

## **Glossar zum Lehrbuch „Empirische Methoden in der Psychologie“ von Markus Pospeschill**

**a-posteriori-Aussage (*posterior assertion*):** eine Aussage, die zwar auf empirischen Beobachtungen beruht, aber erst nach Eintritt eines Ereignisses angestellt wird.

**A-B-Design (*AB design*):** ein Phasendesign mit einer  
→ Ausgangslage bzw. Baseline-Phase (A) und einer Interventions- bzw. Treatmentphase (B).

**A-B-A-Design (*ABA design*):** ein Phasendesign mit einer  
→ Ausgangslage bzw. Baseline-Phase (A), einer Interventions- bzw. Treatmentphase (B) und einer erneuten Phase A.

**adaptiver Algorithmus (*adaptive algorithm*):** ein Regelsystem, welches die Itemauswahl zu Beginn und während des Tests regelt sowie Kriterien der Testbeendigung spezifiziert.

**adaptives Testen (*adaptive testing*):** ein spezielles Vorgehen bei der Messung individueller Ausprägungen von Personmerkmalen, bei dem sich die Auswahl der zur Bearbeitung vorgelegten Items an der Leistungsfähigkeit des untersuchten Probanden orientiert.

**Aggregat (*aggregate*):** Untersuchungseinheiten, die anhand gemeinsamer Merkmale zusammengefasst werden.

**Akaike's Information Criterion (*AIC*):** ein Maß für die Anpassungsgüte des geschätzten Modells an die vorliegenden empirischen Daten (Stichprobe) und die Komplexität des Modells. Daraus hervorgegangen sind weitere Maße wie → *BIC* und *CAIC*.

**Akquieszenz (*acquiescent response style*):** gelegentlich auftretende Neigung von Personen, Fragen oder Statements unabhängig von Inhalt und Kontext zuzustimmen (Zustimmungstendenz, Ja-Sager-Tendenz). → soziale Erwünschtheit

**Algorithmus (*algorithm*):** **1)** eine Berechnungsformel; **2)** ein Regelwerk von Schritten zur Lösung eines Problems.

**ALM (*GLM*):** Abkürzung für Allgemeines Lineares Modell (*General Linear Model*). Hierbei handelt es sich um Grundannahmen der

Regressions- und Korrelationsberechnung sowie der Varianzanalyse (ANOVA), um lineare Beziehungen zwischen abhängigen und unabhängigen Variablen untersuchen zu können.

**Alpha (*alpha*): 1)** Kurzbezeichnung für → Cronbachs Alpha; **2)**

Bezeichnung für den → Alpha-Fehler ( $\alpha$ ).

**Alpha-Fehler (*alpha error*):** Wahrscheinlichkeit, bei der statistischen Hypothesenprüfung eine zutreffende → Nullhypothese zu verwerfen. Der Fehler tritt auf, wenn ein Effekt zwischen Variablen behauptet wird, aber faktisch (im Sinne der Nullhypothese) nicht vorliegt. Dieser Fehler 1. Art wird mit Festlegung des → Signifikanzniveaus bei einem statistischen Test kontrolliert. Je kleiner das Signifikanzniveau gewählt wird, umso geringer die Wahrscheinlichkeit, dass ein Ergebnis per Zufall zustande gekommen ist.

**Alternativhypothese (*alternative hypothesis*):** die (eigentlich interessierende) Forschungshypothese, in der der erwartete Effekt behauptet wird. Bei Gültigkeit der Alternativhypothese resultiert ein statistisch signifikantes Ergebnis, das zur Zurückweisung der → Nullhypothese ( $H_0$ ) und Annahme der → Alternativhypothese ( $H_1$ ) führt.

**Alternativmerkmal (*dichotomous variable*):** eine Variable mit zwei (dichotomen) Merkmalsausprägungen.

**Antwortformat, freies (*free format*):** Bei den Aufgaben sind keine Antwortalternativen vorgegeben. Die Antwort wird von der Person selbst formuliert bzw. produziert.

**Arbeitshypothese (*working hypothesis*):** eine noch weiter zu spezifizierende Annahme, die am Anfang eines Forschungsprojektes steht.

**Artefakt (*artifact*):** ein durch eine Untersuchungs- oder Auswertungsmethode produziertes Untersuchungsergebnis, das nur bedingt etwas über den interessierenden Sachverhalt aussagt bzw. in wesentlichen Teilen Produkt der eingesetzten Methoden ist.

**Asymmetrie (*asymmetry*):** eine Abweichung von der Symmetrie (z.B. bei Verteilungen).

**Asymptote (*asymptote*):** ein theoretischer Grenzwert, der niemals erreicht wird.

**Aufgabenanalyse (*item analysis*):** spezifische Phase in der Test- und Skalenkonstruktion, in der die Items (zumeist anhand von Itemstatistiken) ausgewählt werden, die zur Endform des Tests gehören sollen. → Itemanalyse

**Auftragsforschung (*contract research*):** eine von einem Auftraggeber vorgegebene Fragestellung, der i.d.R. auch über die Verwertung der Ergebnisse entscheidet.

**Augenscheinvalidität (*face validity*):** die auch für Laien offensichtliche (logische und sachliche) Gültigkeit eines Verfahrens.

**Außenkriterien (*external criterion*):** eine Variable, mit der die → prognostische Validität eines Tests oder Verfahrens festgestellt wird.

**Ausgangslage (*baseline*):** die Phase vor einer Intervention oder einem → Treatment, die eine Ausgangsmessung liefern soll.

**Ausreißer (*outlier*):** auffällige Messwerte oder Messwertkombinationen, die außerhalb eines definierten Bereiches einer Verteilung liegen und daher einer gesonderten Berücksichtigung bedürfen; bei besonderer Auffälligkeit auch als Extremwerte bezeichnet.

**Ausschöpfungsquote (*response rate*):** zumeist prozentual angegebener Anteil einer Zielgruppe in einer empirischen Untersuchung, die tatsächlich beteiligt ist.

**Auswahlaufgaben (*multiple choice*):** eine spezifische Aufgabenform, aus mehreren vorgegebenen Antwortalternativen die richtige bzw. zutreffende Antwort zu identifizieren.

**Auswahlgrundgesamtheit (*sampling frame*):** der Teil der Grundgesamtheit, der – da bekannt und zugänglich – als Erhebungsgrundgesamtheit verwendet wird.

**Auswertungsobjektivität (*evaluation objectivity*):** ein Test entspricht diesem → Gütekriterium, wenn das Testergebnis unabhängig davon ist, wer den Test auswertet.

**Autokorrelation (*auto-correlation*):** Eine Korrelation von Variablen über Messzeitpunkte hinweg, die eine Interdependenz aufeinander folgender Beobachtungen aufzeigt.

**Axiom (*axiom*):** theoretische Grundannahme, die ohne weitere Überprüfung als gegeben angesehen wird.

**Badness of fit (*badness of fit*):** Maße der Anpassungsgüte, die den Grad der Nicht-Übereinstimmung zwischen einem theoretischen und einem empirischen Modell anzeigen. Hohe Maßzahlen sprechen für eine schlechte Anpassung. ↔ Goodness of fit

**Baseline-Phase (*baseline phase*):** → Ausgangslage

**Basisrate (*base rate*):** Ausgangshäufigkeit eines Verhaltens vor Beginn einer Maßnahme.

**Bayesian information criterion (*BIC*):** Darunter wird ein dem → *AIC* ähnliches Kriterium der Anpassungsgüte des Modells an die Daten verstanden, das im Unterschied zum *AIC* die Nicht-Sparsamkeit von Modellparametern stärker bestraft.

**Bedingung (*condition*):** ein → Treatment oder eine Kombination von Treatments, die einer Versuchsperson durch den Versuchsleiter in einer experimentellen Bedingung zugewiesen wird. In einem faktoriellen Design definiert jede Faktorstufe bzw. Kombination aus Faktorstufen eine Bedingung.

**Befragung (*survey*):** Form der Datenerhebung, z.B. in Form von Fragebogen oder Interview, in standardisierter, teilstandardisierter und unstandardisierter Form.

**Beobachtung (*observation*):** Form der Datenerhebung, z.B. in Form eines Rollenspiels, einer Einzelfall- oder Selbstbeobachtung.

**Beobachtungswert (*score*):** eine Einzelinformation über eine interessierende Variable.

**Beta-Fehler (*beta error*):** Fehlerrisiko bei der statistischen Hypothesenprüfung, eine  $\rightarrow$  Nullhypothese zu akzeptieren, obwohl tatsächlich ein Effekt besteht. Der Fehler tritt auf, wenn die (nicht zutreffende) Nullhypothese nicht zurückgewiesen wird, obwohl die  $\rightarrow$  Alternativhypothese gilt.

**Beurteiler (*rater*):** allg. eine befragte Person oder auch ein Experte, der ein Urteil abgibt.

**Beurteilungsaufgabe (*rating task*):** Aufgaben, bei denen der individuelle Zustimmungs- oder Ablehnungsgrad zu einer vorgelegten Aussage ( $\rightarrow$  Statement) erfasst wird.

**Bias (*bias*):** Einflüsse in einer Untersuchung, die eine systematische Verzerrung der Untersuchungsergebnisse bewirken können. Unterschieden werden Selektionsfehler (*selection bias*) durch nicht zufällige Stichproben, Informationsfehler durch fehlerhafte Angaben oder invalide Indikatoren sowie die  $\rightarrow$  Konfundierung (*confounding bias*) durch ungenügende Kontrollen von Störvariablen. Schließlich fällt unter diesen Begriff auch der Beurteilungsfehler, wenn Verzerrungseffekte bei der Beurteilung auftreten.

**bivariat (*bivariate*):** zwei Variablen betreffend.

**Blindversuch (*blinding*):** ein  $\rightarrow$  Experiment, bei dem der Versuchsperson verborgen bleibt, welches  $\rightarrow$  Treatment (z.B. ein Medikament oder ein  $\rightarrow$  Placebo) verabreicht wird. Im Doppel-Blindversuch bleibt das Treatment zusätzlich auch dem Versuchsleiter verborgen.

**Bootstrap (*resampling*):** Unter der Annahme, dass eine gezogene  $\rightarrow$  Zufallsstichprobe die gleiche Verteilung wie die  $\rightarrow$  Population besitzt, wird eine Approximation an die Population dadurch erreicht, indem zahlreiche Kopien der Zufallsstichprobe erzeugt werden, aus denen die gewünschte Verteilungsstatistik ermittelt wird.

- Compliance (*compliance*):** die Bereitschaft von → Probanden, mit einem Untersuchungsleiter zusammenzuarbeiten bzw. seinen Anweisungen zu folgen.
- Confounder (*confounder*):** eine → Störvariable, die den untersuchten Zusammenhang zwischen Variablen beeinflusst, so dass unterschiedliche Einflüsse vermengt werden.
- Consequential Validity (*consequential validity*):** ein spezielles → Gütekriterium, das sich mit der Frage beschäftigt, ob mit dem Einsatz eines Testverfahrens das damit in der Praxis verfolgte Ziel erreicht wird.
- Corrected-Information-Criterion (*CIC*):** eine Abwandlung des → *AIC*, bei dem der Stichprobenumfang berücksichtigt wird.
- Cronbachs Alpha (*Cronbach's alpha*):** Koeffizient der → internen Konsistenz als Reliabilitätsschätzung; beruht auf dem Verhältnis zwischen der Summe aus Varianzen und Kovarianzen der Items eines Tests und der Varianz der Testwertvariablen. Je höher die Kovarianzen zwischen den Testitems sind, desto höher werden die interne Konsistenz und damit die → Reliabilität.
- Cut-off-Punkt (*cut-off-point*):** der kritische Punkt einer Verteilung, der bei Klassifikationen die Zugehörigkeit zu unterschiedlichen Kategorien angrenzt.
- Daten (*data*):** quantitative (numerische) und qualitative (textliche, bildliche) Informationen, die im Zuge eines Forschungsprozesses erhoben werden. Sie stellen im Kontext der Forschungsfrage den interessierenden Realitätsausschnitt dar.
- Daten, objektive (*objective data*):** **1)** Daten, die anhand genauer Messoperationen erhoben werden; **2)** Daten, die aufgrund ihres Messverfahrens als unverfälschbar gelten (z.B. physiologische Messungen).
- Daten, qualitative (*qualitative data*):** nicht-numerische Informationen (Texte, Bilder), die zur Beschreibung von Ausschnitten der Realität verwendet werden.

- Daten, quantitative (*quantitative data*):** numerische Größen, die zur Beschreibung von Merkmalsausprägungen verwendet werden.
- Datenanalyse (*data analysis*):** der in Abhängigkeit der Daten eingeleitete Prozess, unter Verwendung spezifischer Auswertungsverfahren die Forschungshypothesen zu überprüfen.
- Datenanalyse, explorative (*exploratory data analysis*):** eine Sammlung statistischer Maßzahlen und graphischer Darstellungsverfahren zur Beschreibung und Analyse von spezifischen Charakteristika – Verteilungen, Streuungen, → Ausreißern, → Residuen etc. – in Datensätzen.
- Datenbereinigung (*data cleaning, data screening, debugging*):** vor der eigentlichen Datenauswertung stattfindende Fehlerüberprüfung und → Plausibilitätsanalyse des Datenmaterials.
- Datenerhebung (*data collection*):** der Prozess der Informationsgewinnung (Datengenerierung) unter Verwendung geeigneter Erhebungsmethoden.
- Debriefing (*debriefing*):** beschreibt die Qualitätssicherungsmaßnahme, nach der Testung den Testleiter nach Besonderheiten während der Testung zu befragen.
- Deckeneffekt (*ceiling-effect*):** ein Effekt der entsteht, wenn der Maximalwert wegen einem zu geringeren Schwierigkeitsgrad von Aufgaben von vielen Versuchspersonen erreicht wird. Eine Varianzeinschränkung der Daten und eine geringe Trennschärfe der Items ist die Folge.
- Definition (*definition*):** Begriffsbestimmung durch Benennen zentraler Merkmale und wesentlicher Eigenschaften und ggf. angrenzende Merkmale zu anderen Begriffen.
- Definition, analytische (*analytical definition*):** eine Begriffsbestimmung auf der Basis einer Bedeutungsanalyse, die Semantik und die Verwendung eines Begriffes versucht zu klären.
- Definition, konzeptuelle (*conceptual definition*):** eine präzise Begriffsbestimmung zur Vorbereitung der → Operationalisierung.

- Definition, operationale (*operational definition*):** eine Begriffsdefinition auf der Basis von Anweisungen zur Erfassung im Kontext einer Datenerhebung (z.B. erhebbaren → Indikatoren).
- Delphi-Methode (*delphi-method*):** mehrphasige Expertenbefragung zur Vorhersage voraussichtlicher Entwicklungen.
- DIN 33430:** eine verbindliche Norm von Qualitätsstandards bei standardisierten Tests für die berufsbezogene Eignungsbeurteilung und die einzelnen Ablaufschritte.
- dichotom (*dichotomous*):** ein Merkmal mit nur zwei Ausprägungen. Merkmale mit kontinuierlichen Ausprägungen lassen sich prinzipiell auch dichotomisieren.
- disjunkt (*disjunct*):** Eigenschaft von Antwortalternativen, die vorliegt, wenn zwischen den Antwortalternativen keine Überlappungen bestehen.
- diskret (*discrete*):** Merkmalsausprägungen (z.B. kategoriale Variablen) mit Abstufungen.
- Diskriminationsindex (*index of discrimination*):** Kennwert zur Identifizierung „nicht trennscharfer“ Items bei der → Latent-Class-Analyse.
- Distraktoren (*distractors*):** plausibel erscheinende, aber nicht zutreffende Antwortalternativen bei Auswahlaufgaben.
- Dummy-Kodierung (*dummy coding*):** die Rekodierung von Variablen in ein 0-1-Format, das eine statistische Behandlung wie intervallskalierte Größen erlaubt.
- Durchführungsobjektivität (*implementation objectivity*):** ein Test entspricht diesem → Gütekriterium, wenn das Testergebnis unabhängig davon ist, wer den Test vorgibt.
- EDA (*exploratory data analysis*):** Abkürzung für → Explorative Datenanalyse.
- Effektgröße (*effect size*):** **1**) in der → Meta-Analyse ein Index zur Kennzeichnung eines mittleren Effektes; **2**) in der Poweranalyse ein Index für die Abweichung von einem Parameter.



- Eichstichprobe (*standardization sample*):** eine repräsentative Stichprobe von → Probanden, die zur → Normierung eines Testverfahrens eingesetzt wird.
- Eigenwert (*eigenvalue*):** der Eigenwert eines Faktors gibt an, wie viel Varianz von allen Items (Variablen) durch diesen Faktor erklärt wird. → Faktorenanalyse, exploratorische
- Eindimensionalität (*one-dimensionality*):** liegt dann vor, wenn Items einer gemeinsamen Skala hoch korrelieren und lässt sich faktorenanalytisch nachweisen. → Faktorenanalyse, exploratorische
- Einzelfallstudie (*single-case study*):** → Fallstudie
- empirisch (*empirical*):** ein Weg der Erkenntnisfindung und des Zugangs zur Welt auf der Grundlage von Erfahrungen, Beobachtungen und Experimenten.
- Erhebungseinheit (*unit of analysis*):** → Merkmalsträger einer Untersuchung.
- Erwünschtheit, soziale (*social desirability*):** Tendenz von Versuchspersonen zur Verfälschung von Antworten, insbesondere in Richtung eines norm- oder erwartungsgerechten Verhaltens.
- Evaluation (*evaluation*):** die Einschätzung, Wirksamkeitsprüfung und Bewertung einer Maßnahme (eines Interventionsprogramms) durch systematische Anwendung empirischer Forschungsmethoden.
- Evaluation, formative (*formative evaluation*):** Prüfung und Optimierung eines Forschungsplanes und der entwickelten Interventionsmaßnahme.
- Evaluation, summative (*summative evaluation*):** Prüfung der Wirkung und des Eintretens der gewünschten Veränderungen einer Maßnahme.
- Exhaustivität (*exhaustive*):** Merkmal von Antwortalternativen; liegt vor, wenn alle möglichen Antworten auf den vorgegebenen Antwortalternativen abgebildet werden können.

**Ex-post-facto-Design (*ex-post-facto design*):** Forschungsdesign, bei dem Einflussfaktoren bereits einen Effekt bewirkt haben, wenn die Daten erhoben werden, und damit eine kausale Zuschreibung erschweren, wenn nicht verhindern.

**Experiment (*experiment*):** eine spezifische Untersuchungsform, die die Untersuchung einer → abhängigen Variable durch gezielte Manipulation der → unabhängigen Variablen erlaubt.

**Exploration (*exploration*).** **1)** Erkundungsstudie zur Generierung neuer → Hypothesen; **2)** im klinisch-psychologischen Kontext durchgeführte Befragung von Patienten oder Klienten.

**Extension (*extension*):** Menge von Objekten, die unter eine → Nominaldefinition fallen. ↔ Intension

**Fairness (*fairness*):** ein Test erfüllt dieses → Gütekriterium, wenn die resultierenden Testwerte zu keiner systematischen Benachteiligung bestimmter Personen aufgrund ihrer Zugehörigkeit zu ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppen führen.

**Faktorenanalyse, exploratorische (*exploratory factor analysis*):** ein strukturierendes datenreduzierendes Verfahren, das typischerweise dann zur Anwendung kommt, wenn keine → Hypothesen über die Anzahl der zugrundeliegenden Faktoren und über die Zuordnung der beobachteten Variablen zu den Faktoren vorliegen.

**Faktorenanalyse, konfirmatorische (*confirmatory factor analysis*):** Verfahren zur Datenreduktion, mit dem → Hypothesen über die Anzahl der Faktoren, die Korrelationen zwischen den Faktoren und die Zuordnung der beobachteten Variablen zu den einzelnen Faktoren theoriegeleitet überprüft werden. Bildet das → Messmodell der Verfahrensgruppe der Strukturgleichungsmodelle.

**Faktorladung (*factor loading*):** Die Gewichtungszahl  $\lambda_{jk}$  einer beobachteten Variablen  $j$  auf dem latenten Faktor  $k$  heißt

Faktorladung und beschreibt die Stärke des Zusammenhangs zwischen Faktor und Variable (meist  $\rightarrow$  Item). Sie kann bei orthogonal rotierten Faktoren als Korrelation interpretiert werden.

**Faktorwert (*factor score*):** Der Faktorwert  $f_{ki}$  gibt an, wie stark ein Faktor  $F_k$  bei der  $i$ -ten Person ausgeprägt ist.

**Fallstudie (*case study*):** eine unabhängig von der Erhebungstechnik auf eine Einzelperson ausgerichtete Forschungsstudie, die sich über einen definierten Beobachtungszeitraum erstreckt und zumeist Messwiederholungen einschließt.

**Fehlervarianz (*error variance*):** die Varianz der Fehlerwerte  $Var(\epsilon)$  der Personen stellt in der Klassischen Testtheorie den unerklärten Anteil der Testwertvarianz  $Var(x)$  dar.

**Feldexperiment (*field experiment*):** ein Experiment in der natürlichen Umgebung der Probanden zur Steigerung der  $\rightarrow$  externen Validität, allerdings ggf. durch eine geringere Bedingungskontrolle unter Einbußen bei der  $\rightarrow$  internen Validität.

**Follow-up-Studie (*follow-up study*):** eine Form der Nachuntersuchung, um Daten zu gewinnen, die mit Erhebungen früherer Zeitpunkte in Beziehung gesetzt werden können.

**Forced-Choice-Antwortformat (*forced choice format*):** Es werden nur zwei Antwortalternativen vorgegeben, so dass eine Entscheidung erzwungen wird.

**Forschungsdesign (*research design*):** ein Untersuchungsplan über verschiedene Entwicklungsstadien hinweg, der die Erhebung möglichst reliabler und valider Daten gewährleisten soll.

**Frage, geschlossene (*closed question*):** Frage mit vorgegebenen Antwortalternativen.

**Frage, offene (*free-answer question*):** Frage ohne Antwortvorgaben.

**Funktion, itemcharakteristische (*item characteristic function*):** beschreibt die Beziehung zwischen dem latenten Merkmal und dem Reaktionsverhalten der Probanden auf ein Item in Form einer Wahrscheinlichkeitsaussage (abgekürzt: IC-Funktion).

**Funktion, logistische (*logistic function*):** eine nicht-lineare Funktion mit S-förmigem Verlauf, aus der Aussagen über Wahrscheinlichkeiten abgeleitet werden. Der Verlauf eignet sich besonders gut zur Modellierung von Schwellenwertunter- oder -überschreitungen (z.B. im Kontext der → Item-Response-Theorie).

**Goodness of fit (*goodness of fit*):** Maße der Anpassungsgüte, die den Grad der Übereinstimmung zwischen einem theoretischen und einem empirischen Modell anzeigen. Hohe Maßzahlen sprechen für eine gute Anpassung. ↔ Badness of fit

**Grundgesamtheit (*target population*):** Gesamtheit aller Elemente bzw. Fälle, über die Aussagen generiert bzw. Parameter geschätzt werden sollen (abgekürzt: GG). Gelegentlich wird eine Erhebungsgrundgesamtheit (*frame population*) davon unterschieden, aus der faktisch die Elemente gezogen werden.  
→ Population

**Gütekriterien (*criteria for measurement quality*):** Kriterien zur Beurteilung der wissenschaftlichen Qualität von Messverfahren (dann auch als psychometrische Tests bezeichnet). Zu den Hauptgütekriterien gehören die → Objektivität, die → Reliabilität und die → Validität eines Tests. → Testgütekriterien

**Guttman-Skala (*Guttman scaling*):** Methode zur Konstruktion einer monoton steigenden, eindimensionalen Einstellungsskala.

**Halo-Effekt (*halo effect*):** Beurteilungsfehler (Attribuierungsfehler) durch eine zeitlich gekoppelte oder merkmalsübergreifende Übergeneralisierung von Antworttendenzen, die durch ein (positives oder negatives) Leit- oder Initialmerkmal gesteuert wird.

**Hauptachsenanalyse (*principal axis analysis*):** Methode der → exploratorischen Faktorenanalyse, mit der versucht wird, das Beziehungsmuster zwischen den → manifesten Variablen mit dahinterliegenden Faktoren zu erklären.

**Hauptkomponentenanalyse (*principal components analysis, PCA*):**

Methode der → exploratorischen Faktorenanalyse, mit der versucht wird, möglichst viel Varianz der beobachteten Variablen durch sog. Hauptkomponenten zu erklären.

**Hawthorne-Effekt (*Hawthorne effect*):** eine Verhaltensänderung, bedingt durch das Wissen als Versuchsperson an einer Studie teilzunehmen. Der Begriff nimmt Bezug auf eine Studie in den zwanziger Jahren in den gleichnamigen Hawthorne-Werken, bei der die Anwesenheit der Beobachter im Betrieb zu einem veränderten Verhalten der Arbeiter führte.

**Heteroskedastizität (*heteroscedasticity*):** die unerwünschte Auffälligkeit ungleichmäßiger Streuung der → Residuen um die Regressionsgerade in Teilabschnitten der Verteilung. In der Folge ist die Regressionsgerade nicht in allen Abschnitten optimal angepasst und durch einzelne Kennwerte nicht korrekt zu beschreiben (da die Streuungsverhältnisse in verschiedenen Teilabschnitten der Verteilung unterschiedlich sind).

↔ Homoskedastizität

**Homogenität (*homogeneity*):** liegt vor, wenn verschiedene Items eines (Sub-)Tests dasselbe Merkmal messen und entsprechende Korrelationen aufweisen.

**homomorph (*homomorph*):** nicht umkehrbare Entsprechung der funktionalen Beziehungen in zwei Objektmengen.

**Homoskedastizität (*homoscedasticity*):** die Grundvoraussetzung des linearen Modells einer gleichmäßigen Streuung der → Residuen um die Regressionsgerade in allen Teilabschnitten einer Verteilung. ↔ Heteroskedastizität

**Hypothese (*hypothesis*):** **1**) allg.: eine noch unbestätigte Behauptung über einen Sachverhalt; **2**) in der empirischen Forschung: eine vorläufige Aussage über die Beziehung zwischen zwei oder mehr Variablen, die anhand geeigneter Beobachtungen überprüft werden soll; **3**) in der Statistik: eine statistische Hypothese mit der Unterscheidung zwischen → Null- und → Alternativhypothese.

**Index (*index*):** ein Kennwert für eine komplexe Variable, der auf der kombinierten Messung verschiedener Merkmale beruht und diese in spezifischer Weise zu einem singulären Datum verrechnet.

↔ Skala

**Indikator (*indicator*):** Variablen, die einen Rückschluss auf nicht oder nur sehr aufwendig direkt beobachtbare Merkmale ermöglichen sollen. Die Beziehung zum Merkmal kann dabei enger (definitiv) oder weniger eng (korrelierend oder schlussfolgernd) sein. Komplexe Begriffe erfordern dabei den Einsatz multipler Indikatoren.

**Informationskriterien (*information criteria*):** Maße zur deskriptiven, relativen Beurteilung der Güte eines Modells. Häufig verwendete Informationskriterien sind der  $\rightarrow AIC$ , der  $\rightarrow BIC$  und der  $CAIC$ .

**Intension (*intension*):** Menge von Eigenschaften eines Begriffes, die unter eine  $\rightarrow$  Nominaldefinition fallen. ↔ Extension

**Interpretationsobjektivität (*interpretation objectivity*):** Ein Test entspricht diesem  $\rightarrow$  Gütekriterium, wenn bezüglich der Interpretation der Testwerte eindeutige Richtlinien (z.B. Normentabellen) vorliegen.

**Interventionsstudie (*intervention study*):** eine Untersuchungsform, in der ein gezielter Eingriff dazu dient, den Effekt einer Maßnahme abzuschätzen.

**Inzidenz (*incidence*):** der Neuauftritt eines bestimmten Merkmals (in der Epidemiologie z.B. Neuerkrankungen) in einem räumlich und zeitlich definierten Bezugsrahmen.

**Inzidenzrate (*incidence rate*):** ein standardisiertes Maß aus dem Verhältnis der Anzahl der Neuerkrankungen ( $\rightarrow$  Inzidenz) dividiert durch den Zeitraum und die Individuenzahl der definierten Risikobevölkerung (das entspricht dem absoluten Risiko, also dem Erkrankungsrisiko in einer bestimmten  $\rightarrow$  Population).

**Irrtumswahrscheinlichkeit (*significance level*):** bezeichnet das verbleibende Restrisiko (in einem Prozentwert angegeben), die

→ Nullhypothese zurückzuweisen und die → Alternativhypothese anzunehmen. Dieses ist vor der Datenerhebung festzulegen.

**isomorph (*isomorph*):** ähnlich; umkehrbar eineindeutige

Entsprechung der funktionalen Beziehungen zweier Objektmengen.

**Item (*item*):** eine einzelne Aussage (→ Statement), Frage oder Aufgabe in einem Fragebogen(test), (psychometrischen) Test oder in einer → Skala.

**Items, invertierte (*inverted items*):** „umgepolte“ Items, bei denen nicht die Bejahung, sondern die Verneinung symptomatisch für eine hohe Merkmalsausprägung ist.

**Item-Response-Theorie (*item-response theory*):** auch

→ Probabilistische Testtheorie; beschreibt den Zusammenhang zwischen beobachtbarem Antwortverhalten und dem dahinterstehenden Persönlichkeitsmerkmal auf Grundlage eines wahrscheinlichkeitstheoretischen → Modells. Dabei wird die → Wahrscheinlichkeit für das beobachtbare (gezeigte) Antwortverhalten als von der latenten Merkmalsausprägung abhängig modelliert.

**Itemanalyse (*item analysis*):** Prozess der (statistischen und inhaltlichen) Beurteilung der Güte von → Items aus einem → Itempool, mit dem Ziel, eine Endfassung für eine Skala zu erstellen. Dabei spielen Aspekte der → Trennschärfe, der → Itemschwierigkeit, der → Homogenität und der Dimensionalität eine zentrale Rolle.

**Itemhomogenität (*item homogeneity*):** Verschiedene Items sind bezüglich einer latenten Dimension  $\xi$  dann homogen, wenn das Antwortverhalten auf die Items nur von diesem Merkmal (der latenten Dimension) und keinem anderen systematisch beeinflusst wird. → Homogenität

**Itempool (*item pool*):** eine Menge von → Items, für die mit einem IRT-Modell → Itemhomogenität festgestellt wurde; beim adaptiven

Testen können daraus beliebige Items zur Vorgabe ausgewählt werden.

**Itemschwierigkeit (*item difficulty*):** beschreibt in der Klassischen Testtheorie das mit 100 multiplizierte Verhältnis des durchschnittlich erreichten Itempunktwertes zum maximal möglichen Itempunktwert. Je größer der → Schwierigkeitsindex, desto leichter ist das Item.

**Itemselektion (*item selection*):** beschreibt den Prozess, → Items hinsichtlich ihrer Eignung zur Erfassung des interessierenden Merkmals auszuwählen. Neben der Betrachtung deskriptivstatistisch gewonnener Kennwerte (z.B. → Itemschwierigkeit, → Itemtrennschärfe und → Itemvarianz) fließen auch inhaltliche und modelltheoretische Überlegungen in den Selektionsprozess ein.

**Itemtrennschärfe (*corrected item-total correlation*):** gibt an, wie stark die mit dem jeweiligen Item erzielte Differenzierung zwischen den Probanden mit der Differenzierung durch den Gesamttest übereinstimmt.

**Itemvarianz (*item variance*):** Maß für die Differenzierungsfähigkeit des Items; es gibt an, wie unterschiedlich die Probanden auf das Item antworten.

**Itemzwillinge (*item pairs*):** Methode der Aufteilung eines Tests in Testhälften zur Bestimmung der → Splithalf-Reliabilität. → Items werden in Paare zu zwei Items mit möglichst gleicher Schwierigkeit und Trennschärfe gruppiert. Jeder Testhälfte wird zufällig jeweils eines der beiden Items zugeordnet.

**Ja-Sager-Tendenz (*yeah-saying*):** eine Zustimmungstendenz, auf beliebige Fragen mit „Ja“ zu antworten, wobei logische Widersprüche meistens unentdeckt bleiben.

**John-Henry-Effekt (*John-Henry effect*):** Die → Probanden einer → Kontrollgruppe finden sich aufgrund ihrer Sonderstellung herausgefordert und ändern daraufhin ihr Verhalten, um sich den



Leistungen der Experimentalgruppe anzugleichen. Zur Aufdeckung eines solchen Effektes kann es erforderlich sein, eine Messung der → Ausgangslage vorzunehmen.

**Kategorien (*categories*):** basale Ordnungseinheiten zur Unterscheidung von Fällen oder Objekten anhand eines oder mehrerer Merkmale. Die Kategorienbildung kann theoriegeleitet und merkmalsgeleitet (deskriptiv) erfolgen.

**Kategoriensystem (*categorical system*):** System von Zuordnungsregeln für die Zuordnung von Fällen, Objekten oder Einheiten zu Kategorien. Ätiologische Kategorien stützen ihre Unterscheidung dabei auf Aspekte der Entstehung und Ursachen, während morphologische Kategorien formale, beobachtbare (und damit intersubjektiv besser überprüfbare) Unterschiede heranziehen.

**Klassifikation (*classification*):** die Zuordnung von Beobachtungswerten, Variablenausprägungen oder Fällen zu Klassen oder Kategorien. Die Klassen müssen eindeutig definiert, disjunkt und suffizient sein.

**Klassifikationsvariable (*classification variable*):** Variablen, die wie Nationalität, Geschlecht oder Zugehörigkeit zu einer Partei eine Person charakterisieren und daher durch den Experimentator nicht randomisiert (manipuliert) werden können.

**kognitives Vortesten (*cognitive pretest*):** Bei diesem Vorgehen legt der Testleiter in Erprobung befindliche → Items vor und bittet die → Probanden alle Überlegungen, die zur Beantwortung der Frage führen, zu formulieren. Diese Äußerungen werden meist auf Video aufgenommen.

**Kohorte (*cohort*):** eine Gruppe von Menschen mit spezifischen Erfahrungen oder (demografischen) Merkmalen (z.B. Alterskohorte).

**Kollinearität (*collinearity*):** linearer Zusammenhang mehrerer → **Prädiktoren** bei Aufstellung einer multiplen

Regressionsgleichung. Die Aufnahme einer kollinearen Variablen in eine Regressionsgleichung erbringt nur eine geringe spezifische zusätzliche Varianzaufklärung.

**Kommunalität (*communality*):** Beitrag einer Variablen zu einer bestimmten Faktorlösung. Die Kommunalität einer Variablen  $i$  gibt an, in welchem Ausmaß die wahre Varianz der Variablen durch die extrahierten  $q$  Faktoren erklärt wird. → Faktorenanalyse, exploratorische

**Konfidenzintervall (*confidence interval*):** kennzeichnet denjenigen Bereich um einen empirisch ermittelten individuellen → Testwert  $x_v$ , in dem sich 95% oder 99% aller möglichen wahren Testwerte  $\tau_v$  befinden, die den Testwert  $x_v$  erzeugt haben können.  
→ Vertrauensintervall

**Konfundierung (*confounding*):** Uneindeutigkeit bei der Zurechenbarkeit von Effekten zu erklärenden Variablen.

**Konkordanz (*concordance*):** Maße der Übereinstimmung bzw. Entsprechung von Urteilen.

**Konsistenz (*consistency*):** Die Konsistenz einer Messvariablen beschreibt das Ausmaß der durch ein → Trait erklärten Varianz relativiert an der Gesamtvarianz der Messvariablen.

**Konsistenz, interne (*internal consistency*):** eine Methode der Reliabilitätsschätzung. Dabei werden die Kovarianzen zwischen den Items eines Tests als wahre Varianz angesehen und zur Bestimmung der → Reliabilität verwendet.

**Konsistenzeffekte (*consistency effects*):** treten dann auf, wenn → Probanden versuchen, „stimmige Antworten“ aufgrund ihrer Antworten auf vorangegangene Items zu geben.

**Konstrukt (*construct*):** Bezeichnung für ein nicht direkt beobachtbares, aber empirisch verankerbares latentes Persönlichkeitsmerkmal (auch: hypothetisches Konstrukt).

**Konstruktäquivalenz (*construct equivalence*):** empirisch bestätigte Äquivalenz eines psychologischen → Konstrukts über Sprachen und Kulturen hinweg.

**Konstruktvalidität (*construct validity*):** liegt vor, wenn ein Test tatsächlich das → Konstrukt erfasst, das er erfassen soll.

**Kontingenz (*contingency*):** das gemeinsame Auftreten von Variablen im Verbund als Ausdruck eines Zusammenhangs zwischen ihnen (z.B. dargestellt in einer Kontingenztabelle).

**Kontrollgruppe (*control group*):** Gruppe ohne Behandlung bzw. → Treatment, die Vergleichswerte für die Experimentalgruppe (mit Behandlung bzw. Treatment) liefern soll.

**Kontrollvariable (*control variable*):** eine zusätzlich erhobene Variable, die zwar nicht Bestandteil der eigentlichen Fragestellung ist, aber im Kontext der Fragestellung einen spezifischen Einfluss auf die beobachteten Effekte ausüben kann und daher mit erhoben wird. → Moderatorvariable

**Kovariate (*covariate*):** eine Drittvariable, die den Zusammenhang zwischen anderen Variablen beeinflusst. Ihr (möglicher) Einfluss wird zumeist statistisch kontrolliert. Gelingt dies nicht, resultiert ggf. eine Störvariable.

**Kreuzvalidierung (*cross-validation*):** eine Form der Validitätsprüfung, bei der gewonnene Befunde an einer Stichprobe an einer neuen Stichprobe repliziert werden.

**Kriterium (*criterion*):** ein definierter (empirisch valide messbarer) Sachverhalt, an dem sich überprüfen lässt, ob ein erwarteter bzw. vorhergesagter Effekt eingetreten ist oder nicht.

**Kriteriumsgruppe (*criterion group*):** Zielgruppe, mit der ein Testverfahren validiert wird.

**kriteriumsorientierte Testwertinterpretation (*criterion based interpretation of test scores*):** Bei dieser Form der Testwertinterpretation erfolgt die Interpretation des → Testwertes nicht in Bezug zur Testwerteverteilung einer Bezugsgruppe, sondern in Bezug auf ein spezifisches inhaltliches Kriterium. Es wird vorab festgelegt, welches Testverhalten zum Erreichen des Kriteriums führt.

**Kriteriumsvalidität (*criterion validity*):** liegt vor, wenn von einem Testergebnis auf ein für diagnostische Entscheidungen praktisch relevantes Kriterium außerhalb der Testsituation geschlossen werden kann. Kriteriumsvalidität kann durch empirische Zusammenhänge zwischen dem  $\rightarrow$  Testwert und möglichen Außenkriterien ( $\rightarrow$  externe Validität) belegt werden.

**Kriteriumsvariable (*criterion variable*):** Alternative Bezeichnung für  $\rightarrow$  abhängige Variable im Kontext der Regressionsanalyse (auch: Prognosevariable oder Regressand). Es handelt sich dabei um die Variable, die vorhergesagt bzw. erklärt werden soll.

**kurvilinear (*curvilinear*):** Sammelbezeichnung für alle Formen nicht linearer Zusammenhänge zwischen Variablen.

**Ladung (*factor loading*):** die Korrelation einer Variablen mit einem Faktor bei der  $\rightarrow$  Faktoren- und  $\rightarrow$  Hauptkomponentenanalyse. Je höher die (Faktor-)Ladung, desto eher trifft ein Item die faktorielle Dimension.

**Längsschnittstudie (*longitudinal study*):** Sammelbegriff für Untersuchungsdesigns mit wiederholten Messzeitpunkten, zumeist an denselben  $\rightarrow$  Probanden.

**Lateinisches-Quadrat-Design (*latin square design*):** ein experimenteller Versuchsplan.

**Latent-Class-Analyse, LCA (*latent class analysis*):** probabilistisches Verfahren zur Kategorisierung von Personen (Objekten) in qualitative latente Klassen.

**Latent State-Trait-Theorie (*latent state-trait theory*):** formale Erweiterung der Klassischen Testtheorie, die neben der Aufteilung der Messvariable  $X_{it}$  einer Messung  $i$  zu Messgelegenheit  $t$  in eine Messfehlervariable  $\epsilon_{it}$  und in eine Variable der wahren Werte  $\tau_{it}$  auch eine Trennung von situationalen und dispositionellen Einflüssen erlaubt. Dazu wird die Variable der wahren Werte  $\tau_{it}$  einer Messung  $X_{it}$  zusätzlich in eine Trait-Variable  $\xi_{it}$  und in eine State-Residuums-Variable  $\zeta_{it}$  zerlegt:  $X_{it} = \tau_{it} + \epsilon_{it} = \xi_{it} + \zeta_{it} + \epsilon_{it}$

**Latent-Class-Modelle (*latent class models*):** Bezeichnung für → IRT-Modelle, die davon ausgehen, dass das latente Persönlichkeitsmerkmal zur Charakterisierung von Personenunterschieden aus qualitativen kategorialen latenten Klassen besteht.

**Latent-Trait-Modelle (*latent trait models*):** Bezeichnung für → IRT-Modelle, die davon ausgehen, dass es sich bei dem latenten Persönlichkeitsmerkmal zur Charakterisierung von Personenunterschieden um eine quantitative kontinuierliche → latente Dimension handelt.

**latente Dimension (*latent dimension*):** nicht direkt beobachtbare Variable zur Erfassung von Merkmalsausprägungen in Leistungs-, Einstellungs- oder Persönlichkeitsmerkmalen, von welchen das manifeste Verhalten als abhängig angesehen wird.

**latentes State-Residuum (*latent state-residuum*):** Teil eines → States, der ausschließlich die Situation und die Interaktion zwischen Person und Situation repräsentiert.

**Leistungstests (*achievement tests*):** Tests zur Erfassung der individuellen kognitiven Leistungsfähigkeit in Problemlösesituationen (z.B. Intelligenztests, Konzentrationsleistungstests). → Speed-Tests

**Likelihood (*likelihood*):** In der → Latent-Class-Analyse (LCA) ist die Likelihood das Anpassungskriterium bei der Parameterschätzung. Es ist dort definiert als das Produkt der unbedingten Antwortmusterwahrscheinlichkeiten  $P(a_v)$  über alle Antwortmuster in der Stichprobe ( $N_a$ ) hinweg.

**Likelihood-Ratio-Test (*likelihood ratio test*):** Möglichkeit zur inferenzstatistischen Absicherung der Güte eines LCA-Modells: Der Test wird zur inferenzstatistischen Absicherung des Unterschieds zweier → geschachtelter Modelle (*nested models*) verwendet. → Latent-Class-Analyse

**Likert-Skala (*Likert scale*):** ein spezifischer Itemsatz mit einer einheitlichen 5-stufigen Antwortskala, die als Summenskala verrechnet wird. → Ratingskala

**Lizenzprüfung nach DIN 33430:** Nachweis einschlägiger Kenntnisse von Auftragnehmern (Lizenz A), bzw. Mitwirkenden an Verhaltensbeobachtungen (Lizenz MV) und von Mitwirkenden an Eignungsinterviews (Lizenz ME) gemäß DIN 33430.

**lokale stochastische Unabhängigkeit (*local stochastic independence*):** Bedingung, die erfüllt sein muss, um die Korrelation zwischen zwei Testitems auf eine dahinterliegende latente Persönlichkeitsvariable zurückführen zu können. Die lokale stochastische Unabhängigkeit liegt dann vor, wenn die Korrelation zwischen den → Items verschwindet, wenn man sie auf den einzelnen (lokalen) Stufen des latenten Persönlichkeitsmerkmals untersucht.

**Matching (*matching*):** Verfahren der Zuordnung von Personen anhand eines oder mehrerer Merkmale zu einer → Experimental- und → Kontrollgruppe, sodass Untersuchungszwillinge entstehen. Die resultierenden Stichproben sind abhängig.

**Mediatorvariable (*mediator variable*):** eine intervenierende Drittvariable, die zwischen einer Beziehung von → Prädiktor und → Kriterium wirkt. Entsprechende Zusammenhänge werden nicht als unmittelbar (direkt) angenommen, sondern durch die Mediatorvariable vermittelt (mediert). Die Prädiktorvariable sollte die Mediatorvariable direkt und das Kriterium indirekt beeinflussen.

**Mehrdimensionalität (*multi-dimensionality*):** die Aufspaltung komplexer → Konstrukte in verschiedene (abhängige oder unabhängige) Aspekte, Facetten oder Dimensionen. Die gezielte Zerlegung eines komplexen Konstruktes soll deren Messbarkeit erleichtern.

**Merkmalsausprägung (*variable attribution*):** die im Rahmen eines Maßsystems und auf der Basis eines Messinstrumentes festgestellte individuelle (Werte-)Ausprägung einer Variablen oder Variablendimension.

**Merkmalsraum (*property space*):** der Bereich theoretisch möglicher Merkmalsausprägungen (Minimal- und Maximalwert).

**Merkmalsträger (*unit*):** Untersuchungseinheit oder Fall. → Vp

**Messeffizienz (*measurement efficiency*):** Parameter eines Tests, der sich durch den Quotienten aus Messpräzision und Testlänge berechnet, wobei letztere häufig durch die Anzahl der präsentierten → Items quantifiziert wird.

**Messmodell (*measurement model*):** Im Rahmen von Strukturgleichungsmodellen sind dies Teilmodelle, in denen die Zuordnungen der beobachteten Variablen zu den Faktoren erfolgt.

**Messpräzision (*precision of measurement*):** Grad der Übereinstimmung von wahren Merkmalsausprägungen und den Testwerten. Auf Skalenebene oft durch die mittlere quadratische Abweichung von wahrer und geschätzter Merkmalsausprägung bestimmt.

**Meta-Analyse (*meta-analysis*):** statistische Verfahren zur formalen Bewertung von → Effektgrößen einschlägiger Einzelstudien. Nach Bereinigung verschiedener Unterschiede in den berücksichtigten Studien wird die mittlere, wahre → Validität geschätzt.

**Methode (*method*):** das planmäßige Vorgehen in der Wissenschaft, um zu gesicherten Erkenntnissen zu gelangen. Spezifische Methoden sind dabei an bestimmte wissenschaftstheoretische Positionen verbunden. Die geordnete Abfolge von Einzelschritten umfasst Erhebungs-, Auswertungs- und Interpretationsmethoden. Insbesondere wissenschaftliche Methoden sollten dabei kommunizierbar, lehrbar und intersubjektiv überprüfbar sein.

**Methode der kleinsten Quadrate (*least squared method*):** ein bei der Regression verwendetes Verfahren, um bei der Aufstellung der Regressionsgleichung den Vorhersagefehler, die Summe der

Abweichungsquadrate zwischen beobachtetem und über die Regressionsgerade geschätztem Wert der abhängigen Variable, zu minimieren.

**Methodeneffekte (*method effects*):** ein Sammelbegriff für verschiedene systematische Varianzquellen bei der → Multitrait-Multimethod-Analyse, die sich über den → Trait hinausgehend auf die → Validität der Messung auswirken können. Hierbei kann es sich u.a. um Charakteristika der eingesetzten Messinstrumente, der Beurteiler oder der Situationen, in der eine Messung erfolgt, handeln.

**Methodik (*methodology*):** Allg. wird darunter die Lehre von den → Methoden verstanden. Einer Methodik zu folgen bedeutet, bestimmte Methoden anzuwenden.

**Methodologie (*methodology*):** das Nachdenken über → Methoden, die allgemein oder fachspezifisch erfolgen kann. Methodologien klassifizieren und bewerten Methoden.

**Modell (*model*):** **1)** Vorbild; **2)** verkleinertes Abbild; **3)** die Repräsentation eines umschriebenen Realitätsausschnittes einschließlich bestimmter Annahmen und Vorhersagen über die Funktionsmechanismen (die sich empirisch überprüfen lassen).

**Modell, statistisches (*statistical model*):** eine in Gleichungen und Funktionen definierte Repräsentation eines Realitätsausschnittes, einschließlich bestimmter Hypothesen über die Funktionsmechanismen.

**Modelle, deterministische (*deterministic models*):** Diese Modelle nehmen an, dass die → Wahrscheinlichkeit, ein → Item zu lösen, nur 0 oder 1 betragen kann; die Wahrscheinlichkeit, ein Item zu lösen bzw. ihm zuzustimmen, „springt“ ab einem bestimmten Punkt auf der Merkmalsdimension ( $\xi$ ) von 0 auf 1. Die dazugehörige → itemcharakteristische Funktion entspricht einer Treppenfunktion.

**Modelle, genestete (*nested models*):** (geschachtelte) Modelle, die durch Parameterrestriktionen ineinander überführbar sind.



**Modelle, hierarchisch geschachtelte (*hierarchical nested models*):**

Mehrere Modelle der → konfirmatorischen Faktorenanalyse werden als hierarchisch geschachtelt bezeichnet, wenn sie dieselbe Modellstruktur aufweisen, sich jedoch in der Anzahl der fixierten oder freigesetzten Parameter unterscheiden. Sie heißen hierarchisch geschachtelt, weil in den verschiedenen Modellen zunehmend mehr Parameter fixiert oder freigesetzt werden, so dass sie auseinander hervorgehen, während die Modellstruktur ansonsten erhalten bleibt.

**Modelle, linear-logistische (*linear-logistic models*):** zerlegen die Schwierigkeitsparameter der → Items in für die Bearbeitung des Items erforderliche Basisoperationen. Jeder der Schwierigkeitsparameter wird als Linearkombination einer möglichst geringen Anzahl von Basisparametern ausgedrückt.

**Modelle, probabilistische (*probabilistic models*):** gehen im Unterschied zu → deterministischen Modellen davon aus, dass die → Wahrscheinlichkeit, ein → Item zu lösen bzw. ihm zuzustimmen, jeden Wert zwischen 0 und 1 annehmen kann; die → itemcharakteristische Funktion entspricht einer monoton steigenden Funktion.

**Moderatorvariable (*moderator variable*):** eine Drittvariable, die die Stärke einer Wirkbeziehung zwischen zwei Variablen beeinflusst, ohne (im Idealfall) selber mit dem → Prädiktor oder dem → Kriterium zu korrelieren. Der Moderator begründet sich aus theoretischen Annahmen über Wirkungszusammenhänge und soll erklären, unter welchen Bedingungen ein bestimmter Effekt auftritt.

**Monte-Carlo-Studie (*Monte-Carlo method*):** die Prüfung der Angemessenheit statistischer Modelle durch Untersuchung künstlich erzeugter Datensätze. Die Methode wird verwendet, um bspw. die → Robustheit von statistischen Verfahren bei Verletzung notwendiger Voraussetzungen zu untersuchen.

**Multidimensionales adaptives Testen (*multidimensional adaptive testing*):**

eine spezielle Form des adaptiven Testens bei der mehrere latente Dimensionen als ursächlich für das beobachtete Antwortverhalten angesehen werden; aus den Antworten wird simultan auf mehrere latente Merkmale geschlossen.

**Multikollinearität (*multicollinearity*):**

Vorliegen einer hohen Korrelation zwischen zwei oder mehreren → unabhängigen Variablen. Dies deutet auf eine spezifische Abhängigkeit und gemeinsame Varianz hin. Damit lassen sich die spezifischen Beiträge einzelner unabhängiger Variablen bei der Varianzaufklärung der → abhängigen Variable nicht mehr differenzieren. Entsprechend steigt der → Standardschätzfehler der Regressionskoeffizienten und die Schätzung wird unzuverlässiger. Das Bestimmtheitsmaß  $r^2$  kann unter diesen Umständen signifikant werden, obwohl alle Koeffizienten der Funktion es nicht sind. → Kollinearität

**Multiple Regressionsanalyse (*multiple regression analysis*):**

eine Sammlung multivariater Analysetechniken, die es erlauben, den gewichteten Einfluss mehrerer → unabhängiger Variablen auf eine → abhängige Variable vorherzusagen.

**Multitrait-Multimethod-Analyse (*multitrait-multimethod analysis*):**

ein korrelatives Verfahren zum Nachweis der → Konstruktvalidität unter Berücksichtigung einer systematischen Kombination von mehreren Merkmalen (→ Traits) und mehreren Messmethoden.

**Nominaldefinition (*nominal definition*):**

Begriffsbestimmung durch Feststellen von Bedeutungsgleichheit. ↔ Realdefinition

**nomologisches Netz (*nomological network*):**

Beziehungsgeflecht zwischen (latenten) Konstrukten und beobachtbaren Testvariablen. Die beiden Ebenen werden mit → Axiomen bzw. empirischen Gesetzen beschrieben und durch Korrespondenzregeln miteinander verbunden.

**nomothetisch (*nomothetic*):** Entdeckung und Beschreibung von Gesetzmäßigkeiten.

**Nonresponse (*nonresponse*):** **1)** Ausfälle von Befragten in Stichprobenerhebungen wegen Nichterreichbarkeit oder Teilnahmeverweigerung; **2)** fehlende Angaben zu einzelnen Fragen. → Werte, fehlende

**Normalisierung (*normalising*):** Bei der Normalisierung wird eine nicht-normalverteilte Testwertvariable zur besseren Interpretierbarkeit so transformiert, dass die Variable danach normalverteilt ist. Die Normalisierung ist von der → Normierung zu unterscheiden, die bei der Testeichung vorgenommen wird.

**Normierung (*standardization*):** dient dazu, zur normorientierten Testwertinterpretation Vergleichswerte zu gewinnen. Dazu werden Testergebnisse von Personen einer Eichstichprobe in Normierungstabellen zusammengestellt (auch: Testeichung).

**Normwert (*standard value*):** ermöglicht es, den → Testwert  $x_v$  einer Testperson hinsichtlich seiner Position in der Testwertverteilung einer bestimmten Bezugsgruppe zu interpretieren (z.B. → Prozentrang, →  $z_v$ -Normwert).

**Nullhypothese (*null hypothesis*):** die der statistischen Hypothesenprüfung zugrundeliegende Behauptung eines Nichtunterschieds oder Nichtzusammenhangs zwischen Gruppen oder Bedingungen. ↔ Alternativhypothese

**Nützlichkeit (*utility*):** Ein Test entspricht diesem → Gütekriterium, wenn die auf seiner Grundlage getroffenen Entscheidungen (Maßnahmen) mehr Nutzen als Schaden erwarten lassen (auch: Utilität).

**Objektivität (*objectivity*):** Ein Test entspricht diesem → Gütekriterium, wenn er dasjenige Merkmal, das er misst, unabhängig von Testleiter, Testauswerter und von der Ergebnisinterpretation misst.

**Objektivität, spezifische (*specific objectivity*):** liegt vor, wenn alle  
 → itemcharakteristischen Funktionen die gleiche Form aufweisen, d.h. lediglich entlang der  $\xi$ -Achse parallel verschoben sind. Ist dies der Fall, kann der Schwierigkeitsunterschied zweier Items ( $\sigma_j - \sigma_i$ ) unabhängig davon festgestellt werden, ob Personen mit niedrigen oder hohen Merkmalsausprägungen  $\xi$  untersucht wurden. In Umkehrung kann auch der Fähigkeitsunterschied zweier Personen ( $\xi_w - \xi_v$ ) unabhängig von den verwendeten Items festgestellt werden.

**Odd-Even-Methode (*odd-even method*):** Methode der Aufteilung eines Tests in Testhälften zur Bestimmung der → Splithalf-Reliabilität. Die → Items des Tests werden nach ihrer Schwierigkeit geordnet und abwechselnd den beiden Halbttests zugewiesen. Ein Halbttest enthält so alle ungeradzahigen Items, der andere die geradzahigen Items des Gesamttests.

**Operationalisierung (*operationalization*):** die Formulierung einer Messvorschrift, die es erlaubt, das interessierende Merkmal zu messen. Für die Erfassung der Merkmalsausprägungen sind valide und reliable Indikatoren zu finden.

**Orthogonalität (*orthogonality*):** Zustand der Unabhängigkeit (Nullkorrelation) von Variablen.

**Outcome-Variable (*outcome variable*):** (abhängige) Variable, an der sich der interessierende Effekte zeigen soll und deren Werte vorhergesagt werden sollen. → Variable, abhängige

**Paarvergleich (*pair comparison*):** ein Skalierungsverfahren, bei dem zwei aus  $N$  Objekten hinsichtlich eines vorgegebenen Kriteriums verglichen werden.

**Panel (*panel*):** eine längsschnittlich wiederholt untersuchte Personengruppe, um Zeitverlaufsdaten zu erhalten.

**Parallelisieren (*matching*):** Bildung einer abhängigen Stichprobe durch Auffinden von Personenpaaren, die jeweils der

Experimental- und → Kontrollgruppe zugeordnet werden.

→ Matching

**parallele Tests (*parallel tests*):** Zwei Tests heißen parallele Tests, wenn sie gleiche wahre Werte und gleiche Fehlervarianzen aufweisen.

**Paralleltest-Reliabilität (*parallel forms reliability*):** Methode der Reliabilitätsschätzung; die → Reliabilität eines Tests, von dem zwei parallele Formen existieren, wird über die Korrelation der → Testwerte der beiden parallelen Testformen geschätzt.

**Partialkorrelation (*partial correlation*):** Verfahren, um eine Korrelation zu bestimmen, die um den Einfluss einer Drittvariablen bereinigt wurde.

**Persönlichkeitsmerkmale (*personality traits*):** mehr oder weniger zeitlich stabile psychische und physische Eigenschaften von Probanden (z.B. Extraversion, Körpergröße). → Trait

**Persönlichkeitstests (*personality tests*):** Persönlichkeitstests dienen der Erfassung von individuell typischem Verhalten als Indikator für die Ausprägung von Persönlichkeitsmerkmalen (Verhaltens- oder Erlebensdispositionen).

**Personenparameter (*person parameter*):** Unter Personenparametern versteht man in der → Item-Response-Theorie die latente Merkmalsausprägung  $\xi_v$  einer Person  $v$ , die im Rahmen eines IRT-Modells geschätzt wird.

**Perzentil (*percentile*):** bezeichnet jenen → Testwert, der einem bestimmten → Prozentrang in der Normierungstichprobe entspricht.

**Placebo (*placebo*):** ein unwirksames Präparat im Sinne einer Scheinbehandlung, dass Mitgliedern der → Kontrollgruppe verabreicht wird. Tritt trotzdem ein Effekt ein, wird von einem Placebo-Effekt gesprochen.

**Plausibilitätsanalyse (*plausibility analysis*):** Arbeitsschritt vor der eigentlichen → Datenanalyse, um die Daten auf Eingabefehler zu

überprüfen, zu bereinigen und ggf. ungewöhnliche Angaben  
(→ Ausreißer, Extremwerte) zu identifizieren.

**Polaritätsprofil (*semantic differential*):** eine standardisierte Erhebungstechnik zur Beurteilung von Objekten oder Sachverhalten.

**Population (*population*):** → Grundgesamtheit

**Powertests (*power test*):** → Leistungstests mit eher schwierigen Aufgaben, wobei erhoben wird, welches Schwierigkeitsniveau der Aufgaben der Proband ohne Zeitbegrenzung bewältigen kann (auch: Niveautests).

**Prädiktor (*predictor*):** andere Bezeichnung für → unabhängige Variable, die zur Vorhersage der Werte einer → abhängigen Variablen verwendet wird.

**Prävalenz (*prevalence*):** Vorkommenshäufigkeit (Zahl der Fälle oder Ereignisse) einer Krankheit zu einem bestimmten Zeitpunkt.

**Prävalenzrate (*prevalence rate*):** Verhältnis der zu einem bestimmten Zeitpunkt Erkrankten zur → Grundgesamtheit aller Personen, die hätten erkranken können.

**Pretest (*pretest*):** Vortestphase bestimmter Arbeitsschritte einer Untersuchung, z.B. zur Erprobung eines Untersuchungsinstrumentes.

**Proband (*subject*):** Teilnehmer/in an einer Untersuchung.  
→ Untersuchungseinheit; →  $V_p$

**projektive Tests (*projective tests*):** bei diesen Tests kommt mehrdeutiges Stimulusmaterial (meist Bilder) zum Einsatz. Es wird angenommen, dass → Probanden unbewusste oder verdrängte Bewusstseinsinhalte in das Bildmaterial hineinprojizieren und dadurch Persönlichkeitsmerkmale ermittelt werden können.

**Proximitätsmaße (*proximity measures*):** Maße zur Quantifizierung von Ähnlichkeiten (z.B. Korrelationsmaße) oder Unähnlichkeiten (z.B. Distanzmaße).

**Prozentrang (*percentile rank*):** erlaubt eine Angabe, wie viel Prozent der Bezugsgruppe bzw. Normierungsstichprobe einen → Testwert

erzielten, der niedriger oder maximal ebenso hoch ist, wie der Testwert  $x_v$  der Testperson  $v$ . → Perzentil

**Pygmalion-Effekt (*Pygmalion effect*):** Effekt der Auswirkungen von Erwartungen des Versuchsleiters auf das Verhalten der → Probanden (auch: Rosenthal-Effekt). Benannt wurde der Effekt in Anlehnung an das Schauspiel von Shaw.

**qualitativ (*qualitative*):** nicht-numerisches (verbales, visuelles) Untersuchungsmaterial. ↔ **quantitativ**

**Quantifizierung (*quantification*):** Kennzeichnung von Merkmalsausprägungen über Zahlen und Auswertung über Auszählung (Häufigkeiten).

**quantitativ (*quantitative*):** numerisches oder numerisch-kodiertes Untersuchungsmaterial. ↔ qualitativ

**Quartil:** Das erste, zweite bzw. dritte Quartil (Q1, Q2, Q3) ist jener → Testwert  $x_v$ , der von 25%, 50% bzw. 75% der Testwerte unterschritten bzw. höchstens erreicht wird. → Perzentil

**Quasi-Experiment (*quasi-experiment*):** Forschungsmethoden mit dem Anspruch der Analyse kausaler Abhängigkeiten, allerdings mit der Einschränkung (gegenüber echten → Experimenten), dass die → unabhängige Variable nicht manipuliert und die Aufteilung der Versuchspersonen auf die experimentellen Bedingungen nicht durch den Forscher kontrolliert werden können.

**Querschnittsuntersuchung (*cross-sectional design*):** Studien mit einem Messzeitpunkt.

**Quotenstichprobe (*quota-sampling*):** ein (nicht zufallsgesteuertes) Verfahren zur Stichprobenziehung, bei dem durch als relevant erachtete Merkmale Quoten vorgegeben werden, die den bekannten Anteilen der Bezugspopulation entsprechen.

**Rangplatz (*rank*):** Ordnungsnummer in einer nach Größe geordneten Reihe von Beobachtungswerten.

**Rangplatzbindungen (*ties*):** Auftreten gleicher

Merkmalsausprägungen, die zu gleichen Werten und damit zu einem identischen Rangplatz führen. Für das Auslösen identischer Rangplätze existieren verschiedene Möglichkeiten.

**Rasch-Modelle (*Rasch models*):** Klasse von spezifisch objektiven Modellen in der → Item-Response-Theorie.

**Ratekorrektur (*guess-correction*):** zieht bei der Testwertbestimmung jene Anzahl an „richtigen“ Lösungen ab, die nur durch Erraten der richtigen Antworten entstanden ist.

**Ratingskala (*rating scale*):** Antwortformat, bei dem ein Urteil oder eine Einschätzung graduell auf einer kategorialen oder analogen Skala markiert wird (auch: Schätzskala).

**Realdefinition (*real definition*):** Begriffsbestimmung durch Angabe wesentlicher Merkmale. ↔ Nominaldefinition

**Receiver-Operating-Characteristics-Analyse (*ROC*):** ermöglicht für eine binäre Klassifikation (z.B. krank vs. nicht krank) den zur Fallunterscheidung verwendeten Schwellenwert optimal in der Weise festzulegen, dass Trefferquote und Quote korrekter Ablehnungen maximiert werden.

**Regressand (*regressand*):** Bezeichnung für → abhängige Variable in der Regressionsanalyse.

**Reliabilität (*reliability*):** bezeichnet die Messgenauigkeit eines Tests (→ Gütekriterium nach der Klassischen Testtheorie). Ein Testverfahren ist perfekt reliabel, wenn die damit erhaltenen → Testwerte frei von zufälligen Messfehlern sind. Je höher die Einflüsse solcher zufälligen Messfehler sind, desto weniger reliabel ist das Testverfahren. In der Klassischen Testtheorie wird Reliabilität als Verhältnis zwischen true score-Varianz  $Var(\tau)$  und Testwertvarianz  $Var(x)$  definiert (auch: Zuverlässigkeit, Genauigkeit).

**Replikation (*replication*):** Wiederholung einer Studie zur Überprüfung vorheriger Studien und zwecks Steigerung der → Validität.



**Repräsentativität (*representativeness*): 1)** das Ausmaß, in dem Stichprobenergebnisse auf die zugehörige → Grundgesamtheit verallgemeinert werden können ( → externe Validität); **2)** eine Stichprobe weist dieses Merkmal dann auf, wenn sie hinsichtlich ihrer Zusammensetzung die jeweilige Zielpopulation möglichst genau abbildet.

**Repräsentativität, globale (*global representativeness*):** eine Situation, bei der eine gezogene Stichprobe in allen Merkmalen der → Grundgesamtheit entspricht.

**Repräsentativität, lokale (*local representativeness*):** eine Situation, bei der eine gezogene Stichprobe nur in spezifischen Merkmalen der → Grundgesamtheit entspricht.

**Residualvarianz (*residual variance*):** der durch die Faktoren nicht erklärte Varianzanteil einer Variablen bei einer → exploratorischen Faktorenanalyse; gilt damit als Indikator der nicht gemeinsamen spezifischen Varianz einer Variablen, die durch die gefundene Faktorenlösung nicht erklärt werden kann.

**Residuum (*residual*):** bezeichnet die Differenz (Restbetrag) zwischen einem beobachteten und einem erwarteten (geschätzten) Wert, z.B. als Fehlerterm bei einer Regressionsrechnung. Aus der Größe des Vorhersagefehlers bemisst sich die Güte des Regressionsmodells.

**Restkategorie (*residual category*):** eine gesonderte Kategorie, die Fälle aufnimmt, die sich nicht im Kategoriensystem einordnen lassen; diese sollte möglichst wenige Fälle enthalten.

**Retest-Reliabilität (*retest reliability*):** Methode der Reliabilitätsschätzung. Ein Test wird zu zwei Messzeitpunkten der gleichen Stichprobe vorgegeben. Die Korrelation der Testwerte beider Messzeitpunkte dient als Maß der → Reliabilität des Tests.

**Robustheit (*robust*):** eine Bezeichnung für statistische Testverfahren, die (relativ) unempfindlich gegenüber Verletzungen der Modellannahmen sind oder nicht auf Extremwerte reagieren.

**Rohdaten (*raw scores*):** die (Ur-)Daten, die direkt mit dem Erhebungsinstrument bzw. dem Messverfahren erzielt werden. Diese Daten müssen für die statistische Verarbeitung ggf. noch vorbereitet (bereinigt, verrechnet, aggregiert) werden.

**Rotation (*rotation*):** ein bei der → exploratorischen Faktorenanalyse angewandtes Verfahren, das Variablen entweder in voneinander unabhängige (orthogonale, rechtwinklige) Faktoren oder in korrelierte (oblique, schiefwinklige) Faktoren gruppiert. Das Verfahren dient der Verbesserung der Eindeutigkeit der Zuordnung und damit der Erleichterung der Ergebnisinterpretation.

**Schichtung (*stratification*):** eine Technik der Stichprobenziehung, bei dem Vorkenntnisse zu relevanten Merkmale der → Grundgesamtheit genutzt werden, um diese als Schichtungsmerkmale bei der Ziehung von Zufallsstichproben zu verwenden.

**Schwellenwert (*quantitative threshold*):** im Rahmen kriteriumsorientierter Testwertinterpretation jener → Testwert, ab dem das Kriterium als zutreffend angenommen wird. Kann z.B. mittels ROC-Analyse empirisch bestimmt werden.

**Schwierigkeitsindex (*discriminative power*):** Konzept der → Itemanalyse zur Beurteilung der Differenzierungsfähigkeit von → Items. Ermittelt sich aus dem Verhältnis richtig gelöster zu bearbeiteten Aufgaben. Der Schwierigkeitsindex schwerer Aufgaben ist also niedrig (minimal 0), der Schwierigkeitsindex leichter Aufgaben hoch (maximal 100).

**Schwierigkeitsparameter (*complexity parameter*):**  $\sigma_i$  ist ein Itemparameter, der durch jene Merkmalsausprägung  $\xi$  definiert ist, bei der die Lösungswahrscheinlichkeit des Items 50% beträgt.

**Screeninginstrument (*screening instrument*):** Bezeichnung für Kurzdiagnoseverfahren, um einfach und schnell über Verdachtsmomente einer Erkrankung entscheiden zu können

(z.B. zur Früherkennung bei Reihenuntersuchungen). Bei Verdacht sollte sich immer eine ausführliche Diagnostik anschließen.

**Sekundäranalyse (*secondary analysis*):** eine Form der Datenanalyse, die auf bereits vorliegende Datenbestände zurückgreift, z.B. um eine frühere Auswertung zu überprüfen, eine neue → Hypothese anhand vorhandener Daten zu explorieren oder neue Auswertungstechniken zu erproben.

**Serendipität (*serendipity*):** bezeichnet den Umstand, wenn im Rahmen einer Untersuchung (mit anderer Zielsetzung) unerwartete und unbeabsichtigte Entdeckungen gemacht werden.

**Signifikanz (*significance*):** die Bezeichnung für einen statistischen Effekt auf der Grundlage eines entsprechenden statistischen → Signifikanztests. Auf der Grundlage wahrscheinlichkeitstheoretischer Überlegungen wird entschieden, ob ein (über das → Signifikanzniveau eingegrenzter) Unterschied zwischen einer empirischen Testgröße und einem Vergleichswert einer Zufallsverteilung besteht.

**Signifikanzniveau (*alpha level*):** ein inferenzstatistisches Konzept für die → Wahrscheinlichkeit, sich bei dem Schluss von Stichprobenstatistiken auf Populationsparameter zu irren (→ Irrtumswahrscheinlichkeit). Über eine vorab festgelegte Wahrscheinlichkeit wird ein adäquater → Alpha-Fehler definiert, der begangen wird, wenn ein Effekt zwischen Variablen behauptet wird, aber in Wirklichkeit keiner besteht.

**Signifikanztest (*significance test*):** Prüfverfahren, das eine Prüfgröße zur Beurteilung der statistischen → Signifikanz liefert (auch: Hypothesentest).

**Skala (*scale*):** **1)** eine eindimensionale Folge von Positionen, die unterschiedliche Ausprägungsgrade eines Merkmals anzeigt; **2)** Zusammenfassung mehrerer → Items, die eine gemeinsame Merkmalsdimension messen sollen. ↔ Index

**Skalierung (*scaling*):** ein Test erfüllt dieses → Gütekriterium, wenn die laut Verrechnungsregel resultierenden Testwerte die empirischen Merkmalsrelationen adäquat abbilden.

**soziale Erwünschtheit (*social desirability*):** beinhaltet die Antworttendenz eines → Probanden, sich selbst so darzustellen, wie es soziale Normen seiner Wahrnehmung nach erfordern (auch: soziale Desirabilität). → Akquieszenz

**Spearman-Brown-Korrektur (*Spearman-Brown correction*):** Formel zur Schätzung der → Reliabilität eines Tests bei Verlängerung des Tests um homogene Testteile; findet auch bei der → Splithalf-Reliabilität Verwendung, bei der die Halbttest-Reliabilität auf die Reliabilität des Gesamttests aufgewertet wird.

**Speedtest (*speed test*):** → Leistungstest mit meist einfachen Aufgaben, wobei erhoben wird, wie viele Aufgaben unter Zeitdruck gelöst werden können (auch: Geschwindigkeitstest).

**Spezifität (*specificity*):** beschreibt das Ausmaß der durch die Situation und die Person-Situation-Interaktion erklärten Varianz relativiert an der Gesamtvarianz einer Messvariablen.

**Splithalf-Reliabilität (*split-half reliability*):** Methode der Reliabilitätsschätzung. Aus den → Items eines Tests werden zwei möglichst parallele Testhälften gebildet. Aus der Korrelation der Testwerte der Halbttests wird mittels Spearman-Brown-Korrektur die → Reliabilität des Gesamttests geschätzt (auch: Testhalbierungs-Reliabilität).

**Standardmessfehler (*standard error of measurement*):**  $SD(\epsilon)$  eines Tests resultiert aus der Unreliabilität des Tests und errechnet sich als Wurzel aus der Fehlervarianz eines Tests. Der Standardmessfehler ist bei höherer → Reliabilität kleiner und bei niedrigerer Reliabilität größer.

**Standardschätzfehler (*standard error of estimate*):** die bei mehreren Schätzungen eines Parameters gegebene Standardabweichung der Fehler zwischen Schätzwert und wahren Wert.

**Standardnorm (*standard norm*):** die z-Norm sowie weitere durch Lineartransformationen gewonnene Normen (z.B. IQ- oder *T*-Norm). →  $z_{\nu}$ -Normwert

**Standardwerte (*standard values*):** Hierbei handelt es sich um  $z_{\nu}$ -Werte, die unter der Annahme der Normalverteilung auch Angaben zu prozentualen Häufigkeiten innerhalb beliebiger Flächenanteile erlauben. →  $z_{\nu}$ -Normwert

**State (*state*):** Ein zeitlich begrenzter biologischer, emotionaler und kognitiver Zustand, in dem sich eine Person befinden kann. Er kennzeichnet sich durch personenbedingte, situativ bedingte und durch die Interaktion zwischen Person und Situation bedingte Einflüsse. ↔ Trait

**Statement (*statement*):** Aussagen oder Behauptungen in einem schriftlichen Untersuchungsinstrument, die von den Befragten (mittels eines freien oder gebundenen Antwortformates) einzuschätzen sind.

**Statistiken, erschöpfende (*sufficient statistics*):** die Zeilen- und Spaltensummenscores einer (0/1)-Datenmatrix werden als solche bezeichnet, wenn die → Wahrscheinlichkeit der Daten nicht davon abhängt, welche Personen welche → Items gelöst haben, sondern lediglich davon, wie viele Personen ein Item gelöst haben (Schwierigkeit des Items), bzw. wie viele Items eine Person lösen konnte (Fähigkeit der Person).

**Stichprobe (*sample*):** Auswahl aus einer → Grundgesamtheit nach einem Zufallsprinzip oder nicht zufälligen Ziehungsverfahren. Echte → Zufallsstichproben erlauben dabei Rückschlüsse auf Parameter der → Grundgesamtheit.

**Stichprobe, unabhängige (*independent sample*):** resultiert aus einem Zuweisungsverfahren, bei dem die Zuordnung von Untersuchungseinheiten einer Gruppe nicht durch die Zuordnung zu einer anderen Gruppe beeinflusst wird.

**Stichprobenfehler (*sample error*):** bezeichnet die (unvermeidliche) Abweichung einer Statistik von einem Parameter, wenn diese auf

Stichprobendaten basiert. Eine Schätzung kann über den  
→ Standardschätzfehler erfolgen.

**Stichprobenunabhängigkeit (*sample independence*):** bedeutet, dass in IRT-Modellen die Itemparameter unabhängig von den Personen und die Personenparameter unabhängig von den Items geschätzt werden können. → Item-Response-Theorie

**Suppressorvariable (*suppressor variable*):** dient dem Effekt, die Genauigkeit einer Prognose der geschätzten Werte bei einer Regression durch Berücksichtigung einer Variablen zu verbessern, die selbst nur schwach mit der → abhängigen Variable korreliert. Der Suppressor erhöht die Vorhersageleistung einer (oder mehrerer) anderer Variablen, indem er irrelevante Varianz in der/den Prädiktorvariable/n unterdrückt. Dazu muss dieser mit einer anderen → unabhängigen Variable korreliert sein, die wiederum mit der abhängigen Variable korreliert ist.

**Taxonomie (*taxonomy*):** ein hierarchisches Ordnungssystem zur systematischen Klassifikation von Fällen, Objekten, Sachverhalten etc. nach festen Regeln.

**Tendenz zur Mitte (*bias to the middle*):** eine Antworttendenz, die extreme Antworten eher vermeidet und mittlere Antwortkategorien eher bevorzugt.

**Testadaptation (*test adaptation*):** bezeichnet den Prozess einer qualitativ hochwertigen Übertragung (Übersetzung unter Berücksichtigung von Konstruktäquivalenz) und empirischen Evaluation psychologischer Tests aus anderen Sprachen und in andere Sprachen unter Beachtung der kulturellen Unterschiede.

**Testeichung (*test standardization*):** dient dazu, → Normwerte zur normorientierten Testwertinterpretation zu gewinnen. Dazu wird der Test an Personen einer Normierungsstichprobe, welche hinsichtlich einer definierten Bezugsgruppe repräsentativ ist, durchgeführt.

**Testgütekriterien (*psychometric properties*):** stellen ein System zur Qualitätsbeurteilung psychologischer Tests dar. Dabei werden hauptsächlich unterschieden: → Objektivität, → Reliabilität, → Validität, → Skalierung, → Normierung, → Testökonomie, → Nützlichkeit, → Zumutbarkeit, Unverfälschbarkeit, → Fairness und Transparenz. → Gütekriterien

**Testitem (*item*):** Zu beantwortende Aufgabe (Frage, Aussage, → Statement etc.) eines Tests.

**Testökonomie (*test economy*):** Ein Test entspricht diesem → Gütekriterium, wenn er, gemessen am diagnostischen Erkenntnisgewinn, relativ wenig Ressourcen wie Zeit, Geld oder andere Formen beansprucht.

**Teststandards (*test standards*):** vereinheitlichte Leitlinien, in denen sich allgemein anerkannte Zielsetzungen zur Entwicklung, Adaptation, Anwendung und Qualitätsbeurteilung psychologischer Tests widerspiegeln.

**Teststärke (*power*):** das Leistungsvermögen eines statistischen → Signifikanztests, bestehende Unterschiede nachzuweisen. Eine Analyse der Teststärke erlaubt es, zwischen fehlenden Effekten und Problemen in der Anlage der Untersuchung (Stichprobengröße, → Effektgröße, → Signifikanzniveau, Stichprobenvarianz) aufzudecken.

**Testwert (*test value*):** das individuelle numerische Testresultat; wird aus den registrierten Antworten einer Testperson durch Anwendung definierter Regeln unmittelbar gebildet.

**Testwertestreuung (*standard deviation*):** sagt aus, wie breit die empirisch gewonnenen → Testwerte einer Stichprobe um den Mittelwert der Testwerte verteilt sind. Die Streuung der Testwerte wird meist als Standardabweichung  $SD(x)$  angegeben; man gewinnt sie als Wurzel aus der → Testwertestreuung  $Var(x)$ .

**Testwertestreuung (*variance*):**  $Var(x)$  ist die Varianz der beobachteten → Testwerte. Sie setzt sich aus der wahren Varianz  $Var(\tau)$  und der Fehlervarianz  $Var(\epsilon)$  zusammen.

- Thurstone-Skala (*Thurstone scale*):** eine Einstellungsskala, bei der  
 → Statements zunächst von Beurteilern auf einer Elf-Punkte-Skala bewertet werden. Danach wird die → Skala aus den eindeutig einer Position zugeordneten Statements gebildet.
- Trait (*trait*):** mehr oder weniger zeitlich stabiles Merkmal (Disposition), das personeninhärent und transsituativ überdauernd ist. ↔ State
- Transformation (*transformation*):** Vorgang der Erzeugung einer neuen Variablen durch bestimmte mathematische Operationen (z.B. Erzeugung von → Standardwerten durch z-Transformation).  
 →  $z_v$ -Normwert
- Treatment (*treatment*):** **1)** die Maßnahme, Bedingung oder Einwirkung in einem experimentellen Forschungsdesign, denen die Versuchspersonen in der Experimentalgruppe ausgesetzt werden; **2)** die Behandlung (Therapie, Rehabilitationsmaßnahme), die ein Patient (in einer Einzelfallstudie) erhält.
- Trennschärfe (*selectivity, discriminability*):** ein statistisches Maß das angibt, wie gut ein → Item das Gesamtergebnis eines Tests (der übrigen Items) repräsentiert. Das Maß ist definiert als Korrelation der Beantwortung diese Items mit dem Gesamtestwert (daher auch: *corrected item-total correction*).  
 → Itemtrennschärfe
- Triangulation (*triangulation*):** ein Verfahren zur Sicherung der  
 → Reliabilität einer Messung, bei der ein Messpunkt von unterschiedlichen Punkten her bestimmt wird (z.B. durch Einsatz verschiedener Erhebungsmethoden, unabhängiger Beobachter, unterschiedlicher oder kombinierter Datenquellen etc.).
- True Score (*true score*):** der True Score bzw. wahre Wert  $\tau_v$  ist die wahre Ausprägung des → Probanden  $v$  in dem von einem Test gemessenen Merkmal. Da Messungen in der Regel fehlerbehaftet sind, stimmen → Testwert  $x_v$  und wahrer Wert  $\tau_v$  nicht völlig überein. Ein → Konfidenzintervall für  $\tau_v$  kann mit Hilfe des  
 → Standardmessfehlers bestimmt werden.



**Typologie (*typology*):** ein Zuordnungssystem, das unter Verwendung eines Klassifikationsverfahrens von definierten Merkmalen entsteht.

**Untersuchungseinheit (*unit of analysis*):** → Merkmalsträger;  
→ Proband; → Vp

**Urliste (*list of raw scores*):** eine erste, noch unbereinigte Zusammenstellung der → Rohdaten einer Untersuchung.

**Validität (*validity*): 1)** Ein Test entspricht diesem → Gütekriterium, wenn er das Merkmal, das er messen soll, auch wirklich misst und nicht irgendein anderes. **2)** Bezeichnet die Menge der zutreffenden Schlussfolgerungen, die aus einem Testergebnis gezogen werden können.

**Validität, curriculare (*curricular validity*):** bezeichnet die Übereinstimmung von Inhalten eines Tests, der zur Überprüfung der Erreichung eines Lernziels dienen soll, mit den Inhalten des Lehrplans.

**Validität, diskriminante (*discriminant validity*):** gilt im Rahmen der Konstruktvalidierung als nachgewiesen, wenn Messungen verschiedener Konstrukte mit derselben Methode nicht oder nur gering miteinander korrelieren.

**Validität, empirische (*empirical validity*):** ein Konzept für die Angabe, ob mit Hilfe der → Operationalisierung auch wirklich das gemessen wird, was mit dem Begriff gemeint ist.

**Validität, externe (*external validity*):** eine graduelle Größe der zulässigen Verallgemeinerung von Stichprobenergebnissen auf andere Personen, Situationen oder Zeitpunkte.

**Validität, inkrementelle (*incremental validity*):** bezeichnet das Ausmaß, in dem die Vorhersage eines externen → Kriteriums verbessert werden kann, wenn zusätzliche Testaufgaben oder (Sub-)Tests zu den bereits eingesetzten Verfahren hinzugenommen werden.

**Validität, interne (*internal validity*):** Grad in einem Untersuchungsdesign, die Veränderungen der → abhängigen Variable ausschließlich auf Variationen der → unabhängigen Variablen zurückzuführen. Dies erfordert die Kontrolle potenzieller → Störvariablen, um alternative Erklärungen von Effekten auszuschließen.

**Validität, konvergente (*convergent validity*):** Im Rahmen der Konstruktvalidierung gilt die konvergente Validität als nachgewiesen, wenn Messungen eines Konstrukts, das mit verschiedenen Methoden erfasst wird, hoch miteinander korrelieren.

**Validität, kriteriumsbezogene (*criterion-related validity*):** eine graduelle Größe der Treffsicherheit eines Verfahrens bei einer Prognose anhand definierter Außenkriterien (auch: prädiktive Validität). → Konstruktvalidität

**Validität, logische (*logical validity*):** ein Konzept für die Angabe, ob die operationalen Vorschriften den gleichen Bedeutungsumfang besitzen wie die Begriffsdefinition.

**Validität, statistische (*statistical validity*):** ein Konzept für die Angabe, ob eine empirische Aussage durch eine ausreichende große Stichprobe, ein genaues Messinstrument und die Anwendung angemessener statistischer Verfahren gerechtfertigt ist.

**Variable, abhängige (*dependent variable*):** die Variable, deren Werte durch mehrere andere Variablen systematisch beeinflusst werden. Besteht diese Abhängigkeit, können Werte der abhängigen Variable auf Grund der Kenntnis der Werte der → unabhängigen Variable vorhergesagt werden. → Kriteriumsvariable (im nicht-experimentellen Forschungskontext auch *response variable* oder → Outcome-Variable genannt).

**Variable, latente (*latent variable*):** ein Merkmal, das nicht direkt beobachtbar ist, dessen Berücksichtigung aber im Kontext theoretischer Überlegungen hilfreich ist.

**Variable, manifeste (*manifest variable*):** Variable zur Erfassung des beobachtbaren Antwortverhaltens mit verschiedenen → Items, die → Indikatoren für die latente Dimensionen darstellen.

**Variable, unabhängige (*independent variable*):** eine Variable, die zur Erklärung der Variation der Werte einer → abhängigen Variablen herangezogen wird. In einem experimentellen Forschungsdesign wird dieser Effekt als ursächlich angenommen, wobei allerdings unkontrollierte Störvariablen den Effekt beeinflussen können.  
→ Prädiktor

**Vertrauensgrenzen (*confidence limits*):** obere und untere Grenze eines → Vertrauensintervalls.

**Vertrauensintervall (*confidence interval*):** → Konfidenzintervall

**Verweigerer (*non-respondent*):** Personen, die im Stichprobenplan als → Probanden vorgesehen sind, sich aber trotz (mehrmaliger) Aufforderung nicht oder nur unzureichend an einer Untersuchung beteiligen.

**Vollerhebung (*census*):** eine solche Erhebung liegt vor, wenn alle Elemente der → Grundgesamtheit erfasst werden. In diesem Fall reichen zur Auswertung deskriptive Maße und Modellprüfungen.

**Vp (*subject*):** Kurzform für Versuchsperson → Proband;  
→ Untersuchungseinheit

**Wahrscheinlichkeit (*probability*):** ein Wert  $p$  zwischen den theoretischen Grenzen 0 (Unmöglichkeit eines Ereignisses) und 1 (Ereignis tritt mit Sicherheit auf). In der Statistik wird mit steigender (Auftretens-)Wahrscheinlichkeit die Zufälligkeit für das Zustandekommen eines Ereignisses ausgeschlossen.

**wahre Varianz (*true variance*):** Die wahre Varianz  $\text{Var}(\tau)$  ist die Varianz der wahren Werte  $\tau_v$  in einem Test. Sie ist meistens niedriger als die Testwertvarianz  $\text{Var}(x)$ . Aus dem Verhältnis beider Varianzanteile resultiert in der Klassischen Testtheorie die → Reliabilität.

**Werte, fehlende (*missing values*):** leere Zellen im Datensatz, die aufgrund einer Nicht-Beantwortung oder Beantwortung außerhalb vorgesehener Bewertungskategorien entstehen. Fehlende Angaben bilden grundsätzlich eine gesonderte Kategorie („keine Antwort“) und sind von Angaben wie „trifft nicht zu“ oder „Sonstiges“ zu trennen.

**x-Achse (*x-axis*):** waagerechte Achse in einem Koordinatensystem (auch: Abszisse).

**x-Variable (*x-variable*):** andere Bezeichnung für → unabhängige Variable.

**y-Achse (*y-axis*):** senkrechte Achse in einem Koordinatensystem (auch: Ordinate).

**y-Variable (*y-variable*):** andere Bezeichnung für → abhängige Variable.

**z-Transformation (*z-score*):** → Standardwerte einer Verteilung mit Mittelwert 0 und Standardabweichung 1.

**z<sub>v</sub>-Normwert (*z-norm-value*):** gibt an, wie stark der → Testwert  $x_v$  einer Testperson  $v$  vom Mittelwert  $x_{\text{quer}}$  der Verteilung der Bezugsgruppe in Einheiten der Standardabweichung  $SD(x)$  abweicht.

**Zeitpartitionierungsmethode (*time-partitioning method*):** Methode der Aufteilung eines Tests in Testhälften zur Bestimmung der → Splithalf-Reliabilität, wobei die Testhälften aus den Items von zwei jeweils gleich langen Bearbeitungsabschnitten gebildet werden.

**Zielpopulation (*target population*):** Im Rahmen der → Testeichung diejenige Bezugsgruppe, für welche die zu erstellenden Testnormen gelten sollen und aus der entsprechend die Normierungsstichprobe zu ziehen ist.

**zufällig (*random*):** Ereignisse, deren Auftreten sich nicht sicher vorhersagen lässt.

**Zufallsexperiment (*random experiment*):** ein in der Wahrscheinlichkeitstheorie und Inferenzstatistik zentrales Konzept, demzufolge realisierte Merkmalsausprägungen als Ergebnis eines nicht-deterministischen, durch zahlreiche Einflüsse bestimmten Zufallsprozesses interpretiert werden können.

**Zufallsstichprobe (*random sample*):** eine Stichprobe, bei der per Zufall aus einer hinsichtlich ihrer Elemente bekannten und abzählbaren → Grundgesamtheit ein bestimmter Umfang an Elementen gezogen wird.

**Zufallsvariable (*random variable*):** eine Variable, bei der zufälligen Ereignissen per Zuordnungsvorschrift reelle Zahlen zugeordnet werden.

**Zumutbarkeit (*reasonability*):** liegt vor, wenn ein Test absolut sowie relativ zu dem aus seiner Anwendung resultierenden Nutzen die zu testende Person in zeitlicher, psychischer sowie körperlicher Hinsicht nicht über Gebühr belastet.