

Werner Stangl

Tests und Experimente in der Psychologie

1 Test und Experiment als paradigmatische Methoden der Psychologie

Test und Experiment sind geradezu typisch für das naturwissenschaftliche Paradigma der Psychologie. Beide stehen für eine vereinfachte bzw. reduzierte Welt. Diese Entwicklung der einzelwissenschaftlichen Psychologie in der Nachfolge der Naturwissenschaften hat seinen Grund in den epistemologischen Zielsetzungen, die hinter diesem Paradigma stehen: gesichertes, objektives Wissen über die Welt zu erlangen, d.h. letztlich sie so zu beschreiben bzw. zu erklären, wie und warum sie so und nicht anders ist.

Test und Experiment als Methoden entsprechen daher dem heute in der Psychologie nach wie vor dominierenden empiristisch-nomologisch-naturwissenschaftlichen Wissenschaftsverständnis, dessen wesentliche Wurzeln im logischen Empirismus zu finden sind. Historisch gesehen hat besonders in der experimentellen Psychologie das Modell der Naturwissenschaften eine zentrale Rolle gespielt. Dieses reduktionistische Modell macht die Gültigkeit ihrer Ergebnisse und damit die Erfüllung ihres Wissenschaftlichkeitskriteriums unter anderem abhängig von der Übernahme und Anwendung der naturwissenschaftlich erprobten, wiederholbaren und jederzeit kontrollierbaren Strategien (vgl. Stangl 1989).

Viele Denkansätze und Methoden der psychologischen Diagnostik sind auf die experimentelle Psychologie zurückzuführen. Schwemmer (1983, 66) schreibt in Bezug auf das Experiment: „Fragt man nach dem entscheidenden Schritt zu unserer neuzeitlichen und gegenwärtigen Wissenschaft, so wird man das Experiment - und seine mathematisch-technische Bewältigung - als das Kriterium nennen können, durch das die Grenzen zwischen vorwissenschaftlichen Spekulationen und wissenschaftlichen Theorien markiert werden. In der Tat hat die Einrichtung von Experimenten und die Ausrichtung aller Beobachtungen an dem Muster des Experiments die - so scheint es: alles - entscheidende Berufsinstanzen geschaffen, die das endgültige Urteil im Streit der Meinungsparteien zu fällen hat“.

1.1 Die historische Entwicklung der Methoden aus der sozialen Situation

Obwohl es eine lange Vorgeschichte zur Entwicklung einer facheinschlägigen Methodologie, insbesondere zu Test und Experiment, in der Psychologie gibt, beherrschte bis zum Beginn des 19. Jahrhunderts ein nicht-naturwissenschaftliches Paradigma die Tätigkeit der ForscherInnen. War bis etwa 1920 die Introspektion immer noch das Hauptmittel der psychologischen Forschung, so setzte sich ab dieser Zeit eine radikal neue Sicht der psychologischen Forschung durch: Die Behavioristen, also rein verhaltensorientierte Forscher, proklamierten eine Psychologie ohne Introspektion, die bis zu diesem Zeitpunkt der bevorzugte Zugang zum Psychischen war. Ihr Vorreiter war John B. Watson, der 1913 ein behavioristisches Manifest veröffentlichte. Hauptmerkmal behavioristischer Psychologie ist es, keinerlei Annahmen über Struktur oder Funktionsweise des Geistes zu machen, sondern ausschließlich die Beobachtungen aus Experimenten zu sammeln und zu kategorisieren.

Experiment wie Test stellen in der Psychologie zunächst eine soziale Situation dar, also jenen Kontext, in der „Psychologie“ stattfindet. Innerhalb einer Wissenschaft ist der Ort der Erforschung weitgehend institutionalisiert, d.h., dass die allgemeine Verteilung der Rollen bzw. der Erwartungen hinsichtlich dieser Rollen und damit die Regeln des Interagierens innerhalb dieses sozialen Prozesses festgelegt sind. Während sich beim Experiment Versuchsperson und Versuchsleiter gegenüberstehen, sind es beim Test in der Regel der psychologische Praktiker und der Klient.

Danziger (1985) gibt einen historischen Überblick über die Entwicklung dieser sozialen Situation, wobei er sich auf die Anfänge der Psychologie bezieht. Er zeigt, dass es damals noch keine festgelegten Rollen gab, denn diese ersten empirisch-experimentellen Ansätze entwickelten sich unabhängig und waren weitgehend von individuell-wissenschaftlichen Interessen getragen. Drei differierende Ansätze unterscheiden sich grundlegend in Bezug auf die Fragestellungen, den sozialen Kontext und somit hinsichtlich des „context of discovery“.

Fehler! Textmarke nicht definiert. Im **Leipziger Modell** (Wundt, Lorenz, Merkel, Mehner, Estel, Kollert u.a.) gab es keine - wie wir heute sagen würden - strenge Trennung zwischen Forschungssubjekt und Forschungsobjekt. Man könnte diese Form des experimentellen Kontextes idealisierenderweise als Forschungsgemeinschaft bezeichnen, die allgemein typisch und kennzeichnend für die deutsche Universität im späten 19. Jahrhundert war. Dabei waren diese Gruppen relativ unabhängig vom Lehrbetrieb und auf Grund der strengen Zugangsbedingungen zu solchen Zirkeln ziemlich isoliert. Eine solche Isolation bestand natürlich auch weitgehend vom gesamtgesellschaftlichen Kontext. Eine Konsequenz dieses Modells liegt in der hohen Bewertung des Forschungsobjektes, so dass Experimente unter dem Namen der Studenten publiziert wurden und der Name des Experimentators oft nicht einmal erwähnt wurde.

Dieses Modell wurde später unverändert auf andere Fragestellungen übertragen, etwa bei den berühmten Untersuchungen Binets an Kindern. Dieser sollte gemeinsam mit Simon im Auftrag des französischen Unterrichtsministeriums einen Intelligenztest entwickeln („Echelle metrique de l'intelligence“), der helfen sollte, lernbehinderte von normalbegabten Kindern zu unterscheiden.

In diesem **Pariser Modell** (Richet, Beaunis, Binet, Féré, Deleboef u.a.) wird das Subjekt als Erkenntnisquelle „benutzt“, der Erkenntnisgewinn liegt mehr beim Forscher bzw. innerhalb der scientific community. In diesem Modell liegt auch eine der Wurzeln für den differenzialpsychologischen Ansatz der Psychologie, denn dieses Design legt eher eine Suche nach Unterschieden zwischen den untersuchten Objekten nahe als die Suche nach Gemeinsamkeiten. Ein genereller Unterschied zwischen Leipziger und Pariser Modell liegt auch in den Inhalten: das Leipziger Modell untersucht die Aspekte der normalen menschlichen Kognition, während im Pariser Modell das Hauptgewicht auf abnormem und deviantem Funktionieren des Menschen liegt. Kennzeichnend für diese beiden ersten Modelle ist auch die Bezeichnung der Rollen, wobei vor allem die Rolle der Versuchsperson in unserem Zusammenhang von Interesse ist. Während es im Leipziger Modell auf Grund der Egalität der Personen keine diesbezüglichen Standardisierungen gab, bürgerte sich im Pariser Modell der Begriff des „sujet“ ein, der dann als „subject“ im englischsprachigen Bereich der Psychologie (zuerst bei Cattell) übernommen wurde, nicht zuletzt auf Grund der ähnlichen medizinischen Konnotationen. Folgerichtig entwickelte sich aus diesem am medizinisch orientierten sozialen Modell der Test als Anamnese- bzw. Diagnoseinstrumentarium. Das amerikanische **Clark Modell** unterscheidet sich fundamental von den europäischen Modellen durch die Einführung der Population von Individuen als Forschungsobjekt. Diese Fokussierung hängt teilweise mit den pädagogisch-psychologischen Inhalten aber auch mit den Methoden der Forschung (insbesondere bei Hall) zusammen. Dadurch wird die Asymmetrie zwischen Forscher und Subjekt im Pariser Modell noch stärker verankert. Das Hauptgewicht liegt nicht mehr auf der Analyse psychologischer Prozesse bei einem Individuum, sondern auf der Verteilung von psychologischen Phänomenen in einer Population. Dadurch kommt es zu einer Anonymisierung, die den Beitrag des Individuums zur Forschung eliminiert. Die Interaktionen zwischen Forscher und Forschungsobjekt werden auf ein Minimum reduziert (**Fehler! Textmarke nicht definiert.**Objektivität), da sie in diesem Modell bloß eine unerwünschte Störquelle darstellen. Die Reduktion des Forschungsobjektes auf die austauschbare Datenquelle ist vollzogen, sie hat sich bis heute erhalten und dominiert die akademische psychologische Forschung - vor allem im Bereich der Differenziellen und Sozialpsychologie - vollkommen. Dieses soziale Muster äußert sich schließlich sowohl in der Methode des Experiments als auch des Tests: in beiden Fällen erweisen sich große Stichproben bzw. Vergleichspopulationen als günstig.

1.2 Test und Experiment als zentrale Tätigkeitsfelder der Psychologie

Wenn man von der durchaus nicht trivialen Begriffsbestimmung der Psychologie abieht, dass diese dadurch definiert sei, was PsychologInnen eben so tun, versteht man allgemein darunter jene Wissenschaft, die sich mit dem Erleben und Verhalten von Lebewesen, insbesondere von Menschen, beschäftigt. Vom „Alltag“ psychologischer Tätigkeit - und zwar nicht nur vom wissenschaftlichen sondern auch praktischen psychologischen Handeln - ausgehend kann man vier zentrale Tätigkeitsfelder unter-

scheiden (vgl. Schneewind 1977, 16): Beschreiben, Erklären, Vorhersagen und Verändern.

Setzt man diese vier Tätigkeitsfelder mit den in dieser Arbeit zu beschreibenden Methoden der Psychologie in Beziehung, so ergibt sich folgendes Schema, in welches exemplarisch typische psychologische Handlungsweisen eingesetzt wurden:

Methoden	Beschreiben	Erklären	Vorhersagen	Verändern
psychologischer Test	Diagnose, Anamnese in der psychologischen Praxis, Einstufung	Zurückführen menschlichen Verhaltens auf Merkmale der Person	Leistungsprognose, z.B. Schul- oder Berufserfolg	Therapie- und Behandlungsvorschläge auf Grund von Testergebnissen, Selektion
psychologisches Experiment	Definition von Variablen	Kontrolle der Bedingungen bzw. Variablen, Kausalität	Hypothesen für neue Experimente generieren	Veränderung der abhängigen Variablen in Bezug auf die Variation der unabhängigen

Wie man sieht, finden sich zwar in dieser Übersicht in allen Feldern Überschneidungen, jedoch sind die durch Fettdruck hervorgehobenen Felder zentral bzw. sogar methodenkonstituierend. Psychologische Tests sind dadurch gekennzeichnet, dass mit ihnen eine genaue Beschreibung einer Person möglich sein soll, während das Experiment dadurch charakterisiert ist, Kausalbeziehungen zur Erklärung von psychologischen Phänomenen zu erhellen. Ebenfalls methodentypisch sind die Anwendungsfelder: während das Experiment in den meisten Fällen in der wissenschaftlichen bzw. forschungsorientierten Psychologie zum Einsatz kommt, ist der Test - nach seiner wissenschaftlichen Entwicklung - hauptsächlich in der psychologischen Praxis zu finden.

1.3 Gemeinsamkeiten von Experiment und Test

Die psychologische Diagnostik hat viele Anregungen und Vorgehensweisen von der experimentellen Psychologie übernommen: philosophische und methodologische Grundeinstellungen, allgemeine Untersuchungsansätze, spezifische Messverfahren, für die Diagnostik inhaltlich relevante Forschungsergebnisse. Diagnostischen Modellen liegen in der Regel bestimmte Vorstellungen des Menschen zu Grunde. Nach Jäger & Petermann (1995) sind solche Menschenbilder unterschiedlich und wandelbar, denn sie stehen unter dem Einfluss von Zeitgeist, Biografie, Weltbild, Bildungsniveau u.v.m. So gibt es zum Beispiel Theorien zeitübergreifender Persönlichkeitsdispositionen, aber auch Auffassungen, die Persönlichkeitseigenschaften als situationsspezifisch beschreiben, sowie die interaktionistische Sichtweise, die zwischen diesen Standpunkten gewissermaßen eine Brücke geschlagen hat.

Die Übernahme von Ideen und Methoden des Experiments durch die Testpsychologie wurde dadurch begünstigt, dass viele der ersten diagnostisch arbeitenden Wissenschaftler Schüler von **Fehler! Textmarke nicht definiert.** Wilhelm Wundt, dem Begründer der Experimentalpsychologie, bzw. Schüler anderer Experimentalpsychologen waren, wie z. B. James McKeen Cattell und Hugo Münsterberg. Sie lernten bei

den Experimentalpsychologen in einer bestimmten Weise zu denken und zu forschen und übertragen das dann auch auf ihre diagnostische Arbeit.

Sowohl psychologische Diagnostik als auch experimentelle Psychologie berufen sich auf den **Fehler! Textmarke nicht definiert.** Empirismus. Eine empiristische Einstellung wurde schon von den frühen Experimentalpsychologen übernommen. So bezeichnete schon Wundt Experimente und Beobachtungen als Mittel zur Erkenntnis psychischer Sachverhalte und auch die Diagnostik geht davon aus, denn sie fordert, dass das Vorliegen von Merkmalen einer Person mittels der Empirie zu überprüfen sei. Für die interessierenden Eigenschaften müssen entsprechende empirische Indikatoren gewonnen werden, die das Vorliegen der Eigenschaft indizieren bzw. überprüfen, inwieweit Individuen, Institutionen oder Sachverhalte diese empirischen Indikatoren erfüllen.

Manchmal werden psychologische Tests als Experimente definiert: „Ein psychodiagnostischer Test kann als ein spezifisches psychologisches Experiment gekennzeichnet werden, das der Erkundung und Beschreibung individueller psychischer Merkmale dient. Es besteht im Wesentlichen darin, dass unter standardisierten Bedingungen eine Informationsstichprobe über den Probanden (Pb) erhoben wird, die einen wissenschaftlich begründeten Rückschluss auf die Ausprägung eines oder mehrerer psychischer Merkmale des Pb gestattet“ (Michel & Conrad 1982, 1).

Die Diagnostik hat in ihren psychodiagnostischen Verfahren zwei Kernideen des Experiments übernommen:

- **Manipulation:** durch Items, Fragen bzw. Anforderungen werden bestimmte Verhaltensweisen oder psychische Vorgänge ausgelöst.
- **Kontrolle:** bei Testverfahren sollen durch die Standardisierung von Durchführung, Auswertung und Interpretation mögliche Störvariablen ausgeklammert werden.

Wie die experimentelle Psychologie zeichnet sich auch die psychologische Testtheorie durch eine quantitative Grundeinstellung aus. Sie kommt unter anderem in Messung, Fehlertheorien, statistischen Methoden und mathematischen Formulierungen von Gesetzen zum Tragen. Die experimentelle Psychologie wendet zur Erfassung der sie interessierenden theoretischen Konstrukte bestimmte Messverfahren an. Die experimentelle Psychologie benutzt zur Auswertung ihrer Daten ein ganzes Bündel verwandter statistischer Methoden. Historisch wichtige Arbeiten zu den statistischen Methoden stammen z. B. von Carl Friedrich Gauß, der die Normalverteilung entwickelte, von Francis Galton, der die Korrelationsrechnung begründete und von Ronald A. Fisher, auf den die klassische Statistik und die Varianzanalyse zurückgehen. Statistische Verfahren finden in der gesamten Diagnostik Verwendung: Normwerte, wie Prozentränge oder Z-Werte, werden herangezogen, um die relative Position des Probanden zu beschreiben. Korrelationen sind für die Quantifizierung von **Fehler! Textmarke nicht definiert.** Gütekriterien (Objektivität, Reliabilität, Validität) von Bedeutung, und Konfidenzintervalle geben Bereiche an, in denen z.B. Entwicklung stattfindet.

Die gegenseitige Befruchtung von Experiment und Test kann man auch an methodischen Details erkennen, wie etwa bestimmten Aufgabenarten: Herrmann Ebbinghaus entwickelte für seinen Intelligenztest Ergänzungsaufgaben, die er als „Lückentest“ bezeichnete, weil in einem vorgegebenen Text Wortlücken zu ergänzen waren. Auch Binet bezog solche Ergänzungsaufgaben in seine **Fehler! Textmarke nicht definiert**. Intelligenztests ein, und diese Form des Lückentests ist auch heute noch in vielen Verfahren zu finden.

Die Diagnostik hat auch eine Reihe spezieller Messinstrumente aus der Experimentellen Psychologie übernommen bzw. weiterentwickelt. Das elementare Inventar der frühen psychologischen Laboratorien bildeten Tachistoskop, Perimeter (Gesichtsfeldmesser), Farbtafeln, Hörschärfemesser, Tastzirkel und Ergograph. Diese Messverfahren und Geräte, die sensorische und motorische Funktionen maßen, waren besonders für die Anfänge der Testentwicklung von Bedeutung (Galton, Cattell). Einige davon finden sich heute in abgewandelter Form bei verschiedenen **Fehler! Textmarke nicht definiert**. Leistungstests.

2 Das psychologische Experiment

2.1 Stellung des Experiments innerhalb der Psychologie

Im Gegensatz zur Alltagspsychologie versucht die wissenschaftliche Psychologie, ihre Aussagen mit geeigneten Methoden daraufhin kritisch zu überprüfen, ob sie wahr oder falsch sind. Diese Überprüfung erfolgt systematisch und methodisch kontrolliert. Damit soll neben der Sammlung von Tatsachenwissen (beschreibend) auch die Erforschung von Gesetzmäßigkeiten mit dem Ziel der Erklärung und Vorhersage von Ereignissen betrieben werden. Das Experiment ist daher ein wichtiges methodisches Hilfsmittel der Psychologie und wurde in größerem Umfang Ende des 19. Jahrhunderts eingeführt und setzte sich mit der Blüte des Behaviorismus vorerst in der amerikanischen und später auch in der europäischen wissenschaftlichen Psychologie durch. Vor der Einführung des Experiments musste man sich auf die persönlichen Erfahrungen einzelner Menschen stützen. Das Experiment erweitert diese „natürlichen“ Einsichten durch „synthetische“. Der Vorteil dabei ist, dass man sie planmäßig und unabhängig von der nur subjektiven Geltung erarbeiten kann.

Auch wenn bei der Definition des Experiments bisweilen unterschiedliche Akzente gesetzt werden, so besteht doch allgemein Einigkeit darin, dass die aktive Manipulation der Versuchsbedingungen durch den Experimentator und damit die Möglichkeit Ursache und Wirkung zu unterscheiden, das Wesentliche am Experiment ausmacht.

Das Experiment stellt die einzige Forschungsform dar, die es erlaubt Kausalbeziehungen zwischen Variablen zu überprüfen: Zwei oder mehr Variablen sind kausal verbunden, wenn sie in einem empirisch nicht umkehrbaren, asymmetrischen Zusammenhang stehen. X erzeugt Y, aber nicht umgekehrt. X ist dabei die unabhängige und Y die abhängige Variable.

2.2 Was ist ein Experiment?

Ein Experiment ist durch folgende Bedingungen gekennzeichnet:

- **Variation:** Der Experimentator variiert systematisch mindestens eine unabhängige Variable und registriert, welchen Effekt diese aktive Veränderung auf die abhängige Variable hat.
- **Kontrolle:** Gleichzeitig schaltet er die Wirkung von anderen (Stör)Variablen aus. Darüber hinaus sind folgende eher formale Bedingungen - sie betreffen u.a. die Wissenschaftlichkeit eines Experiments - einzuhalten:
 - **Planmäßigkeit:** Experimentelle Bedingungen müssen nachher genau beschrieben werden, folglich ist eine Vorausplanung der Versuchsbedingungen unerlässlich.
 - **Replikation:** Ein Experiment muss so geplant sein, dass es jedermann bei Einhaltung der gleichen Versuchsbedingungen wiederholen und überprüfen kann.
 - **Kontrollverlauf:** Bei jedem Experiment ist die Versuchsperson als ganzer Mensch, u. a. mit ihren Motivationen, beteiligt. Folglich muss während des Versuchs kontrolliert werden, ob sich bei ihr nicht psychische Veränderungen ereignen.
 - **Auswertung:** Die Ergebnisse sind "Rohdaten", erst ihre Verarbeitung (mithilfe statistischer Mittel) lässt die Verwertung zu.

Noch weitgehend ungeklärt ist etwa das Problem der Versuchspersonenmotivation, d.h. die Auswirkung von Motivationen aufseiten der Versuchspersonen auf die Versuchsergebnisse. Diese wird nur in den seltensten Fällen überprüft. Relativ gut erforscht ist der „Versuchsleitereffekt“, dessen Wirksamkeit in zahlreichen Experimenten nachgewiesen konnte, z.B. in Form des „Rosenthal-Effekts“. Der amerikanische Psychologe Robert Rosenthal analysierte, welchen Einfluss der Versuchsleiter in psychologischen Experimenten auf das Verhalten von "Versuchspersonen" oder auch Versuchstieren hat.

Dieser Effekt kann z.B. ausgeschaltet werden durch einen **Doppelblindversuch**, damit die Hypothesen des Versuchsleiters bzw. seine Annahmen über den Versuch nicht von der Versuchsperson unbewusst übernommen werden, belässt man bei der Verteilung der Bedingungen (z.B. Medikament und wirkungsloses Präparat, „Placebo“) auch ihn in Unkenntnis. Bei der nicht-experimentellen Forschung finden nur Beobachtungen statt, ohne in das Geschehen einzugreifen (oft ist dies auch aus praktischen oder ethischen Gründen so).

2.3 Phasen eines Experiments

Die Durchführung von Experimenten lässt sich in mehrere Phasen untergliedern, die nacheinander durchlaufen werden:

- **Fragestellung**, die im Laufe der Untersuchung weiter präzisiert wird.
- **Hypothesen**, die auf bisherigen Forschungsergebnissen, theoretischen Überlegungen oder empirischen Beobachtungen aufbauen.
- **Operationalisierung**, also Zuordnung von Begriffen der Hypothese zu beobachtbaren Phänomenen: Man sucht einen empirisch beobachtbaren Indikator für

die Begriffe. Dazu gehört auch die Auswahl einer leicht realisierbaren Situation, in der die Versuchspersonen bisher noch keine Erfahrungen sammeln konnten. Die Information, die mithilfe einer Operationalisierung gewonnen wird, wird als Datum bezeichnet. Die Güte der Operationalisierung wird oft auch als Konstruktvalidität bezeichnet. Die theoretischen Begriffe einer Hypothese können mehr oder weniger direkt operationalisiert werden, so wie Variablen mehr oder weniger beobachtungsnah sind.

Bei der Planung von Experimenten verfolgt der Experimentator eine Strategie, die man als **MAX-KON-MIN-Regel** charakterisieren kann: Die Wirkung der unabhängigen Variablen auf die abhängige Variable zu MAXimieren, die Einflüsse von unerwünschten systematischen (Stör-) Variablen auf die abhängige Variable zu KONtrollieren und die Effekte von unsystematischen (Zufalls)Variablen zu MINimieren.

Bei der experimentellen Forschung ist die Planungsphase die eigentlich entscheidende Phase. Durchführung und Auswertung sind dann weitgehend festgelegt. Die Planung des Experimentes erfordert, dass der Experimentator zunächst eine Versuchssituation findet, in der die unabhängige Variable manipulierbar ist, und für die abhängige Variable ein geeignetes Messinstrument gefunden werden kann. Wenn eine geeignete Versuchsanordnung gefunden ist, muss als Nächstes der Versuchsplan im engeren Sinn (Versuchsgruppen-Plan) festgelegt werden: Dabei wird entschieden, welche experimentellen Bedingungen einbezogen werden, d.h. welche unabhängigen Variablen in wie viel Stufen variiert werden, und wie die Versuchspersonen den experimentellen Bedingungen zugeordnet werden.

2.4 Grenzen des Experiments

Das Experiment gilt als Grundmethode der auf das physikalischen Vorbild ausgerichteten Psychologie. Analog zur Physik wird eine feste kausale Beziehung zwischen Messwert und Erscheinung angenommen und vorausgesetzt. Dem gegenüber betonen Kritiker, dass psychische Erscheinungen eine grundsätzliche Inhaltsmehrdeutigkeit (Ambiguität) aufweisen, die man mit der Experimentalanordnung kategorial verfehlen kann. Das Experiment in der Psychologie ist daher in der epistemologischen Diskussion nach wie vor umstritten. Folgende zentrale Fehlerquellen können bei der experimentellen Forschung häufig beobachtet werden:

- **Bedeutungsfehler:** Die experimentelle Planung engt die Untersuchungsmöglichkeiten auf „machbare“ (operationalisierbare) Anordnungen ein. Aus bedeutenden Problemen werden deshalb oft nichtige, die zu einer „leblosen Psychologie“ führen, weil eher eine Methode demonstriert als Inhalte ermittelt werden.
- **Planungsfehler:** Auch die vorausschauendste Planung wird nicht alle möglichen Gründe für ein Scheitern der Versuche verhindern können; man sollte also vor Versuchsbeginn eine „taktische“ Überlegungspause einlegen.
- **Instruktionsfehler:** Fehler entstehen häufig durch mangelhafte Unterweisung der Versuchspersonen. Auf die Standardisierung muss daher großer Wert gelegt werden.

- **Isolationsfehler:** Durch die unnatürliche Versuchssituation können „Isolationseffekte“ bei der Versuchsperson entstehen, die die Ergebnisse verzerren.
- **Übergeneralisation** ergibt sich, wenn die Ergebnisse über die (zumeist eng begrenzte) Gültigkeit verallgemeinert werden.

Bei bestimmten Fragestellungen verbieten sich Experimente manchmal aus Gründen der Ethik oder der Moral; z.B. kann man experimentell nicht nachprüfen, wie sich die Prügelstrafe auf kleine Kinder auswirkt. Hier bleibt die Psychologie auf die Beobachtung angewiesen, die durch Kontrolle zusätzliche Genauigkeit gewinnen kann (z.B. durch Erfassung möglichst aller beteiligten Einflüsse, Tonbandaufzeichnungen usw.).

In einem frühen Verhaltensexperiment (Watsons „Little Albert“) wurde ein Kleinkind immer durch Geräusche erschreckt, wenn es eine zahme Ratte sah. Dadurch konnte eine vorher nicht vorhandene Rattenphobie ausgelöst werden. Viele Kritiker sahen darin die Überschreitung ethischer Grenzen. Man kann und darf nicht alles experimentell untersuchen, was sich nur planen lässt.

Ethische Probleme treten sowohl bei der Anwendung des Wissens als auch bei der Forschung auf. Der Versuchsleiter trägt eine besondere Verantwortung gegenüber den Versuchspersonen, die sich insbesondere aus der Undurchschaubarkeit der experimentellen Situation ergibt. Folgende ethischen Probleme können auftreten:

- **Schädigung** der Versuchsperson (auch durch Verletzung von deren Selbstwertgefühl durch zu schwierige Aufgaben o. Ä.),
- **Täuschung** hinsichtlich des Zwecks der Untersuchung,
- **Manipulation** von Probandeneigenschaften ,
- Verletzungen der **Vertraulichkeit** oder des Datenschutzes.

Oft lassen sich derartige ethische Probleme beseitigen, indem man die Untersuchung entsprechend abändert. In jedem Fall sollte die Versuchsperson über alle möglichen negativen Aspekte informiert werden und dann entscheiden können, ob sie am Experiment teilnehmen will; sie sollte aber auch jederzeit die Untersuchung abbrechen dürfen. Nach dem Experiment sollte eine Aufklärung über dessen Ziele stattfinden. Die negativen Aspekte des Versuches sollten durch positive Aspekte aufgehoben werden (finanzielle oder andere Belohnung). Durch Kosten-Nutzen-Rechnung sollte entschieden werden, ob die Verletzung von ethischen Prinzipien durch den möglichen Wissenszuwachs gerechtfertigt ist. Ein kritisches Experiment in diesem Zusammenhang ist die berühmte Milgram-Studie zum Gehorsam gegenüber Autoritäten.

2.5 Welche Experimente gibt es?

Experimente lassen sich auf verschiedene Arten klassifizieren:

Einteilung nach dem Ziel:

- (Hypothesen)**Prüfexperiment:** das „klassische“ Experiment, bei dem Hypothesen kausal geprüft werden; z. B.: fördert die Aggressionsdarstellung im Film (unabhängige Variable) die Aggressivität im nachfolgenden Spiel (abhängige Variable)

- **Erkundungsexperiment** (exploratives Experiment, pilot study): zielt auf die Verbreiterung der Hypothesenbasis oder auf einen Zugewinn an allgemeiner Erfahrung, die durch weitere Experimente untermauert werden soll
- **Demonstrationsexperiment**: Die Vorführung der Müller-Lyerschen Täuschung soll nicht zu neuen Erkenntnissen führen, sondern den Lernenden eine wohl bekannte Erscheinung in experimenteller Vorführung nahe bringen.

Einteilung nach der Zahl der unabhängigen Variablen: einfaktorielles oder mehrfaktorielles Experiment.

Einteilung nach der Zahl der abhängigen Variablen: univariates oder multivariates Experiment.

Einteilung nach dem Ort:

- **Laborexperiment** (spezieller Untersuchungsraum): klassische Laborexperimente sind etwa die Tierversuche zum Konditionieren.
- **Feldexperiment** (natürliches Umfeld): dieses wird in natürlicher Umgebung (z. B. in einem Kaufhaus, Schule, Familie, vor einem Kino) durchgeführt, wobei es besonders schwierig ist, „konfundierende“ (nebenwirkende) Variablen auszuschalten. Diese sind die klassischen Grundformen des Experiments, denn beide sind gegenüber dem einfachen Beobachten aktiv eingreifende Formen des Erkenntnisgewinns. Das Feldexperiment hat auch für die psychologische Diagnostik größere Bedeutung, weil das Vorgehen in der Diagnostik dem der Feldforschung ähnelt.

Einteilung nach der Versuchsplanung:

- **Echte Experimente** (aktives Variieren der unabhängigen Variable und Kontrolle der Störvariablen): in der Regel nur im Labor in einer künstlichen Situation möglich.
- **Quasi-Experimente** (Interdependenz-Analyse, keine Kontrolle der Störvariablen): Versuchssituation mit der natürlich vorgefundenen Variiertheit z. B. wenn zu einer Experimentalgruppe von drei Personen eine vierte hinzutritt, wobei Änderungen des Gruppengeschehens nicht kausal (als Bedingung für die Veränderung) interpretiert werden.

Häufig wird auch die „**ex post facto-Forschung**“ zu den Experimenten gerechnet, allerdings ist das nicht angemessen, denn beim experimentellen Ansatz macht man von einem kontrollierten X eine Vorhersage auf Y. Bei der ex post facto-Forschung wird hingegen nur Y systematisch beobachtet. Dann folgt rückblickend eine Suche nach X. Der wichtigste Unterschied zwischen experimenteller und ex post facto-Forschung liegt also in der Möglichkeit der Kontrolle. Bei einem Experiment verfügt man über die Möglichkeit der gezielten Variation der Versuchsbedingungen. Bei einem ex post facto-Design muss man die Dinge so hinnehmen, wie sie vorliegen. Ein großer Teil erziehungswissenschaftlicher aber auch soziologischer Studien sind ex post facto-Untersuchungen.

2.6 Stichprobe

Wenn für eine experimentelle Fragestellung die geeignete Versuchsanordnung gefunden ist, muss als Nächstes der Versuchsplan im engeren Sinn, der Versuchsgruppen-Plan festgelegt werden. Dabei wird entschieden, welche experimentellen Bedingungen einbezogen werden, d.h. welche unabhängigen Variablen in wie viel Stufen variiert werden, und wie die Versuchspersonen den experimentellen Bedingungen zugeordnet werden. Auf die Behandlung von Versuchsplänen mit mehreren unabhängigen Variablen soll hier nicht eingegangen werden, da diese im Prinzip demselben Schema folgen, jedoch komplexere statistische Modelle als Grundlage haben. Komplexe Designs finden sich in Bernhard Jacobs (1998) Tutorium „Einführung in die Versuchsplanung“.

2.6.1 Unabhängige Gruppen

Die einfachsten Versuchspläne enthalten nur eine unabhängige Variable. Zwei oder mehr Bedingungen, z.B. zwei Therapievarianten und eine unbehandelte Kontrollgruppe sollen in ihren Mittelwerten hinsichtlich der vegetativen Labilität am Ende einer Behandlungsperiode verglichen werden. Die Hauptvorteile eines solchen Designs mit nur einer unabhängigen Variablen sind:

- Der Versuchsplan ist einfach und robust.
- Die Zahl der Versuchspersonen in den einzelnen Versuchsgruppen muss nicht gleich groß sein.
- Gehen zufällig Messwerte verloren, so können die übrigen durch einfache statistische Korrekturen verwertet werden.

Der Nachteil eines Versuchsplanes mit unabhängigen Gruppen, besteht darin, dass der erforderliche Stichprobenumfang recht groß werden kann. Das gilt besonders dann, wenn keine gerichteten Hypothesen bestehen, die Varianz innerhalb der Gruppen groß und die Effektstärke mittel oder klein ist.

2.6.2 Mehrere parallelisierte Gruppen

Durch Parallelisierung der Versuchsgruppen soll erreicht werden, dass die Versuchsgruppen sich in einem oder in mehreren relevanten Merkmalen genau entsprechen. In dem oben genannten Beispiel des Vergleiches von Therapiebedingungen (zwei Therapiearten, eine Kontrollgruppe) wäre es z.B. sinnvoll, die Gruppen nach der Ausgangslage (gemessen mit verschiedenen Fragebögen) zu parallelisieren. Bei einem Lernexperiment zum Vergleich von Unterrichtsmethoden könnte man Vorkenntnisse, Schulleistungen in bestimmten Fächern, aber auch Motivationsvariablen heranziehen. Kurz: Alles, was mit der abhängigen Variablen (Therapieerfolg, Lernerfolg) korreliert, kann eine sinnvolle Parallelisierungsvariable sein.

Die Parallelisierung wird durchgeführt, indem man zunächst von allen Versuchspersonen die Variable erhebt, nach der parallelisiert werden soll. Das erfordert in der Regel eine eigene Vortest-Sitzung. Für k experimentelle Bedingungen benötigt man k parallelisierte Gruppen. Man sucht aus dem Datenmaterial jeweils k Versuchspersonen mit gleichen Vortestwerten heraus. Diese k Versuchspersonen werden dann nach

dem Zufall auf die k experimentellen Bedingungen verteilt. Danach wird der eigentliche Versuch durchgeführt.

Der Hauptvorteil von parallelisierten Gruppen besteht in der kleineren Stichprobengröße gegenüber einem Versuchsplan mit unabhängigen Gruppen. Dem stehen folgende Nachteile gegenüber: Vor dem Beginn des eigentlichen Versuchs müssen von allen Versuchspersonen Werte für die Parallelisierungsvariablen erhoben und ausgewertet werden. Das erfordert in der Regel nicht nur eine eigene Vortest-Sitzung, sondern führt bei zeitlich lang erstreckter Versuchspersonen-Anwerbung auch zu organisatorischen Problemen.

2.6.3 Meßwiederholung an derselben Versuchspersonengruppe

Versuchspersonen zu beschaffen ist meist schwierig (Kosten, Zeitaufwand). Daher liegt es nahe, eine Versuchsperson für mehr als eine experimentelle Bedingung heranzuziehen. Wird z.B. der Einfluss der Rückmeldungsart (richtig, falsch und keine Rückmeldung) auf das Erlernen von Listen von Wortpaaren untersucht, so könnte man daran denken, jede Versuchsperson drei Listen, je eine unter einer der drei Rückmeldungsarten lernen zu lassen. Die Vorteile: Es werden wesentlich weniger Versuchspersonen benötigt als bei den beiden vorangegangenen Formen der Versuchsplanung. Es wird keine zusätzliche Sitzung benötigt (wie bei parallelisierten Gruppen). Die Nachteile: Im Laufe länger dauernder Versuchssitzungen verändert sich die Versuchsperson: Sie gewöhnt sich an die Situation, an die Aufgabe, zeigt Übungsfortschritte, Ermüdungserscheinungen. Im Versuchsplan ist also dafür zu sorgen, dass diese Effekte nicht mit den experimentellen Bedingungen kollidieren.

2.7 Variablen

Ein wesentlicher Bestandteil von Hypothesen und Gesetzen sind Variablen, die mindestens in zwei Abstufungen vorliegen müssen, von denen immer nur eine realisiert werden kann. Variablen mit nur zwei Ausprägungen nennt man auch qualitativ, Variablen mit vielen Ausprägungen werden dagegen als quantitativ bezeichnet; eine unter auswertungsmethodischen Gesichtspunkten wichtige Unterscheidung der Variablen ist nach dem Skalenniveau notwendig:

- **Nominalskala:** Hier wird nur zwischen Gleichheit und Verschiedenheit unterschieden (qualitative Variablen).
- **Ordinalskala:** Unterscheidung von größer und kleiner (komparative Variablen).
- **Intervallskala:** Auch Aussagen über das Verhältnis von Intervallen zwischen Skalenwerten sind möglich, jede lineare Transformation ist zulässig.
- **Verhältnisskala:** Auch Aussagen über das Verhältnis von Skalenwerten sind möglich; die Einheit ist willkürlich, aber der Nullpunkt ist fest. Daher sind beliebige Multiplikationen mit einer positiven Konstante möglich.
- **Absolutskala:** Es existiert ein natürlicher Nullpunkt und eine natürliche Einheit.

Die Grundidee einer empirischen Prüfung ist die, dass man aus der Hypothese eine empirische Vorhersage formuliert und diese mit der Wirklichkeit vergleicht. Bei einem Experiment greift der Forscher aktiv in das Geschehen ein, während bei einer nicht-experimentellen Untersuchung nur Ereignisse beobachtet werden. Ein Experiment ist also allgemein formuliert die systematische Beobachtung von veränderlichen Merkmalen unter planmäßig kontrollierten oder künstlich geschaffenen Bedingungen. Kontrolle wird beim Experiment im Sinne der Überprüfung bzw. Berücksichtigung von Variablen verwendet. Die Maßnahmen bei der Planung und Durchführung eines Experiments werden kontrolliert, um Veränderungen der Variablen gesichert und unbeeinflusst feststellen zu können.

Beim Experimentieren unterscheidet man **unabhängige Variablen**, die aktiv vom Experimentator verändert werden (oft auch als Faktor bezeichnet) von den **abhängigen Variablen**, die letztlich vorhergesagt werden sollen.

Störvariablen beeinflussen die abhängige Variable ebenfalls, sodass deren Wirkung im Experiment entweder neutralisiert oder aktiv in die Untersuchung einbezogen (als zusätzliche unabhängige Variable) werden soll, da sie den Effekt der unabhängigen Variable stören würde. Eine Störvariable kontrollieren heißt, ihre Wirkung auszuschalten. Dazu sind mehrere Wege möglich:

- **Elimination:** Falls eine bestimmte Variable in einem Experiment als Störvariable auftritt, ist es nahe liegend, dass man versucht, sie durch geeignete Versuchsbedingungen auszuschalten.
- **Konstanthalten:** Gelingt es nicht, eine Störvariable zu eliminieren (z.B. die unterschiedliche Schulbildung von Versuchspersonen), oder tritt eine Störvariable versuchsbedingt auf (Hawthorne-Studien), so muss man bestrebt sein, diese Einflussgröße konstant zu halten. Nach einer vorexperimentellen Befragung werden nur Versuchspersonen mit gleicher Schulbildung zum Versuch zugelassen u. ä.. Dadurch, dass man die Störvariablen zwar nicht eliminiert, aber doch konstant gehalten hat, wird angenommen, dass sie in allen Versuchsgruppen die Messwerte nur um denselben (konstanten) Betrag verändern und folglich die Vergleichbarkeit der Gruppen gewährleistet ist.
- **Parallelisierung:** Wenn beispielsweise Versuchspersonen mit sehr unterschiedlicher Schulbildung an einem Problemlösungsexperiment teilnehmen und es sich als unökonomisch erweist, alle Volksschüler aus dem Experiment auszuschließen, so behilft man sich, indem man in allen Versuchsgruppen dieselbe Verteilung der Versuchspersonen nach den vorexperimentell erhobenen Schulabschlüssen sicherstellt.
- **Randomisierung:** Die auch als Zufallszuweisung bezeichnete Technik gilt als das wirksamste methodische Hilfsmittel zur Kontrolle störender Bedingungen. Hierbei werden die Versuchspersonen nach einem Zufallsprinzip den verschiedenen Versuchsgruppen bzw. -bedingungen zugeordnet und man unterstellt dabei, dass sich die verschiedenen vorexperimentellen Störvariablen ebenfalls zufällig hinsichtlich ihrer Wirkung auf die Versuchsbedingungen verteilen und folglich keine systematischen Unterschiede zwischen den Versuchsgruppen bestehen. Da

die Störvariablen nicht eliminiert werden können, ist man zumindest bestrebt, die durch sie verursachten systematischen Fehler in unsystematische Fehler zu verwandeln.

Ein wesentlicher Schritt der Hypothesenprüfung ist die **Operationalisierung** der Variablen, d.h. die Zuordnung von beobachtbaren Phänomenen zu diesen Variablen. Aus den praktisch unendlich vielen Variablen müssen einige wenige selegiert werden, auf die man sich konzentrieren will. Wissenschaft besteht auch darin, die für eine bestimmte Fragestellung bedeutsamen Variablen zu entdecken.

Ein einfaches Beispiel: Um in einem Experiment die Wirkung von Koffeingenuss auf die Konzentrationsleistung zu untersuchen, werden die Versuchspersonen auf zwei Gruppen aufgeteilt und der einen Gruppe Koffein verabreicht und der anderen nicht. Unabhängige Variable: Koffeingenuss. Abhängige Variable: Konzentrationsleistung. Störvariablen: Alter der Versuchspersonen, Tageszeit bei der Durchführung des Experimentes u. Ä..

3 Der psychologische Test

3.1 Stellung des Tests innerhalb der Psychologie

Tests sind innerhalb der Psychologie dem Zweig der Differenziellen Psychologie zuzuordnen, jenem Teilgebiet, das von dem deutschen Psychologen und Philosophen William Stern (1871-1938) eingeführt wurde. Sie beschäftigt sich mit den Unterschieden im Erleben und Verhalten zwischen einzelnen Menschen bzw. zwischen Gruppen von Menschen, versucht diese zu beschreiben und auf ihre Bedingungen zurückzuführen. Die Differenzielle Psychologie wurde in ihrer Bedeutung umso größer, je mehr Psychologie im Alltag angewendet wurde (z.B. in Schulen, in Kliniken, bei Gericht, in der Personalauslese usw.).

Menschen verhalten sich auf Grund ihrer unterschiedlichen Ausprägung von Fähigkeiten, Bedürfnissen und Emotionen in gleichen Situationen recht unterschiedlich. Um ein Verhalten vorhersagen zu können, müssen daher allgemeine Gesetzmäßigkeiten und individuelle Persönlichkeitsstrukturen berücksichtigt werden. Die Differenzielle Psychologie versucht daher, die typische Ausprägung und Wechselwirkung zwischen psychischen Merkmalen (z.B. Intelligenz, Gefühle, Bedürfnisse usw.) für bestimmte Altersstufen, Geschlechter, Berufe usw. festzustellen.

PsychologInnen in Wissenschaft wie Praxis haben es daher bei ihrer Tätigkeit häufig mit differenziellen Fragestellungen zu tun, wobei diese Tätigkeit unter dem Oberbegriff der Psychodiagnostik subsumiert werden kann. Unter Psychodiagnostik versteht man dabei die Gesamtheit jener psychologischen Verfahren, mit deren Hilfe die Persönlichkeit eines Menschen erfasst werden kann. Außer den verschiedenen Testverfahren gehören zum methodischen Inventar die Anamnese, die Exploration, die Verhaltensbeobachtung und die Verhaltensanalyse. Diese Methoden werden eingesetzt bei psychiatrischen und somatisch-psychosomatisch erkrankten Personen, in der Erziehungs- und Eheberatung, in schulpsychologischen Diensten, bei forensischen Ange-

legenheiten, in der Berufsberatung oder bei Eignungsuntersuchungen. Dabei erfolgt die Untersuchung jeweils mit standardisiertem Material, in den meisten Fällen eben mit psychologischen Testverfahren.

3.2 Was ist ein Test?

Ein psychodiagnostischer Test ist vereinfacht gesprochen ein standardisiertes wissenschaftliches Routineverfahren zur Messung von psychologisch wichtigen, abgrenzbaren Merkmalen einer Person. Sie dienen vornehmlich zur quantitativen Bestimmung des relativen Grades von individuellen Merkmalsausprägungen, können aber auch qualitative Aussagen über individuelle Ausprägungen von Merkmalen ermöglichen. Psychodiagnostische Tests können als Mittel zur Querschnittsdiagnose, zur Längsschnittsdiagnose und als Forschungsverfahren eingesetzt werden

Brauchbare Testverfahren müssen bestimmte Anforderungen erfüllen, die üblicherweise als **Fehler! Textmarke nicht definiert.** Testgütekriterien bezeichnet werden. Nach Lienert (1989) unterscheidet man Haupt- und Nebengütekriterien. Hauptkriterien sind die Objektivität, die Reliabilität und die Validität, Nebengütekriterien sind die Ökonomie (Wirtschaftlichkeit), Nützlichkeit, Normierung und Vergleichbarkeit von Testverfahren. Weist ein Test diese Gütekriterien nicht auf, kann man im eigentlichen Sinne nicht von einem Test sprechen, da ihm die wissenschaftlich überprüften Grundlagen und notwendigen Kontrolluntersuchungen fehlen.

- **Objektivität** ist das Ausmaß, in dem ein Testergebnis in Durchführung, Auswertung und Interpretation vom Testleiter nicht beeinflusst werden kann, bzw. wenn mehrere Testauswerter zu übereinstimmenden Ergebnissen kommen.
- **Reliabilität** (Zuverlässigkeit) gibt die Zuverlässigkeit einer Messmethode an. Ein Test wird dann als reliabel bezeichnet, wenn es bei einer Wiederholung der Messung unter denselben Bedingungen und an denselben Gegenständen zu demselben Ergebnis kommt. Sie lässt sich u.a. durch eine Testwiederholung (Retest-Methode) oder einen anderen, gleichwertigen Test ermitteln (Paralleltest). Das Maß ist der Reliabilitätskoeffizient und definiert sich aus der Korrelation der beiden Testungen.
- **Validität** (Gültigkeit) ist das wichtigste Testgütekriterium, denn es gibt den Grad der Genauigkeit an, mit dem ein Testverfahren das misst, was es messen soll. Die Überprüfung der Gültigkeit wird mithilfe der Korrelation mit einem Außenkriterium vorgenommen. Beispiel: Ein Test soll Depression messen. Diesen Test wendet man bei Personen an, die nachweislich eine Depression haben. Dann wird geprüft, wie genau diese Testergebnisse mit anderweitig ermittelten Bewertungen übereinstimmen (z.B. mit der Einschätzung durch Psychotherapeuten).

Psychologische Tests müssen auch **standardisiert** sein, d.h. sie enthalten eine Testanweisung, die vorschreibt, wie der Test vorgenommen und durchgeführt werden muss. Das Gleiche gilt für die Auswertung eines Tests, denn auch diese muss feste Regeln enthalten, sodass verschiedene Auswerter zum gleichen Ergebnis kommen (Objektivität).

3.3 Wozu werden Tests verwendet?

Psychologische Diagnostik findet ihre Umsetzung in verschiedenen Zusammenhängen und damit auch mit unterschiedlichen Zielsetzungen. Die wichtigsten Anwendungsbereiche sind

- **Persönlichkeitsdiagnostik** z.B. im Zusammenhang mit therapeutischen Fragestellungen,
- **differenzielle Fragestellungen** z.B. bei der Personalauswahl,
- **Diagnose** der Folgen von Hirnschädigungen z.B. in der klinischen Praxis,
- **Leistungs- und Eignungsabklärungen** z.B. in der schulischen oder beruflichen Einstufung,
- **Forschung** z.B. als Erklärungsvariable für Forschungsergebnisse.

Wer Psychodiagnostik betreibt, braucht Grundkenntnisse zu Testtheorie, Anamneseerhebung, Gesprächsführung und zu Regeln der Beobachtung. Testverfahren müssen daher inhaltlich und in der Durchführung beherrscht werden. Hierzu gehört auch eine kritische Bewertung der Verfahren und eine sichere Interpretation der Resultate. Der Diagnostiker muss alle Informationen, die sich im Verlauf eines Diagnoseprozesses ansammeln, zu einer Synthese integrieren können. Ein Teil davon lässt sich während einer einschlägigen (meist universitären) Ausbildung theoretisch erlernen, entscheidend ist aber die praktische Erfahrung. Der Erwerb und die Verwendung psychologischer Verfahren ist in der Regel an eine entsprechende Ausbildung bzw. zertifizierte Qualifikation gebunden.

3.4 Testtheorien

Die psychologische Testtheorie beschäftigt sich mit grundsätzlichen Fragen der Möglichkeit von Messung psychologischer Phänomene wie: Kann ein Merkmal überhaupt gemessen werden? Welche Fehler treten bei der Messung auf? Wie können diese Fehler vermieden bzw. mit welchen Methoden kontrolliert werden?

Nach Amelang & Zielinski (1997, 34) war der Ausgangspunkt für die Entwicklung einer Testtheorie die Feststellung von Spearman (1910), dass messfehlerbehaftete Variablen miteinander niedriger korrelieren, als sie es ohne Fehlerbehaftetheit tun würden. Die Testtheorie versucht also, auf Grund von Testergebnissen wahre Ausprägungen von menschlichen Merkmalen festzustellen. Sie geht davon aus, dass das Ergebnis eines Tests mit dem wirklichen Merkmal übereinstimmt, dass es aber dabei zu Messfehlern kommen kann. Diese Messfehler sind in dieser Vorstellung einzige Fehlerquelle, sodass diese bloß bestimmt werden müssen, damit wahre Aussagen über Merkmale getroffen werden können.

Die **klassische Testtheorie** hat daher drei grundlegende Probleme (Amelang & Zielinski 1997): Die Skalendignität der untersuchten Merkmale kann nicht genau angegeben werden, die Kennwerte sind stichprobenabhängig, die Zulässigkeit von Aussagen bleibt fraglich, und die Merkmale können nur operational definiert werden, da nicht überprüft werden kann, ob die Items des Tests bezüglich der Merkmale homogen sind.

Zur Umgehung dieser Probleme der „klassischen Testtheorie“ wurden mehrere Ansätze entwickelt, daher gibt es heute mehrere Testtheorien. Der wichtigste Ansatz neben der klassischen ist die **probabilistische Testtheorie** (Fischer 1995). Die meisten in der Praxis eingesetzten psychologischen Tests beruhen allerdings noch immer auf der klassischen Testtheorie. Diese geht davon aus, dass es einen wahren (deterministischen) Zusammenhang zwischen Messung und Merkmal gibt, wobei nur der (Mess)Fehler gering gehalten werden muss, aber nie ganz ausgeschaltet werden kann. Genau diesen Nachteil der klassischen Testtheorie versucht der Ansatz der probabilistischen Testtheorie (auch als stochastische bezeichnet) zu umgehen. Stochastische Modelle „gehen davon aus, dass zwischen dem Testergebnis und der zu messenden latenten Dimension von Natur aus ein Wahrscheinlichkeitszusammenhang besteht, nicht ein deterministischer, der bloß durch das Hinzukommen von Fehlern verwischt ist“ (Dorsch et al. 1994, 796).

Die probabilistische Testtheorie sieht Itembeantwortungen (beobachtete Variablen) als Indikatoren für latente Variablen. Diese latenten Variablen steuern in dieser Vorstellung das Verhalten im Test. Der Zusammenhang zwischen einer Itembeantwortung und latenten Merkmalen ist nicht deterministisch sondern probabilistisch, d.h. die Wahrscheinlichkeit einer Antwort auf ein Item (die Wahrscheinlichkeit der beobachteten Reaktionen) hängt von latenten Merkmalen ab. Es gibt zwei Arten latenter Variablen: **Verhaltensdispositionen** (Fähigkeiten, Einstellungen, Persönlichkeitsmerkmale) und **Itemmerkmale** (Schwierigkeit, Trennschärfe, Anreizmerkmale).

Mithilfe komplexer statistischer Modelle versucht die probabilistische Testtheorie, diese beiden Variablen möglichst genau zu bestimmen. In der Praxis unterscheiden sich allerdings die auf Grund dieser unterschiedlichen theoretischen Konzepte entwickelten Testverfahren kaum.

3.5 Welche Tests gibt es?

Es gibt verschiedene Ordnungsgesichtspunkte für Tests, die sich sowohl an inhaltlichen (z.B. Fragestellung, etwa Entwicklung, Leistung) als auch formalen Kriterien (z.B. Einsatzbereich, etwa klinische, schul- oder berufspsychologische Verfahren) orientieren können. Eine klassische Gruppierung ist jene hier gewählte in

- Fehler! Textmarke nicht definiert. **Intelligenztests**: Messen die allgemeine Intelligenz bzw. eine spezielle Begabung im Bereich der Intelligenz.
- Fehler! Textmarke nicht definiert. **Leistungstests**: Gemessen werden motorische, sensorische oder intellektuelle Leistungen.
- Fehler! Textmarke nicht definiert. **Persönlichkeitstests**: Erfassen Eigenschaften, Einstellungen und Interessen eines Menschen, z.B. die Kontaktfreude.

Die Zentralstelle für Psychologische Information und Dokumentation (1994) führt ein Verzeichnis psychologischer und pädagogischer Testverfahren aus der Datenbank PSYTKOM, die folgende Hauptkategorien umfasst:

1. Entwicklungstests
2. Intelligenztests

3. Kreativitätstests
4. Leistungstests
5. Fähigkeits- und Eignungstests
6. Schulleistungstests
7. Einstellungstests
8. Interessentests
9. Persönlichkeitstests
10. Projektive Verfahren
11. Verfahren im Bereich der Klinischen Psychologie
12. Verhaltensskalen
13. Sonstige Verfahren

3.5.1 Intelligenztests

Fehler! Textmarke nicht definiert. Der Begriff der Intelligenz wird innerhalb der wissenschaftlichen Psychologie sehr unterschiedlich definiert. Allen Erklärungsversuchen gemeinsam ist, dass sie als wichtigstes Moment jene Fähigkeit bezeichnen, sich an neue Situationen anzupassen bzw. Aufgaben mithilfe des Denkens zu lösen. Sie umfasst aber auch Faktoren wie Flexibilität des Denkens und die Fähigkeit zur Anpassung an die Erfordernisse neuer Situationen, die rasche Erfassung und Herstellung von Beziehungen und Sinnzusammenhängen.

William Stern prägte 1912 als Erster den Begriff „**Intelligenzquotient**“. Ausgangspunkt war das Problem, dass es recht wenig über die Intelligenz aussagt, wenn man nur weiß, um wie viele Jahre jemand vor- oder zurückliegt. Er verwendete daher für jede Altersstufe Aufgaben, welche Personen des entsprechenden Alters im Allgemeinen lösen konnten. Im endgültigen Test begannen die Getesteten dann mit den Aufgaben für die unterste Altersstufe und arbeiteten sich so lange hoch, bis sie die Aufgaben nicht mehr lösen konnten. Kam ein 18-jähriger nur bis zu den Aufgaben der 16-jährigen, war er für sein Alter unterdurchschnittlich, kam er bis zu denen der 21-jährigen, überdurchschnittlich intelligent. Es ist sicherlich nichts Besonderes, wenn ein 12-jähriger schon ein Intelligenzalter von 14 Jahren hat. Ein 6 Monate alter Säugling mit einem Intelligenzalter von 2 Jahren dagegen ist eine geradezu unglaubliche Sensation. Stern suchte ein altersunabhängiges Maß für die menschliche Intelligenz, teilte zu diesem Zweck Intelligenzalter durch das Lebensalter. Dieser Intelligenzquotient war aber nur bei Kindern sinnvoll, weil sich erfahrungsgemäß die Intelligenzentwicklung etwa ab dem zwanzigsten Lebensjahr stark verlangsamt und sich im Alter im Durchschnitt sogar wieder allmählich zurückentwickelt. Spätere Forscher multiplizierten das Ergebnis dann mit 100, um Nachkommastellen zu vermeiden.

Mit den Jahren wurden immer mehr methodische Probleme dieser Berechnung offenbar, sodass man bald dazu überging, den IQ als Abweichung einer Person vom Mittelwert ihrer Altersgruppe zu definieren und so etwas wie das Intelligenzalter zu vermeiden; der Begriff IQ blieb dennoch bis heute. Diese neue Sichtweise des IQ brachte aber ein anderes Problem: Abweichungen von einem Mittelwert kann man in der Statistik vielfältig ausdrücken. Die meisten IQ-Skalen haben ihren Mittelwert bei

100, aber sie sind verschieden gestreckt. So markiert ein IQ von 130 in einer Skala vielleicht die Grenze zu den oberen 2%, in einer anderen liegen u.U. 10% der Bevölkerung noch darüber. Deshalb sagt die Nennung eines IQ-Wertes allein überhaupt nichts aus; insbesondere in den Vereinigten Staaten werden Tests eingesetzt, deren Skalen bis weit über 200 hinausgehen, weswegen man Boulevard-Meldungen von Rekord-IQs mit Vorsicht begegnen sollte.

Intelligenztests sind ungeachtet ihrer messmethodischen Problematik ein wichtiges Hilfsmittel der Forschung und der psychologischen Beratung. Sie werden in fast allen Bereichen des öffentlichen Lebens eingesetzt (Schulberatung, Berufsberatung usw.). Ihr Nachteil besteht darin, dass sie kein richtiges Bild der gesamten Begabungen des Menschen vermitteln können, sondern immer nur einen Ausschnitt. Kreativität, schöpferisches Erfinden, soziale Fertigkeiten und künstlerische Fähigkeiten werden häufig nicht mit einbezogen, obwohl sie für den Lebens- oder Berufserfolg von größter Bedeutung sind. Zwar wurden auch Tests für diese Bereiche entwickelt, allerdings sind diese in vielen Fällen in ihrer Interpretation genauso schwierig wie die Definition der damit gemessenen Merkmale.

Fehler! Textmarke nicht definiert. Die meisten Intelligenztests sind so angelegt, dass einzelne Intelligenzfaktoren durch eine Reihe von Aufgaben verschiedener Schwierigkeitsgrade gemessen werden. Intelligenzfaktoren sind z.B. Sprachverständnis, logisches Denken, Raumvorstellung, Wahrnehmungsgeschwindigkeit, Rechenfähigkeit und Gedächtnisleistungen. Je mehr Aufgaben innerhalb einer festgelegten Zeit gelöst werden, desto höher ist die Ausprägung des betreffenden Intelligenzfaktors. Auf diese Art lassen sich zwei Dinge erkennen: der Gesamtintelligenzquotient (IQ), der angibt, welchen Rangplatz die Person innerhalb der Vergleichsgruppe einnimmt und das Intelligenzprofil, das einzelne Stärken und Schwächen der Intelligenz einer Person erkennen lässt. Da viele Testverfahren sprach- und kulturabhängig sind, wurden für interkulturelle Vergleiche so genannte sprachfreie Verfahren entwickelt, am bekanntesten ist der Progressive Matrizentest von Raven (1956).

3.5.2 Leistungstests

Leistungstests erfassen die Leistungsfähigkeit einer Person. Allgemeine Leistungstests erfassen Funktionsbereiche wie Aufmerksamkeit, Konzentration, Willensanstrengung, spezielle Leistungstests prüfen Funktionen und Fähigkeiten im motorischen und/oder sensorischen Bereich, erfassen räumliches Vorstellungsvermögen, Gedächtnisleistungen usw. Die meisten dieser Tests sind daraufhin konstruiert, den reaktiven Grad der Güte einer Fertigkeit zu bestimmen, die durch Lernen oder Üben erworben wurde. Die Testwerte (Ergebnisse) ergeben sich aus der Zahl der richtigen Lösungen (oder Fehler), durch die Lösungszeit oder durch die Qualität der Arbeit. Bei diesen Tests soll die Konzentrationsfähigkeit und in der Regel eine spezielle Eignung für eine bestimmte Aufgabe gemessen werden. Erreicht wird dies durch die Stellung von Routineaufgaben, die dann so schnell wie möglich gelöst werden müssen (Speed-Faktor).

3.5.3 Persönlichkeitstests

Diese Testverfahren sollen jene Merkmale einer Person erfassen, die nicht Intelligenz und Leistungsvermögen des Menschen zugeordnet werden können, sondern bestimmten Eigenschaften, Einstellungen, Neigungen, Interessen usw. Die ersten wissenschaftlichen Persönlichkeitstests wurden 1919 von R.S. Woodworth und Mitarbeitern entwickelt, die anhand eines Fragebogens US-Soldaten im ersten Weltkrieg untersuchten. Zu den klassischen Persönlichkeitstests gehören die Persönlichkeitsfragebögen, die in der Forschung am häufigsten eingesetzt werden. Typische Beispiele für Persönlichkeitstests sind **Fehler! Textmarke nicht definiert.** der 16-Persönlichkeitsfaktoren-Test - 16PF (Schneewind, Schröder & Cattell 1994) und das Freiburger Persönlichkeitsinventar – FPI (Fahrenberg, Hampel & Selg 1994).

Neben den mit Fragen und Statements arbeitenden Testverfahren gibt es unter den Persönlichkeitstests die verbalen Ergänzungsverfahren, die thematischen Apperzeptionsverfahren (z.B. Thematic Apperception Test - TAT), die Formdeutetests (z.B. Rorschach-Test), die spielerischen und zeichnerischen Gestaltungstests usw.

3.6. Praktische Tips für TestteilnehmerInnen

Für die Vorbereitung zur Teilnahme an einem psychologischen Test gelten ähnliche „Regeln“ wie bei jeder anderen Form einer Prüfung oder Untersuchung. Das gilt insbesondere für Eignungs- und Aufnahmetests. Einige wichtige Aspekte sollen hier erwähnt werden:

- Vor Testbeginn sollten alle TeilnehmerInnen von der Testleitung über den Ablauf des Test, über die Art des Test und über die Bekanntgabe der Ergebnisse informiert werden.
- Ein psychologischer Eignungstest darf nur von DiplompsychologInnen durchgeführt werden. Ihr Fachwissen ist Voraussetzung dafür, dass ein solcher Test richtig beurteilt werden kann und somit auch aussagefähig ist. Sind diese Voraussetzungen nicht gegeben, sollte die Teilnahme am Test abgelehnt werden.
- Wer an einem Test teilnimmt, muss auch über das Ergebnis des Tests informiert werden. Wenn er dies möchte, müssen ihm auch die Beurteilungskriterien erläutert werden.
- PsychologInnen, die einen Persönlichkeitstest durchführen, sind bei der Weitergabe der Untersuchungsergebnisse an den Auftraggeber oder dritte Personen an ihre gesetzliche Schweigepflicht gebunden.
- Bei den Fragen und Aufgaben des Test muss ein Zusammenhang zwischen den getesteten Merkmalen und den Anforderungen des Arbeitsplatzes bestehen. Fragen, die die Privatsphäre verletzen sind unzulässig. Aus diesem Grund sind Persönlichkeitstests meistens rechtlich bedenklich, weil sie häufig Fragen zum Privatleben enthalten. Unzulässige Fragen werden von TestteilnehmerInnen häufig nicht bemerkt, da sie zwischen den anderen Fragen nicht sofort auffallen. Außerdem ist die Testperson meistens aufgeregt und unter Zeitdruck, sodass ein Widerspruch gegen bedenkliche Fragen häufig ausbleibt.

- Wer zu einem Einstellungstest eingeladen wird, sollte sich am besten durch einschlägige Lektüre oder durch Gespräche mit Testerfahrenen vorbereiten. Insbesondere, wenn es das erste Mal ist, ist eine gute Vorbereitung sinnvoll, damit bei ungewöhnlichen Aufgaben nicht die Ruhe verloren geht.
- Den Abend vor dem Test sollte man nicht mit Testvorbereitungen verbringen, um nicht am Testtag blockiert zu sein.

Literatur- und Medienverzeichnis

- Amelang, Manfred & Zielinski, Werner (1997). Psychologische Diagnostik und Intervention. Berlin: Springer.
- Bortz, Jürgen (1984). Lehrbuch der empirischen Forschung für Sozialwissenschaftler. Berlin: Springer.
- Bortz, Jürgen (1993). Statistik für Sozialwissenschaftler. Berlin: Springer.
- Brickenkamp, Rolf (1997). Handbuch psychologischer und pädagogischer Tests. Göttingen: Hogrefe.
- Danziger, Kurt (1985). The origins of the psychological experiment as a social institution. *American Psychologist*, 40, 133-140.
- Davies-Oesterkamp, Susanne & Heigl-Evers, Annelise (1988). Psychodynamische Theorie. In Reinhold S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (61-65). München, Weinheim: PVU.
- Dorsch, Friedrich, Häcker, Hartmut & Stapf, Kurt H. (Hrsg.) (1994). *Psychologisches Wörterbuch*. Bern: Huber.
- Fahrenberg, J., Hampel, R. & Selg, H. (1989, 1994). *Freiburger Persönlichkeitsinventar (FPI) Revidierte Fassung FPI-R und teilweise geänderte Fassung FPI-A1*. Göttingen: Hogrefe.
- Fischer, Gerhard H. (1995). *Rasch models: foundations, recent developments, and applications*. New York: Springer.
- Fisseni, Herrmann-Josef (1990). *Lehrbuch der psychologischen Diagnostik*, Göttingen: Hogrefe.
- Funke, Joachim & Krüger, Thomas (1997). Let's test! Psychologische Datenerhebungen im Netz. *PSYCHOLOGIE HEUTE*, 9, 70-71.
- Graumann, Carl F. & Métraux, Alexandre (1977). Die phänomenologische Orientierung in der Psychologie (27-53). In Kurt A. Schneewind (Hrsg.), *Wissenschaftstheoretische Grundlagen der Psychologie*. München: Reinhardt.
- Grubitzsch, Siegfried (1991). *Testtheorie - Testpraxis*. Reinbeck bei Hamburg: Rowohlt.
- Guthke, Jürgen & Wiedl, Karl Heinz (1996). *Dynamisches Testen. Zur Psychodiagnostik der intraindividuellen Variabilität*. Göttingen: Hogrefe.
- Hager, Willi (1992):. *Jenseits von Experiment und Quasi-Experiment. Zur Struktur psychologischer Versuche und zur Ableitung von Vorhersagen*. Göttingen: Hogrefe.
- Hager, Willi & Westermann, R. (1983). *Planung und Auswertung von Experimenten*. In Jürgen Bredenkamp & Hubert Feger (Hrsg.), *Forschungsmethoden der Psychologie*, Band 5: *Hypothesenprüfung*. (24-238). Göttingen: Hogrefe.
- Häcker, Hartmut (1988). *Faktoretheorie: Cattell*. In Reinhold S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (67-76). München, Weinheim: PsychologieVerlagsUnion.
- Jacobs, Bernhard (1998). *Einführung in die Versuchsplanung*. WWW: <http://www.phil.uni-sb.de/~jakobs/seminar/vpl/index.htm> (00-11-17)
- Jäger, Reinhold S. & Petermann, Franz (1995). *Psychologische Diagnostik*. Weinheim: PsychologieVerlagsUnion.

- Kallus, Konrad Wolfgang. & Janke, W. (1988). In Reinhold S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (131-144). München, Weinheim: PsychologieVerlagsUnion.
- Kluge, Norbert & Reichel, Hans (Hrsg.) (1979). *Das Experiment in der Erziehungswissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Krauth, Joachim (1995). *Testkonstruktion und Testtheorie*. Weinheim: PsychologieVerlagsUnion.
- Legewie, Heiner & Ehlers, Wolfram (1992). *Knaurs moderne Psychologie*. München: Droemersch Verlaganstalt.
- Lehmann, Günter (1983). *Testtheorie: Eine systematische Übersicht*. In Jürgen Bredenkamp & Hubert Feger (Hrsg.), *Forschungsmethoden der Psychologie, Band 3: Messen und Testen* (427-543). Göttingen: Hogrefe.
- Lienert, Gustav A. (1989). *Testaufbau und Testanalyse*. München: PsychologieVerlagsUnion.
- Lienert, Gustav A. & Raatz, Ulrich (1994). *Testaufbau und Testanalyse*. Weinheim: PsychologieVerlagsUnion.
- Michel, Lothar & Conrad, Wolfgang (1982). *Theoretische Grundlagen psychometrischer Tests*. In Karl-Josef Groffmann & Lothar Michel, *Psychologische Diagnostik. Grundlagen psychologischer Diagnostik* (1-129). Göttingen: Hogrefe.
- Raven, John C. (1956). *Standard progressive matrices*. London: Watt.
- Reips, Ulf-Dieter (1999). *Warum Experimente im Web?*
WWW: <http://www.psych.unizh.ch/genpsy/Ulf/Lab/WebExpMethode.html> (99-12-06)
- Rost, Jürgen (1996). *Testtheorie Testkonstruktion*. Bern: Huber.
- Scheurer, Harald (1988). *Diagnostik als Testung*. In Reinhold S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (207-211). München, Weinheim: PsychologieVerlagsUnion.
- Schneewind, Kurt A. (1977). *Zum Verhältnis von Psychologie und Wissenschaftstheorie* (11-25). In Kurt A. Schneewind (Hrsg.), *Wissenschaftstheoretische Grundlagen der Psychologie*. München: Reinhardt.
- Schneewind, Kurt A., Schröder, G. & Cattell, R.B. (1994). *Der 16-Persönlichkeits-Faktoren-Test (16PF)*. Bern: Huber.
- Schwemmer, Oswald (1983): *Empirie ohne Experiment. Plädoyer für einen methodischen „Zwischenschritt“ bei unserer wissenschaftlichen Erfahrungsbildung* (66-99). In Gerd Jüttemann (Hrsg.), *Neue Aspekte klinisch-psychologischer Diagnostik*. Göttingen: Hogrefe.
- Sixtl, Friedrich (1982). *Messmethoden der Psychologie. Theoretische Grundlagen und Probleme*. Weinheim: Beltz.
- Stangl, Werner (1989). *Das neue Paradigma der Psychologie. Die Psychologie im Diskurs des Radikalen Konstruktivismus*. Braunschweig: Vieweg.
- Stangl, Werner (1993). *Handlungsforschung*.
WWW: <http://paedpsych.jk.unilinz.ac.at/INTERNET/ARBEITSBLAETTERORD/ERZIEHUNGSWISSENSCHAFT/Handlungsforschung.html> (00-01-29)
- Stangl, Werner (1997). *Kurzüberblick: Psychologische Schulen*.
WWW: <http://paedpsych.jk.unilinz.ac.at/INTERNET/ARBEITSBLAETTERORD/PSYCHOLOGIEORD/PsychologieSchulen.html> (00-01-29)
- Stangl, Werner (1997a). *Empirische Erziehungswissenschaft*.
WWW: <http://paedpsych.jk.unilinz.ac.at/INTERNET/ARBEITSBLAETTERORD/ERZWISSENSCHAFTEMPIRISCH/Erziehungswissenempirisch.html> (00-01-30)

-
- Stangl, Werner (1997). Grundbegriffe zu Theorie und Theoriekonstruktion.
WWW: <http://paedpsych.jk.uni-linz.ac.at/INTERNET/ARBEITSBLAETTERORD/ERZIEHUNGSWISSENSCHAFT/Theoriekonstruktion.html> (00-01-31)
- Tack, Werner H. (1994). Ziele und Aufgaben einer Allgemeinen Methodenlehre der Psychologie. In Theo Herrmann & Werner Tack (Hrsg.), *Forschungsmethoden der Psychologie*, Band 1: Methodologische Grundlagen der Psychologie. Göttingen: Hogrefe.
- Watson, John B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20, 158-177.

Angaben zur Person :

1947 in Wien geboren, Studium an der Universität Wien, seit 1975 an der Sozial- und Wirtschaftswissenschaftlichen Fakultät der Johannes Kepler Universität Linz am Institut für Pädagogik und Psychologie. Forschungsschwerpunkte sind u.a. psychologische Testverfahren, elterliche und schulische Erziehung, Wissenschaftstheorie der Psychologie, neue Lehr- und Lernmethoden. Seit 1996 Aufbau, Entwicklung und Betreuung des Abteilungsservers (<http://paedpsych.jk.uni-linz.ac.at/>), Herausgabe von p@psych, dem ersten pädagogisch-psychologischen e-zine im internet (ISSN 1561-2503), praktische Nutzung des internet im Rahmen der Entwicklung neuer Lehr- und Lernformen für die universitäre Lehre, Untersuchung der Einsatzmöglichkeiten des internet in Bildungseinrichtungen (internet @ Schule. Insiderwissen für LehrerInnen. Innsbruck: Studienverlag 2000).

Zusammenfassung

Test und Experiment werden als paradigmatische Methoden der Psychologie aus ihrem historischen und methodischen Kontext entwickelt und einander vergleichend gegenübergestellt. Das psychologische Experiment verkörpert in seiner strengen Form die Übernahme eines reduktionistisch-naturwissenschaftlichen Weltbildes und dominiert auch heute noch wesentliche Bereiche der wissenschaftlichen Psychologie. Der psychologische Test spielt eine wichtige Rolle in der psychologischen Praxis und ist in seinen vielfältigen Formen idealtypisch für das Handeln des Psychologen im Alltag.

Abstract

Tests and experiments are fundamental methods in psychology. They are compared with respect to their paradigmatic foundations, and are explained in their historical and methodological context. The psychological experiment represents the reduced scientific concept of the human world and dominates most of the research in this field. The psychological test, although emerging from scientific roots, is important in daily psychological practice and shows the breadth of everyday work.

Stichwörter

Test, Experiment, Psychologie, Gütekriterien, Messung, klassische Testtheorie, probabilistische Testtheorie, Differenzielle Psychologie, Online-Testen, Persönlichkeits-test, Leistungstest, Intelligenztest, Projektiver Test, Untersuchungsdesign, Stichprobe, Variable, Laborexperiment, Feldexperiment, ex post facto-Forschung, Reliabilität, Validität, Objektivität, Paradigma.