

1. Einleitung	1
1.1 Tests als Datenerhebungsmethode.....	2
1.2. Definitionen von Tests.....	5
1.3. Klassifikation von Tests	7
1.4. Vorurteile gegenüber Tests	8
1.5. Anwendungen für Tests	9
1.6. Fragen zur Einleitung.....	13

1. Einleitung

Die Psychologie als Lehre vom Erleben und Verhalten des Menschen befasste sich anfangs vorwiegend mit der Erarbeitung von Gesetzmässigkeiten, die für die meisten Menschen eines bestimmten Kulturkreises Gültigkeit beanspruchen können. Die Frage "Wie kommt es, dass sich Menschen gleich verhalten?" stand und steht im Blickpunkt der "Allgemeinen Psychologie". Eine weitere Frage erscheint aber ebenso interessant, nämlich "Wie kommt es, dass sich Menschen unterschiedlich verhalten?". Zur Beantwortung dieser Frage reichen "durchschnittliche", "allgemeine" Aussagen über das menschliche Erleben und Verhalten nicht aus. Vielmehr stehen die Aspekte der Unterschiede zwischen Personen, also die "interindividuellen Differenzen" im Mittelpunkt. Untersucht werden diese Fragen von der "Differentiellen Psychologie". Solche Untersuchungen sind zum Beispiel:

Feststellung der Schulreife

Eignungsbeurteilung für einen bestimmten Beruf

Auswirkung einer Werbekampagne auf das Kaufverhalten

Auswahl einer geeigneten therapeutischen Interventionsmethode

Die Unterschiede zwischen Menschen - zum Beispiel bezüglich ihres Aussehens oder Körperbaus - begegnen uns jeden Tag. Es sind Merkmale, die uns ins Auge stechen, also leicht zu beobachten sind. Doch wie kann man Unterschiede in psychischen Merkmalen feststellen? Die Psychologie als Sozialwissenschaft hat hierzu verschiedene Methoden entwickelt. Eine dieser Methoden wurde als "Test" bezeichnet und entwickelte sich zunächst im Blick auf praktische Fragestellungen (Feststellung der Schulreife von Kindern, Auswahl von Rekruten) und wurde darauf aufbauend auch für zahlreiche Fragen der Grundlagenforschung (insbesondere im Bereich der Persönlichkeits- und der Intelligenzforschung) eingesetzt.

1.1 Tests als Datenerhebungsmethode

Tests sind nur eine Methode zur Gewinnung von Daten in den Sozialwissenschaften. Weitere Datenerhebungsmethoden, die von Kerlinger (1979, S. 733-931) detailliert beschrieben werden sind in Tabelle 1 zusammengefasst.

Tabelle 1: Sozialwissenschaftliche Datenerhebungsmethoden

Methode:	Charakteristik der Methode:	Varianten der Methode:
Interview	Interpersonelle Rollensituation mit Befrager* und Befragtem	- standardisiert - halb-standardisiert - unstandardisiert
Fragebogenerhebung	Papier- und Bleistiftverfahren mit Einleitung (was wird untersucht?), demographischen Fragen und den inhaltlichen Fragen	- Fragen mit festgelegten Antwortalternativen - offene Fragen
objektive Tests und Skalen	Verfahren, das nach wissenschaftlichen Kriterien konstruiert wurde	- Intelligenz- & Fähigkeitstests - Tests zur Prüfung von Kenntnissen und Fertigkeiten - Persönlichkeitstests - Einstellungs- und Wertskalen
Verhaltensbeobachtung	Menschliches Verhalten steht im Blickpunkt	- Selbstbeobachtung - Fremdbeobachtung
Projektive Verfahren	Verfahren zur Untersuchung der Reaktionen auf unstrukturiertes Reizmaterial	- Assoziationstechniken - Konstruktionsverfahren - Ergänzungsverfahren - Expressive Verfahren
Inhaltsanalyse	Systematische, objektive und quantitative Analyse von sprachlichem oder geschriebenem Material, zur Erfassung von Merkmalen	- sprachliches Material - geschriebenes Material
Soziometrie	Messung zwischenmenschlicher Präferenzen	- Soziometrische Matrizen - Soziogramme - soziometrische Indizes
Semantisches Differential	Erfassung der psychologischen Bedeutung von Begriffen	
Q-Methodologie	Sortieren von Reizobjekten nach einer vorgegebenen Verteilung (Normalverteilung)	

* Obwohl im weiteren Text nur die männliche Form angegeben wird, sind beide Geschlechter angesprochen.

Interviews unterscheiden sich danach, wie stark der Befragter* an einen vorgegebenen Fragenkatalog gebunden ist. Mittels Interview können riesige Datenmengen erhoben werden. Entsprechendes gilt für Fragebögen, die mit offenen Fragen arbeiten. Die Fragen mit festgelegten Alternativen sind demgegenüber einfacher auszuwerten.

Objektive Tests sollten den wissenschaftlichen Gütekriterien genügen. Fahrenberg (1964) fasst die Definition von objektiven Tests noch enger: Es dürfen keine subjektiven Verfälschungen durch die untersuchten Personen möglich sein. Diese Tests sind für den Untersuchten völlig intransparent: Fragen werden zum Beispiel nicht danach ausgewertet, was man inhaltlich darauf antwortet, sondern wieviel Zeit man für die Entscheidung verwendet hat.

Durch Selbst- oder Fremdbeobachtung kann Verhalten registriert werden. Technologien wie die Videoregistrierung vereinfachen diese komplexe Methode. Bei den projektiven Verfahren steht ebenfalls sehr komplexes Verhalten im Mittelpunkt. Beim Rorschachtest (Rorschach, 1921) muss man zu Tintenklebsen assoziieren, beim Thematischen Apperzeptionstest (Murray, 1935, 1943) zu Bildern Geschichten konstruieren, beim Picture Frustration-Test (Rosenzweig, 1957) muss man die Sprechblasen von Comicbildern ergänzen, und bei expressiven Verfahren wie beim Spielen oder Zeichnen ist Ausdrucksfähigkeit gefragt.

Verschiedene Arten von Material können durch Inhaltsanalysen untersucht werden. So kann zum Beispiel politische Propaganda in Radiosendungen analysiert werden. Wahlsituationen stehen im Mittelpunkt der Soziometrie. Der "Star" einer Schulklasse kann ermittelt werden, indem man die Schüler einzeln befragt, neben wem sie sitzen möchten oder wer Klassensprecher werden soll.

Im Semantischen Differential (Osgood, Suici & Tannenbaum, 1953) wird ein Reizobjekt, zum Beispiel das Wort "Vater", anhand von Adjektivpaaren eingestuft. Man erhält so die konnotative Bedeutung des Begriffes "Vater". Bei der Q-Methodologie (Stephenson, 1953) werden mehrere Reizobjekte, zum Beispiel Politiker, auf Kärtchen in Kästchen geordnet. Die Ordnung soll danach erfolgen, wer am besten für das Land sorgen wird. Zudem soll die Anzahl der Kärtchen in den einzelnen Kästchen eine Normalverteilung beschreiben.

Die verschiedenen Datenerhebungsmethoden müssen anhand wissenschaftlicher Kriterien bewertet und verglichen werden. Die verschiedenen Modelle der Testtheorie bieten den Rahmen für einen solchen Vergleich. Die Testtheorie ist damit ein theoretisches Rahmenmodell, mit dessen Hilfe neben Tests viele andere Datenerhebungsmethoden betrachtet werden können.

* Obwohl im weiteren Text nur die männliche Form angegeben wird, sind beide Geschlechter angesprochen.

1.2. Definitionen von Tests

Im Alltagsverständnis wird die Anwendung psychologischer Tests oft mit der Psychologischen Diagnostik gleichgesetzt. Tatsächlich ist die Psychologische Diagnostik eine wissenschaftliche Disziplin, deren Funktion darin besteht, Methoden zu entwickeln und anzuwenden, a) zur Gewinnung psychologisch relevanter Charakteristika von Personen oder Bedingungen, b) zur Integration der Daten zu einem Urteil sowie c) zur Vorbereitung und Evaluation von Entscheidungen (Jäger, 1988). Psychologische Tests sind eine Teilmenge der Datenerhebungsmethoden und diese sind wiederum eine Teilmenge der Psychologischen Diagnostik.

Der Begriff "Test" hat sich aus dem Englischen eingebürgert und findet im alltäglichen Sprachgebrauch immer häufiger Verwendung. Er wird im Sinne von Prüfung oder Probe verwendet und bezeichnet die Untersuchung einer Sache. In der Schule werden Klassenarbeiten in Tests umbenannt. Im technischen Bereich werden Qualitätsprüfungen als Tests bezeichnet. Im wissenschaftlichen Bereich taucht der Begriff in zwei Gebieten auf: Erstens wird beim Hypothesentesten geprüft, ob eine Hypothese angenommen werden kann, oder ob sie verworfen werden muss. Zur Prüfung einer wissenschaftlichen Hypothese muss man diese in eine statistische Hypothese überführen. Ein theoretischer, möglichst aus einer Theorie abgeleiteter Satz soll hierbei in eine statistische Aussage überführt werden, die mittels statistischer Tests geprüft werden kann. Das zweite Gebiet definiert Tests als Messinstrumente, die bestimmte Kriterien erfüllen. Wottawa (1980) stellt folgendes fest:

Tests können aus methodischer Sicht als Verfahren charakterisiert werden, mit denen über einzelne Personen (aber auch über Objekte, Sachverhalte und dergleichen) Information gewonnen werden. Typisch für Tests ist, dass man sich bemüht, über die ausgewählten Personen (Objekte, Sachverhalte) vergleichbare Erkenntnisse zu sammeln. Dies setzt in der Regel voraus, dass für die Testung eine mehr oder weniger stringent vereinbarte Durchführungsform festgelegt wird. Ein weiteres Kennzeichen ist, dass im Regelfall die Durchführung eines Tests nicht ohne bestimmte Absicht, ohne eine Zielsetzung des Testers erfolgt. (S. 11)

Lienert (1967) gibt folgende Definition eines Tests:

"Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung." (S. 7)

Die Betrachtung von Wottawa erlaubt es unterschiedliche Informationserhebungsmethoden als Tests zu bezeichnen. Crash-Tests in der Automobil-Industrie, Krankheitsdiagnosen aufgrund ärztlicher Untersuchungen, Qualitätsprüfung von Waren u.s.w. sind nach dieser Kennzeichnung Tests. Wottawa weist aber darauf hin, dass man im üblichen psychologischen Sprachgebrauch unter Tests nur ein erprobtes und veröffentlichtes Erhebungsinstrument versteht. Lienerts Definition ist deutlich formaler. Der Test ist ein wissenschaftliches Verfahren, das bestimmte Gütekriterien erfüllen muss. Es werden Persönlichkeitsmerkmale untersucht, die man auch als individuelle psychische Merkmale auffassen kann. Zudem soll eine möglichst quantitative Aussage gemacht werden, d.h. Aussagen, die Messungen beinhalten und die in Form von Zahlen ausgedrückt werden. Der relative Grad der Merkmalsausprägung beinhaltet zudem ein Vergleich zu einer Untersuchungsgesamtheit bezüglich des gemessenen Merkmals.

Die Psychologie als Wissenschaft vom Erleben und Verhalten des Menschen muss natürlich Informationen über Menschen sammeln. Dabei entstehen methodische Schwierigkeiten, die in anderen Bereichen nicht in dem Masse auftreten. Wenn wir beispielsweise die physikalische Grösse eines Menschen feststellen wollen, so können wir ihn mit der Länge eines Meterstabes vergleichen. Wir erhalten so einen vergleichbaren Wert für die Länge des Menschen. Wollen wir hingegen eine psychologische Eigenschaft eines Menschen erfassen, wie zum Beispiel die Stärke der Neigung sich in Krisensituationen unkontrolliert zu verhalten, so stehen wir zunächst vor dem Problem, dass wir diese Eigenschaft nicht direkt beobachten können. Solche Kennzeichen, die wir nicht direkt beobachten können, die wir aber als grundlegend ansehen nennt man "latente Eigenschaften". Wir können somit nur durch indirektes Erfragen dieses Charakteristikum erfassen. Bei der Feststellung eines physikalischen Kennzeichens, wie der Grösse eines Menschen, könnte es nun vorkommen, dass wir nach der Messung unsicher werden, ob wir den Meterstab richtig angelegt haben. Um uns zu vergewissern, wollen wir ein zweites Mal messen. Dabei kommt eine Differenz zustande, und wir werden verunsichert, was nun die exaktere Messung war. Eine Lösung könnte sein, die Messung öfter zu wiederholen, und uns für den Wert zu entscheiden, der am häufigsten aufgetreten ist. Analog können wir bei der Erfassung der psychologischen Eigenschaft nicht sicher sein, ob wir durch eine Frage die latente Eigenschaft adäquat erfasst haben. Wir sollten folglich mehrere Fragen verwenden, um ein Merkmal zu erfassen. Die Fragen, die wir dabei stellen, müssen allerdings in etwa denselben Merkmalsbereich erfragen. Die Fragen müssen in sich "homogen" sein. Wenn die Items wirklich homogen sind, so können sie zu einer "Skala" zusammengefasst werden.

1.3. Klassifikation von Tests

Brickenkamp (1975) teilt Tests nach inhaltlichen Kriterien ein. Er unterscheidet Leistungstests und Persönlichkeitstests.

Unter Leistungstests werden folgende Testarten subsumiert:

- Entwicklungstests
- Intelligenztests
- Schultests
- Allgemeine Leistungstests
- Spezielle Funktions- und Eignungstests

Persönlichkeitstests werden untergliedert in:

- Einstellungs- und Interessentests
- Persönlichkeitstests
- Klinische Tests

Entwicklungstests sollen lebenslaufbezogene Veränderungen oder Entwicklungsstufen feststellen, die sich im Erleben und Verhalten eines Menschen von den ersten vorgeburtlichen Anfängen bis zum Lebensende ergeben. Intelligenztests sind Tests zur Erfassung des Konstruktes "Intelligenz". Sie spielen bei der historischen Entwicklung von Testverfahren eine besondere Rolle. Schultests prüfen die Schulleistung zum Beispiel in einem bestimmten Fach, oder sie stellen die Schulfähigkeit (Schulreife bzw. Umschulungsfähigkeit) fest. Allgemeine Leistungstests sollen die allgemeinen Voraussetzungen für Leistungsfähigkeit erfassen. Leistungsverhalten wird in den Tests meist dadurch erzeugt, dass die Versuchsperson schon in der Instruktion aufgefordert wird ihr Bestes zu geben. Spezielle Funktions- und Eignungstests ist der Ueberbegriff für eine Klasse von Tests, die spezifische Funktionen wie zum Beispiel Händigkeit oder Psychomotorik testen. Diese Tests stehen oft in engem Zusammenhang mit Berufseignung. Die Eignung für einen bestimmten Beruf wird ansonsten in Eignungstests festgestellt.

Einstellungstests messen Einstellungen oder Meinungen über bestimmte Bezugsobjekte, Interessentests eher Vorlieben und Abneigungen gegenüber dem Objekt. Persönlichkeitstests erfassen Persönlichkeitsmerkmale im Bereich der sogenannten "normalen" Persönlichkeit, was sie von den Klinischen Tests unterscheidet, die eher "anormale", "psychopathologische" Verhaltensweisen zu erfassen versuchen. Für die Unterscheidung Klinische Tests und Persönlichkeitstests gilt, was für die Klassifikation im allgemeinen zutrifft, nämlich dass die Grenzen zwischen ihnen fließend sind.

1.4. Vorurteile gegenüber Tests

Im Alltagsleben begegnen wir Tests vor allem dann, wenn wichtige Entscheidungen anstehen. Die Tests bleiben häufig nicht ohne Konsequenzen. Im Einzelfall kann dies bedeuten, dass man beispielsweise einen erwünschten Beruf nicht erlernen kann. Viele Menschen sind bei solchen Prüfungen emotional sehr stark beteiligt. Wottawa (1980) schildert ein Beispiel aus einem anderen Bereich. Viele Personen sind schon bei der technischen Überprüfung ihres Fahrzeugs sehr erregt. Diese emotionale Beteiligung zusammen mit der standardisierten, unpersönlichen Situation der Testerhebung, den psychologischen Testaufgaben, die oft für den Laien nichts mit dem Untersuchungsziel zu tun haben und schliesslich die Erfassung der Ergebnisse in Zahlen, können zu Vorurteilen gegenüber psychologischen Tests führen. "Hier werden meine geheimsten Wünsche und Absichten erfasst", "Tests sind Instrumente zur unberechtigten Selektion", "Man wird an seinen Absichten gehindert" und "Der Mensch wird auf Zahlen reduziert" können solche Vorurteile sein. Diese Vorurteile können durchaus ihre Berechtigung haben, wenn Tests beispielsweise von Berufsgruppen angewendet werden, die keine Ausbildung in Testtheorie und Testkonstruktion haben. Aber auch ethische Probleme treten auf, nämlich immer dann, wenn Individuen etwas tun, was Einfluss auf das Leben eines anderen hat. Psychologische Berufsverbände haben ethische Richtlinien aufgestellt, die die Versuchsperson oder den Klienten davor schützen sollen, dass gegen moralische Massstäbe verstossen wird. Für die Anwendung psychologischer Tests sollte u.a. eine Richtlinie sein, dass keine Anwendung erfolgt, ohne anschliessendes psychologisches Gespräch über die Ergebnisse.

In einem persönlichen Gespräch kann geklärt werden, was es mit den Zahlen als Testergebnisse auf sich hat. Zahlen stellen eine vereinfachte Form der Informationsdarstellung dar. Zahlen sind nichts anderes als Symbole, die auch in Form von Begriffen aus dem Alltagsleben ausgedrückt werden könnten. Zahlen sind aber deutlich einfacher darzustellen. Wottawa (1980) gibt folgendes Rechenbeispiel. Es liegt ein Test mit 20 Fragen (Items) vor, so könnte man die Testleistung eines jeden Probanden verbal umschreiben. Dies würde aber dazu führen, dass für jedes mögliche Testergebnis - bei 20 Aufgaben gibt es $2^{20}=1048576$ verschiedene Möglichkeiten der

Testbeantwortung - eine eigene Beschreibung erstellt werden müsste. Selbst wenn man nur die unterschiedlichen Häufigkeiten von "richtigen" Lösungen betrachtet, für jedes der 21 verschiedenen möglichen Resultate (von 0 bis 20 richtigen Antworten), müsste man sehr umständliche, verbale Bezeichnungen einführen, etwa von "ganz besonders extrem wenig Items gelöst" bis zu der entsprechend umständlichen Formulierung am anderen Ende der Beschreibungsskala. Man könnte auch anstelle solcher verbaler Bezeichnungen Symbole verwenden, wie zum Beispiel jedes der 21 Resultate mit einem griechischen Buchstaben bezeichnen. Die Reihenfolge der Buchstaben im Alphabet könnte dazu verwendet werden, um die Unterscheidung zwischen "besseren" und "schlechteren" Ergebnissen wiederzugeben. Es ist aber sicher am einfachsten Personen ohne richtigen Antwort die 0, Personen mit einer richtigen Antwort die 1 usw. zuzuweisen.

1.5. Anwendungen für Tests

Am Anfang der Entwicklung psychologischer Tests (Binet & Simon 1905, Terman 1937) stand das Bestreben, Menschen hinsichtlich ihrer Leistungsfähigkeit in Gruppen einzuteilen. Binet und Simon bemühten sich, ein Messinstrument zu konstruieren, das zu einem frühen Zeitpunkt die Feststellung der Schulreife von Kindern (und damit eine eventuell erforderliche Zuweisung in eine Sonderschule) ermöglichen sollte. Bei dem "Army-Alpha-Test" von Terman ging es um das Aussortieren von ungeeigneten Rekruten. Im Vordergrund dieser Arbeiten stand also das Bestreben nach Selektion, das auch heute noch vielfach das Ziel der Konstruktion bzw. der Anwendung von Tests ist (Personalauslese, Aufnahme in bestimmte Ausbildungsrichtungen etc.).

Selektion ist damit der erste Anwendungsfall für Tests. Die Selektion kann man für die Berufsauswahl mit der amerikanischen Devise "The right man on the right job" umschreiben. Sie hat meist eine negative konnotative Bedeutung, denn hierin wird das Verwehren von Chancen gesehen. Man muss allerdings Merkmale der Person und Merkmale der Situation auseinanderhalten. Wenn die Situation so ist, dass aus "Platzmangel" (zu wenig Stellen bzw. Ausbildungsplätze) ein Bewerber abgelehnt werden muss, so kann man sicher von unberechtigter Selektion sprechen. Wenn die Merkmale der Person es aber erfordern, dass sie abgelehnt werden muss, ist Selektion sicher berechtigt. In Deutschland werden zum Beispiel Personen, die im Strassenverkehr zweimal durch Trunkenheitsfahrten aufgefallen sind, einer medizinisch-psychologischen Untersuchung unterzogen. Neben anderen Daten werden hier auch psychologische Tests herangezogen, um Aussagen über die momentane Alkoholgefährdung der Person zu machen. Kommt nun die Untersuchung zum Schluss, dass eine erhöhte Alkoholgefährdung festzustellen ist, kann dem Untersuchten bis auf weiteres seine Fahrerlaubnis entzogen werden. Die Entscheidung, dass die Person mit ihrem Fahrzeug nicht mehr am öffentlichen Strassenverkehr teilnehmen darf, wird allerdings nicht vom Psychologen oder

Mediziner getroffen, sondern von der Verkehrsbehörde, die sich die beiden Berufsgruppen als Ratgeber heranzieht. Die Entscheidung hat wichtige Konsequenzen für den Betroffenen, aber auch für die Gesellschaft. Man denke nur an mögliche Unfallopfer bei einer falschen Entscheidung.

Sichere Entscheidungen sind in solchen und anderen Fällen der Selektion nicht möglich. Die Entscheidungen erfolgen auf der Grundlage unvollständiger Information. Die Aussagen über die geeigneten und ungeeigneten Bewerber können nur mit einer gewissen Wahrscheinlichkeit erfolgen. Viele Entscheidungen in unserer Gesellschaft geschehen eher zufällig, beziehungsweise auf einer nicht näher erläuterten Datenbasis. Tests machen die Entscheidung aber sicherer. Je sorgfältiger getestet wird, je besser die herangezogenen Verfahren sind und je gründlicher zuvor eine Klärung der inhaltlichen Zusammenhänge zwischen Testverfahren und den späteren Anforderungen erfolgt ist, um so seltener wird es zu Fehlentscheidungen kommen.

Dies sind Tatsachen, die im Prozess der Entscheidungsfindung berücksichtigt werden müssen, genauso wie letztlich beim Urteil reflektiert werden muss, wie es zustandekam. Die sozialpsychologische Forschung, insbesondere zur sozialen Urteilsbildung, hat die möglichen Fehlerquellen bei der Urteilsbildung aufgezeigt (Upmeyer, 1985).

Neben der Selektion werden Tests auch zur Beschreibung von Personen herangezogen. In vielen Zeitschriften findet man heute "Tests", die dem Leser gewissen Eigenschaften zuschreiben, ihm sagen "wie er ist", "ob er die richtigen Freunde hat" und ähnliches. Meist sind diese Tests so konstruiert, dass sie nicht zwischen Personen differenzieren; es kommen alle zu demselben Ergebnis. Wenn die Ergebnisse unterschiedlich sind, so sind deren Interpretationen oft so allgemeingültig, dass sie nichts aussagen. Dennoch besteht die Gefahr, dass Einzelpersonen zu extremen Ergebnissen gelangen. Sie sind danach mit ihrem Ergebnis alleingelassen.

Psychologen benötigen Beschreibungen von Personen auf der Basis von psychologischen Tests vor allem im Bereich der Beratung. In der psychologischen Beratung ist aber immer eine Situation gegeben, in der Proband und der Berater interagieren können. Dies bedeutet, dass sie gemeinsam über Ergebnisse sprechen und gemeinsam Lösungen finden können.

Die Bewertung von Personen ist ebenfalls ein Anwendungsgebiet von Tests. Im schulischen Bereich wird oft nach einer Bewertung der Leistungsfähigkeit von Schülern gestrebt. Ziel ist es, den Kenntnisstand zu prüfen, um vielleicht geeignete Massnahmen zu ergreifen.

In den drei bisher genannten Anwendungsbieten von Tests (Selektion, Beratung und Bewertung) wird versucht Personen sinnvoll in Gruppen einzuteilen, sei es, dass man für jede Einzelgruppe eine spezielle Beratung vorschlägt, sie mit einer bestimmten Bewertung versieht, oder dass man auf der Basis der Gruppeneinteilung verschiedene praktische Massnahmen (Zulassung bzw. Ablehnung, Auswahl einer speziellen Therapieform) einleitet.

Ein weiteres wissenschaftliches Anwendungsgebiet von Tests ist die Operationalisierung. Begriffe, die nicht unmittelbar beobachtbar sind, müssen in konkret beobachtbare Indikatoren entschlüsselt werden. Solch ein Begriff, auch "Konstrukt" genannt, ist beispielsweise die "Handgeschicklichkeit". Was können nun Indikatoren sein, die darauf schliessen lassen, dass eine Person über hohes Handgeschick verfügt? In der "motorischen Leistungsserie nach Schoppe" (MLS), einem apparativen Testverfahren, wird dies unter anderem dadurch geprüft, dass die Versuchsperson auf einer Metallplatte mit einem Metallstift durch eine geschwungene Rille fahren muss, ohne die Ränder der Rille zu berühren. Dies ist ein Operationalisierungsvorschlag für das Konstrukt "Handgeschicklichkeit". Für zahlreiche psychologisch relevante Variablen liegen fertige Operationalisierungsvorschläge in Form von Tests vor, auf die man bei entsprechendem Bedarf zurückgreifen kann.

Bei Anwendungen von Tests für diese Fragestellung steht nicht die Beschreibung der Einzelperson im Vordergrund, sondern der Versuch, Zusammenhänge zwischen verschiedenen operationalisierten Grössen an Gruppen von Personen zu prüfen. Dies ist nicht nur ein Problem der Grundlagenforschung; auch viele berufsnahe, angewandte Forschungstätigkeiten können sich mit grossem Vorteil der vorhandenen Tests bedienen; oder es ist möglich, für das vorliegende Problem einen neuen, speziell auf die konkrete Situation abgestellten Test zu konstruieren (etwa bei Operationalisierung von Begriffen wie "Behandlungserfolg", "Fahreignung", "Berufserfolg"). In den meisten Fällen liegen auch den Tests, die vorwiegend zu Zwecken der Selektion oder der Beschreibung eingesetzt werden, ein Operationalisierungskonzept zugrunde.

In der psychologischen Forschung werden entsprechend Zusammenhänge zwischen verschiedenen operationalisierten Grössen an Gruppen von Personen geprüft. Das Zusammenspiel verschiedener Konstrukte kann so erschlossen werden. Dies ist eine andere Fragestellung als die bisher betrachtete. Sie ist die Grundlage für die Konstruktion neuer Tests. Neue Tests sollten zeigen können, dass sie zu den bisher bestehenden wirklich neue Erkenntnisse beinhalten. Dabei ist nicht entscheidend, dass immer neue verbale Konstrukte entstehen, sondern dass sich die neuen Tests durch vielfältige Untersuchungen in bereits bestehende Konstrukte einordnen lassen.

Ob Tests eher "normorientiert" oder "kriterienorientiert" konstruiert wurden ist ein weiteres Unterscheidungskriterium für ihren Anwendungsbereich. Bei normorientierten Tests werden die Ergebnisse einer Person in bezug zu Gruppen von Personen gestellt, die quasi als Referenz für die Person gelten soll. Die Gruppen von Personen gelten als Norm, mit der die Person verglichen wird. Die kriterienorientierten Tests zeigen demgegenüber wo die Person mit ihren Ergebnissen in bezug zu einem Kriterium, zum Beispiel einem Lernziel steht. Hier bietet also das Kriterium die Referenz, und nicht andere Gruppen von Personen. Die "kriterienorientierte Messung" (Fricke, 1974; Klauer, 1987) ist ein neuer interessanter Ansatz zur Konstruktion von Tests. Reicherts (1985) zeigt für die Klinische Psychologie die Entwicklung eines kriterienorientierten Test.

Nach dieser Einleitung, die sich sehr stark an Wottawa (1980) anlehnt, wird ein kurzer historischer Abriss zur Entwicklung und Anwendung von Tests in Anlehnung an Drenth (1969) gegeben. Die Grundlagen der Messung und Skalierung wurde aus Ahrens (1974) und diversen anderen Autoren wie z. B. Kerlinger (1974) zusammengetragen. Die klassische Testtheorie wurde aus einem unveröffentlichten Lehrveranstaltungsskriptum von Spada und Rost unter Rückgriff auf Unterlagen von Scheiblechner sowie aus dem neuen Buch von Diehl & Kohr (1989) kombiniert. Zusätzlich habe ich starken Wert auf die graphische Veranschaulichung gelegt. Für die kurze Darstellung der probabilistischen Testtheorie wurde ebenfalls eine grösstenteils graphische Veranschaulichung gewählt, um einen Einblick in die "modernen" Testmodelle in Anlehnung an Rost und Spada (1978) zu geben. Die Faktorenanalyse von Tests wurde der sehr guten Darstellung in Diehl und Kohr (1989) entnommen. Im Blickpunkt stehen der Zusammenhang zur Testtheorie und das Verständnis der Faktorenanalyse und nicht die mathematische Begründung des Verfahrens.

Der Testkonstruktionsteil gibt eindeutige Richtlinien zum Erstellen eigener Tests und enthält zudem eine Einführung in die computerunterstützte Auswertung der Tests mittels Personal Computer oder Grossrechner.

Da Tests in vielen Anwendungsfeldern Hilfe leisten sollen, um bessere Entscheidungen zu treffen, wurde zum Abschluss ein Kapitel zur Entscheidungstheorie angefügt, das sich sehr stark an Wiggins (1973) anlehnt, aber ebenfalls mit Graphiken aufgelockert wurde.

1.6. Fragen zur Einleitung

1. Ergänze die Leerstellen in der Definition eines Tests nach Lienert (1967):

Ein Test ist ein Routineverfahren zur Untersuchung eines oder mehrerer
..... abgrenzbarer Persönlichkeitsmerkmale, mit dem Ziel einer
möglichst Aussage über den Grad der
..... Merkmalsausprägung.

2. Wie unterscheidet sich die normative (normorientierte) von der kriterienorientierte Messung?

3. Was bedeutet "Operationalisierung"?

4. In sozialwissenschaftlichen Untersuchungen werden oft auch physiologische Messungen verwendet, wie zum Beispiel die Messung der Herzfrequenz, des Blutdrucks oder der Hautleitfähigkeit. In welche Kategorie der Datenerhebungsmethoden nach Kerlinger (1979) würdest Du diese Messungen einordnen?