

Bernd Marcus

Grundlagen der Testkonstruktion

kultur- und
sozialwissenschaften

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Inhaltsverzeichnis

GRUNDLAGEN DER TESTKONSTRUKTION

Bernd Marcus & Markus Bühner

Einführung in den Kurs	5
1 Grundlagen psychologischer Testverfahren.....	8
1.1 Anwendungsbereiche und Klassifikation von Testverfahren	10
1.2 Tests und Testtheorie	14
1.3 Grundlagen der Testtheorie.....	16
1.3.1 Einführung in die Klassische Testtheorie (KTT).....	17
1.3.2 Kritische Anmerkungen zur Klassischen Testtheorie.....	22
1.3.3 Grundideen der Probabilistischen Testtheorie (PTT)	24
2 Testkonstruktion: Von der Planung zum ersten Entwurf.....	27
2.1 Merkmalsbereich und Geltungsbereich	27
2.2 Festlegung des generellen Konstruktionsprinzips.....	30
2.3 Generierung eines Itempools	38
2.4 Sprachliche Formulierungsregeln und Antwortformate	42
3 Itemanalyse: Kürzung und Revision des Entwurfs	58
3.1 Itemanalyse nach Klassischer Testtheorie (KTT).....	58
3.1.1 Rekodierung, Deskriptive Itemanalyse und Itemschwierigkeiten	59
3.1.2 Trennschärfeanalyse.....	68
3.2 Überblick über die Probabilistische Testtheorie	79
3.2.1 Grundlagen des Rasch-Modells	80
3.2.2 Weitere probabilistische Testmodelle für dichotome Daten.....	85
3.2.3 Probabilistische Modelle für Ratingskalen	89
4 Skalenbildung: Faktorenanalyse, Aggregation von Items und Normierung.....	95
4.1 Explorative Faktorenanalyse (EFA).....	95
4.1.1 Grundlagen der Faktorenanalyse	95
4.1.2 Varianten und Vorgehensweise bei der Explorativen Faktorenanalyse	98
4.1.3 Beispiel für die Durchführung einer EFA mit SPSS.....	113
4.1.4 Bewertung der Explorativen Faktorenanalyse	127
4.2 Konfirmatorische Faktorenanalyse (CFA).....	129
4.2.1 Grundbegriffe Linearer Strukturgleichungsmodelle (SEM).....	130
4.2.2 Vorgehensweise bei einer Konfirmatorischen Faktorenanalyse.....	135
4.3 Aggregation zu Skalenwerten, Normierung und Interpretation.....	153
4.4 Zusammenfassung der Schritte der Testkonstruktion	161
5 Evaluation psychologischer Testverfahren	165
5.1 Richtlinien und Beurteilungssysteme für Tests	165

5.2	Psychometrische Gütekriterien	169
5.2.1	Objektivität und Reliabilität	169
5.2.2	Validität	180
5.3	Weitere allgemeine Gütekriterien	198
5.3.1	Kriterien mit vorwiegend praktischem Anwendungsbezug	198
5.3.2	Kriterien mit Bezug zu Rechten und Reaktionen der Teilnehmer	200
5.3.3	Kriterien mit Bezug zum Verhalten der Teilnehmer	203
5.4	Gütekriterien für die Einzelfalldiagnostik	206
5.4.1	Anforderungen an die Normierung	206
5.4.2	Kennwerte mit Bezug zur Messgenauigkeit	208
	Literaturverzeichnis	218
	Stichwortverzeichnis	232

Einführung in den Kurs

„Was Ihre Frisur über Sie verrät“. Eine Online-Frauenzeitschrift beantwortet diese Frage, indem einfach aus sechs verschiedenen Frisuren ausgewählt werden kann. Der Erstautor dieses Studienbriefs hat diesen „psychologischen Test“ nach bestem Wissen unter Auswahl der subjektiv ähnlichsten Frisur (Stirn frei) bearbeitet, mit folgendem Ergebnis: „Sie begegnen dem Leben auf direkte Weise und Sie haben keine Angst, direkt auf Ihr Ziel zuzugehen. Sie sind offen und ehrlich und haben dynamische Qualitäten. Durch Ihre aktive Energie kann es Ihnen manchmal an Feinheit fehlen, denn Sie kümmern sich nicht unbedingt immer um die Details. Sie gehen schnell zum Wesentlichen über. Sie lieben den Kontakt mit Ihren Mitmenschen und Sie sind optimistisch. Ihre Frisur zeigt, dass Sie der Welt und dem Leben gegenüber offen sind.“ Unter Zuhilfenahme eines Kammes wäre es ihm nach diesem Test auf recht einfache Weise gelungen, seine Persönlichkeit grundlegend zu verändern (Mittelscheitel): „Sie möchten in einer harmonischen und ausgeglichenen Atmosphäre leben. Mit einer gewissen Strenge versuchen Sie, in allen Dingen gerecht zu sein. Sie sind geradeheraus und Sie streben nach einer einfachen, direkten und soliden Existenz und ebensolchen Beziehungen. Sie wünschen sich eine moralische Ordnung, die Ihnen entspricht. Sie brauchen geistreiche Kontakte und ein Leben, das reibungslos und ohne zu viel Aufregung verläuft. Ihre Frisur offenbart Ihren leidenschaftlichen und energischen Charakter.“ (alle Zitate nach Gofeminin.de, n.d.)

Wenn Sie es sich nicht ohnehin schon dachten, sollte Ihnen nach der Lektüre dieses Studienbriefs klar werden, dass es sich bei „Was Ihre Frisur über Sie verrät“ *nicht* um einen psychologischen Test handelt. Zwischen dem, was sich eine Redakteurin vermutlich bei einer Tasse Kaffee ausgedacht hat und dem, was Psychologinnen in jahrelanger intensiver Arbeit entwickeln, bestehen bestenfalls oberflächliche Ähnlichkeiten. Die Konstruktion psychologischer Tests nach psychometrischen Prinzipien gilt als eine der bedeutendsten Leistungen der wissenschaftlichen Psychologie und als ein wesentlicher Beitrag der Verhaltenswissenschaften zur Gesellschaft (z.B. American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999). Die vertiefte Kenntnis dieser Prinzipien gehört zu den Qualifikationen, welche die Psychologie normalerweise von allen anderen Disziplinen unterscheidet und ist damit auch ein Alleinstellungsmerkmal für Psychologen im Wettbewerb auf dem Arbeitsmarkt. In diese Kenntnisse soll der vorliegende Studienbrief einführen. Allerdings werden Tests keineswegs nur von Psychologen angewandt (vgl. Kap. 1.1).

Psychologische Testkonstruktion ist ein im Wesentlichen methodisches Teilgebiet der Psychologie, das eine bestimmte statistisch fundierte Sozialtechnologie bereitstellt. Viele Aspekte werden in der Fachliteratur auf einem mathematisch recht anspruchsvollen Niveau diskutiert. Natürlich kann eine Einführung hier nur die Grundprinzipien vermitteln. Vorausgesetzt werden Kenntnisse zu statistischen

Die Konstruktion psychologischer Tests gilt als eine der bedeutendsten Leistungen der wissenschaftlichen Psychologie.

Psychologische Testkonstruktion als statistisch fundierte Sozialtechnologie

Grundlagen wie z.B. Skalenniveaus oder Korrelationskoeffizienten, die in den Modulen 1 und 2 des Studiengangs vermittelt wurden. Wir haben uns bemüht, hier vor allem das grundsätzliche Verständnis des Vorgehens bei der Testkonstruktion in den Vordergrund zu rücken und mathematisch-technische Details wie z.B. Formeln nur insoweit darzustellen, wie es die Nachvollziehbarkeit des Gesagten erfordert. Auf eine vertiefende Zusatzlektüre kann jedoch in vielen Fällen nicht verzichtet werden. Am Ende der einzelnen Abschnitte des vorliegenden Textes werden dazu jeweils Literaturhinweise gegeben.

Der Studienbrief beruht in geringen Teilen auf dem Lehrbuch „Einführung in die Test- und Fragebogenkonstruktion“ von Markus Bühner in der Auflage von 2006 (aktuell Bühner, 2011). Einige kurze Abschnitte wurden dabei in gekürzter und überarbeiteter Fassung für den vorliegenden Text übernommen, ohne dies, um der besseren Lesbarkeit willen, im Detail durch Zitate kenntlich zu machen. Das Lehrbuch geht jedoch vielfach inhaltlich erheblich über diesen Studienbrief hinaus und wird dementsprechend auch als vertiefende Lektüre empfohlen. Weitere vertiefende und ergänzende Elemente können, wie Sie es aus anderen Modulen gewohnt sind, in Form von Online-Vorlesungen, Texten und Aufgaben in der Lernumgebung Moodle abgerufen und bearbeitet werden.

Struktur des Studienbriefs

Dieser Studienbrief unterscheidet sich außerdem von vielen Lehrtexten zur Testkonstruktion dadurch, dass der Aufbau sich weniger an den statistischen Grundlagen orientiert als an dem idealtypischen Ablauf einer Testkonstruktion. Nach einer generellen Einführung in Kapitel 1 werden in Kapitel 2 zunächst Möglichkeiten besprochen, nach welchen grundsätzlichen Konzepten Tests und ihre einzelnen Elemente (Skalen, Aufgaben bzw. Items) generiert werden können. Das dritte Kapitel beschäftigt sich mit empirischen Informationen, die genutzt werden können, um die in einem so entstandenen Rohentwurf enthaltenen einzelnen Aufgaben oder Fragen zu beurteilen. Dazu zählen insbesondere Verfahren der Itemanalyse nach der Klassischen und Probabilistischen Testtheorie. Im folgenden vierten Kapitel geht es um die Zusammenfassung einzelner Aufgaben zu Skalen mittels verschiedener Varianten der Faktorenanalyse. Im fünften und letzten Hauptteil geht es um Anforderungen an eine anwendungsbezogene Dokumentation und Evaluation (psychometrische und anwendungsbezogene Gütekriterien) von Testverfahren. Dabei ist es gar nicht so erheblich, ob diese Kenntnisse genutzt werden, um ein eigenes Testverfahren zu entwickeln (was eher die Ausnahme sein dürfte) oder ob vorhandene Testverfahren für eine bestimmte Fragestellung bewertet und daraus die bestmögliche Auswahl getroffen werden soll. Auch für eine qualifizierte Bewertung ist es erforderlich, mit den Prinzipien der Testkonstruktion vertraut zu sein.

Im Einzelnen sollen in diesem Kurs die folgenden Ziele und Lernergebnisse erreicht werden. Sie sollen nach erfolgreicher Absolvierung des Kurses...

Ziele und Lernergebnisse

- das Spektrum psychologischer Testverfahren kennen und deren Relation zur zugrunde liegenden Testtheorie einschätzen können. Um übrigens ein mögliches Missverständnis zu vermeiden: Die Begriffe „Test(verfahren)“, „Testkonstruktion“, „Testtheorie“ etc. beziehen sich hier stets auf Tests im Sinne bestimmter psychologischer Messverfahren, nicht auf statistische Signifikanztests (vgl. Modul 2) und deren Theorie.
- mit den grundsätzlichen Möglichkeiten, psychologische Konstrukte mittels Tests zu operationalisieren und die Testitems in einer dem Anwendungszweck und den Teilnehmern angemessenen Weise zu formulieren vertraut sein.
- die wichtigsten Itemkennwerte berechnen und angemessen interpretieren können und in der Lage sein, diese Informationen für die Konstruktion der Endform von Tests und die Beurteilung vorhandener Testverfahren zu nutzen sowie dabei...
- mit den Voraussetzungen und Grenzen der nach der klassischen Testtheorie berechneten Kennwerte vertraut sein und die Möglichkeiten der probabilistischen Testtheorie einschätzen können.
- die wichtigsten Verfahren der exploratorischen Faktorenanalyse kennen und hinsichtlich ihres Stellenwerts für die Skalenbildung beurteilen sowie deren Nutzen vor dem Hintergrund eines Grundverständnisses der konfirmatorischen Faktorenanalyse einschätzen können.
- die Einflussfaktoren und die Möglichkeiten zur Bestimmung der Reliabilität und zur Validierung von Tests kennen und diese Gütekriterien in ihrer Bedeutung einschätzen können. Darüber hinaus besitzen Sie ein Verständnis für die Bedeutung anwendungsbezogener Kriterien und können diese in die Beurteilung der Güte von Testverfahren einbeziehen.
- in der Lage sein, anhand der in Testmanualen und aus anderen Quellen zugänglichen Informationen eine dem Anwendungszweck angemessene Auswahl aus vorhandenen Tests zu treffen.

Wir wünschen Ihnen bei der Erarbeitung des Kursmaterials viel Erfolg und viel Freude.

Bernd Marcus und Markus Bühner

1 Grundlagen psychologischer Testverfahren

Warum ist das eingangs der Einführung in diesen Kurs erwähnte Verfahren „Was Ihre Frisur über Sie verrät“ eigentlich kein psychologischer Test? Hier erhalten Teilnehmerinnen die Möglichkeit, auf standardisierte Vorlagen – die Bilder der Frisuren sind immer gleich – zu reagieren, indem sie eine davon auswählen. Darauf erfolgt eine genau vorgegebene Interpretation, durch die den einzelnen Frisuren Aussagen über psychologische Merkmale der Teilnehmerinnen, nämlich ihre Persönlichkeit, zugeordnet werden. All dies erfüllt tatsächlich einige Bestimmungsstücke der Definition psychologischer Testverfahren. Anderes dagegen, insbesondere wesentliche Kriterien psychometrischer Güte, ist auf so katastrophale Weise verletzt, dass von einem psychologischen Test keine Rede sein kann. Am Ende dieses Kapitels sollten Sie in der Lage sein, diese Aussage konkret und detailliert zu belegen. Wir können nur hoffen, dass derartige „Tests“ – ähnlich wie Horoskope, mit denen sie einiges gemeinsam haben – lediglich als Unterhaltungslektüre dienen und keinesfalls als Grundlage ernsthafter Urteile über die Persönlichkeit.

Den obigen Aussagen über das Verfahren „Was Ihre Frisur über Sie verrät“ dürfte kaum ein Fachkollege widersprechen, dennoch wird die Bezeichnung Test auch in der Fachliteratur keineswegs einheitlich gehandhabt. Nach einem engeren Verständnis wird der Begriff auf solche Verfahren beschränkt, deren Ergebnis frei von jeder subjektiven Beurteilung seitens der Testteilnehmer entsteht, also z.B. Fähigkeits- und Leistungstests, zu deren bekanntesten Vertretern sog. Intelligenztests gehören. Ein solches Verständnis schließt bspw. Persönlichkeitsfragebogen aus, in denen subjektive Selbstbeurteilungen erhoben werden. Wir folgen hier nicht der Unterscheidung in (Leistungs- oder objektive) Test- und (subjektive) Fragebogenverfahren, sondern verwenden ein breiteres Konzept, nach dem diese Unterscheidung lediglich zur Klassifikation von Testverfahren herangezogen wird (s.u.). Allgemein verstehen wir mit Lienert und Raatz (1998, S.1) unter einem psychologischen Test:

Definition psychologischer Test

„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“.

Wissenschaftlichkeit

In dieser Definition stecken verschiedene Bestimmungsstücke. Wissenschaftlich bedeutet in diesem Zusammenhang, dass Tests relativ eindeutig definierte und nachweisbar relevante Merkmale erfassen sollen, ihre Konstruktion auf statistischen Modellen psychologischer Testtheorien aufbaut und sie sich gemäß psychometrischen Gütekriterien empirisch bewähren müssen (dazu mehr in späteren Kapiteln). Außerdem ist ein wichtiger Indikator der wissenschaftlichen Fundierung, dass Entwicklung, empirische Bewährung und Anwendung des Verfahrens in einer für Fachleute im Detail nachvollziehbaren Form dokumentiert

werden. Bei kommerziell vermarkteten Tests, die das Attribut ‚wissenschaftlich‘ verdienen, erfolgt dies regelmäßig in Form sog. Testmanuale oder Handanweisungen. Es gibt aber auch zahlreiche frei zugängliche Tests, die nach durchaus wissenschaftlichen Kriterien entwickelt wurden und für manche Zwecke sinnvoll einsetzbar sind, bei denen die wissenschaftliche Fundierung aber häufig nicht in Form eines handlichen Manuals, sondern auf oft mehrere Zeitschriftenartikel und technische Reports verteilt dokumentiert ist. Das Fehlen jeglicher Dokumentation ist aber auf jeden Fall ein sicherer und für die Anwendungspraxis sehr einfach handhabbarer Indikator mangelnder wissenschaftlicher Fundierung.

Merksatz: Der einfachste Weg, wissenschaftliche von unwissenschaftlichen Testverfahren zu unterscheiden, ist, nach der Existenz eines Testmanuels oder einer äquivalenten Dokumentation zu fragen. Deren Nichtvorliegen ist ein zuverlässiges Ausschlusskriterium; allerdings garantiert das bloße Vorliegen noch nicht die Wissenschaftlichkeit des Verfahrens und schon gar nicht die Eignung für einen bestimmten Zweck.

Routineverfahren sind Tests deshalb, weil Inhalt, Verfahrensablauf und Bewertung der Aussagen genau vorgegeben (standardisiert) und damit routinemäßig wiederholbar sind. Der Begriff Persönlichkeitsmerkmal ist hier weit zu verstehen und schließt im Grunde alle Arten interindividueller Differenzen ein. Testergebnis ist eine quantitative Aussage über das Merkmal, das heißt die Reaktionen (Antworten) auf die einzelnen Items, aus denen sich der Test zusammensetzt (i.d.R. sind dies zahlreiche zu lösende Aufgaben bzw. zu beantwortende Fragen), werden mit Zahlen verknüpft und am Ende nach einer festgelegten Regel zu einem Testwert (Score) zusammengefasst. Dieser Rohwert ist für sich betrachtet noch nicht aussagekräftig, erlaubt aber dann eine relative Aussage über die individuelle Merkmalsausprägung, wenn er mit einer Vergleichsgruppe oder einem Zielkriterium in Beziehung gesetzt wird. Die Begriffe ‚quantitativ‘ und ‚relativer Grad‘ sind dabei insofern vielleicht etwas missverständlich als mit einem Testwert durchaus auch die Zuordnung zu einem qualitativen, kategorial skalierten Merkmal, z.B. einem Persönlichkeitstyp, verknüpft sein kann (vgl. Rost, 2004). Der Begriff ‚qualitativ‘ bezieht sich in diesem Fall auf die Skalierung des zu messenden Merkmals. Qualitative Erhebungsmethoden, z.B. un- oder teilstrukturierte Interviews, sind dagegen keine psychometrischen Tests, weil ihnen u.a. das Merkmal der Standardisierung fehlt.

weitere Bestimmungsstücke der Definition

Obwohl sich die Ursprünge psychologischen Testens bis in die Antike zurückverfolgen lassen – hier ist insbesondere das Auswahlverfahren für Beamte im kaiserlichen China zu nennen – spielten für die Entwicklung der modernen Testtheorie besonders die Fortschritte beim Bemühen um die Messung der menschlichen Intelligenz im 19. und frühen 20. Jahrhundert eine zentrale Rolle. Meilensteine waren hier u.a. Arbeiten von Sir Francis Galton (1822-1911), Alfred Binet (1857-1911) oder die Entwicklung des Army Alpha Tests im 1. Weltkrieg (vgl. für einen kurzen Überblick Amelang, Bartussek, Stemmler & Hagemann, 2006, sowie die



Alfred Binet (1859 - 1911), Pionier psychologischen Testens

Module 1 und 7). Aus diesen Wurzeln in der Intelligenzmessung erklärt sich auch die noch heute gültige Terminologie, die z.B. auch dann von „Aufgaben“ und deren „Schwierigkeit“ spricht, wenn es etwa bei Persönlichkeitstests keine objektiv richtigen oder falschen „Lösungen“, sondern lediglich unterschiedliche Antwortmöglichkeiten gibt.

1.1 Anwendungsbereiche und Klassifikation von Testverfahren

Psychologische Tests werden für zahlreiche Zwecke und an zahlreichen unterschiedlichen Institutionen eingesetzt. Einen unvollständigen Überblick gibt die folgende Aufzählung:

Tabelle 1.1: Anwendungsgebiete psychologischer Tests

Kliniken

- Psychische Störungen
- Psychosomatische Störungen
- Somatische Störungen
- Auf somatische Ursachen zurückgehende psychologische Störungen (z.B. Hirnschäden)
- Belastung und Schmerz
- Krankheitsbewältigung
- Therapieverlaufs- und Erfolgskontrolle

Beratungsstellen

- Allgemeine psychosoziale Beratung
- Familienberatung
- Eheberatung
- Erziehungsberatung
- Suchtberatung

Staatliche Verwaltung

- Berufsberatung
- Auslese
- Jugendhilfe
- Verkehrseignung (TÜV)

Forensischer Bereich

- Strafvollzug (Haftentlassung)
- Im Rahmen von Gerichtsverfahren (zivil- und strafrechtliche Verantwortlichkeit, Glaubwürdigkeit von Zeugen, Sorgerecht)

Betriebe/Personalverwaltung

- Eignung/Auslese
- Arbeitsplatzgestaltung/Ergonomie
- Arbeitsanalyse
- Prozesse innerhalb und zwischen Arbeitsgruppen

Schulen, pädagogischer Bereich

- Entwicklung/Entwicklungsstörungen
- Lernprozesse
- Schulreife
- Sonderschulbedürftigkeit/geistige Behinderung
- Spezifische Lernschwierigkeiten
- Leistungsstörungen
- Hochbegabung
- Schulleistung
- Hochschuleignung
- Intelligenzdiagnostik allgemein

Militär

- Eignungsprüfungen

Marktforschung/Werbepsychologie

- Produktbeurteilung
- Werbung
- Einstellungsmessung

Forschung

- Einstellungen/Interessen
- Eigenschaften
- Momentane Zustände
- Verlaufsprozesse

Viele Testverfahren werden eigens für genau spezifizierte Anwendungsbereiche konstruiert, etwa für die Diagnostik bestimmter klinischer Störungen oder für die Arbeitsanalyse in bestimmten Berufsgruppen. Andere Verfahren, z.B. aus der Intelligenzdiagnostik, können dagegen in vielen verschiedenen Gebieten zum Einsatz kommen. Für den praktischen Einsatz ist es von zentraler Bedeutung, dass die Anwendung in einem spezifischen Gebiet theoretisch und empirisch begründet werden kann. Ein und dasselbe Verfahren kann für einen Zweck sehr nützlich und für einen anderen völlig ungeeignet sein. Diese Aussage mag trivial erscheinen, ist aber notwendig, wenn man bedenkt, dass z.B. klinische Tests zur Diagnostik von Persönlichkeitsstörungen manchmal bedenkenlos für die Personalauswahl eingesetzt werden. Die Güte eines Tests kann daher nicht unabhängig vom Ein-

Tests sollten nur hinsichtlich spezifischer Anwendungen beurteilt werden.

satzzweck beurteilt werden (etwa mit der Aussage: „Dieser Test ist valide.“). Mehr oder weniger gut sind nicht die Tests selbst, sondern deren Eignung als Grundlage für spezifische diagnostische Aussagen.

Klassifikationsmöglichkeiten von Tests

Mit den Anwendungsbereichen ist bereits ein mögliches Kriterium zur Klassifikation von Tests angesprochen. Da, wie erwähnt, viele Tests in mehreren Gebieten eingesetzt werden, ist damit nicht immer eine eindeutige Zuordnung möglich. Gebräuchlich sind daneben auch Einteilungen nach dem *Inhalt*, den *Testungsstrategien* sowie nach *formalen Gesichtspunkten* (vgl. z.B. Amelang & Schmidt-Atzert, 2006; Brähler, Holling, Leutner & Petermann, 2002; Schütz & Marcus, 2005).

Nach dem *Inhalt* unterscheiden z.B. Brähler et al. (2002) folgende Kategorien:

Inhalt: Leistungstests

- *Leistungstests* (Entwicklungstests, Intelligenztests, Allgemeine Leistungstests, Schultests, spezielle Funktionsprüfungs- und Eignungstests) geben Aufschluss über Fähigkeiten und Fertigkeiten eines Individuums. Sie bestehen in der Regel aus einer Vielzahl von Problemlöseaufgaben, deren Beantwortung als richtig oder als falsch eingestuft wird.

Inhalt: Psychometrische Persönlichkeitstests

- *Psychometrische Persönlichkeitstests* (Persönlichkeits-Struktur-Tests, Einstellungs- und Interessentests, Klinische Tests) sind meistens Fragebogen, deren Fragen sich auf Gefühle, Vorlieben, Abneigungen, Interessen und Meinungen der Probanden beziehen. Die Probandin soll dabei eine subjektive Selbstbeurteilung oder Selbstbeschreibung vornehmen; daher gibt es keine objektiv richtigen oder falschen Lösungen bzw. Antworten. Wenn Fragebogen mehrere Merkmale erfassen, also aus verschiedenen Subtests bestehen, bezeichnet man sie auch als *Inventare*.

Inhalt: Persönlichkeits-Entfaltungsverfahren

- *Persönlichkeits-Entfaltungsverfahren* (Formdeuteverfahren, Verbalthematische Verfahren, Zeichnerische und Gestaltungsverfahren) sind sog. projektive Tests. Sie bestehen aus weitgehend unstrukturiertem Material (z. B. Farbkleckse, Satzanfänge, Spielmaterial) oder mehrdeutigen Bildern, die vom Probanden vollendet bzw. gedeutet werden müssen. Dabei wird davon ausgegangen, dass die Testpersonen bei der Strukturierung des Materials ihre unbewussten Motive, Gefühle oder Einstellungen in die Geschichte oder das Bild projizieren. Allerdings sind solche Verfahren häufig nicht nach testtheoretischen Prinzipien konstruiert und überprüft; die Auswertung erfolgt nach Regeln mit weiten subjektiven Interpretationsspielräumen, so dass sie streng genommen nicht den Tests zuzurechnen sind. Anders zu beurteilen sind *semi-projektive Tests*, bei denen das Prinzip der Projektion mit dem der Standardisierung verbunden wird.

Nach der *Testungsstrategie* lässt sich zwischen normorientierten Konstrukttests und kriterienorientierten Tests unterscheiden. *Konstrukttests* erlauben einen wissenschaftlich begründbaren Rückschluss vom Testverhalten auf dahinter liegende Eigenschaften und Fähigkeiten (Konstrukte). Diese Tests sind normorientiert, da das Testergebnis einer Person relativ zur durchschnittlichen Ausprägung (= Norm) einer Bezugsgruppe gesehen wird. Bei den aus dem pädagogischen Bereich stammenden *kriterienorientierten Tests* (lehrzielorientierte Leistungsmessung) wird dagegen die individuelle Testleistung nicht mit einer Bezugsgruppe verglichen, sondern mit einer Idealnorm, dem Lehrziel. Diese Tests gestatten keine Aussagen über die Fähigkeiten und Eigenschaften eines Schülers, sondern informieren darüber, ob das Unterrichtsziel erreicht wurde. Als psychometrisch konstruierte Verfahren sind kriterienorientierte Tests erheblich seltener als Konstrukttests, das kriterienorientierte Prinzip findet sich aber im Grunde auch in einer gewöhnlichen Klausur realisiert. So lassen sich etwa nach dem Mehrfachwahlprinzip (multiple choice) konstruierte Klausuren ohne Schwierigkeiten mit testtheoretischen Mitteln auswerten und beurteilen.

Testungsstrategie: Konstrukt- vs. kriterienorientierte Tests

Daneben lassen sich Tests auch nach verschiedenen *formalen* Gesichtspunkten ordnen. Nach der Durchführungszeit kann man zeitgebundene *Schnelligkeitstests* (Speedtests) von Tests ohne Zeitbeschränkung, sog. *Niveautests* (Powertests), unterscheiden. Diese Unterscheidung ist für die Konstruktion von Leistungstests besonders bedeutsam. In beiden Fällen steigt in der Regel der Schwierigkeitsgrad von Item zu Item an. Während jedoch bei zeitbegrenzten Speedtests alle Items bei unbegrenzter Lösungszeit für die meisten Probanden lösbar wären, erreichen Aufgaben in Powertests ein Niveau, an dem eine im Verlauf zunehmende Anzahl von Teilnehmern scheitert. Für die Konstruktion von Leistungs- wie Persönlichkeitstests außerordentlich wichtig ist die Unterscheidung nach der Anzahl der zu erfassenden Merkmale zwischen *ein-* und *mehrdimensionalen* Tests. Die Durchführung eines Tests kann des Weiteren entweder in einer Gruppe oder nur mit Einzelpersonen erfolgen (*Gruppen-* bzw. *Individualtests*). Nach der Sprachgebundenheit der Items lässt sich zwischen *verbalen* und *nicht-verbalen* Tests unterscheiden. In Abhängigkeit vom Medium, in dem sie durchgeführt werden, ergibt sich ferner durch die Einführung neuer Technologien die Unterscheidung in *Papier- und Bleistift-, PC-gestützte* und *Online-Tests*. Die Einführung der Computertechnologie hat dabei u.a. die Entwicklung sog. *adaptiver* Tests erst ermöglicht, bei denen die Schwierigkeit der Aufgaben dynamisch dem individuellen Leistungsniveau angepasst wird, was eine erheblich kürzere und weniger belastende Durchführung zulässt. Jeder Test kann natürlich mehreren der vorgenannten Kategorien zugeordnet sein.

Formale Kriterien:
Speed vs. power, ein- vs. mehrdimensional, Gruppe vs. Einzel, mediale Unterstützung