmale des interessierenden Gegenstandes identifiziert und es wird auf dieser Grundlage ein erster Entwurf eines Leitfadens formuliert. Dieser Leitfaden wird in einer Pilotphase erprobt und ggf. modifiziert; die Pilotphase dient zugleich der Schulung der Interviewer/innen. An die Pilotphase schließt sich die eigentliche (Haupt-)Untersuchung an, die Interviewphase. Im weiteren Verlauf sind die auf Kassette aufgezeichneten Interviews zu verschriftlichen (transkribieren) und auszuwerten (s. dazu unten II.2.).

Formen des Leitfadeninterviews

Auch der Begriff des Leitfadeninterviews stellt seinerseits einen Sammelbegriff für verschiedene konkrete Interviewarten dar. Zu den wichtigsten Formen des Leitfadeninterviews zählen das *problemzentrierte Interview* (das an einem konkreten gesellschaftlichen Problem ansetzt; entwickelt von Witzel) und das *fokussierte Interview* (entwickelt von Merton & Kendall), bei dem es darum geht, die Reaktionen der Untersuchungsteilnehmer/innen auf ein ‚fokussiertes Objekt‘ zu ermitteln (z.B. einen Spielfilm). Der Begriff des Leitfadeninterviews wird aber

auch ganz generell zur Bezeichnung verschiedenster Formen halbstandardisierter Interviews herangezogen, die auf einem Leitfaden basieren. Die verschiedenen Arten des Leitfadeninterviews unterscheiden sich in erster Linie im Hinblick auf die Art der Fragen, die in dem Interview vorgesehen sind.

Fragen im problemzentrierten Interview

Das problemzentrierte Interview enthält beispielsweise die folgenden Arten von Fragen:

- *Sondierungsfragen*: Diese werden zu Beginn des Interviews gestellt. Mit den Fragen soll die Bedeutung des interessierenden Problems für die befragte Person erfasst werden.
- *Leitfadenfragen*: Diese Fragen machen den eigentlichen, theoriegeleitet erstellten Leitfaden aus. Dieser Punkt stellt zugleich den zentralen Unterschied zwischen dem problemzentrierten Interview (und anderen Formen des Leitfadeninterviews) einerseits und nonstandardisierten Formen des Interviews andererseits (s.u. II.2.1.3. dar).
- *Ad hoc-Fragen*: Ad hoc-Fragen dienen dazu, interessante Aspekte des Gegenstandes, die von der befragten Person im Interview thematisiert werden, im Leitfaden jedoch nicht vorgesehen sind, vertiefend zu explorieren. In diesem Punkt unterscheidet sich das problemzentrierte Interview wesentlich von Formen des standardisierten Interviews, in denen solche zusätzlichen, spontanen Fragen nicht zulässig sind.

Anwendungsbereiche

Da das Leitfadeninterview auf theoriegeleiteten Vorarbeiten aufbaut, ist es in erster Linie für solche Fragestellungen geeignet, die bereits in einem gewissen Maß psychologisch erforscht und aufgearbeitet sind. Da das Leitfadeninterview außerdem einen Vergleich verschiedener Personen erlaubt, ist es außerdem dann gut einsetzbar, wenn das Forschungsinteresse sich nicht auf die Beschreibung von Einzelfällen beschränkt, sondern darüber hinaus auch ein Fallvergleich intendiert ist.

Prüfungsliteratur: Bortz & Döring 2002, Kap. 5.2.1.; Mayring 2002, Kap. 4.2.1.

II.2.3. Das nonstandardisierte Interview (Schwerpunkt: Narratives Interview)

Überblick

Das nonstandardisierte Interview ist diejenige Form des Interviews, die sowohl den Forschern/innen als auch den befragten Personen den größten Spielraum lässt. Anders als beim Leitfadeninterview werden keine Fragen vorformuliert; allenfalls werden im Vorfeld relevante Themenbereiche festgelegt, zu denen – falls die befragte Person sie nicht spontan thematisiert – im Interviewverlauf Fragen gestellt werden. Bei einer Untersuchung zum Thema ‚In Rente gehen‘ könnte beispielsweise der Gesichtspunkt, ob die Pensionierung von einem Tag auf den anderen oder allmählich erfolgte, einen solchen Themenbereich darstellen; andere relevante Themenbereiche in diesem Zusammenhang wären etwa: Art der Arbeit, Freude an der Arbeit, Hobbies, soziales Umfeld usw. Im Wesentlichen wird der Interviewverlauf jedoch von der befragten Person gesteuert und weniger von der Person des Forschers oder der Forscherin.

Formen des nonstandardisierten Interviews

Auch beim nonstandardisierten Interview handelt es sich wieder um einen Sammelbegriff für verschiedene konkrete Interviewformen. Zu den bekanntesten Formen des nonstandardisierten Interviews zählen das *Tiefeninterview* (auch *Intensivinterview* genannt), das häufig mit dem Ziel verbunden ist, unbewusste Prozesse und Motive der befragten Personen aufzuklären, sowie das *narrative Interview*, das dazu dient, Erzählungen der gesamten Lebensgeschichte oder

auch einzelner relevanter Episoden aus dem Leben der befragten Personen zu rekonstruieren. Sowohl beim Tiefen- als auch beim narrativen Interview steht oft nicht die Rekonstruktion des subjektiv gemeinten, sondern eines ‚objektiven Sinns‘ im Mittelpunkt des Forschungsinteresses. Darüber hinaus existieren wiederum unspezifische Formen des nonstandardisierten, offenen Interviews, die auch durchaus auf den Nachvollzug subjektiv gemeinten Sinns abzielen können. Im Folgenden soll als Beispiel genauer auf das narrative Interview eingegangen werden.

Das narrative Interview: Ablauf

Das narrative Interview (entwickelt von Schütze) basiert auf der Annahme, dass Erzählungen eine inhärente, quasi natürliche Struktur aufweisen, die sich nur dann entfalten kann, wenn man einen Menschen die entsprechende Geschichte erzählen lässt. Nachfragen, so wird angenommen, stören diesen Erzählfluss eher. Entsprechend wird das narrative Interview stark von der befragten Person gesteuert (s. auch oben).

- Das Interview beginnt mit einem sog. *Erzählanstoss*, einer Aufforderung zum Erzählen (z.B.: „Wie war das denn damals, als Sie in Rente gegangen sind?“).
- Daran schließt sich die Phase der *Haupterzählung* an: Hier erzählt die befragte Person, wobei der Forscher oder die Forscherin sich weitgehend auf ‚Zuhörersignale‘ beschränkt („mhm“ usw.).
- In der folgenden *Nachfragephase* hat der/die Forscher/in Gelegenheit, eventuelle Verständnisfragen zu stellen. In dieser Phase des Interviews können außerdem vorab festgelegte Gegenstandsbereiche thematisiert werden, die die befragte Person zuvor in ihrer Erzählung nicht spontan thematisiert hat.
- In der *Bilanzierungsphase* bittet die Forscherin oder der Forscher die befragte Person um eine abschließende Bewertung - z.B. dahingehend, wie sie die damaligen Ereignisse und ihre Handlungen heute sieht, ob sie im Nachhinein anders handeln würde usw. Eine solche Bilanzierungsphase ist aber kein notwendiger Bestandteil des narrativen Interviews.

Annahmen über die Struktur von Erzählungen

Die inhärente Struktur von Erzählungen / Geschichten, die dem narrativen Interview als Methode zugrunde liegt, umfasst die folgenden Bestandteile:

- *Abstract bzw. Einführender Überblick*: Hier benennt die befragte Person den Gegenstand, das Thema ihrer Erzählung. Im Forschungsprozess kann diese Benennung auch durch den Forscher oder die Forscherin in Form des Erzählanstoßes erfolgen.
- *Orientierung*: Dies ist die Hinführung auf die eigentliche Geschichte; in diese Phase fallen beispielsweise die Erläuterung von Zeit und Ort oder auch von anderen Personen, die an der Geschichte beteiligt waren oder sind.
- *Komplikation*: Die Komplikation bildet quasi den Kern der Geschichte – das, was passiert ist bzw. das, was erzählt werden soll, erzählenswert ist.
- *Evaluation*: In der Evaluation wird die Komplikation bewertet: Waren die Ereignisse eher positiver, eher negativer Art – oder vielleicht auch eine Mischung aus beidem? Wie hat die betroffene Person sich damals gefühlt?
- *Auflösung*: Die Auflösung stellt sozusagen das Ende der Geschichte dar: Wie wurde ein bestimmtes Problem gelöst, ein Ereignis bewältigt? Die Auflösung markiert den Übergang zu einem ‚nicht mehr erzählenswerten Normalzustand‘.
- *Schlussbetrachtungen*: Die Schlussbetrachtung entspricht im Interviewverlauf der Bilanzierungsphase. Hier müssen die Forscher/innen allerdings darauf achten, dass sie bei der befragten Person keinen Rechtfertigungsdruck erzeugen.

Die Differenz zwischen Ereignis und erzählter Wirklichkeit

Bei der Auswertung narrativer Interviews ist die Unterscheidung zwischen tatsächlichen Ereignissen und im Interview rekonstruierten Ereignissen von zentraler Bedeutung: Nicht die tatsächlichen Ereignisse werden erzählt, sondern die subjektiv verarbeiteten, erinnerten und rekonstruierten Ereignisse. Das Interesse der Forscher/innen bei der Auswertung richtet sich also nicht notwendiger Weise auf die Erfassung vergangener Ereignisse, sondern häufig gerade auf die Differenz zwischen Ereignis (sofern dieses zugänglich ist) und dessen Rekonstruktion, in der die spezifisch subjektive Verarbeitung seitens der befragten Person (ggf. auch im Vergleich mit anderen Personen) zum Ausdruck kommt.

Anwendungsbereiche

Das narrative Interview eignet sich zur Erfassung der subjektiven Repräsentation und Verarbeitung von Ereignissen, die sich am besten in Geschichtenform wiedergeben lassen. Eine besondere Unterform des narrativen Interviews stellt das *biographische* oder *lebensgeschichtliche Interview* dar, das zur Erfassung speziell von Lebensgeschichten eingesetzt wird. Das narrative Interview ist wesentlich eine interdisziplinäre Methode, die keineswegs nur in der Psychologie Anwendung findet; so bildet das narrative Interview beispielsweise den Kern der in der Geschichtsforschung verankerten ‚oral history‘ (‚Geschichte von unten‘).

Prüfungsliteratur: Mayring 2002, Kap. 4.1.2.

III.2.4. Bewertung des Interviews als Forschungsmethode

Vorteile

- *Individuelle Bedeutungsstrukturen:* Der zentrale Vorteil des Interviews besteht darin, dass es den Untersuchungsteilnehmern/innen die Möglichkeit bietet, sich in ihren eigenen Worten zu dem interessierenden Themengebiet zu äußern.
- *Möglichkeit der Verständnissicherung:* Außerdem besteht beim Interview, anders als bei der schriftlichen Befragung, die Möglichkeit einer Verständnissicherung: Wenn Forscher/innen sich unsicher sind, ob sie eine Äußerung der befragten Person auch richtig verstanden haben, können sie jederzeit nachfragen.

Nachteile

- *Geringe Durchführungsobjektivität:* Der zentrale Nachteil des halb- oder nonstandardisierten Interviews aus der Sicht des quantitativen Forschungsansatzes besteht in der mangelnden Durchführungsobjektivität des Verfahrens: Wenn verschiedene Forscher/innen ein Interview mit demselben Untersuchungsteilnehmer führen würden, würden die Interviewäußerungen unterschiedlich ausfallen. Der eine Interviewer fragt an einer bestimmten Stelle vielleicht nach, die andere Interviewerin dagegen nicht usw. Vertreter/innen der qualitativen Sichtweise halten diesem Einwand das Prinzip der ‚inneren Vergleichbarkeit‘ entgegen.
- *Probleme seitens der Forscher/innen:* Aufgrund der geringen Durchführungsobjektivität des Verfahrens kann es dazu kommen, dass Forscher/innen die Antworten der befragten Personen in Richtung auf soziale Erwünschtheit beeinflussen: z.B., indem sie beim Gegenüber eine konsistente Einstellungsstruktur vermuten (z.B.: “Wenn sie vorhin gesagt hat, dass sie den Winter mag, fährt sie sicher im Urlaub nicht gerne nach Italien – also frage ich das gar nicht mehr“) oder indem sie Suggestivfragen stellen.

- *Probleme seitens der befragten Personen:* Auf Seiten der befragten Personen kann es im Interview zu ähnlichen Problemen kommen, wie sie bereits im Kontext der schriftlichen Befragung dargestellt wurden. Dazu zählen z.B.: Nicht-Erreichbarkeit von Personen in der Stichprobe, Antwortverweigerung oder Verfälschen von Antworten (in Richtung auf soziale Erwünschtheit). Diese letzteren Probleme lassen sich durch eine intensive Interviewerschulung jedoch zumindest stark reduzieren.

Bewertung des Interviews aus qualitativer Sicht

Die o.g. Problempunkte des Interviews stellen in erster Linie aus der Sicht des quantitativen Forschungsansatzes ein Problem dar. Innerhalb des qualitativen Ansatzes findet sich dagegen auch die Position, dass es beim Interview nicht darum geht, ‚Fakten‘ hervorzuholen, die im Kopf der befragten Person sozusagen schon vorhanden und ‚abrufbereit‘ sind. Statt dessen wird das Interview wesentlich als eine soziale Situation gesehen, in der Interviewer/in und befragte Person gemeinsam soziales Wissen herstellen. Strittig ist bei dieser Sichtweise, nach welchen Kriterien eine so konzipierte Forschung bewertet werden kann (s. unten III.5. zu Gütekriterien qualitativer Forschung).

Prüfungsliteratur: nur Skript

III.2.5. Gruppendiskussion

Arten der Gruppendiskussion

Während beim Interview in der Regel ein/e Forscher/in eine/n Untersuchungsteilnehmer/in befragt, erlaubt es die Gruppendiskussion, mehrere Personen gleichzeitig zu befragen. Wiederum lassen sich verschiedene Formen des Verfahrens unterscheiden:

- Bei der *Gruppenbefragung* werden die Meinungen und Einstellungen einzelner Untersuchungsteilnehmer/innen erkundet. Es handelt sich hier genau genommen um eine Form des Interviews, wobei aus Ökonomiegründen mehrere Personen gleichzeitig befragt werden. Die spezielle Gruppensituation ist dabei jedoch nicht von Interesse.
- Die Gruppendiskussion kann zweitens dazu dienen, die *Meinungen und Einstellungen einer Gruppe zu erkunden*. Untersuchungseinheit ist hier also im Unterschied zur Gruppenbefragung nicht das einzelne Individuum, sondern die Gruppe selbst.
- Ziel einer weiteren Form der Gruppendiskussion ist es, die *Prozesse zu untersuchen, die der Meinungsbildung in Gruppen zu Grunde liegen*. Hier geht es also weniger um die Meinungen und Einstellungen selbst, als um die Art und Weise, wie diese in der Gruppe zustande kommen.
- Viertens schließlich kann die Gruppendiskussion der *Erkundung öffentlicher Meinungen und Einstellungen* vor allem in solchen Gegenstandsbereichen dienen, bei denen die soziale Erwünschtheit eine Rolle spielt (z.B. Vorurteile, Fremdenfeindlichkeit). Hier sind zwar primär die Meinungen und Einstellungen selbst von Interesse. Zugleich spielen Gruppenprozesse jedoch insofern eine Rolle, als diese Form der Gruppendiskussion auf der Annahme aufbaut, dass psychische Sperren in der Gruppe leichter durchbrochen werden (z.B.: sich in der Hitze des Gesprächs zur Äußerung einer ‚unpopulären‘ Meinung hinreißen lassen).

Im Folgenden steht die letztere (vierte) Art der Gruppendiskussion im Mittelpunkt.

Ablauf der Gruppendiskussion

Die Gruppendiskussion lässt sich in die folgenden Phasen unterteilen.

- Am Anfang stehen, wie bei jeder Form der Datenerhebung, *theoretische Vorüberlegungen*. Diese betreffen Gedanken sowohl zu eventuellen psychischen Sperren und Möglich-

keiten ihrer Überwindung (s. im Folgenden: Grundreiz, Reizargumente) als auch zur Zusammensetzung der Gruppe (s. unten).

- Auf die Begrüßung der Teilnehmer/innen folgt die Präsentation eines sog. *Grundreizes* durch die Untersuchungsleitung. Der Grundreiz dient zum einen dazu, das Thema der Diskussion noch einmal zu umreißen; zum anderen ist der Grundreiz in der Regel provokativ gestaltet, so dass er zum Widerspruch reizt und auf diese Weise die Diskussion in Gang bringt.
- Es schließt sich die eigentliche, *freie Gruppendiskussion* an. In dieser Phase beschränkt sich der/die Forscher/in in der Regel auf die Gesprächssteuerung (das Wort vergeben; dafür sorgen, dass Teilnehmer/innen ausreden können), greift aber selbst nicht durch Äußerung eigener Meinungen in die Diskussion ein (um die Teilnehmer/innen nicht unter Druck zu setzen).
- Für den Fall, dass die Diskussion ins Stocken gerät, sollte der/die Forscher/in möglichst einige *Reizargumente* vorbereitet haben. Diese dienen (ähnlich dem Grundreiz) dazu, die Diskussion wieder in Gang zu bringen. Da der/die Forscher/in jedoch selbst nicht in die Diskussion eingreifen sollte, werden solche Reizargumente häufig durch einen sog. *stooze* (Vertraute/r der Untersuchungsleitung) vorgebracht – durch eine Person also, die nur zum Schein an der Gruppendiskussion teilnimmt.
- Am Ende der Gruppendiskussion kann eine *Metadiskussion* stehen (also eine Diskussion über die Diskussion). Hier haben die Untersuchungsteilnehmer/innen Gelegenheit, zu beschreiben, wie sie die Diskussion empfunden haben, ob sie sich in ihren Äußerungen frei gefühlt haben usw.

Die Gruppendiskussion wird (mit Zustimmung der Teilnehmer/innen) aufgezeichnet.

Faktoren, die die Gruppendiskussion beeinflussen können

Alle Faktoren, die sich auf die Gruppendynamik auswirken, können auf diese Weise grundsätzlich auch den Verlauf einer Gruppendiskussion beeinflussen. Das sind insbesondere:

- *Diskussionsthema*: Ein Thema, von dem die Diskutierenden selbst betroffen sind (z.B. Diskussion über Studiengebühren unter Studierenden), führt vermutlich zu lebhafteren Diskussionen als ein eher abstraktes Thema (z.B. Stadtrandgestaltung). Bei einem eng umrissenen Thema wird die Diskussion vermutlich ebenfalls anders verlaufen als bei einem eher weit definierten Thema, das viele Möglichkeiten zum ‚Abschweifen‘ bietet.
- *Gruppengröße*: Als optimal gilt eine Gruppengröße zwischen 5 und 15 Personen. Je mehr Personen an der Diskussion teilnehmen, desto geringer sind natürlich die Möglichkeiten einzelner Teilnehmer/innen, ihre Meinung auch tatsächlich zum Ausdruck zu bringen. Je kleiner die Gruppe, desto schneller kann sich die Diskussion erschöpfen; zugleich ist hier der Druck auf die Teilnehmer/innen größer, sich auch tatsächlich zu äußern.
- *Zusammensetzung der Gruppe bzgl. soziodemographischer Merkmale*: In Gruppen, die im Hinblick auf soziodemographische Merkmale eher *homogen* zusammengesetzt ist, verläuft die Diskussion vermutlich symmetrischer als in soziodemographisch heterogenen Gruppen. In *heterogenen* Gruppen ist zu befürchten, dass einzelne Mitglieder sich nicht ‚trauen‘, ihre Meinung zu sagen, wenn diese von der Meinung soziodemographisch höher gestellter Mitglieder abweicht. Andererseits ist es in einer heterogenen Gruppe wahrscheinlicher, dass sich unter den Mitgliedern ein breites Spektrum von Meinungen und Einstellungen findet.
- *Bekanntheitsgrad der Mitglieder*: In der Literatur wird zwischen *natürlichen Gruppen* (deren Mitglieder auch im Alltag interagieren) und *ad hoc-Gruppen* (die speziell für die Untersuchung zusammengestellt sind) unterschieden. Natürliche Gruppen benötigen keine ‚Aufwärmphase‘. Ein Durchbrechen psychischer Sperren bei problematischen Themen kann wiederum Fremden gegenüber leichter fallen, da hier keine sozialen Konsequenzen zu befürchten sind.

- *Meinungsverteilung in der Gruppe:* Auch in Bezug auf die in der Gruppe vertretenen Meinungen kann eine Gruppe eher heterogen oder eher homogen sein. In heterogenen Gruppen besteht die Gefahr, dass die Meinungen so weit auseinandergehen, dass die Diskussion in einen Streit ausartet. In einer allzu homogenen Gruppe kann dagegen der Fall eintreten, dass die Mitglieder in so hohem Maß einer Meinung sind, dass die weitere Diskussion sich erübrigt.
- *Verhalten der Diskussionsleitung:* Das Verhalten der Forscher/innen hat ebenfalls einen deutlichen Einfluss auf den Diskussionsverlauf. Unter formalen Gesichtspunkten können Forscher/innen die Funktion der Diskussionsleitung übernehmen und so z.B. weniger durchsetzungsfähigen Gruppenmitgliedern Gehör verschaffen. Wenn Forscher/innen ihre eigene Meinung zu dem Diskussionsthema äußern, kann dies die anderen Teilnehmer/innen unter Druck setzen, die es ggf. nicht wagen, mit ihrer eigenen von der forschersseitig präferierten Meinung abzuweichen.

Bewertung der Gruppendiskussion als Forschungsmethode

Der zentrale *Vorteil* der Gruppendiskussion besteht darin, dass dieses Verfahren es als einziges ermöglicht, Gruppenprozesse, kollektive Einstellungen, deren Verbreitung usw. zu erfassen.

Das Verfahren ist allerdings auch in mehrerer Hinsicht *problematisch*:

- *keine Trennbarkeit von Themen- und Gruppeneffekten:* Forscher/innen wollen sich mit der Gruppendiskussion zwar gruppenspezifische Prozesse zu Nutze machen; eine systematische Erforschung der Art und Weise, wie sich die verschiedenen Einflussfaktoren auf den Verlauf der Gruppendiskussion auswirken, fehlt jedoch. Entsprechend ist bei der Auswertung auch keine Trennung zwischen Themen- und Gruppeneffekten möglich – es ist also im Nachhinein z.B. nicht feststellbar, ob eine bestimmte Äußerung der tatsächlichen Meinung der Diskussionsteilnehmer/in entspricht oder auf gruppenspezifische Faktoren zumindest mit zurückzuführen ist.
- *hoher Aufwand:* Die Gruppendiskussion ist ein sehr aufwändiges, arbeitsintensives Verfahren. Schon die Organisation von Gruppenterminen kann sich schwierig gestalten. Arbeitsaufwändig ist aber vor allem die Auswertung des Materials. So ist z.B. die Verschriftlichung von Gruppendiskussionen, bei denen häufig mehrere Personen durcheinandersprechen, ungleich schwieriger als die von Interviews.
- *Problem der unvollständigen Beteiligung:* Bei Gruppendiskussionen kommt es nicht selten vor, dass einzelne Teilnehmer/innen sich gar nicht an der Diskussion beteiligen („Schweiger“). Eine repräsentative Erfassung öffentlicher Einstellungen und ihrer Verbreitung ist auf dieser Grundlage nicht möglich. Das Problem lässt sich durch eine Reduktion der Gruppengröße abmildern; auch die soziodemographische Zusammensetzung der Gruppe kann hier eine Rolle spielen.

Prüfungsliteratur: Bortz & Döring 2002, Kap. 5.2.1.; Mayring 2002, Kap. 4.1.3.

III.2.6. Verfahren der Transkription

Überblick

Interviews und Gruppendiskussionen werden in der Regel auf Band aufgezeichnet. Bevor solche verbalen Daten weiter ausgewertet werden können, müssen sie zunächst von der auditiven in eine schriftliche Form gebracht, d.h. sie müssen *transkribiert* werden. Verfahren der Transkription lassen sich nach zwei Gesichtspunkten unterteilen: nach der Vollständigkeit und nach dem Umfang der berücksichtigten Information.

Vollständige vs. selektive Transkription

- Bei einer *vollständigen Transkription* wird die gesamte Aufnahme verschriftlicht, also z.B. das gesamte Interview.
- Bei der *selektiven Transkription* werden dagegen nur Ausschnitte des erhobenen Materials in eine schriftliche Form gebracht. Dies ist z.B. dann der Fall, wenn bei einem Interview nur die Antworten der befragten Person verschriftlicht werden, nicht dagegen die Fragen seitens des Forschers/der Forscherin. Eine selektive Transkription liegt auch dann vor, wenn lediglich die Äußerungen der befragten Person verschriftlicht werden, die von unmittelbarem thematischem Interesse sind. Eine solche selektive Transkription beinhaltet notwendiger Weise Relevanzentscheidungen seitens des Forschers / der Forscherin und stellt damit bereits eine Form der Interpretation des Materials dar.

Transkriptionsvarianten unterschiedlichen Umfangs

- *Wörtliche Transkription*: Die Äußerungen werden Wort für Wort verschriftlicht. Es wird der Inhalt der Äußerungen in eine schriftliche Form gebracht; die Art und Weise der Äußerung findet dagegen keine Berücksichtigung.
 - *Übertragung in Schriftdeutsch*: Dies ist die vergleichsweise einfachste und damit zugleich auch informationsärmste Form der Transkription: Das erhobene Material wird in Schriftdeutsch überführt; hier steht die inhaltlich-thematische Ebene der Äußerungen im Vordergrund.
 - *literarische Umschrift*: Die Verschriftlichung erfolgt näher an der gesprochenen Sprache; Einwüfe wie *ne*, Versprecher, Dialekt usw. bleiben erhalten.
 - Transkription mittels phonetischem Alphabet: Die Äußerungen werden in Lautschrift wiedergegeben.
- *Kommentierte Transkription*: Zusätzlich zum Inhalt wird auch die Art und Weise der Äußerung in der Transkription berücksichtigt.
 - *Berücksichtigung paraverbaler Äußerungsmerkmale*: Durch Verwendung von Sonderzeichen werden auch paraverbale Merkmale der Äußerung wie z.B. Sprechpausen, Betonung, Verschleifungen, gleichzeitiges Sprechen mehrerer Personen usw. in die Transkription aufgenommen.
 - *Einfügen einer Kommentarspalte*: In einer Kommentarspalte können sonstige Auffälligkeiten notiert werden, z.B. Lachen der befragten Person, Seufzen usw.
- *Zusammenfassendes Protokoll*: Während die Kommentierte Transkription eine Erweiterung der wörtlichen Transkription darstellt, handelt es sich beim Zusammenfassenden Protokoll um eine Form der Informationsreduktion: Das erhobene Material wird nicht wörtlich verschriftet, sondern nur sinngemäß und in zusammenfassender Form. Damit stellt das Zusammenfassende Protokoll zugleich eine Form des Übergangs zum selektiven Protokoll dar.⁵

Je mehr Transkription, desto besser?

Die verschiedenen Formen der Transkription beinhalten ein je unterschiedliches Maß an Information: Ein zusammenfassendes Protokoll z.B. vermittelt ein Minimum, eine kommentierte Transkription unter Berücksichtigung sowohl paraverbaler als auch anderer Merkmale der Äußerungssituation ein Maximum (im auditiven Bereich). Ein ‚Mehr‘ ist hier aber nicht unbedingt auch besser: Transkriptionen sind zum einen sehr aufwändig anzufertigen – und sie sind natürlich umso aufwändiger, je mehr Information berücksichtigt wird. Zum anderen geht

⁵ Bei Mayring (2002) wird das Zusammenfassende Protokoll als eine Form der Inhaltsanalyse (und damit als Auswertungsverfahren) behandelt.

der Informationsgehalt auf Kosten der Lesbarkeit: Je informationshaltiger, desto unübersichtlicher und schwerer lesbar ist das Transkript.

Der Umfang der Transkription sollte sich daher nach dem Untersuchungszweck richten: Wenn in einer Untersuchung in erster Linie inhaltliche Aspekte von Interesse sind (z.B. Art der genannten Wertorientierungen), dann ist eine wörtliche Transkription völlig ausreichend. Wenn das Forschungsinteresse sich dagegen auf ‚objektive Strukturen‘ hinter dem Gesagten richtet, dann sind z.B. lautliche und paraverbale Information durchaus von Bedeutung. So kann es etwa ein Indikator von Unsicherheit sein, wenn jemand plötzlich in einen Dialekt verfällt, viele Pausen macht usw. In einem solchen Fall sollte das Transkript auch durchaus ausführlich sein. Dies gilt umso mehr, wenn im Rahmen der Auswertung auch Verfahren zur Anwendung kommen, die ihren Ursprung in der Linguistik haben, wie beispielsweise Formen der Konversations- und der Diskursanalyse.

Transkription ist Interpretation

Im Zusammenhang mit dem selektiven Protokoll wurde bereits darauf hingewiesen, dass die Entscheidung darüber, ‚was zum Thema gehört‘ (und daher auch transkribiert wird) und was dagegen ‚nicht so wichtig ist‘ und daher ‚weg gelassen‘ werden kann, immer schon eine Interpretation der jeweiligen Äußerungen voraussetzt. Eine solche Entscheidung stellt genau genommen schon eine Form der Auswertung dar, die mittels Verfahren wie der Inhaltsanalyse (s. unten III.3.3.) stärker systematisiert werden kann.

Aber auch die wörtliche Transkription (ggf. mit Kommentaren) beinhaltet Interpretationsleistungen des Transkribenten oder der Transkribentin. So wird das Transkript aus Gründen der Lesbarkeit häufig mit Satzzeichen versehen – diese ‚sind‘ aber in keiner Weise auf dem Band ‚vorhanden‘, sondern werden im Transkriptionsprozess eben je nach Interpretation eingefügt (und verschiedene Personen würden aus dem Gesagten meist auch unterschiedliche Sätze bilden). Interpretation kommt auch immer dann zum Tragen, wenn die Äußerungen auf Band schwer verständlich sind – sei es aufgrund von Hintergrundgeräuschen oder weil der/die Transkribent/in nicht über relevantes Hintergrundwissen verfügt (z.B. die Namen von Fernsehsendungen in den 60er und 70er Jahren). Die Transkripte sollten daher möglichst von der Interviewerin bzw. dem Interviewer mit dem Band selbst verglichen werden.

Prüfungsliteratur: Bortz & Döring 2002, Kap. 5.2.1.; Mayring 2002, Kap. 4.2.

III.2.7. Verfahren qualitativer Beobachtung

Merkmale der qualitativen Beobachtung

- Vergleichbar der qualitativen Befragung, erfolgt auch die qualitative (im Vergleich zur quantitativen) Beobachtung meist in einer *wenig standardisierten*, offenen Form. Es wird also kein Beobachtungsschema erstellt. Vor der Beobachtung werden allenfalls die interessierenden Aspekte festgelegt; relevante Ausprägungen dieser Aspekte werden jedoch erst im Verlauf der Beobachtung induktiv entwickelt.
- Während der Beobachtung wird ein *Protokoll* im Sinne einer möglichst umfassenden Dokumentation von Ereignisabläufen und deren situativer Einbettung erstellt.
- Das Beobachtungsprotokoll umfasst auch *subjektive Eindrücke* und Reaktionen der Forscher/innen, die allerdings klar von den eigentlichen Beobachtungen abgegrenzt sein sollten.

Formen der qualitativen Beobachtung

Zu den Formen qualitativer Beobachtung zählen:

- *Teilnehmende Beobachtung*: Bei der Teilnehmenden Beobachtung werden die Forscher/innen selbst zu einem Teil des untersuchten Feldes. Diese Form der Beobachtung wird im Rahmen des Ansatzes der Deskriptiven Feldforschung noch genauer dargestellt (s. unten III.4.3.).
- *Einzelfallbeobachtung*: Hier wird das Verhalten und Handeln einer Person im Detail erfasst. Piagets Untersuchung der Wahrnehmungsprozesse bei Kindern anhand von Beobachtungen seines eigenen Sohnes stehen beispielsweise in dieser Tradition.
- *Rollenspiel*: Beim Rollenspiel werden die Untersuchungsteilnehmer/innen aufgefordert, sich in eine bestimmte Situation hineinzusetzen, eine bestimmte Rolle zu übernehmen. Erhebungen mittels Rollenspiel sind valider als z.B. die Erhebung von Selbstbeobachtungen mittels Fragebögen; zugleich vermeidet das Rollenspiel die ethischen Probleme, wie sie bei einer verdeckten Beobachtung auftreten würden.
- *Lautes Denken*: Die Methode des Lauten Denkens wird bei der Erforschung komplexer Kognitionen eingesetzt (Problemlöseprozesse, Metaphernverstehen, Rezeption von Medienprodukten usw.). Die Untersuchungsteilnehmer/innen werden aufgefordert, alles zu verbalisieren, was ihnen bei der Bearbeitung der Untersuchungsaufgabe durch den Kopf geht. Lautes Denken kann kognitionsbegleitend erfolgen (also während der Aufgabenbearbeitung) oder retrospektiv bzw. im Nachhinein.

Prüfungsliteratur: Bortz & Döring 2002, Kap. 5.2.2.

III.3. Qualitative Verfahren der Datenanalyse

Qualitative Verfahren der Datenanalyse

wenig regelgeleitet, induktiv

Hermeneutik
 Objektive Hermeneutik
 Globalauswertung
 Phänomenologische Analyse
 Sozialwissenschaftlich-hermeneutische Paraphrase
 Psychoanalytische Textinterpretation
 Narrative Analyse
 Textanalyse
 Diskursanalyse
 Konversationsanalyse
 Typenbildung.....
 Kodierung
 Qualitative Inhaltsanalyse
 Komplexe Inhaltsanalyse

stärker regelgeleitet, deduktiv-induktiv

Das Spektrum qualitativer Methoden zur Datenanalyse reicht von eher wenig regelgeleiteten, induktiven (z.B. hermeneutischen Verfahren) bis hin zu systematischen, deduktiv-induktiven Verfahren (z.B. Inhaltsanalyse). Dabei werden häufig auch ‚Anleihen‘ bei Nachbardisziplinen gemacht; bei Konversations- und Diskursanalyse z.B. handelt es sich um linguistische Methoden. Im Folgenden werden solche qualitativen Verfahren der Datenanalyse dargestellt, die sich vor allem für die Analyse von verbalen Daten eignen. Die Texte selbst können beliebiger Art sein: Es kann sich um Material handeln, das zuvor im Rahmen von Interviews erhoben

wurde; die Methoden sind aber ebenso auf vorgefundenes Material anwendbar wie z.B. Dokumente, Bücher, Artikel oder gesprochene Sprache (etwa Talkshows im Fernsehen). Aus der Vielzahl qualitativer Methoden der Datenanalyse werden hier näher erläutert: Hermeneutik, Objektive Hermeneutik, verschiedene Verfahren der Inhaltsanalyse sowie Typenbildung.

III.3.1. Hermeneutik

Ursprünge und Entwicklung der Hermeneutik

Bei der Hermeneutik handelt es sich um die älteste, ‚traditionelle‘ Methode zur Auslegung von bedeutungshaltigem Material. Entstanden ist sie zunächst im theologischen und juristischen Kontext als ‚Lehre von der Auslegung verbindlicher Texte‘: Im Rahmen dieser sog. dogmatischen Hermeneutik wurden verbindliche Texte (wie beispielsweise die Bibel) auf ‚Handlungsanweisungen‘ für verschiedene Situationen hin ‚befragt‘; es fand also eine Textinterpretation statt. Im 18. und 19. Jahrhundert begründete Schleiermacher eine allgemeine geisteswissenschaftliche Hermeneutik, die im Folgenden von Dilthey für die Sozialwissenschaften weiterentwickelt wurde. In der zeitgenössischen Philosophie fand außerdem eine Ausarbeitung der Hermeneutik zu einer generellen Theorie des Umgangs mit historisch-gesellschaftlichen Gegenständen statt (z.B. Gadamer). Im Folgenden ist jedoch nicht diese letztere Auffassung von Hermeneutik als Philosophie von Interesse, sondern es steht die geisteswissenschaftlich-sozialwissenschaftliche Auffassung von Hermeneutik als Methode des Verstehens im Mittelpunkt.

Hermeneutik als Methode des Verstehens

Auch in den Sozialwissenschaften verbindet sich mit der Hermeneutik die Zielsetzung, menschliche Lebensäußerungen zu verstehen – wobei ‚Verstehen‘ sich auf die Erfassung der Bedeutung dieser Äußerungen bezieht (nicht z.B. auf Verstehen im Sinne eines empathischen Einfühlens in eine andere Person). Dieser Prozess der Bedeutungserfassung ist als Überwindung der Differenz zwischen dem von den Äußerungsproduzenten/innen Gesagten und dem subjektiv-alltäglichen Verständnis der Äußerung konzipiert. Hermeneutik richtet sich also nicht auf eine Bedeutungserfassung im alltäglichen Sinn; vielmehr soll die Subjektivität alltäglichen Verstehens gerade überwunden werden. Dies wiederum ist nur möglich vor dem Hintergrund historisch-kultureller Gemeinsamkeit, also vor einem Hintergrund bereits geteilter Bedeutung. Ziel ist es, zu einem adäquaten, dem Gegenstand angemessenen Verständnis zu gelangen.

Hermeneutischer Zirkel I und II

Grundprinzip des hermeneutischen Verstehens ist der sog. hermeneutische Zirkel, wobei genau genommen zwei solcher Zirkel anzusetzen sind:

- *Hermeneutischer Zirkel I*: Verstehen vollzieht sich erstens in einem Ineinandergreifen von Vorverständnis und Textverständnis: Ein Verstehen eines Textes ist nur vor dem Hintergrund eines bereits gegebenen Vorverständnisses möglich. Das resultierende Textverständnis modifiziert seinerseits das Vorverständnis, das erneut an den Text herangetragen wird und seinerseits ein vertieftes Textverständnis ermöglicht – usw.
- *Hermeneutischer Zirkel II*: Verstehen vollzieht sich zweitens in einem Ineinandergreifen von Textteil und Textganzem: Ein Verstehen der Textteile ist Voraussetzung für das Verstehen des Textganzen. Aus dem Verstehen des Textganzen resultiert ein neues, vertieftes Verstehen der Textteile, das wiederum in ein erneutes Verstehen der Textteile eingeht – usw.

Der Begriff des ‚Zirkels‘ bezeichnet hier also keinen ‚Teufelskreis‘, sondern ein fortschreitendes, vertieftes Verstehen. Um dieses vertiefte Verstehen abzubilden, das dem Text in zu-

nehmendem Maße gerecht wird, ist manchmal auch von einer hermeneutischen *Spirale* die Rede.

Bewertung der Hermeneutik

Bei der Hermeneutik handelt es sich auch in der Adaptation für die Sozialwissenschaften letztlich um eine *Kunstlehre*: Die Schritte des hermeneutischen Verstehens lassen sich nicht systematisch explizieren; das Verfahren ist – obwohl ja gerade die Subjektivität des individuellen Alltagsverstehens überwunden werden soll – stark von der Subjektivität des Forschers/der Forscherin abhängig. Auch das Kriterium der Angemessenheit, dem das Verstehensresultat genügen soll, bleibt letztlich sehr vage.

Gerade in den sozialwissenschaftlichen Adaptationen des Verfahrens sind solche Kritikpunkte z.T. aufgenommen worden. Dies hat beispielsweise dazu geführt, dass systematischere Regeln für den Verstehensprozess aufgestellt wurden. Auch wird versucht, die Subjektivität individuellen Verstehens dadurch zu mindern, dass mehrere Personen denselben Text zu verstehen versuchen und ihre Verstehensresultate vergleichen.

Prüfungsliteratur: nur Skript

II.3.2. Objektive Hermeneutik

Überblick

Eine Variante der Hermeneutik, die speziell für die Sozialwissenschaften entwickelt wurde (von Oevermann), stellt die Objektive Hermeneutik dar. Das Adjektiv ‚objektiv‘ bezieht sich übrigens nicht auf die Vorgehensweise (die Objektive Hermeneutik ist vom Anspruch her also nicht objektiver als die klassische Hermeneutik), sondern auf den Gegenstand: Zielsetzung der Objektiven Hermeneutik ist es, allgemeine, ‚objektive‘ Strukturen zu identifizieren, die menschlichem Handeln zugrunde liegen.

Grundannahmen

- *Verstehen als Sinnverstehen*: Verstehen ist immer Verstehen von Sinn. Zielsetzung des Verstehens (und damit der Objektiven Hermeneutik als verstehender Methode) ist die Zuschreibung von Sinn.
- *Konstitution von Welt in der Sprache*: Die sinnhafte Welt konstituiert sich ausschließlich in der Sprache und manifestiert sich in Texten. Texte stellen also Manifestationen sinnhafter sozialer Wirklichkeit dar. Texte sind somit der eigentliche Gegenstand der Objektiven Hermeneutik, und das Verstehen von Texten dient der Interpretation der Wirklichkeit.
- *Text als regelerzeugtes Gebilde*: Menschliches Handeln wird durch soziale Strukturen geregelt und gesteuert (auf eine Frage folgt eine Antwort; man sollte nichts versprechen, was man nicht halten kann usw.). Soziale Regeln und Strukturen manifestieren sich auch in Texten; entsprechend kann bei der Interpretation von Texten auch auf solche Regeln zurückgegriffen werden. Regelkompetenz bildet die Basis des Textverstehens bzw. der Interpretation von Wirklichkeit.
- *Zielsetzung der Strukturrekonstruktion*: Regeln geben sozusagen die Optionen einer konkreten Lebenspraxis vor - wie Fäden, entlang derer Handlungen und damit verbundene Sinstrukturen sich entfalten können. Jede Handlung – und damit verbunden die Wahl einer bestimmten Handlungsmöglichkeit - schließt jedoch in der Regel andere Handlungsmöglichkeiten aus (‚Selektivität der konkreten Lebenspraxis‘: Wer z.B. ein Versprechen gibt,

geht damit auch eine gewisse Handlungsverpflichtung ein). Eine solche Realisation bestimmter Handlungsmöglichkeiten unter der Vielzahl potenzieller Möglichkeiten vor dem Hintergrund sozialer Regeln bildet die Sinnstruktur des konkreten Falls. Die Zielsetzung der Hermeneutik besteht darin, unter Herausarbeitung der zugrunde liegenden Regeln diese Sinnstruktur zu rekonstruieren.

- *Differenz zwischen latenter Sinnstruktur und subjektiver Sinnrepräsentanz*: Es wird in der Objektiven Hermeneutik davon ausgegangen, dass Menschen meist keinen Zugang zu den Regeln haben, die ihrem Handeln zugrunde liegen. Es wird also angenommen, dass die latente Sinnstruktur, die menschlichem Handeln zugrunde liegt, und subjektiv gemeinter Handlungssinn vielfach nicht übereinstimmen. Die Objektive Hermeneutik als Verstehensmethode setzt daher nicht am subjektiven Sinn an, sondern an der Differenz zwischen den beiden Sinnstrukturen. Entsprechend ist die Methode auch nur dann anwendbar, wenn eine solche Differenz vorliegt.
- *Prinzip der Fallstruktur-Generalisierung*: Jeder ‚Fall‘, jede Form der Interaktion mit und durch Sprache, ist insofern allgemeiner Natur, als er auf den geltenden sozialen Regeln basiert. Jeder Fall ist aber zugleich auch insofern besonders, als er eine ganz spezifische Fallstruktur aufweist (in Form einer konkreten Realisation von Handlungsmöglichkeiten unter den vielen potenziell verfügbaren). In diesem Sinn ist jede rekonstruierte Fallstruktur auch zugleich verallgemeinerbar und spezifisch. Bei der Objektiven Hermeneutik wird daher meist auch nur mit sehr wenigen Fällen gearbeitet.

Allgemeines Vorgehen

Das Grundprinzip der Objektiven Hermeneutik besteht darin, dass auf der Grundlage eines ersten, sehr kleinen Ausschnitts einer Interaktion *Hypothesen* über dem Fall zugrunde liegende objektive Sinnstrukturen aufgestellt werden. Diese Hypothesen werden dann mit dem konkreten Material *verglichen*. Es folgen mehrere solcher Analysen aufeinander (ggf. auch am selben Textausschnitt), wobei jede Analyse einen potenziellen Falsifikator für die vorausgehende darstellt. Wenn weitere Analysen die aufgestellte Strukturhypothese bestätigen, werden die Analysen zu einer *Strukturgeneralisierung* verdichtet. Datenerhebung und -analyse fallen bei der Objektiven Hermeneutik also zusammen.

Formen der Objektiven Hermeneutik

Es lassen sich verschiedene Verfahren der Objektiven Hermeneutik unterscheiden: die *Feinanalyse*, die *Sequenzanalyse* und die *Interpretation der objektiven Sozialdaten*; weitere Formen, die aber nur selten zur Anwendung kommen, sind die summarische Interpretation sowie die Glosse. Im Folgenden soll das Vorgehen bei der Sequenzanalyse genauer erläutert werden.

Vorgehen bei der Sequenzanalyse

- *Unterscheidung zwischen innerem und äußerem Kontext*: Der äußere Kontext ist durch die Situation und die spezifische Interaktionskonstellation gegeben, in der eine Äußerung gemacht wird. Der innere Kontext ergibt sich dagegen aus dem Zusammenhang, in dem die Äußerungen in einer Interaktionssequenz untereinander stehen; die vorausgehenden bilden den inneren Kontext für die nachfolgenden Äußerungen. Diese Unterscheidung zwischen innerem und äußerem Kontext ist grundlegend für die Sequenzanalyse. Dabei wird der äußere Kontext zunächst überhaupt nicht berücksichtigt. Auch der innere Kontext kommt erst allmählich, im Laufe der Analyse, zum Tragen. Dabei dürfen spätere Äußerungen die Interpretation früherer Äußerungen jedoch nicht beeinflussen.

- *Ermittlung von Handlungsspielräumen:* Es wird zunächst für die erste Äußerung (oder den ersten Äußerungsteil) geprüft, welche sozialen Regeln dieser Äußerung potenziell zugrunde liegen könnten. Es werden sozusagen verschiedene ‚Geschichten‘ zu einer Äußerung generiert, innerhalb derer die Äußerung einen Sinn macht.
- *Lesarten erstellen:* Die Handlungsspielräume werden dann zu ‚Lesarten‘ gruppiert. Jede Lesart beinhaltet Äußerungsinterpretationen, die auf dieselbe soziale Regel, auf dieselbe objektive Struktur zurückgreifen.
- *Kompatibilitätsprüfung:* Im nächsten Schritt wird geprüft, welche Lesarten mit der nachfolgenden Äußerung vereinbar sind und welche nicht. Die unvereinbaren Lesarten werden bei der nachfolgenden Analyse nicht weiter berücksichtigt.
- *Fallstrukturhypothese:* Auf der Grundlage der Ergebnisse der Kompatibilitätsprüfung wird eine Hypothese über die objektive Struktur aufgestellt, die dem Fall zugrunde liegt.
- *Strukturgeneralisierung:* Es wird sukzessive immer mehr vom inneren Kontext in die Kompatibilitätsprüfung einbezogen, bis eine Fallstrukturhypothese resultiert, die es erlaubt, den weiteren Interaktionsverlauf vorherzusagen. Die Strukturhypothese kann dann auf den gesamten vorliegenden Fall verallgemeinert werden.

Bewertung der objektiven Hermeneutik

- Wie die klassische Hermeneutik stellt auch die objektive Hermeneutik wesentlich eine *Kunstlehre* dar, die nicht systematisch vermittelbar ist. Die Ergebnisse der Objektiven Hermeneutik bleiben damit letztlich an die Subjektivität der Forschenden gebunden.
- Die Methode ist von der Durchführung her extrem aufwändig.
- Auch andere, inhaltliche Annahmen werden z.T. kritisch gesehen, so z.B. die Grundannahme, dass sinnhafte Wirklichkeit sich ausschließlich in Texten manifestiert.

Prüfungsliteratur: Mayring 2002, Kap. 4.3.15.

III.3.3. Inhaltsanalytische Verfahren

Überblick

Bei dem Begriff der Inhaltsanalyse handelt es sich wiederum um einen Sammelbegriff, unter den sich verschiedene konkrete Verfahren subsumieren lassen. Diesen Verfahren liegt aber im Wesentlichen dasselbe Ziel und dasselbe Prinzip zugrunde: Sie dienen dazu, die Bedeutung von Texten im Hinblick auf ausgewählte Aspekte zu ermitteln (und nicht ‚die Bedeutung‘ der Texte insgesamt). Dies geschieht, indem die interessierenden Bedeutungsaspekte in Form von Kategorien expliziert werden. Die Kategorien enthalten Anweisungen dahingehend, unter welchen Bedingungen ein Text-(Teil) einer Kategorie zuzuordnen ist. Die Zuordnung der Texte zu den Kategorien (und damit die Ermittlung ihrer Bedeutung) erfolgt entsprechend diesen Anweisungen.

Kriterien zur Unterscheidung verschiedener inhaltsanalytischer Verfahren

- *Manifestheit vs. Latenz der Bedeutung:* In der Geschichte der Inhaltsanalyse wiederholt sich der ‚Richtungsstreit‘ zwischen quantitativen und qualitativen Ansätzen. Eine quantitative Form der Inhaltsanalyse war darauf ausgerichtet, nur solche Textbedeutungen zu erfassen, die auch klar und explizit an der Textoberfläche ersichtlich sind (z.B.: „Das habe ich ja wieder mal großartig hingekriegt!“ → Eigenlob). Mit qualitativen Varianten des Verfahrens wurde dagegen versucht, auch ‚verborgene‘, erst aus dem Äußerungskontext heraus ersichtliche Bedeutungsaspekte zu erfassen (z.B. „Das habe ich ja wieder mal großartig hingekriegt!“, nachdem mir die Lieblings-Kaffeekanne meiner Schwester heruntergefallen ist → Tadel). Zwar existieren auch heute noch Vertreter/innen einer ‚rein‘

quantitativ oder qualitativ orientierten Vorgehensweise. Insgesamt ist das ‚entweder-oder‘ von manifester versus latenter Bedeutungserfassung heute jedoch meist zu Gunsten eines ‚sowohl-als auch‘ überwunden (s. auch unten: Komplexe Inhaltsanalyse).

- *Deduktive vs. induktive Vorgehensweise:* ‚Quantitativ‘ und ‚qualitativ‘ orientierte Verfahren unterscheiden sich auch hinsichtlich ihres Vorgehens bei der Herleitung der Bedeutungskategorien. Eher quantitativ orientierte Verfahren sehen (in Übereinstimmung mit der linearen Vorgehensweise der Hypothesentestung) eine theoriegeleitet-deduktive Herleitung der Kategorien vor. Bei einer eher qualitativ orientierten Vorgehensweise werden die Kategorien dagegen meist induktiv an dem analysierten Material selbst entwickelt und zunehmend verfeinert (in Übereinstimmung mit einer zyklischen Vorgehensweise des Ineinandergreifens von Datenerhebung und –auswertung). Es existieren auch ‚Zwischenformen‘, in denen zunächst deduktiv ein Kategoriengerüst erstellt und dann an einem Teil des Materials induktiv weiter verfeinert wird.
- *Intersubjektivität vs. Subjektivität der Kodierung:* Eher quantitativ orientierte inhaltsanalytische Verfahren sehen eine Absicherung der Objektivität und Reliabilität der Kodierungen (d.h. Zuordnungen von Textteilen zu Kategorien) vor, und zwar, indem die Kodierungen unabhängig voneinander durch mehrere Personen vorgenommen werden; deren Übereinstimmung wird quantifiziert und hinsichtlich ihrer Güte bewertet. In qualitativ orientierten Varianten des Verfahrens wird die Kodierung u.U. von einer Einzelperson vorgenommen.
- *Systematik vs. Flexibilität:* Insgesamt lassen sich quantitativ orientierte Formen der Inhaltsanalyse als in hohem Maß systematisch und regelgeleitet charakterisieren; am qualitativen Paradigma orientierte Formen sind weniger regelgeleitet, weisen dafür aber ein höheres Maß an Flexibilität auf und können ggf. der Vielfalt des Textmaterials besser gerecht werden.
- *Datenreduktion vs. –erweiterung:* In der Regel handelt es sich bei der Inhaltsanalyse um ein datenreduzierendes Verfahren, d.h. die Vielfalt des Textmaterials wird einer begrenzten Zahl von Kategorien zugeordnet und damit auch überschaubarer gemacht. Es existiert aber auch eine Form der explikativen Inhaltsanalyse, bei der im Gegenteil zusätzliches Material an den Text herangetragen wird, um dessen Bedeutung aufzuklären. Diese letztere explikative Form der Inhaltsanalyse wird im Folgenden jedoch nicht weiter berücksichtigt (s. Mayring, 2002, S. 115).

Kodierung

Überblick

Die Kodierung stellt eines der am häufigsten eingesetzten Auswertungsverfahren in der qualitativen Forschung dar (so z.B. in der Gegenstandsbezogenen Theorienbildung; s.u. II.4.2.). Es handelt sich dabei um die vergleichsweise am wenigsten systematisierte Variante der inhaltsanalytischen Vorgehensweise. In einschlägigen qualitativen Lehrbüchern wird das Verfahren häufig nicht als Form der Inhaltsanalyse aufgeführt, sondern nur unspezifisch als ‚Kodierung‘ bezeichnet.

Ablauf

Das Prinzip der Kodierung besteht darin, einen Text sukzessive durchzuarbeiten und interessierende Bedeutungsaspekte als Kategorien am Rand zu vermerken. Die Kategorien werden bei dieser Vorgehensweise häufig auch als ‚Codes‘ bezeichnet, die dem Text zugewiesen werden. Ziel ist es, ein Gesamt von Kategorien für das Gesamt der vorliegenden Texte zu

entwickeln (z.B. für das Gesamt der Interviews, die in einer Untersuchung erhoben werden). Die Kategorien, wie sie zu Beginn der Analyse vergeben werden, werden im Verlauf der Auswertung aller Interviews in der Regel noch erheblich verändert. Die Kodierung beinhaltet entsprechend mehrere Materialdurchläufe und mehrere Veränderungen der Kategorien, bis Kategorien und Material optimal aufeinander abgestimmt sind bzw. bis die Kategorien das Material unter den interessierenden Gesichtspunkten optimal erfassen.

Formen der Kodierung

Je nach Zielsetzung der Analyse lassen sich zwei Arten von Kodierungen unterscheiden. die thematische und die konzeptuelle Kodierung.

- Die *thematische Kodierung* dient dazu, das Äußerungsthema zu identifizieren. Diese Form der Kodierung lässt sich als eine Form des ‚Aufklebens von Etiketten‘ verstehen. Sie wird vor allem dann eingesetzt, wenn das erhobene Material sehr umfangreich ist und wenn die Äußerungen verschiedener Personen zu einem Thema verglichen werden sollen. Die thematische Kodierung hilft, die relevanten Äußerungen schnell zu lokalisieren.
- Bei der *konzeptuellen Kodierung* werden Kategorien an die Texte herangetragen, die dazu dienen, das Material unter theoretischen Gesichtspunkten weiter auszuwerten und zu strukturieren. Solche Konzepte können ganz unterschiedlicher Art sein; sie ergeben sich in der Regel aus der Fragestellung.

Diese beiden Formen der Kodierung schließen einander nicht aus; die thematische kann als Vorstufe zu einer konzeptuellen Kodierung eingesetzt werden.

Auswertung von Kodierungen

An die Kodierung kann sich eine weitere Auswertung anschließen, z.B.:

- die Bildung einer Theorie (s.u. Gegenstandsbezogene Theoriebildung) oder
- ein Häufigkeitsvergleich zwischen verschiedenen Fällen oder Fallgruppen.

Abgrenzung der Kodierung gegenüber anderen inhaltsanalytischen Verfahren

Gegenüber anderen inhaltsanalytischen Verfahren zeichnet sich die Kodierung dadurch aus:

- dass die Kategorienerstellung praktisch rein *induktiv* erfolgt;
- dass sie meist nur von einer Person durchgeführt wird; die Kodierung ist daher (ähnlich wie die hermeneutischen Verfahren) vergleichsweise *subjektiv*;
- dass eine Textstelle auch mehreren Kategorien zugeordnet werden kann (*Nicht-Exklusivität der Kategorien*);
- dass eine *Definition* der Kategorien bzw. Codes *nicht immer vorgenommen* wird, so dass die Objektivität, Reliabilität und Nachvollziehbarkeit des Verfahrens stark eingeschränkt sind. Wenn eine Kategoriendefinition erfolgt, bestehen fließende Übergänge zwischen der Kodierung und der Qualitativen Inhaltsanalyse nach Mayring.

Prüfungsliteratur: nur Skript

Qualitative Inhaltsanalyse (nach Mayring)

Die Qualitative Inhaltsanalyse nach Mayring (2002, S. 119ff.: Strukturierende Inhaltsanalyse) steht hinsichtlich Systematik und Regelgeleitetheit ‚zwischen‘ der weniger regelgeleiteten Kodierung und der stark regelgeleiteten Komplexen Inhaltsanalyse (nach Rustemeyer; s.u.).

Ablauf:

- Es werden zunächst theoriegeleitet-deduktiv relevante *Strukturierungsdimensionen* im Sinne von Kategorien *festgelegt*.
- Es erfolgt eine *Definition der Kategorien*. Diese beinhaltet:

- *Kategorienexplikation*: Es werden die Bedingungen angegeben, unter denen eine Textstelle der Kategorie zuzuordnen ist.
- *Ankerbeispiele*: Die Kategorien werden durch besonders prägnante Beispiele illustriert.
- *Abgrenzungen*: Dort, wo zwischen den Kategorien Abgrenzungsprobleme bestehen können (wo also unklar ist, welcher Kategorie eine Textstelle im Zweifelsfall zuzuordnen ist), werden explizite Regeln formuliert.

Ziel der Kategoriendefinition ist es, eine *eindeutige Zuordnung* von Textstellen zu Kategorien zu ermöglichen.

- Das Material wird zeilenweise bearbeitet; solche *Fundstellen*, in denen die im Kategoriensystem zusammengestellten Gesichtspunkte thematisch sind, werden am Rand markiert.
- Diese Fundstellen werden den Kategorien zugeordnet; dabei werden die Kategorien weiter verfeinert und dem Material angepasst.
- Nachdem bis zu 50% des Materials auf diese Weise bearbeitet wurden, werden die verbleibenden Fundstellen dem nun nicht weiter modifizierbaren Kategoriensystem zugeordnet.

Abgrenzung der qualitativen Inhaltsanalyse gegenüber anderen inhaltsanalytischen Verfahren: Von der Kodierung unterscheidet sich die Qualitative Inhaltsanalyse nach Mayring in den folgenden Punkten:

- Die Erstellung des Kategoriensystems erfolgt teils deduktiv (teils induktiv).
- Es sind Kategoriendefinitionen zu erarbeiten, die eine möglichst eindeutige Zuordnung von Textstellen zu den Kategorien ermöglichen sollen.
- Die Kategorien schließen einander wechselseitig aus.

Prüfungsliteratur: Mayring 2002, Kap. 4.3.14.

Komplexe Inhaltsanalyse (nach Rustemeyer)

Die Komplexe Inhaltsanalyse stellt das vergleichsweise systematischste, am stärksten regelgeleitete inhaltsanalytische Verfahren dar (unter den drei Formen, die hier aufgeführt sind).

Ablauf:

- Als erstes werden die zu analysierenden Texte festgelegt; es wird also eine *Stichprobe* gezogen, die für das Material, über das eine Aussage gemacht werden soll, möglichst repräsentativ ist. Es findet also eine Stichprobenziehung nach dem quantitativen Ansatz statt (vgl. oben II.1.2.).
- Es werden deduktiv-induktiv (unter Rückgriff auf einen Teil des Textmaterials) relevante *Strukturierungsdimensionen* im Sinne von Kategorien *festgelegt*.
- Es erfolgt eine *Definition der Kategorien*. Wie bei der Qualitativen Inhaltsanalyse umfasst die Definition auch hier den Dreischritt von Explikation, Ankerbeispielen und Abgrenzungsregeln.
- Die Texte werden vor der inhaltsanalytischen Kodierung in *Analyseeinheiten* unterteilt. Eine Analyseeinheit entspricht demjenigen Textteil, der jeweils einer Kategorie zugeordnet werden kann.
- Das Kategoriensystem wird an einem Teil des Materials *erprobt*. Die Zuordnung der Analyseeinheiten zu den Kategorien erfolgt durch mindestens zwei Kodierer/innen (*intersubjektiv*).
- Es wird die *Höhe der Übereinstimmung* zwischen den Kodierern/innen berechnet. Dafür existieren spezielle Maße (z.B. Fleiss' *kappa*). In der Regel ist die Übereinstimmung nach der Probekodierung noch nicht optimal.

- Eine unzureichende Übereinstimmung der Kodierer/innen kann erstens an einer ungenauen Explikation der Kategorien liegen. In diesem Fall ist eine *Modifikation des Kategoriensystems* erforderlich.
- Eine unzureichende Übereinstimmung kann zweitens auf Kodierfehler, Missverständnisse seitens der Kodierer/innen zurückgehen. In diesem Fall ist eine *Kodierschulung* erforderlich. Meist schließen sich an die Probekodierung sowohl eine Überarbeitung des Kategoriensystems als auch eine Kodierschulung an.
- Darauf kann erneut eine Probekodierung (anhand weiteren Textmaterials) mit anschließender Berechnung der Interkodier-Übereinstimmung erfolgen. Kodierschulung und Überarbeitung des Kategoriensystems sollten jedoch nicht häufiger als zweimal wiederholt werden.
- Wenn bei der Probekodierung eine hinreichende Übereinstimmung zwischen den Kodierern/innen erzielt wurde, folgt die *Hauptkodierung*. Die Hauptkodierung wird nur für Textmaterial durchgeführt, das nicht in die Erstellung des Kategoriensystems eingegangen ist und das bei den Probekodierungen noch nicht verwendet wurde.
- An die Hauptkodierung kann sich eine *statistische Auswertung* anschließen, z.B. mittels Häufigkeitsanalysen, aber auch mittels komplexerer Verfahren. Mit dieser Variante der Inhaltsanalyse können auch Hypothesen getestet werden.

Anforderungen an die inhaltsanalytischen Kategorien

Bei der komplexen Inhaltsanalyse sollten die Kategorien den folgenden Kriterien genügen:

- *Exklusivität*: Jede Analyseeinheit sollte nur einer Kategorie zuordenbar sein, d.h. die Kategorien sollten sich gegenseitig ausschließen. Zu dieser Regel gibt es allerdings auch Ausnahmen (etwa dann, wenn verschiedene Kategorien sich auf verschiedene Bedeutungsdimensionen des Materials beziehen).
In einer politischen Diskussion wären z.B. Art des Arguments und Partei des Sprechers/der Sprecherin in diesem Sinne voneinander unabhängige Dimensionen, so dass jede Äußerung auch beiden Kategorien zuzuordnen wäre. Die Kategorien innerhalb jeder der Dimensionen sollten sich dagegen wechselseitig ausschließen; es ist z.B. nicht möglich, dass dieselbe Analyseeinheit dem Sprecher Schröder und der Sprecherin Merkel zugeordnet wird.
- *Exhaustion*: Jede Analyseeinheit muss durch das Kategoriensystem erfassbar sein. Diese Anforderung wird in der Regel durch die Aufstellung einer ‚Restkategorie‘ erfüllt (der all diejenigen Analyseeinheiten zuzuordnen sind, die durch die übrigen Kategorien nicht erfasst werden).
- *Sättigung*: Jeder Kategorie sollte mindestens eine Analyseeinheit zugeordnet sein. Diese Anforderung ist jedoch auf deduktiv-theoriegeleitet erstellte Kategorien nicht anwendbar. In einem solchen Fall kann es im Gegenteil eine sehr wertvolle Information sein, wenn eine Kategorie nicht besetzt ist (wenn z.B. in einer politischen Debatte keine der Äußerungen als ‚persönlicher Angriff‘ kodiert wird).

Abgrenzung der quantitativen Inhaltsanalyse gegenüber anderen inhaltsanalytischen Verfahren:

Von der Qualitativen Inhaltsanalyse nach Mayring unterscheidet sich die Komplexe Inhaltsanalyse nach Rustemeyer vor allem in den folgenden Punkten:

- Durch die Unterteilung des Textmaterials in Analyseeinheiten wird vor der Kodierung genau festgelegt, welche Textteile jeweils den Kategorien zuzuordnen sind.
- Die Kodierung erfolgt intersubjektiv durch mindestens zwei Kodierer/innen. Durch die Berechnung ihrer Übereinstimmung lassen sich die Reliabilität und die Objektivität von Kategoriensystem und Kodierungen sichern.

- Erstellung und Erprobung des Kategoriensystems einerseits und Hauptkodierung andererseits erfolgen anhand von je anderem Material. Durch diese Trennung wird eine Hypothesentestung möglich.

Prüfungsliteratur: Rustemeyer 1992

Vergleichende Bewertung der drei inhaltsanalytischen Verfahren

- *Kodierung:* Aufgrund der textnahen, induktiven Vorgehensweise verbunden mit wiederholten Modifikationen des Kategoriensystems ist die Kodierung dasjenige unter den drei Verfahren, das es am ehesten erlaubt, interindividuell sehr unterschiedliche Bedeutungsaspekte zu erfassen. Allerdings stellt die Kodierung zugleich ein sehr subjektives und potenziell willkürliches Verfahren dar.
- *Qualitative Inhaltsanalyse:* Auch die Qualitative Inhaltsanalyse erlaubt es, interindividuell unterschiedliche Bedeutungsaspekte zu erfassen, ist zugleich aber von der Vorgehensweise her deutlich systematischer angelegt als die Kodierung. Allerdings handelt es sich im Vergleich zur Komplexen Inhaltsanalyse um ein subjektives Verfahren; Hypothesentestungen sind aufgrund des Ineinandergreifens von Kategoriendefinition und Kodierung nicht möglich. Die Qualitative Inhaltsanalyse eignet sich somit vor allem dann als Auswertungsverfahren, wenn keine Hypothesentestung angestrebt wird und wenn der Bedeutungsgehalt des Materials voraussichtlich interindividuell stark variiert.
- *Komplexe Inhaltsanalyse:* Die Anwendungsbedingungen der Komplexen sind komplexer als die der Qualitativen Inhaltsanalyse. Die Komplexe Inhaltsanalyse ist ein systematisches und intersubjektives Verfahren zur Ermittlung der Bedeutung von Texten. Es ist vor allem dann anwendbar, wenn interindividuell vergleichbare Bedeutungsaspekte im Mittelpunkt des Interesses stehen. Wenn eine Hypothese bezüglich des zu analysierenden Materials geprüft werden soll, dann ist sie als einziges unter den drei Verfahren anwendbar. Wenn dagegen gerade die Vielfalt der interindividuellen Bedeutungen exploriert werden soll, dann ist eher die Qualitative Inhaltsanalyse die Methode der Wahl.

Prüfungsliteratur: nur Skript

Computergestützte Auswertung verbaler Daten

Im Laufe der letzten ca. zehn, fünfzehn Jahre sind eine Reihe von Programmen entwickelt worden, die Forscher/innen bei der inhaltsanalytischen Auswertung verbaler Daten unterstützen sollen. Entsprechend den zwei ‚Strängen‘ inhaltsanalytischer Verfahren (quantitativ, qualitativ) lassen sich zwei Arten solcher Programme unterscheiden.

‚Quantitative‘ Programme

Quantitative Programme sind darauf ausgerichtet, Analysen manifester Textbedeutung zu automatisieren (z.B. General Inquirer, Textpack). Je nach Fragestellung wird ein sog. Dictionary erstellt. Dieses Dictionary erfüllt die Rolle des Kategoriensystems bei der ‚Kodierung per Hand‘: Für jede Kategorie werden (im Sinne einer Kategorienexplikation) Suchbegriffe definiert. Bei einer Analyse von Leitartikeln deutschsprachiger Tageszeitungen wäre etwa eine Kategorie ‚Wirtschaft‘ vorstellbar; diese würde Suchbegriffe enthalten wie beispielsweise: Wirtschaft, Finanz*, Kapital*, Börse usw. Das Programm durchsucht die zu analysierenden Texte nach diesen Suchbegriffen; für jeden ‚Fund‘ wird dem Textteil die Kategorie zu-

gewiesen, die durch den Suchbegriff spezifiziert wird. Die Ergebnisse können häufig unmittelbar statistisch weiterverarbeitet werden.

Programme dieses Typs realisieren also eine ‚Arbeitsteilung‘ zwischen Mensch und Computer: Der Mensch muss das Dictionary spezifizieren; die eigentliche Kodierung wird vom Computer übernommen (kann aber ihrerseits immer nur so gut sein wie das Dictionary). Diese Programme stellen gerade bei der Analyse großer Textmengen eine erhebliche Arbeitserleichterung dar, sind aber in der Anwendung weitgehend auf die Analyse manifester Textbedeutungen beschränkt.

‚Qualitative‘ Programme

Qualitative Programme dienen dazu, Verfahren der Kodierung überschaubarer und damit auch systematischer zu machen (z.B. Ethnograph, Winmax, Atlas.ti). Die Programme umfassen meist drei Funktionen: Markierung der Analyseeinheiten, Kodierung und ‚Retrieval‘ (d.h. Auflistung derjenigen Textsegmente, die derselben Kategorie zugeordnet wurden). Manche Programme erlauben es außerdem, Kategorien in Schaubildern untereinander zu verbinden; diese Funktion dient der Visualisierung bei der Theoriebildung.

Im Gegensatz zu den quantitativen Programmen bleiben bei den qualitativen Programmen alle Arbeitsschritte in der Hand der Forscher/innen. Die computergestützte Auswertung mittels qualitativer Programme stellt somit zwar keine Arbeitserleichterung dar, macht den per Hand so fehleranfälligen Kodiervorgang jedoch deutlich systematischer und nachvollziehbarer. Durch Auflistung der Textsegmente, die einer Kategorie zugeordnet wurden, lässt sich z.B. im Analyseverlauf immer wieder überprüfen, ob die Bedeutung der Kategorie sich ‚schleichend‘ verändert hat. Gerade bei umfangreichen Kategoriensystemen wird ein besserer Überblick über schon erstellte Kategorien gewährleistet usw.

Qualitative Programme lassen sich natürlich auch im Rahmen Qualitativer und Komplexer Inhaltsanalysen latenter Bedeutungsaspekte einsetzen.

Prüfungsliteratur: Mayring 2002, Kap. 5

III.3.4. Datenaufbereitende Verfahren: Typenbildung

Datenaufbereitende Verfahren

Die Ermittlung der Bedeutung von Texten mit Hilfe von Verfahren, wie sie in den vorausgehenden Abschnitten beschrieben sind, stellen häufig nur den ersten Schritt bei der Auswertung qualitativen Materials dar. Ein zweiter Schritt besteht in der weitergehenden Aufbereitung der Textbedeutungen.

Für eine solche Aufbereitung stehen verschiedene Verfahren zur Verfügung:

- *Beschreibung von Einzelfällen:* Der einzelne Fall wird ausführlich beschreibend dargestellt. Da die Einzelfallorientierung allgemein für qualitative Forschung charakteristisch ist, findet sich diese Aufbereitungsform sehr häufig in qualitativen Untersuchungen.
- *Fallvergleich:* Auch qualitative Forscher/innen bleiben jedoch häufig nicht bei der Beschreibung des einzelnen Falls stehen, sondern vergleichen darüber hinaus unterschiedliche Fälle miteinander.
- *Typenbildung:* Bei der Typenbildung sind die Beschreibung von Einzelfällen und der Fallvergleich in einer spezifischen Weise kombiniert. Fälle, die einander ähnlich sind, werden zu Gruppen (‚Typen‘) zusammengefasst (s. ausführlich unten).
- *Darstellung der Ergebnisse in Matrizenform:* Auch die Ergebnisse qualitativer Untersuchungen können in Form von Tabellen (Matrizen) übersichtlich zusammengestellt wer-

den. Anstelle von Zahlen stehen in den Zellen der Tabellen Zitate aus den analysierten Texten.

- *Graphische Darstellung der Ergebnisse:* Für die Darstellung komplexer Zusammenhänge eignen sich graphische Formen der Datenaufbereitung (s. auch unten III.4.5. Subjektive Theorien).
- *Häufigkeitsanalysen:* Auf der Grundlage eines Fallvergleichs sind im Anschluss an eine qualitative Analyse auch Häufigkeitsauszählungen möglich.
- *Statistische Auswertung:* An Häufigkeitsanalysen können sich außerdem weitergehende statistische Auswertungen mittels Verfahren für nominalskalierte Daten anschließen.

Typenbildung

Überblick

Bei der Typenbildung handelt es sich um ein datenaufbereitendes Verfahren, das dazu dient, einander ähnliche Fälle zu Gruppen bzw. Typen zusammenzufassen. Kriterium der Typenbildung ist es, die Typen so zu erstellen, dass die Fälle innerhalb eines Typus einander möglichst ähnlich sind (*interne Homogenität auf der Ebene des Typus*), die Typen sich dagegen möglichst stark voneinander unterscheiden (*externe Heterogenität auf der Ebene der Typologie*). Unter einem *Typus* versteht man also diejenigen Fälle, die sich aufgrund von Ähnlichkeit zusammenfassen lassen. Die Typen in ihrer Relation untereinander bilden zusammengenommen eine *Typologie*, d.h. eine bestimmte Art und Weise der Unterteilung und Gruppierung der untersuchten Fälle.

Zentraler Arbeitsschritt bei der Typenbildung ist die *Identifikation von Merkmalen*, anhand derer sich eine Gruppierung der Fälle vornehmen lässt. Hierfür stehen sowohl quantitative als auch qualitative Verfahren zur Verfügung. In der qualitativen Forschung erfolgt die Typenbildung jedoch häufig ‚intuitiv‘, d.h. Forscher/innen nehmen nach eigenem Ermessen eine Unterteilung der Fälle vor, die jedoch für andere nicht mehr nachvollziehbar ist.

Empirisch begründete Typenbildung nach Kluge: Ablauf

In den letzten Jahren hat man sich auch in der qualitativen Forschung zunehmend um eine Systematisierung des Prozesses der Typenbildung bemüht. Maßgeblich ist hier Kluges Konzeption einer empirisch begründeten Typenbildung in vier Schritten:

- *Erarbeitung relevanter Vergleichsdimensionen:* Der erste Schritt besteht in der Identifikation relevanter Merkmale bzw. Vergleichsdimensionen. Dieser Schritt kann sowohl eher theoriegeleitet erfolgen (z.B. unter Rückgriff auf relevante Fragen aus dem Interviewleitfaden) als auch eher induktiv unter Durchsicht des Materials. Als relevante Vergleichsdimensionen eignen sich nur solche Dimensionen, auf denen die untersuchten Fälle auch unterschiedliche Ausprägungen aufweisen. Wenn beispielsweise alle Interviewten aus einer Untersuchung eine Frage in derselben Richtung beantworten, dann stellen diese Antworten keine geeignete Gruppierungsdimension dar.
- *Gruppierung der Fälle und Analyse empirischer Regelmäßigkeiten:* Im zweiten Schritt werden die Fälle auf der Grundlage der in Schritt 1 identifizierten Vergleichsdimensionen zu Typen zusammengefasst. Es werden sowohl die einzelnen Typen beschrieben als auch die Typen untereinander verglichen. Von besonderer Bedeutung sind dabei (Un-)Regelmäßigkeiten im Material: Vielleicht zeigt sich, dass die Fälle in einer Gruppe einander nicht nur hinsichtlich der zuvor identifizierten Merkmale ähnlich sind, sondern auch in Bezug auf weitere Aspekte. Ebenso kann sich natürlich herausstellen, dass die gemeinsam gruppierten Fälle einander hinsichtlich anderer Variablen dagegen gerade unähnlich sind.

- *Analyse inhaltlicher Zusammenhänge und Typenbildung:* Im dritten Schritt wird versucht, die in Schritt 2 ermittelten Zusammenhänge sowohl innerhalb als auch zwischen den Gruppen genauer zu analysieren. Das Ziel besteht darin, die empirisch ermittelten Zusammenhänge auch theoretisch zu rekonstruieren und zu verstehen. Gerade bei diesem Schritt kann sich herausstellen, dass neben den in Schritt 1 identifizierten Vergleichsdimensionen noch weitere (oder ggf. überhaupt andere) heranzuziehen sind. In diesem Fall geht man nach Schritt 3 erneut zu Schritt 1 über.
- *Charakterisierung der gebildeten Typen:* Im letzten Schritt erfolgt eine detaillierte Beschreibung der einzelnen Typen und ihrer Zusammenhänge untereinander unter Rückgriff sowohl auf das empirische Material als auch auf theoretische Gesichtspunkte.

Die Typenbildung wird nach Kluge somit nicht als linearer, sondern als zyklischer Prozess aufgefasst: Die Schritte 1 bis 3 werden in der Regel mehrfach durchlaufen, wobei das Material jeweils neu und nach (etwas) anderen Gesichtspunkten zusammengefasst und unterteilt wird. Kriterium für die schlussendlich ausgewählte Form der Gruppierung ist die theoretische Stimmigkeit in Schritt 3. Auch dieses Kriterium bleibt letztendlich vage. Allerdings stellt die Systematisierung des Prozesses der qualitativen Typenbildung einen deutlichen Fortschritt gegenüber rein ‚intuitiven‘ Formen der Typenbildung dar.

Prüfungsliteratur: Mayring 2002, Kap. 4.3.17.

III.4. Qualitative Forschungsansätze (Komplexe Verfahren)

Im Gegensatz zu Methoden, die primär der Datenerhebung oder der Datenanalyse dienen, sind Qualitative Forschungsansätze (wie dies auch für die Forschungsansätze im quantitativen Bereich gilt) durch eine bestimmte Abfolge von Schritten bei der Untersuchungsdurchführung definiert; es wird also ein bestimmtes Vorgehen im Forschungsprozess spezifiziert. In der qualitativen Forschung wird dieser Prozess typischerweise (aber nicht immer) als zyklisch aufgefasst, d.h. Datenerhebung und –auswertung greifen meist ineinander.

Während in der quantitativen Forschung ein bestimmter Ansatz dominiert (nämlich das Experiment), lassen sich in der qualitativen Forschung verschiedenste Ansätze unterscheiden, die zudem auch im Hinblick auf ihren Geltungsbereich erhebliche Unterschiede aufweisen. Manche Ansätze sind so generell konzipiert, dass sie eigentlich eher einen methodologischen Rahmen für qualitative Forschung insgesamt darstellen; dies gilt insbesondere für die Einzelfallanalyse. Andere Ansätze sind primär auf bestimmte Gegenstände ausgerichtet (z.B. Dokumentenanalyse, Biographische Forschung); wieder andere zeichnen sich durch eine bestimmte Werthaltung aus (Handlungsforschung) usw. Qualitative Forschungsansätze sind also ausgesprochen heterogen.

Trotz dieser Heterogenität sehen sich praktisch alle qualitativen Ansätze mit demselben Problem konfrontiert: Weil qualitative im Vergleich zu quantitativer Forschung ungleich aufwändiger ist, können Untersuchungen in der Regel nur an kleinen Stichproben durchgeführt werden (N=20 bis 30 – manchmal mehr, manchmal aber auch weniger). Zufallsstichproben, wie sie in der quantitativen Forschung dominieren, ‚greifen‘ bei diesen kleinen Zahlen nicht. Es stellt sich also bei jeder qualitativen Untersuchung die Frage, nach welchen Kriterien die Stichprobenziehung erfolgen soll.

Im Folgenden werden zunächst Verfahren der qualitativen Stichprobenziehung beschrieben; anschließend werden die Ansätze der Gegenstandsbezogenen Theorienbildung, der Deskriptiven Feldforschung, der Handlungsforschung sowie der Erhebung und Rekonstruktion Subjektiver Theorien genauer dargestellt.

III.4.1. Stichprobenziehung in der qualitativen Forschung

In der qualitativen Forschung lassen sich zwei Strategien der Stichprobenziehung unterscheiden: das selektive sampling und das theoretische sampling.

Selektives Sampling

Das selektive Sampling ist an einem linearen Forschungsprozess ausgerichtet und ähnelt der Quotenstichprobe innerhalb des quantitativen Ansatzes (vgl. oben II.1.2.). Forscher/innen machen sich vor Beginn der Untersuchung Gedanken darüber, welche Faktoren den interessierenden Gegenstandsbereich beeinflussen können. Das Ziel besteht darin, innerhalb der Stichprobe eine Integration aller relevanten Merkmalsausprägungen zu realisieren (was allerdings aufgrund der Vielzahl potenziell relevanter Merkmale nicht immer möglich ist). Wenn ein/e Forscher/in beispielsweise eine Studie zur Situation pflegender Angehöriger planen würde, dann wären beim selektiven Sampling Gesichtspunkte wie Pflegestufe, Verwandtschaftsgrad, Nähe der persönlichen Beziehung, Wohnfläche, Einkommen usw. zu berücksichtigen. Zentrales Merkmal des selektiven Sampling ist es, dass die Kriterien für die Zusammenstellung der Stichprobe bereits vor Untersuchungsbeginn festgelegt werden.

Theoretisches Sampling

Beim theoretischen Sampling handelt es sich um eine Form der Stichprobenziehung, die im Rahmen des qualitativen Forschungsansatzes der Gegenstandsbezogenen Theorienbildung entwickelt wurde (s. III.4.2.). Anders als beim selektiven Sampling werden relevante Merkmale des Gegenstandsbereichs, die die Zusammensetzung der Stichprobe bestimmen, nicht vor Untersuchungsbeginn deduktiv festgelegt, sondern induktiv aus den Daten heraus entwickelt. Die Untersuchung beginnt mit einem beliebigen Fall aus dem Gegenstandsbereich (bei einer Untersuchung zum persönlichen Erleben von Obdachlosigkeit beispielsweise mit einem älteren Mann, der arbeitslos und schon seit längerer Zeit obdachlos ist). Anhand der Auswertung für diesen einen Fall werden hypothetisch Merkmale identifiziert, die das persönliche Erleben von Obdachlosigkeit beeinflussen könnten – beispielsweise die Dauer der Obdachlosigkeit, ob der/die Obdachlose eine Arbeit hat, das Geschlecht usw. Zur Prüfung der Relevanz dieser Merkmale erfolgt die weitere Stichprobenziehung nach dem *Prinzip der minimalen und der maximalen Ähnlichkeit*. Es werden nach dem Prinzip der minimalen Ähnlichkeit solche Fälle in die Stichprobe einbezogen, die dem zuvor untersuchten Fall möglichst unähnlich sind (z.B. eine jüngere Frau, die eine Arbeit hat und erst seit wenigen Wochen obdachlos ist); nach dem Prinzip der maximalen Ähnlichkeit werden außerdem möglichst ähnliche Fälle berücksichtigt. Ziel dieser Vorgehensweise ist es, die Variation im Gegenstandsbereich auszuschöpfen.

Prüfungsliteratur: nur Skript

III.4.2. Gegenstandsbezogene Theorienbildung (Grounded Theory)

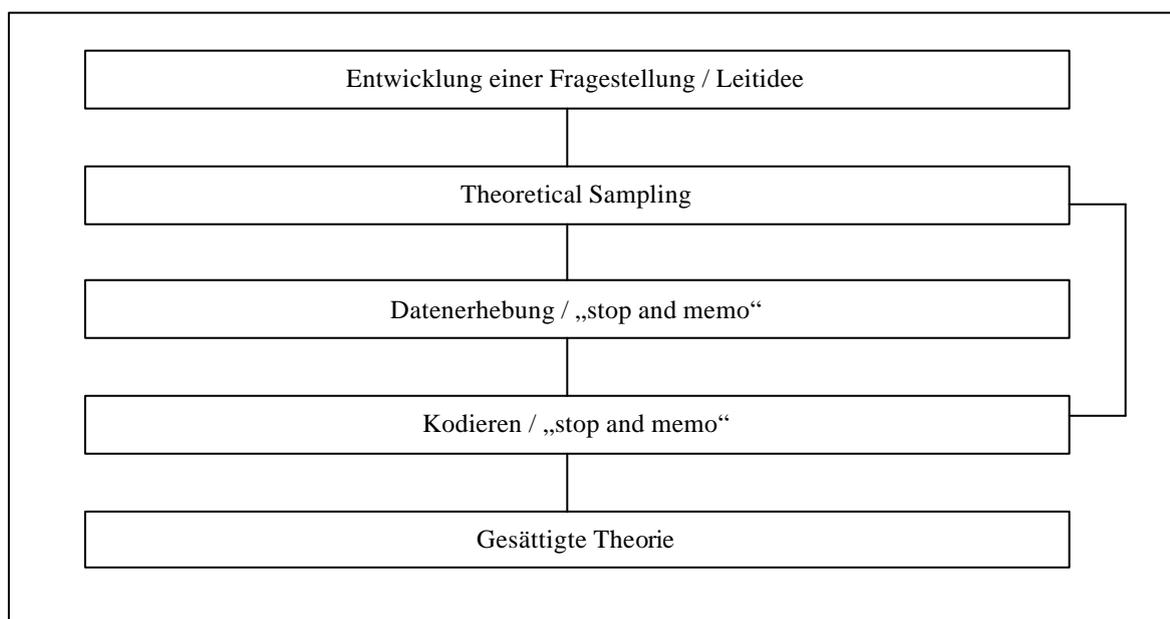
Grundannahmen und Zielsetzung

Das Verfahren der Gegenstandsbezogenen Theorienbildung wurde von Glaser & Strauss (1967) im Rahmen einer Untersuchung zur Interaktion von Klinikpersonal mit Sterbenden entwickelt. Die beiden Forscher wenden sich mit dem Ansatz ganz bewusst gegen das hypothesenprüfende Vorgehen in der quantitativen Forschung.

Sie gehen erstens davon aus, dass Forscher/innen sich auch während der Datenerhebung Gedanken zu ihrem Gegenstand machen und in diesem Zusammenhang ständig neue Hypothesen entwickeln. Nach Ansicht von Glaser & Strauss ist es nicht sinnvoll, die Überprüfung einer a priori festgelegte Hypothese abzuschließen, bevor aufgrund weiterer Überlegungen und Ergebnisse eine neue Hypothese aufgestellt und ihrerseits überprüft wird. Stattdessen gehen sie davon aus, dass Überlegungen bei der Datenerhebung und einer ersten Auswertung unmittelbar zu einer Modifikation von Hypothesen führen müssten. Datenerhebung und –auswertung bei der Gegenstandsbezogenen Theorienbildung überschneiden sich also.

Zweitens gehen Glaser & Strauss davon aus, dass Theorien möglichst ‚datennah‘, möglichst gut in den Daten ‚verankert‘ sein sollten (daher auch die Bezeichnung: Gegenstandsbezogene Theorienbildung). Es findet eine permanente Überprüfung von Hypothesen statt, die immer wieder anhand neuer Daten modifiziert werden. Ziel ist es, auf der Grundlage einer *dichten Beschreibung* zu einer *gesättigten Theorie* über den interessierenden Gegenstandsbereich zu gelangen.

Vorgehen



Ausgehend von einer Fragestellung bzw. *Leitidee* (z.B.: Umgang mit Obdachlosigkeit aus der Sicht Betroffener) wird zunächst ein beliebiger Fall ausgewählt und es werden Daten erhoben. Die Datenerhebung kann mittels verschiedener Methoden erfolgen; typisch ist vor allem die halb- oder nonstandardisierte Befragung.

Die *Auswertung* erfolgt unmittelbar im Anschluss an die Datenerhebung mittels Kodierung und umfasst drei Schritte:

- *Offenes Kodieren*: Hier werden zentrale Themen in der Begrifflichkeit der Befragten festgehalten (z.B. ‚Platte machen‘). Diese erste Form der Kodierung soll möglichst theoriefrei erfolgen.
- *Axiales Kodieren*: Hier werden die offenen Kodierungen in einen theoretischen Zusammenhang gebracht. Die Kodierungen basieren zwar auf den Ausführungen der Befragten, orientieren sich jedoch nicht mehr notwendig an deren Begrifflichkeit. Axiales Kodieren ist abstrakter als das offene Kodieren und bezieht sich meist auf die Äußerungen mehrerer Befragter (z.B.: Sorge um einen Übernachtungsplatz).
- *Theoretisches Kodieren*: Auf dieser Stufe werden alle axialen Codes zu einem Gesamtmodell bzw. einer Theorie integriert. Leitend ist dabei die Basis- bzw. Kernkategorie, die im Sinne einer Leitidee den Ausgangspunkt der Untersuchung darstellt. Hier findet also die eigentliche Theorienbildung statt.

Zu Untersuchungsbeginn wird sich die Kodierung meist auf das offene Kodieren beschränken.

Sowohl bei der Datenerhebung als auch beim Kodieren sind die Forscher/innen angehalten, immer dann, wenn ihnen zu ihrem Vorgehen, ihren Daten usw. etwas auf- oder einfällt, kurz innezuhalten und eine entsprechende Notiz (ein Memo) anzufertigen. Diese Anweisung wird von Glaser & Strauss in dem Prinzip des *stop and memo* zusammengefasst. Die Memos gehen ebenso wie die Codes in die Erarbeitung der Theorie ein.

Auf der Grundlage der offenen Kodierung des ersten Falles erfolgt nach dem Prinzip des *theoretical sampling* eine erneute Fallauswahl und eine erneute Datenerhebung (minimale und maximale Ähnlichkeit). Mit zunehmender Anzahl der Fälle in der Stichprobe gewinnt bei der Auswertung neben dem offenen auch das axiale und schließlich das theoretische Kodieren an Bedeutung.

Abbruchkriterium der Gesättigten Theorie

Diese Vorgehensweise wird so lange fortgesetzt, bis eine Einbeziehung neuer Fälle nach dem Prinzip des *theoretical sampling* nicht mehr zu einer Veränderung der Theorie führt. Neue Fälle führen also beispielsweise nicht mehr dazu, dass neue Bedeutungsaspekte thematisiert werden und eine entsprechende neue Kategorie eingerichtet (oder eine bereits vorhandene weiter unterteilt) werden muss. Eine solche gegenstandsbezogene Theorie, die die gesamte Variation in ihrem Gegenstandsbereich abbildet, gilt als *theoretisch gesättigt*.

Prüfungsliteratur: Bortz & Döring 2002, Kap. 5.3.2.; Mayring 2002, Kap. 4.3.11.

III.4.3. Deskriptive Feldforschung

Überblick

Der Ansatz der Deskriptiven Feldforschung stammt ursprünglich aus der Ethnologie (z.B. Malinowskis Feldforschung zu den Tobriandern, einem Volk aus der Südsee) und der qualitativ orientierten Soziologie (z.B. die Untersuchungen von White zur Obdachlosigkeit).

Zielsetzung ist es, die Innenperspektive einer anderen Kultur kennenzulernen. Dabei sollen Verzerrungen des Gegenstandes möglichst vermieden werden, indem der Gegenstand in seinem natürlichen Umfeld belassen wird und die Forscher/innen den Gegenstand möglichst nicht durch Eingriffe verändern.

Zentrale *Methode* der Deskriptiven Feldforschung ist die Teilnehmende Beobachtung.

Abgrenzung: Die Deskriptive Feldforschung als qualitativer Ansatz ist – trotz der Ähnlichkeit der Bezeichnungen – etwas völlig anderes als das Feldexperiment oder die Feldstudie in der

quantitativen Forschung. Bei Feldstudie oder –experiment ist das Feld lediglich der Ort, an dem eine Untersuchung stattfindet; bei der Deskriptiven Feldforschung ist das Feld dagegen selbst Teil des Untersuchungsgegenstandes.

Vorgehen

Die Deskriptive Feldforschung gliedert sich in die folgenden Phasen:

- *Festlegen der Fragestellung:* Prinzipiell kann eine Kultur unter vielen verschiedenen Gesichtspunkten untersucht werden. Damit die Untersuchung nicht ‚ausufert‘, sind vor Untersuchungsbeginn thematische Schwerpunkte zu setzen. In diese Schwerpunktsetzung gehen auch Überlegungen zur Realisierbarkeit der geplanten Untersuchung ein. Dies betrifft insbesondere Überlegungen zur Zugänglichkeit des Feldes und zur Rolle, die ein/e Forscher/in hier potenziell einnehmen kann.
- *Herstellen des Feldkontakts:* Dies gilt als die schwierigste und sensibelste Phase der Feldforschung. Die Möglichkeiten zur Herstellung des Feldkontakts sind nicht zuletzt davon abhängig, ob es sich bei dem zu untersuchenden Feld um einen geschlossenen oder einen offenen Schauplatz handelt. Ein offener Schauplatz ist prinzipiell für Außenstehende zugänglich (z.B. Fußballplatz, Kinderspielplatz, Friedhof), ein geschlossener Schauplatz dagegen zumindest nicht ohne weiteres (z.B. Sekte, Gefängnis). Wenn die geplante Untersuchung an einem geschlossenen Schauplatz stattfinden soll, ist in der Regel eine Einführung der Forscherin oder des Forschers durch einen *Türhüter* erforderlich. Bei einem Türhüter handelt es sich um eine Person, die selbst Teil des Feldes ist. Der Türhüter vermittelt den Forschern/innen Informationen über das Feld und vermittelt erste Kontakte; dies gelingt umso besser, je höher das Ansehen ist, das der Türhüter im Feld genießt. Die Forscher/innn selbst sollten Interesse an den Aktivitäten im Feld zeigen, Personen im Feld ansprechen und auf diese Weise Feldkontakte anbahnen und aufbauen.
- *Materialsammlung:* Bei der Materialsammlung im Feld wird meist eine Kombination verschiedener Methoden eingesetzt, wie beispielsweise (teilnehmende) Beobachtung, Interviews, Dokumentenanalyse usw. Gemeinsam ist diesen Methoden, dass eher halb- und non-standardisierte Varianten der Verfahren zur Anwendung kommen. Die Datenerhebung erfolgt zunächst breit gestreut, dann zunehmend fokussierter. Das tagsüber gesammelte Material wird abends zusammenfassend protokolliert. Insbesondere halten Forscher/innen im Anschluss an jeden Tag im Feld ihre persönlichen Eindrücke fest (gab es Spannungen mit einzelnen Personen; ergeben sich für die Forscher/innen Wertkonflikte usw.). Es wird davon ausgegangen, dass die Forscher/innen mit zunehmender Länge ihres Aufenthalts im Feld allmählich ‚unsichtbar‘ werden und ihre Anwesenheit sich folglich nicht negativ auf die Validität der Untersuchungsergebnisse auswirkt. Zentrale Methode zur Datenerhebung bei der Deskriptiven Feldforschung ist die *teilnehmende Beobachtung*. Im Unterschied zu anderen Beobachtungsverfahren wird der/die Forscher/in hier selbst Teil des Feldes, nimmt also im Feld eine aktive Rolle ein. Diese Methode ist es, die in erster Linie ein Kennenlernen des Feldes aus der Innenperspektive ermöglicht. Wie auch bei anderen Formen der nonstandardisierten Beobachtung erfolgt auch die teilnehmende Beobachtung nicht anhand eines Kategoriensystems; es werden lediglich Beobachtungsdimensionen vorgegeben. Die teilnehmende Beobachtung kann offen oder verdeckt erfolgen. Im Fall einer verdeckten Beobachtung sind Forscher/innen jedoch verpflichtet, die betroffenen Personen im Feld vor einer Veröffentlichung über das gesammelte Material zu informieren und ihre Zustimmung zur Veröffentlichung einzuholen.

- *Ausstieg aus dem Feld:* Nach Abschluss der Materialsammlung erfolgt der Ausstieg aus dem Feld. Dieser vollzieht sich ebenso allmählich und sukzessive wie zuvor der Einstieg.
- *Auswertung:* Die Auswertung umfasst zwei Teilbereiche: die Protokollierung und die Aufbereitung des gesammelten Materials.
Wie bei anderen qualitativen Verfahren auch, verlaufen Datenerhebung und –auswertung nicht getrennt, sondern greifen ineinander. Im Kontext der Materialsammlung wurde bereits erwähnt, dass im Anschluss an jeden Forschungstag ein *Protokoll* angefertigt wird. Dieses hat mehrere Funktionen: Es werden potenzielle Schwerpunkte für die weitere Datenerhebung sichtbar; Forscher/innen lernen, zunehmend präzise wahrzunehmen; das Protokoll hilft den Forschern/innen, die Eindrücke des Tages zu verarbeiten und über ihre eigenen Reaktionen und Gefühle Klarheit zu gewinnen.
Die Aufbereitung schließt sich an die eigentliche Feldphase an. Hier wird das Material zusammengefasst und es werden relevante Teile für eine Veröffentlichung ausgewählt.

Probleme

- *Einstieg ins Feld:* Ein erstes Problem Deskriptiver Feldforschung wurde bereits erwähnt: Der Einstieg ins Feld ist in der Regel schwierig.
- *Reaktionen auf das Feld:* Da der Untersuchungsgegenstand Deskriptiver Feldforschung eine andere als die eigene Kultur ist, ist zu erwarten, dass die Wertvorstellungen dieser anderen Kultur von der eigenen abweichen. In solchen Fällen ergibt sich das Problem, dass die Forscher/innen zu Zeugen von Handlungsweisen werden können, die sie selbst missbilligen, oder sogar aufgefordert sind, sich an entsprechenden Handlungen zu beteiligen (z.B. Feldforschung zu rechtsextremen Gruppen; Umgang mit geistig und körperlich behinderten Menschen usw.).
- *Gefahr des Distanzverlusts (,going native‘):* Da Forscher/innen im Rahmen der Methode der teilnehmenden Beobachtung selbst zu einem Teil des Feldes werden und über längere Zeit in diesem Feld agieren, besteht die Gefahr, dass sie die Distanz zu ihrem ‚Forschungsgegenstand‘ verlieren, ganz im Feld aufgehen und sich mit dessen Werten identifizieren; die andere Kultur wird dann zur eigenen. Wichtige Gegenmaßnahmen bestehen zum einen darin, die Forschung nicht als Einzelperson durchzuführen, sondern in einem Team. Zum anderen kann durch regelmäßige Supervision einem solchen Distanzverlust gegengesteuert werden
- *Ausstieg aus dem Feld:* Je besser die Integration der Forscher/innen ins Feld, desto schwieriger gestaltet sich (für beide Seiten) der Ausstieg. Dabei ist insbesondere der Doppelcharakter der im Feld aufgebauten Beziehungen von Bedeutung: Was für die Menschen im Feld Teil ihrer alltäglichen Interaktionen ist, ist für die Forscher/innen Materialsammlung. Forscher/innen sollten den Personen im Feld nicht das Gefühl geben, lediglich ‚Materiallieferanten‘ zu sein.
- *Auswertung – Aufwand und Materialfülle:* Die Auswertung der im Feld gesammelten Daten ist sehr aufwändig: Pro Stunde im Feld sind ungefähr vier Stunden Aufbereitungsarbeit anzusetzen. Dieses aufbereitete Material ist wiederum so umfangreich, dass es nur in Ausschnitten für eine Publikation der Ergebnisse herangezogen werden kann. Die Ergebnisdarstellung ist also notwendig subjektiv und selektiv.

Anwendungsbedingungen Deskriptiver Feldforschung

Mayring (2002) nennt die folgenden Anwendungsbedingungen Deskriptiver Feldforschung:

- Das Feld muss zugänglich sein.
- Die Forscher/innen müssen im Feld eine sinnvolle Funktion einnehmen können.
- Die Forscher/innen sollten darin geschult sein, die Balance zwischen Anteilnahme und kritischer Distanz aufrechtzuerhalten.

- Das Vorhaben sollte ethisch gerechtfertigt sein; bloße Neugier auf eine andere Lebensweise stellt keine hinreichende Rechtfertigung dafür dar, andere Menschen dazu zu veranlassen, einen so tiefen Einblick in ihr Leben zu geben.

Prüfungsliteratur: Bortz & Döring, Kap. 5.4.1.; Mayring 2002, Kap. 3.4.

III.4.4. Handlungsforschung (Aktionsforschung)

Zentrale Merkmale

Die Handlungsforschung geht auf Lewin zurück, der in den 1940ern die Diskriminierung von Minderheiten ‚vor Ort‘ (z.B. in Fabriken) untersuchte und dabei zugleich auch Veränderungsstrategien entwickelte.

Zentrale Merkmale der Handlungsforschung sind:

- *Problembezug:* Handlungsforschung ist immer sozial- und gesellschaftskritisch und setzt an konkreten, sozialen Problemen an. Daraus ergeben sich zugleich die zentralen Unterschiede zur quantitativ orientierten Forschung: Handlungsforschung berücksichtigt explizit die Subjektivität und historische Gebundenheit sowohl der Forscher/innen als auch der erforschten Personen. Die Werturteilsfreiheit herkömmlicher Forschung wird kritisiert, weil sie aus der Sicht der Handlungsforschung geltenden Herrschaftsstrukturen in die Hände spielt. Dahinter steht die Idee einer kritischen Sozialwissenschaft, die sich an der Lösung gesellschaftlicher Probleme beteiligt.
- *Praxisveränderung:* Die Ergebnisse von Handlungsforschung werden noch während des Forschungsprozesses in die Praxis umgesetzt. Forschung wird als Lern- und Veränderungsprozess sowohl für die Forscher/innen als auch die erforschten Personen konzipiert. Schlussendliches Ziel ist es, die Kompetenzen der untersuchten Personen so zu erweitern, dass sie ihr gesellschaftliches Interesse selbst vertreten können (z.B. durch Gründung von Selbsthilfegruppen, Anschluss an soziale Bewegungen usw.).
- *Gleichberechtigter Diskurs:* Für die Handlungsforschung ist weiterhin charakteristisch, dass die erforschten Personen als gleichberechtigte Partner/innen der Forscher/innen angesehen werden und an allen Phasen des Forschungsprozesses mit beteiligt sind.

Vorgehen

Es werden zunächst von den Forschern/innen und den erforschten Personen gemeinsam eine Problem- und eine Zieldefinition erarbeitet.

Daran schließt sich der Projektablauf als ein Pendeln zwischen Informationssammlung, Diskurs und praktischem Handeln an. Im Zentrum des Verfahrens steht der Diskurs, für den die zuvor gesammelten Informationen als Material dienen. Im Diskurs werden diese Informationen problematisiert und mit anderem Wissen konfrontiert. Ziel des Diskurses ist die Ausarbeitung von Handlungsorientierungen, die wiederum die Basis für die folgenden praktischen Handlungen darstellen. Über diese Handlungen, deren Wahrnehmung, Folgen usw. werden erneut Informationen gesammelt usw.

Im Rahmen der Informationssammlung können unterschiedliche Methoden zur Anwendung kommen, z.B. offene teilnehmende Beobachtung, Gruppendiskussion, Dokumentenanalyse usw. Ausgeschlossen sind nur solche Methoden, die Distanz zwischen den Forschern/innen und den erforschten Personen schaffen oder die die zu erforschende Situation in irgendeiner Weise festschreiben.

Probleme

- Die gesellschaftliche Einbettung sozialer Probleme setzt den Veränderungsmöglichkeiten von Handlungsforschung Grenzen (Gesetze, Regeln in Institutionen usw.).

- Eine Zusammenarbeit von Forschern/innen und erforschten Personen stößt dort an ihre Grenzen, wo ein Problembewusstsein nur seitens der Forscher/innen vorhanden ist, nicht aber seitens der erforschten Personen.
- Es werden hohe Anforderungen an die erforschten Personen gestellt; die Güte der Forschung ist wesentlich von deren Kompetenz abhängig. Hier stellt sich die Frage, inwieweit Forschungsergebnisse grundsätzlich von der Zustimmung der erforschten Personen abhängig gemacht werden können und sollen.
- Handlungsforschung ist auf ‚sympathische Benachteiligte‘ beschränkt (die meisten Forscher/innen würden sich beispielsweise nicht auf einen gleichberechtigten Diskurs mit neonazistischen Gruppierungen einlassen).

Fazit

Forschungsprojekte, die ausschließlich auf Handlungsforschung basieren, sind in Europa selten. In Südamerika oder Australien finden sich aber durchaus entsprechende Projekte (z.B. zur Unterstützung der Organisation der Aborigines in Australien).

In Europa und den USA werden in der Regel lediglich einzelne Elemente aus der Handlungsforschung in andere Projekte integriert; z.B. wurde der Gedanke der Rückmeldung an die erforschten Personen in der angewandten Arbeits-, Organisations- und Betriebspsychologie aufgenommen.

Prüfungsliteratur: Bortz & Döring 2002, Kap. 5.4.2.; Mayring 2002, Kap. 3.3.3.

III.4.5. Forschungsprogramm Subjektive Theorien (FST)

Überblick

Das Forschungsprogramm Subjektive Theorien wurde in den 1970ern von Groeben & Scheele entwickelt.

In dem Ansatz wird davon ausgegangen, dass Menschen im Alltag ebenso wie Wissenschaftler/innen bemüht sind, sich selbst und die Welt um sich herum zu verstehen, erklären und ggf. auch zu verändern. Solche Gedanken über das Selbst oder die eigene Welt werden als Subjektive Theorien bezeichnet (s. genauer unten).

Gegenstand des Forschungsprogramms ist entsprechend die Erhebung, Rekonstruktion und Validierung solcher Subjektiver Theorien von Menschen im Alltag.

Der Ansatz umfasst zwei Forschungsphasen, eine eher ‚qualitative‘ und eine eher ‚quantitative‘. Das FST ist daher weder dem qualitativen noch dem quantitativen Ansatz eindeutig zuzuordnen.

Subjektive Theorien (STs): Definition

Subjektive Theorien sind definiert als:

- *Kognitionen der Selbst- und Weltsicht:* Gedanken, die sich der Mensch im Alltag über sich selbst, andere Menschen und Ereignisse in der Welt macht;
- *die im Dialog-Konsens aktualisier- und rekonstruierbar sind:* Dieses Definitionsmerkmal bezieht sich auf die erste Forschungsphase der Erhebung und Rekonstruktion Subjektiver Theorien, die im Folgenden noch genauer erläutert wird. Damit eine Erhebung und Rekonstruktion von STs möglich ist, müssen diese kognitiv zugänglich sein; außerdem muss es der befragten Person möglich sein, die Gedanken in Worte zu fassen. Es wird allerdings nicht davon ausgegangen, dass Menschen vor Untersuchungsbeginn ihre Subjektive Theorien sozusagen ‚fertig im Kopf‘ haben; die Rekonstruktion erfolgt vielmehr gemeinsam mit dem/r Forscher/in im Verlauf der Untersuchung.

- *als komplexe Aggregate mit zumindest impliziter Argumentationsstruktur*: Mit STs sind nicht beliebige Gedanken gemeint, sondern solche, die untereinander in einem Zusammenhang stehen, z.B. durch Kausalannahmen, Definitionen usw. miteinander verbunden sind.
- *das auch die zu objektiven wissenschaftlichen Theorien parallelen Funktionen*: Es wird angenommen, dass Subjektive Theorien sich nicht grundsätzlich von objektiven wissenschaftlichen Theorien unterscheiden.
- *der Erklärung, Prognose, Technologie erfüllt*: Die Analogie zwischen Subjektiven und objektiven Theorien zeigt sich beispielsweise darin, dass Subjektive Theorien für Menschen im Alltag vergleichbare Funktionen erfüllen wie objektiv-wissenschaftliche Theorien: Sie erlauben eine Erklärung von Sachverhalten; sie erlauben es, Sachverhalte vorherzusagen und auch zu verändern.
- *und deren Eignung als objektive Theorie zu prüfen ist*: Subjektive Theorien können, ebenso wie wissenschaftliche Theorien, allerdings auch falsch sein (z.B. die Hypothese, dass das Tragen eines Magnetarmbandes vor Gesundheitsproblemen schützt). Subjektive sind daher ebenso wie objektive Theorien auf ihre Gültigkeit zu prüfen. Dieses Merkmal bezieht sich auf die zweite Forschungsphase.

Subjektive Theorien: Arten

- Subjektive Theorien unterscheiden sich erstens nach ihrer *Reichweite*. STs kurzer Reichweite beziehen sich auf Handlungen und Handlungssequenzen (z.B. Lehrerhandlungen bei Aggression seitens der Schüler/innen). STs mittlerer Reichweite beziehen sich beispielsweise auf Konzepte wie Einsamkeit, Ironie, Zivilcourage. In STs großer Reichweite sind umfassende Fragen wie z.B. Gott, Sinn des Lebens o.ä. thematisch.
- Subjektive Theorien unterscheiden sich zweitens nach der *Art des Wissens* bzw. der Zusammenhänge zwischen den verschiedenen Wissensinhalten. STs können sich beispielsweise auf definitorische Zusammenhänge beziehen, auf empirische, auf wertende oder auch auf Handlungswissen.

Die zwei Forschungsphasen

- Die erste (,qualitative‘) Phase der *Kommunikativen Validierung* dient der Erhebung und Rekonstruktion Subjektiver Theorien. Zentrales Gütekriterium ist das adäquate Verstehen der ST durch den/die Forscher/in. Diese Phase ist vom Forschungsablauf her die erste. Sie ist daher der zweiten Phase vorgeordnet, ihr geltungstechnisch jedoch untergeordnet.
- Die zweite (,quantitative‘) Phase der *explanativen Validierung* dient der Prüfung, ob die Subjektive als objektiv-wissenschaftliche Theorie akzeptierbar ist, ob die Annahmen der Theorie sich also auch empirisch bestätigen lassen. Zentrales Gütekriterium ist das der Falsifikation. Die explanative ist der kommunikativen Validierung zeitlich nachgeordnet (die ST ist zu rekonstruieren, bevor sie geprüft werden kann), geltungstechnisch jedoch nachgeordnet.

Untersuchungen innerhalb des FST realisieren nicht notwendig beide Forschungsphasen. Wenn das Forschungsinteresse sich auf die Rekonstruktion Subjektiver Theorien zu einem bestimmten Gegenstandsbereich richtet, ohne dass die Gültigkeit dieser Theorien als objektive thematisch ist, dann beschränkt sich die Untersuchung auf die erste Phase. Im Folgenden wird nur auf die Phase der Kommunikativen Validierung genauer eingegangen.

Die Phase der Kommunikativen Validierung: Ablauf

Die Phase der Kommunikativen Validierung gliedert sich ihrerseits in zwei Schritte:

- In einem ersten Schritt werden die Inhalte der Subjektiven Theorie im Rahmen eines halbstandardisierten Interviews erhoben und von der Forscherin oder dem Forscher auf Kärtchen geschrieben.

- In einem zweiten Schritt wird die Theiestruktur rekonstruiert und im Dialog zwischen Forscher/in und erforschter Person validiert. Die Rekonstruktion erfolgt unter Rückgriff auf sog. Struktur-Lege-Verfahren. Ergebnis dieses zweiten Schrittes ist ein Struktur-Bild. Dieser Schritt stellt den zentralen Unterschied zwischen der Erhebung und Rekonstruktion Subjektiver Theorien einerseits und dem halbstandardisierten Interview andererseits dar.

Struktur-Lege-Verfahren

Struktur-Lege-Verfahren beinhalten erstens Relationen, mit denen sich Theorieinhalte untereinander verbinden lassen (z.B. *das ist / das heißt, führt zu, indem, z.B. / so wie, soll sein*) und zweitens Regeln zu graphischen Darstellung der Theiestruktur. Es existieren verschiedene Struktur-Lege-Verfahren für unterschiedliche Wissensbereiche: Verfahren zur Rekonstruktion von Definitions- und empirischem Wissen (z.B. Heidelberger Struktur Lege-Technik: SLT), von Handlungswissen (z.B. Weingartener Appraisal Legetechnik WAL), von Bewertungen (Ziel-Mittel-Argumentation: ZMA). Weiterhin existiert eine Alltagssprachliche Version, in der Relationen zur Rekonstruktion verschiedener Wissensarten flexibel miteinander kombiniert werden können.

Kommunikative Validierung: Dialog-Konsens

Bei einem Befragungsverfahren, das der Erhebung der Innensicht, dem Verstehen des von der befragten Person subjektiv Gemeinten gilt, besteht grundsätzlich die Gefahr, dass Forscher/innen meinen, ihr Gegenüber adäquat verstanden zu haben, ohne dass dies auch tatsächlich der Fall ist. Dieser Gefahr begegnet das FST, indem im Rahmen der Erhebung und Rekonstruktion Subjektiver Theorien explizit eine Absicherung des Verständnisses seitens der Forscher/innen vorgesehen ist: die kommunikative Validierung (nach der die erste Phase des FST auch benannt ist).

Der Begriff der Kommunikativen Validierung wurde von Lechler (1982) geprägt. Lechler bezeichnet damit ein Verfahren, in dem sich Forscher/in und befragte Person über die Interpretation des Gesagten einigen.

Im FST erfolgt eine Präzisierung des Kriteriums, wann das Verständnis des Forschers oder der Forscherin als adäquat gelten kann. Kriterium für die Angemessenheit des Verstehens ist der *Dialog-Konsens*. Dahinter steht die Überlegung, dass die Kommunikation zwischen Forscher/in und erforschter Person meist eine asymmetrische ist. Es ist daher grundsätzlich nicht auszuschließen, dass befragte Personen nur vorgeben, die Interpretation des von ihnen Gesagten durch den/die Forscher/in sei adäquat – während sie in der Tat etwas ganz anderes gemeint haben (sich aber beispielsweise nicht trauen, dies in der Forschungssituation auch offen zu sagen).

Im FST wird daher versucht, solche Kommunikationszwänge, die zu ‚falscher Zustimmung‘ seitens der befragten Person führen können, möglichst auszuschließen. Dies geschieht unter Rückgriff auf Habermas‘ Konzept der *idealen Sprechsituation des Diskurses*. In einer idealen Sprechsituation sind keine Systemzwänge, keine Asymmetrien wirksam; es resultiert eine wahrhaftige Kommunikation der Beteiligten. Ein Konsens im Dialog zwischen Forscher/in und erforschter Person kann also in dem Maß als Wahrheitskriterium gelten, in dem der Dialog sich einer idealen Sprechsituation annähert.

Habermas geht selbst davon aus, dass eine solche ideale Sprechsituation nicht vollständig realisierbar ist – es sind immer nur Annäherungen möglich. Entsprechend betrachtet Habermas die ideale Sprechsituation als eine *regulative Zielidee*. Als solche wird sie auch im FST konzipiert. Es wurden Maßnahmen entwickelt, die eine möglichst große Annäherung der Kommunikation zwischen Forscher/in und erforschter Person über deren Subjektive Theorie an eine ideale Sprechsituation ermöglichen sollen. Dies sind insbesondere: *Transparenz* (der Fragestellung, der Zielsetzung der Untersuchung usw.) und *Metakommunikation* (Kommuni-

kation über die Kommunikation, z.B. die Atmosphäre, eventuelle Unsicherheiten seitens der befragten Person usw.).

Phase der Kommunikativen Validierung: Zusammenfassung

In der Phase der Kommunikativen Validierung werden zunächst in einem halbstandardisierten Interview die Inhalte der Subjektiven Theorie einer befragten Person erhoben. Anschließend wird der befragten Person ein Struktur-lege-Leitfaden ausgehändigt, in dem die Relationen beschrieben sind, die für die Rekonstruktion der ST maßgeblich sind; die befragte Person wird gebeten, sich mit dem Leitfaden vertraut zu machen.

Die Rekonstruktion der Theoriestructur erfolgt ca. eine Woche nach dem Leitfadeninterview. Der/Die Forscher/in bereitet sich vor, indem er/sie die zentralen Konzepte aus dem Interview auf Kärtchen schreibt und einen Vorschlag zur Rekonstruktion der Theoriestructur vorbereitet.

In der Rekonstruktions-Sitzung legt der/die Forscher/in der befragten Person zunächst die vorbereiteten Konzept-Kärtchen vor; diese werden ggf. nach Maßgabe der befragten Person noch verändert. Dann wird die befragte Person gebeten, selbst einen Vorschlag zur Rekonstruktion ihrer ST zu legen; dabei wird sie von der Forscherin oder dem Forscher unterstützt. Forscher/in und befragte Person vergleichen dann die beiden Theorie-Versionen (vorbereitete Version von Forscher/in und in der Sitzung erstellte Version der befragten Person) und modifizieren sie im Dialog so lange, bis eine konsensuale Version resultiert, die auch die befragte Person als adäquate Rekonstruktion ihrer Subjektiven Theorie empfindet (Dialog-Konsens). Maßgeblich für die Rekonstruktion ist die möglichst freie Zustimmung der befragten Person.

Verfahren der explanativen Validierung

Im FST sind verschiedene Untersuchungsdesigns entwickelt worden, die eine Überprüfung von STs über das eigene Handeln erlauben (Handlungsleitung):

- *Korrelationsstudien*: Es wird der Zusammenhang zwischen der ST und den tatsächlichen Handlungsweisen der befragten Person berechnet. Bei Untersuchungen dieser Art besteht jedoch das Problem, dass die befragte Person u.U. bemüht ist, im Sinne der Selbstdarstellung ganz bewusst Konsistenz zwischen Theorie und Handeln zu schaffen.
- *Prognosestudien*: Auf der Grundlage der ST werden Voraussagen über Handlungsklassen in bestimmten Situationsklassen getroffen. Es wird geprüft, inwieweit die tatsächlichen Handlungsweisen den prognostizierten entsprechen.
- *Modifikationsstudien*: Bei dieser komplexesten Form der Validierungsstudie wird die ST nach der Erhebung und Rekonstruktion ganz gezielt verändert. In dem Maß, in dem die ST handlungsleitend ist, müssten sich auch die Handlungen der befragten Personen in spezifischer Weise ändern.

FST: Voraussetzungen

- *Bewusstseinsfähigkeit*: Groeben & Scheele gehen nicht davon aus, dass das FST für sämtliche Gegenstandsbereiche der Psychologie Geltung besitzt. Wie bereits in der Definition von STs deutlich wird, wird hier vorausgesetzt, dass die zu erhebenden Kognitionen bewusstseinsfähig sind. Dies gilt beispielsweise für Reflexe überhaupt nicht, für automatisiertes und/oder hochgradig verdichtetes Wissen nur in begrenztem Maß.
- *Verbale Kompetenz*: Zweitens ist eine Erhebung und Rekonstruktion von STs nur dann möglich, wenn die befragten Personen ihre ST auch verbalisieren können.
- *Abstraktes Denken*: Schließlich sind die meisten Personen im Alltag es nicht gewohnt, ihre Kognitionen in eine graphische Struktur zu bringen. Hier stellt die alltagssprachliche

- Version der Heidelberger SLT bereits eine Erleichterung gegenüber früheren Formen des Verfahrens dar. Außerdem ist es wichtig, dass die Forscher/innen die befragten Personen beim Legen ihrer Strukturen unterstützen.

Prüfungsliteratur: Dann 1992; Scheele 1995

III.5. Gütekriterien qualitativer Forschung (einschließlich Ethik)

Was die Beurteilung der Qualität qualitativer Forschung betrifft, so finden sich unter den (qualitativen) Forschern/innen zwei verschiedene Positionen:

- Die einen sind der Ansicht, dass die Gütekriterien des quantitativen Ansatzes prinzipiell auch auf qualitative Forschung übertragbar sind; es sind lediglich gewisse Anpassungen erforderlich.
- Die anderen vertreten dagegen die Meinung, dass qualitative Forschung sich hinsichtlich Zielsetzung und Vorgehen so stark von der quantitativen Forschung unterscheiden, dass es erforderlich ist, eigenständige Gütekriterien qualitativer Forschung zu entwickeln.

III.5.1. Übertragbarkeit quantitativer Kriterien auf qualitative Forschung

- *Objektivität:* Da die Forschungssituation in der qualitativen Forschung wesentlich als soziale Situation konzipiert ist, ist die quantitative Auffassung von Objektivität (Unabhängigkeit von der Person des Forschers/der Forscherin) in der qualitativen Forschung nur schwer realisierbar (von Ausnahmen wie der Inhaltsanalyse abgesehen). Es finden sich jedoch durchaus auch Überlegungen zur Annäherung an Objektivität auch in der qualitativen Forschung:
 - *innere Vergleichbarkeit:* Manche qualitativen Forscher/innen vertreten die Auffassung, dass bei der Datenerhebung nicht die äußere, sondern die innere Vergleichbarkeit der Situationen ausschlaggebend sei. Da verschiedene Menschen verschiedene Situationen je anders erleben, könne man Objektivität nicht dadurch herstellen, dass Forscher/innen sich verschiedenen Personen gegenüber in identischer Weise verhalten. Erforderlich sei vielmehr ein personenspezifisch je unterschiedliches Verhalten der Forscher/innen, um auf Seiten der Befragten dieselbe innere Situation zu erzeugen (z.B. eine Atmosphäre des Vertrauens). Dieser Auffassung zu Folge ‚emergiert‘ Objektivität gerade aus der Subjektivität der Interaktionsrelationen. Diese Konzeption der inneren Vergleichbarkeit verschiedener Untersuchungssituationen entspricht (hinsichtlich der Zielsetzung) der quantitativen Konzeption von Durchführungsobjektivität.
 - *Transparenz und Nachvollziehbarkeit:* Qualitative Forscher/innen sind zunehmend bemüht, die Auswertung und Interpretation ihrer Forschung so zu gestalten, dass sie für andere nachvollziehbar ist (s. ausführlich unten). Dies entspricht einer Annäherung an Durchführungs- und Interpretationsobjektivität.
- *Reliabilität:* In der qualitativen Forschung wird die Einzigartigkeit jeder (Forschungs-)Situation betont; dabei wird z.T. auch in Kauf genommen oder sogar angestrebt, dass die erforschten Personen sich im Verlauf des Forschungsprozesses verändern. Vor diesem Hintergrund wird die quantitative Konzeption der Wiederholungsreliabilität aus qualitativer Sicht abgelehnt. Das bedeutet jedoch nicht, dass die Ergebnisse qualitativer Forschung beliebig sein können. Überzeugende Konzepte von Reliabilität aus qualitativer Sicht wurden bisher jedoch kaum entwickelt. Eine Ausnahme stellt lediglich die Interrater-Übereinstimmung aus der Inhaltsanalyse dar, die eine Übertragung des quantitativen Reliabilitätsbegriffs auf ein gemischt qualitativ-quantitatives Verfahren beinhaltet.

- *Validität*: Dem Gütekriterium der Validität kommt im Rahmen qualitativer Forschung zentrale Bedeutung zu. Allerdings wird der Validitätsbegriff dabei sehr unspezifisch verwendet, z.B. im Sinne der generellen ‚Richtigkeit‘ der Forschungsergebnisse, der externen oder selbst der internen Validität⁶; abgelehnt wird lediglich der Begriff der Kriteriumsvalidität.

Im Kontext der quantitativen Forschung wurde zwischen Konzepten der Validität unterschieden, die sich auf Forschungsansätze beziehen (interne und externe Validität), und solchen, die sich auf einzelne Verfahren beziehen (Wird das gemessen, was gemessen werden soll?). Darunter sind in der qualitativen Forschung von besonderer Bedeutung:

- *Validität von Forschungsansätzen*: In der qualitativen Forschung spielt die externe Validität einer Untersuchung eine zentrale Rolle. Das Gütekriterium findet sich in Form der ‚Gegenstandsnahe und -adäquanz‘ bereits in den Postulaten qualitativer Forschung wieder.
- *Validität von Methoden der Datenerhebung*: In der qualitativen Forschung wird bewusst versucht, Verzerrungen bei der Datenerhebung entgegenzuwirken, und zwar sowohl Verzerrungen seitens der erforschten Person als auch seitens der Forscher/innen. Wichtigste Gegenmaßnahme ist die Schulung von Forscher/innen, die lernen, z.B. keine Suggestivfragen zu stellen, eine vertrauensvolle Gesprächsatmosphäre herzustellen, ihre eigenen Wertungen nicht in die Datenerhebung einfließen zu lassen usw.
- *Validität von Auswertung und Interpretation*: Wie dies bereits im Kontext der qualitativen Auswertungsverfahren erläutert wurde, sind diese vielfach subjektiv und damit ggf. auch nicht valide (Interpretationen werden beispielsweise dem tatsächlich Gesagten und Gemeinten nicht gerecht). Hier wurden von qualitativer Seite in erster Linie zwei Strategien zur Validitätssicherung entwickelt:
 - *Konsens der Forscher/innen*: Die Auswertung erfolgt nicht durch eine/n einzelne/n Forscher/in, sondern in der Gruppe. Das genaue Vorgehen kann von einem informellen Gespräch bis hin zur Berechnung von Interrater-Übereinstimmungen reichen.
 - *Kommunikative Validierung*: Die Ergebnisse werden den untersuchten Personen rückgemeldet; es wird geprüft, ob die forscherseitige Rekonstruktion dem Gemeinten entspricht (z.B. im Dialog-Konsens). Diese Form der Validierung ist allerdings nur dann sinnvoll, wenn das Ziel der Untersuchung darin besteht, die Innensicht der Befragten zu rekonstruieren. Wenn die Untersuchung dagegen auf den Handlungen zugrunde liegende Strukturen, unbewusste Motive usw. abzielt, ist eine kommunikative Validierung nicht weiterführend.

Prüfungsliteratur: nur Skript

III.5.2. Eigenständige Gütekriterien qualitativer Forschung

Im vorausgehenden Abschnitt wurde bereits deutlich, dass quantitative Gütekriterien von der qualitativen Forschung meist nicht unverändert übernommen wurden. Neben Adaptationen quantitativer Kriterien für die qualitative Forschung existieren eine Reihe von Versuchen, eigenständige Gütekriterien qualitativer Forschung zu entwickeln. Ein verbindlicher Kanon

⁶ Genau genommen ist es nicht sinnvoll, in der qualitativen Forschung von interner Validität zu sprechen: Interne Validität ist definiert als das Ausmaß, in dem es im Rahmen einer experimentellen Anordnung gelingt, potenzielle Störvariablen zu kontrollieren. Das Experiment im engeren Sinn stellt jedoch einen quantitativen Ansatz dar, keinen qualitativen; der Begriff der internen Validität ist folglich in der qualitativen Forschung nicht anwendbar.

qualitativer Gütekriterien existiert jedoch nicht. Im Folgenden sind die qualitativen Gütekriterien nach Mayring (2002) aufgeführt; diese überschneiden sich z.T. mit den schon genannten Adaptationen quantitativer Kriterien:

- *Verfahrensdokumentation*: Die einzelnen Schritte einer qualitativen Untersuchung sind genau zu dokumentieren, um das Vorgehen für andere nachvollziehbar zu machen. Dies betrifft sämtliche Untersuchungsschritte, von der Explikation des Vorverständnisses über die Datenerhebung (einschließlich eventueller Änderungen der Erhebungsinstrumente im Forschungsprozess) bis hin zur Auswertung.
- *Interpretative Argumentationsabsicherung*: Interpretationen spielen bei der Auswertung verbalen Materials eine wesentliche Rolle. Solche Interpretationen sind genau zu begründen. Insbesondere müssen sie in sich schlüssig sein; außerdem sollte das Material auf ‚Negativfälle‘ geprüft werden, d.h. auf Fälle, die geeignet wären, die vorgenommene Interpretation zu widerlegen.
- *Regelgeleitetheit*: Qualitative Forschung soll zwar ihrem Gegenstand gegenüber offen, zugleich jedoch nicht unsystematisch sein. Konkret bedeutet das, dass auch qualitative Forschung zuvor explizierten Verfahrensschritten folgen sollte, wie sie sich für verschiedene qualitative Methoden herausarbeiten lassen.
- *Nähe zum Gegenstand*: Nähe zum Gegenstand wird erreicht, indem Forscher/innen diesen möglichst in seiner natürlichen Umgebung untersuchen; die Umgebung (das Feld) ist selbst Teil des Gegenstandes.
- *Kommunikative Validierung* (Lechler): Bei der Kommunikativen Validierung werden die Ergebnisse an die befragte Person rückgemeldet; die Ergebnisse werden in dem Maß als gültig betrachtet, in dem die befragte Person zustimmt, dass die forscherseitigen Rekonstruktionen dem von ihr Gemeinten entsprechen. Im FST wird die Kommunikative Validierung durch das Kriterium des Dialog-Konsenses präzisiert.
- *Triangulation* (Denzin): Es werden mit verschiedenen Methoden Daten erhoben (Methoden-Triangulation) oder Daten zu verschiedenen Aspekten des interessierenden Gegenstandsbereiches (Daten-Triangulation). Es gilt allerdings gerade in der qualitativen Forschung als umstritten, inwieweit die Triangulation als Gütekriterium gelten kann. Denn wenn die Daten untereinander nicht übereinstimmen, muss dies nicht gleichbedeutend damit sein, dass eines der Verfahren (bei der Methoden-Triangulation) ‚falsche‘ Daten geliefert hat. Es ist auch möglich, dass die verschiedenen Methoden je unterschiedliche Sichtweisen auf den Gegenstand eröffnen.

Neben diesen generellen Gütekriterien lassen sich weiterhin je verfahrensspezifische Gütekriterien identifizieren. (s. hierzu Mayring, 2002, Kap. 6.2.).

Prüfungsliteratur: Mayring 2002, Kap. 6

III.5.3. Ethische Probleme qualitativer Forschung

Zwar werden ethische Fragen in der Psychologie in erster Linie im Kontext quantitativer Forschung diskutiert; aber auch qualitative Forschung kann ethische Probleme aufwerfen.

- *Gewährleistung von Anonymität:* Während sich das Interesse in der quantitativen Forschung in erster Linie auf aggregierte Daten richtet (Mittelwerte usw.), steht in der qualitativen Forschung der Einzelfall in seiner ganzen Komplexität im Mittelpunkt. Gerade die vielen Einzelheiten, die einen spezifischen Fall ausmachen, können andererseits aber auch dazu führen, dass die betreffende Person für andere identifizierbar wird. Es ist daher in der qualitativen Forschung deutlich schwieriger als in der quantitativen, die Anonymität der Untersuchungsteilnehmer/innen zu gewährleisten. Entsprechend sind Gegenmaßnahmen erforderlich, die eine Identifikation der untersuchten Personen erschweren, insbesondere die Verfremdung von persönlichen Daten und Informationen. Andererseits gehen mit einer Verfremdung jedoch immer auch Aspekte verloren, die gerade diesen Fall ausmachen.
- *Persönlicher Charakter von Informationen:* In der qualitativen Forschung geben die Untersuchungsteilnehmer/innen den Forschern/innen häufig einen tiefen Einblick in ihr persönliches Leben und geben auch Informationen preis, die ihnen im Fall einer Publikation schaden könnten (oder die für andere verletzend sein können). Für Forscher/innen stellt sich daher die Frage, ob es erforderlich ist, auch wirklich alle gewonnenen Informationen zu veröffentlichen; andererseits wird durch eine solche bewusste Selektion die Gültigkeit der Ergebnisse beeinträchtigt.
- *Ungleichheit der Machtverhältnisse:* In der Forschungssituation gelten die Forscher/innen meist als Experten/innen, ihnen kommt häufig ein höheres Prestige zu als den Untersuchungsteilnehmern/innen, die zudem durch die ungewohnte Forschungssituation verunsichert sein können. Diese Ungleichheit der Machtverhältnisse zwischen Forscher/in und erforschter Person kann dazu führen, dass die erforschten Personen mehr Informationen von sich preisgeben, als sie dies bei nachträglicher nüchterner Betrachtung eigentlich möchten: z.B., um alles ‚richtig zu machen‘ – vielleicht auch, um bei persönlichen Problemen von den Forschern/innen Hilfe zu erhalten. Forscher/innen sollten dieses Ungleichgewicht der Machtverhältnisse nicht ausnutzen, um zusätzliche Informationen zu erhalten. Sie sollten es auch von vornherein klarstellen, wenn es nicht Teil ihrer Forschungsbemühungen ist, in Problemsituationen Hilfe anzubieten (z.B. bei einer Untersuchung über Partnerschaftsprobleme, über den Umgang mit chronischen Krankheiten usw.).
- *Interpretationen können verletzend sein:* In der Auswertungsphase qualitativer Forschung wird häufig über das hinausgegangen, was von den Untersuchungsteilnehmern/innen gesagt wurde; häufig werden auch Wertungen vorgenommen (z.B. ‚adaptive‘ und ‚maladaptive‘ Formen des Umgangs mit Trauer usw.). Wenn die Untersuchungsergebnisse den Befragten zugänglich gemacht werden, kann dies für die betreffenden Personen sehr verletzend sein. Die Ergebnisse sollten den Teilnehmern/innen in einem solchen Fall nicht unkommentiert zur Verfügung gestellt werden. Insgesamt handelt es sich hier jedoch um ein Problem, für das noch keine zufriedenstellende Lösung existiert.
- *Probleme einzelner Verfahren:* Über die genannten generellen ethischen Probleme qualitativer Forschung hinaus weisen einzelne Verfahren spezielle ethische Probleme auf. Dies gilt beispielsweise für die verdeckte Beobachtung, die mit Täuschung und häufig auch mit einem Eindringen in persönliche Bereiche anderer Menschen verbunden ist. Auch die Gruppendiskussion ist insofern ethisch problematisch, als die Anwesenheit eines ‚stooge‘ ebenfalls eine Täuschung der übrigen Teilnehmer/innen beinhaltet; außerdem stellt sich die Frage, inwieweit es vertretbar ist, die Teilnehmer/innen dazu zu ‚verleiten‘, Dinge zu sagen, die sie in einer weniger ‚hitzigen‘ Atmosphäre vielleicht nicht äußern würden.

III.6. Zur Vereinbarkeit von qualitativer und quantitativer Forschung

Qualitative und quantitative Forschungsansätze sind zwar in verschiedenen Forschungstraditionen verankert und lassen sich idealtypisch durch je verschiedene Merkmale charakterisieren. In der Praxis sind die Unterschiede zwischen den Ansätzen jedoch eher graduell, und die Ansätze lassen sich durchaus auch miteinander kombinieren. Qualitative und quantitative Forschung schließen einander also keineswegs aus. Im Folgenden sind einige ‚Kombinationsmöglichkeiten‘ aufgeführt:

- *Qualitative gefolgt von einer quantitativen Forschungsphase:* Dies ist die (aus der quantitativen Sicht) ‚klassische‘ Relation von qualitativer und quantitativer Forschung: Der Forschungsprozess beginnt mit einer qualitativen Untersuchung; diese mündet in eine Hypothese, die im nächsten Schritt quantitativ überprüft wird. Qualitative Forschung hat hier also eine ‚Zubringerfunktion‘ für quantitative Forschung.
- *Quantitative gefolgt von einer qualitativen Forschungsphase:* Hier ist die Reihenfolge von quantitativer und qualitativer Forschung im Vergleich zur ersten Kombinationsmöglichkeit ‚vertauscht‘: Der Forschungsprozess beginnt mit einer quantitativen Untersuchung. Auf der Grundlage der Ergebnisse werden besonders interessante Fälle ausgewählt (das können ‚durchschnittliche‘ Fälle sein oder auch besonders unübliche); diese werden in einer anschließenden qualitativen Phase genauer untersucht. Qualitative Forschung dient hier dazu, quantitative Forschung ‚auszufüllen‘ und Sinnzusammenhänge verständlich zu machen.
- *Methoden-Triangulation:* Hier werden qualitative und quantitative Verfahren innerhalb einer Untersuchung ‚gleichberechtigt‘ eingesetzt und kombiniert, um so ein umfassendes Bild des Gegenstandes zu erhalten.
- *(Inferenz)Statistische Auswertung qualitativer Daten:* An eine qualitative Phase der Datenerhebung und Auswertung (z.B. mittels Inhaltsanalyse) kann sich eine weitergehende statistische Aufbereitung anschließen. Die Möglichkeiten reichen von einer einfachen Häufigkeitsauflistung bis hin zu inferenzstatistischen Gruppenvergleichen.
- *Verfahren, die qualitative und quantitative Anteile enthalten:* Schließlich existieren Verfahren, die weder dem qualitativen noch dem quantitativen Ansatz eindeutig zugeordnet werden können; sie enthalten vielmehr Elemente beider Ansätze gleichermaßen. Hierzu zählen z.B. das FST sowie die Inhaltsanalyse.

Sofern sie den Anforderungen an eine wissenschaftliche Methode genügen, sind qualitative, quantitative oder auch kombinierte Verfahren für die Forschung zunächst einmal gleichermaßen gut geeignet. Keine Methode ist per se ‚gut‘ oder ‚schlecht‘; Methoden sind vielmehr für bestimmte Gegenstände mehr oder weniger gut geeignet. Letztlich kommt es bei der Forschungsplanung also nicht auf die Methode als solche an, sondern auf deren Gegenstandsangemessenheit: Die Methode muss in dem Sinn zu dem Gegenstand passen, dass sie geeignet ist, die interessierenden Merkmale auf tatsächlich abzuheben.

Prüfungsliteratur: nur Skript

IV. Im Skript verwendete Prüfungsliteratur

- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. Aufl.). Berlin etc.: Springer.
- Dann, H.-D. (1992). Variationen von Lege-Strukturen zur Wissensrepräsentation. In B. Scheele (Hg.), *Struktur-Lege-Verfahren als Dialog-Konsens-Methodik* (S. 2-41). Münster: Aschendorff.
- Huber, O. (1987). *Das psychologische Experiment*. Stuttgart: Huber.
- Hussy, W. & Jain, A. (2002). *Experimentelle Hypothesenprüfung in der Psychologie*. Göttingen: Hogrefe.
- Mayring, P. (2002). *Einführung in die qualitative Sozialforschung* (5. Aufl.). Weinheim u. Basel: Beltz.
- Nolting, H.-P. & Paulus, P. (1999). *Psychologie lernen* (5. Aufl.). München: PVU
- Rustemeyer, R. (1992). *Praktisch-methodische Schritte der Inhaltsanalyse. Eine Einführung am Beispiel der Analyse von Interviewtexten*. Münster: Aschendorff.
- Scheele, B. (1995). Dialogische Hermeneutik. In U. Flick et al. (Hg.), *Handbuch Qualitative Sozialforschung* (S. 274-278). Weinheim: Beltz.
- Selg, H., Klapprott, J. & Kamenz, R. (1992). *Forschungsmethoden der Psychologie*. Stuttgart: Kohlhammer.