

Anhang

Glossar – 390

Sachverzeichnis – 403

Glossar

A

Adaptiver Algorithmus Ein adaptiver Algorithmus ist ein Regelsystem, welches beim adaptiven Testen die Itemauswahl zu Beginn und während des Tests regelt sowie Kriterien der Testbeendigung spezifiziert.

Adaptives Testen 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.

AIC Unter dem AIC (Akaike's Information Criterion) versteht man ein Maß für die Anpassungsgüte des geschätzten Modells an die vorliegenden empirischen Daten (Stichprobe) unter Berücksichtigung der Komplexität des Modells. Daraus hervorgegangen sind BIC und CAIC.

Akquieszenz Mit Akquieszenz bezeichnet man die Antworttendenz, auf Aussagen (Statements) unabhängig vom Inhalt eher mit Zustimmung zu reagieren.

Austauschbare Methoden Austauschbare Methoden in MTMM Modellen sind solche Methoden, die einer Zufallsauswahl aus einer Menge gleichberechtigter (gleichadäquater) Methoden entsprechen. Beispielsweise wären verschiedene Messgelegenheiten austauschbar, wenn sie einer Zufallsauswahl entsprechen und keine der Messgelegenheiten sich von den anderen Messgelegenheiten strukturell unterscheidet.

Auswahlaufgaben Aufgabentyp, bei dem die Probanden vor die Anforderung gestellt werden, aus mehreren vorgegebenen Antwortalternativen die richtige bzw. zutreffende Antwort zu identifizieren.

Auswertungsobjektivität (Gütekriterium) Ein Test ist dann auswertungsobjektiv, wenn das Testergebnis unabhängig davon ist, wer den Test auswertet.

Autokorrelationseffekt In longitudinalen Strukturgleichungsmodellen korrelieren Indikatoren oft stärker über die Messgelegenheiten hinweg als mit anderen Indikatoren derselben Messgelegenheit, die dasselbe Konstrukt messen.

Axiom Axiome sind theoretische Grundannahmen, die ohne weitere Überprüfung als gegeben angesehen werden.

B

Bedingte Antwortmusterwahrscheinlichkeit $P(a_v|g)$ Bei der dichotomen LCA: Wahrscheinlichkeit eines Antwortmusters a_v unter der Bedingung, dass die Person v zur Klasse g gehört.

Bedingte Itembejahungswahrscheinlichkeit $P(x_{vi}=1|g)$ Bei der dichotomen LCA: Wahrscheinlichkeit, mit der ein Item i bejaht wird, wenn die entsprechende Person v zur Klasse g gehört.

Bedingte Kategorienwahrscheinlichkeit $P(x_{vi}=k|g)$ Bei der polytomen LCA: Wahrscheinlichkeit, mit der ein Item i mit der Antwortkategorie k beantwortet wird, wenn die entsprechende Person v zur Klasse g gehört.

Bedingte Klassenzuordnungswahrscheinlichkeit $P(g|a_v)$ Bei der dichotomen LCA: Wahrscheinlichkeit, mit der eine Person v mit dem Antwortmuster a_v zur Klasse g gehört.

Beurteilungsaufgaben Aufgabentyp, bei dem der individuelle Zustimmung- oder Ablehnungsgrad zu einer vorgelegten Aussage (Statement) erfasst wird.

BIC Unter dem BIC (Bayesian Information Criterion) 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.

C

CAIC Das Corrected-Information-Criterion (CAIC) ist eine Abwandlung des AIC, bei dem der Stichprobenumfang berücksichtigt wird.

Cognitive Lab »Cognitive Lab« beschreibt eine explorative Methode zur Untersuchung von Antwortprozessen. Dabei werden Probanden unmittelbar nach der Bearbeitung der Items zu ihrem Lösungsverhalten interviewt oder bei der Bearbeitung der Items gebeten, laut zu denken.

Consequential Validity Consequential Validity beschäftigt sich mit der Frage, ob mit dem Einsatz eines Testverfahrens das damit in der Praxis verfolgte Ziel erreicht wird.

Cronbachs α Koeffizient der internen Konsistenz als Reliabilitätsschätzung. Cronbachs α 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 wird die interne Konsistenz und damit die Reliabilität.

Curriculare Validität Curriculare Validität bezeichnet die Übereinstimmung von Inhalten eines Tests, der zur Überprüfung der Erreichung eines Lernziels dienen soll, mit den Inhalten des Lehrplans.

D

Debriefing Debriefing beschreibt die Qualitätssicherungsmaßnahme, nach der Testung den Testleiter nach Besonderheiten während der Testung zu befragen.

Deterministische Modelle Deterministische Modelle nehmen an, dass die Wahrscheinlichkeit, ein Item zu lösen, nur 0 oder 1 betragen kann; die Wahrscheinlichkeit, ein Items zu lösen bzw. ihm zuzustimmen, »springt« ab einem bestimmten Punkt auf der Merkmalsdimension ξ von 0 auf 1. Die IC-Funktion entspricht einer Sprungfunktion/Treppenfunktion.

DIN 33430 Die DIN 33430 ist eine verbindliche Norm von Qualitätsstandards für die berufsbezogene Eignungsbeurteilung, die verwendeten Tests und die einzelnen Ablaufschritte.

Disjunktheit von Antwortalternativen Disjunktheit von Antwortalternativen liegt vor, wenn zwischen den Antwortalternativen keine Überlappungen bestehen.

Diskriminante Validität Im Rahmen der Konstruktvalidierung gilt die diskriminante Validität als nachgewiesen, wenn Messungen verschiedener Konstrukte mit derselben Methode nicht oder nur gering miteinander korrelieren.

Diskriminationsindex Unter dem Diskriminationsindex versteht man einen Kennwert zur Identifizierung »nicht trennscharfer« Items bei der LCA.

Distraktoren Als Distraktoren bezeichnet man plausibel erscheinende, aber nicht zutreffende Antwortalternativen bei Auswahlaufgaben.

Dreiparameter-Logistisches-Modell (3PL-Modell) Im 3PL-Modell wird zusätzlich zu den Parametern des 2PL-Modells noch die Ratewahrscheinlichkeit als Parameter ρ_1 in das Modell aufgenommen und geschätzt.

Durchführungsobjektivität (Gütekriterium) Ein Test ist dann durchführungsobjektiv, wenn das Testergebnis unabhängig davon ist, wer den Test vorgibt.

E

Eichstichprobe Stichprobe, die zur Normierung eines Tests eingesetzt wird. Die Eichstichprobe besteht idealerweise aus einer hinreichend großen, repräsentativen Zufallsstichprobe der Zielpopulation, für die der Test beim späteren Einsatz Gültigkeit haben soll.

Eichung (Gütekriterium) ► Normierung

Eigenwert Der Eigenwert eines Faktors gibt an, wie viel Varianz von allen Items (Variablen) durch diesen Faktor erklärt wird.

Einparameter-Logistisches-Modell (1PL-Modell)

Das 1PL-Modell der Item-Response-Theorie beschreibt den Zusammenhang zwischen dem beobachtbaren dichotomen Antwortverhalten und dem dahinter stehenden latenten Merkmal auf Grundlage einer Wahrscheinlichkeitsfunktion mit einem Itemparameter, nämlich dem Schwierigkeitsparameter σ_i .

Erschöpfende (suffiziente) Statistiken 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).

Exhaustivität von Antwortalternativen Exhaustivität von Antwortalternativen liegt vor, wenn alle möglichen Antworten auf den vorgegebenen Antwortalternativen abgebildet werden können.

Exploratorische Faktorenanalyse Die exploratorische Faktorenanalyse (EFA) ist 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.

Exposure Control Strategie zur Vermeidung der öffentlichen Bekanntheit von Items durch unerwünscht häufige Vorgabe der Items oder der Itemgruppen. Beim adaptiven Testen kann Exposure Control leichter erzielt werden.

F

Fairness (Gütekriterium) Ein Test erfüllt das Gütekriterium der Fairness, wenn die resultierenden Testwerte zu keiner systematischen Benachteiligung bestimmter Personen aufgrund ihrer Zugehörigkeit zu ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppen führen.

Faktorladung Die Gewichtungszahl λ_{jk} einer beobachteten Variablen j auf dem latenten Faktor k heißt

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

Faktorwert Der Faktorwert f_{kv} gibt an, wie stark ein Faktor F_k bei der v -ten Person ausgeprägt ist.

Fehlervarianz $\text{Var}(\epsilon)$ Die Varianz der Fehlerwerte $\text{Var}(\epsilon)$ der Personen stellt in der KTT den unerklärten Anteil der Testwertvarianz $\text{Var}(x)$ dar.

Freies Antwortformat Bei Aufgaben mit einem freien Antwortformat sind keine Antwortalternativen vorgegeben. Die Antwort wird von der Person selbst formuliert bzw. produziert.

G

Geschwindigkeitstests ▶ Speedtests

Gleichwertige Methoden Im Rahmen von MTMM-Modellen sind gleichwertige Methoden solche Methoden, die das zu erfassende Trait gleichwertig repräsentieren. Z.B. sind parallele Tests oder Testhälften gleichwertige Methoden. Im Unterschied zu austauschbaren Methoden, ist die Erklärung der Methodeneffekte für gleichwertige Methoden nachrangig.

Gütekriterien ▶ Testgütekriterien

H

Hauptachsenanalyse Methode der EFA, mit der versucht wird, das Beziehungsmuster zwischen den manifesten Variablen mit dahinterliegenden latenten Faktoren zu erklären.

Hauptkomponentenanalyse Methode der EFA, mit der versucht wird, möglichst viel Varianz der beobachteten Variablen durch möglichst wenige Faktoren (sog. Hauptkomponenten) zu beschreiben.

Hierarchisch geschachtelte Modelle Mehrere CFA-Modelle 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 durch Parameterrestriktionen bzw. Freisetzung ineinander übergeführt werden können, während die Modellstruktur ansonsten erhalten bleibt.

Homogenität Homogenität von Items liegt vor, wenn die verschiedenen Items eines (Sub-)Tests das selbe Merkmal messen.

I

Informationskriterien Maße zur deskriptiven, relativen Beurteilung der Güte eines Modells. Häufig verwendete Informationskriterien sind der AIC, der BIC und der CAIC.

Inkrementelle Validität Inkrementelle Validität 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.

Interne Konsistenz (Konsistenzanalyse) Methode der Reliabilitätsschätzung. Die Kovarianzen zwischen den Items eines Tests werden als wahre Varianz angesehen und zur Bestimmung der Reliabilität verwendet. Siehe auch: Cronbachs α .

Interpretationsobjektivität (Gütekriterium) Ein Test ist dann interpretationsobjektiv, wenn bezüglich der Interpretation der Testwerte eindeutige Richtlinien (z. B. Normentabellen) vorliegen.

Invertierte Items Invertierte Items sind »umgepolte« Items, bei denen nicht die Bejahung, sondern die Verneinung symptomatisch für eine hohe Merkmalsausprägung ist.

Itemcharakteristische Funktion (IC-Funktion) Die IC-Funktion beschreibt die Beziehung zwischen dem latenten Merkmal und dem Reaktionsverhalten der Probanden auf ein Item in Form einer Wahrscheinlichkeitsaussage.

Itemhomogenität Verschiedene Items sind bezüglich einer latenten Dimension ξ dann homogen, wenn das Antwortverhalten auf die Items nur von diesem Merkmal (der latenten Dimension) und keinem anderen systematisch beeinflusst wird.

Iteminformationsfunktion I_i Die Iteminformationsfunktion I_i gibt im Rahmen der Itemselektion in der Item-Response-Theorie an, wie hoch der Informationsgehalt eines Items i bzgl. der Diskrimination zwischen verschiedenen Merkmalsausprägungen ist. Die Iteminformationsfunktion I_i eines Items i ist für die Ausprägung ξ maximal, die mit dem Itemparameter σ_i übereinstimmt.

Itempool Eine Menge von Items für die mit einem IRT-Modell Itemhomogenität festgestellt wurde; beim adaptiven Testen können beliebige Items aus dem Itempool zur Vorgabe ausgewählt werden.

Item-Response-Theorie (IRT) Die Item-Response-Theorie (IRT, 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 (► IC-Funktion).

Itemschwierigkeit/Schwierigkeitsindex Die Schwierigkeit eines Items (Schwierigkeitsindex) beschreibt in der KTT das mit 100 multiplizierte Verhältnis des durchschnittlich erreichten Itempunktwertes zum maximal möglichen Itempunktwert. Je größer der Schwierigkeitsindex, desto leichter ist das Item.

Itemparameter ► Schwierigkeitsparameter

Itemselektion Die Itemselektion 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 Die Trennschärfe eines Items gibt an, wie stark die mit dem jeweiligen Item erzielte Differenzierung zwischen den Probanden mit der Differenzierung durch den Gesamttests übereinstimmt.

Itemvarianz Die Varianz eines Items ist ein Maß für die Differenzierungsfähigkeit des Items; es gibt an, wie unterschiedlich die Probanden auf das Item antworten.

Itemzwillinge oder Itempaarlinge 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 (Zwillinge, Paarlinge) gruppiert. Jeder Testhälfte wird zufällig jeweils eines der beiden Items zugeordnet.

K

Kognitives Vortesten Beim kognitiven Vortesten 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.

Kommunalität Die Kommunalität h_i^2 einer Variablen i gibt an, in welchem Ausmaß die Varianz der Variablen durch die extrahierten q Faktoren erklärt wird.

Konfidenzintervall Das Konfidenzintervall kennzeichnet denjenigen Bereich um einen empirisch ermittelten individuellen Testwert x_v , in dem sich 95% (99%) aller möglichen wahren Testwerte τ_v befinden, die den Testwert x_v erzeugt haben können.

Konfirmatorische Faktorenanalyse Die konfirmatorische Faktorenanalyse (CFA) ist ein 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. Die CFA bildet das sog. Messmodell der Verfahrensgruppe der Strukturgleichungsmodelle.

Konsistenz Die Konsistenz einer Messvariablen beschreibt in der Latent-State-Trait-Theorie das Ausmaß der durch ein Trait erklärten Varianz relativiert an der Gesamtvarianz der Messvariablen.

Konsistenzeffekte Konsistenzeffekte treten dann auf, wenn Probanden versuchen, »stimmige Antworten« aufgrund ihrer Antworten auf vorangegangene Items zu geben.

Konstrukt Bezeichnung für ein nicht direkt beobachtbares, aber empirisch verankertes latentes Persönlichkeitsmerkmal.

Konstruktäquivalenz Die Konstruktäquivalenz ist die empirisch bestätigte Äquivalenz eines psychologischen Konstrukts über Sprachen und Kulturen hinweg.

Konstruktvalidität Konstruktvalidität liegt vor, wenn ein Test tatsächlich das Konstrukt erfasst, das er erfassen soll.

Konvergente Validität Im Rahmen der Konstruktvalidierung gilt die konvergente Validität als nachgewiesen, wenn Messungen eines Konstrukts, das mit verschiedenen Methoden erfasst wird, hoch miteinander korrelieren.

Kriteriumsorientierte Testwertinterpretation Bei der kriteriumsorientierten Testwertinterpretation erfolgt die Interpretation des Testwertes nicht in Bezug zur Testwertverteilung einer Bezugsgruppe (► Normorientierte Testwertinterpretation), sondern in Bezug auf ein spezifisches inhaltliches Kriterium. Es wird vorab festgelegt, welches Testverhalten zum Erreichen des Kriteriums führt.

Kriteriumsvalidität Kriteriumsvalidität 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 Testwert und möglichen Außenkriterien belegt werden.

L

Latent-Class-Analyse (LCA) Probabilistisches Verfahren zur Kategorisierung von Personen (Objekten) in qualitative latente Klassen.

Latent-State-Trait-Theorie (LST-Theorie) Die Latent State-Trait-Theorie ist eine formale Erweiterung der klassischen Testtheorie, die neben der Aufteilung der Messvariable X_{it} einer Messung i zu Messgelegenheit t in eine Messfehlervariable ε_{it} und in eine Variable der wahren Werte τ_{it} eine Trennung von situationalen und dispositionellen Einflüssen erlaubt. Dazu wird die Variable der wahren Werte τ_{it} einer Messung X_{it} zusätzlich in eine Trait-Variable ξ_{it} und in eine State-Residuum-Variable ζ_{it} zerlegt: $X_{it} = \tau_{it} + \varepsilon_{it} = \xi_{it} + \zeta_{it} + \varepsilon_{it}$

Latent-Class-Modelle Bezeichnung für IRT-Modelle, die davon ausgehen, dass das latente Persönlichkeitsmerkmal zur Charakterisierung von Personenunterschieden aus qualitativen kategorialen latenten Klassen besteht.

Latente 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 Das State-Residuum ist der Teil eines States, der ausschließlich die Situation und die Interaktion zwischen Person und Situation repräsentiert.

Latent-Trait-Modelle/IRT 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.

Leistungstests Tests zur Erfassung der individuellen kognitiven Leistungsfähigkeit in Problemlösesituationen. Beispiele: Intelligenztests, Konzentrationsleistungstests etc.

Likelihood/IRT In der IRT ist die Likelihood das Anpassungskriterium bei der Parameterschätzung. Sie ist dort definiert als die Wahrscheinlichkeit aller

beobachteten Daten unter Annahme der Modellgültigkeit. Bei der Modellschätzung werden die Parameter iterativ so lange verändert, bis die Likelihood maximal ist.

Likelihood/LCA In der 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 Möglichkeit zur inferenzstatistischen Absicherung der Güte eines LCA-Modells: Der LRT wird zur inferenzstatistischen Absicherung des Unterschieds zweier geschachtelter Modelle (»nested models«) verwendet.

Linear-logistische Modelle Linear-logistische Modelle 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.

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 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.

M

Manifeste Variablen Variablen zur Erfassung des beobachtbaren Antwortverhaltens mit verschiedenen Items, die Indikatoren für die latente Dimension darstellen.

Messeffizienz Die Effizienz eines Tests berechnet sich durch den Quotienten aus Messpräzision und

Testlänge, wobei letztere häufig durch die Anzahl präsentierten Items quantifiziert wird.

Messmodell Im Rahmen von Strukturgleichungsmodellen werden die Teilmodelle, in denen die Zuordnungen der beobachteten Variablen zu den Faktoren erfolgt, als Messmodelle bezeichnet.

Messpräzision Grad der Übereinstimmung von wahren Merkmalsausprägungen und den Testwerten. Auf Skalenebene oft durch die mittlere quadratische Abweichung von wahrer und geschätzter Merkmalsausprägung bestimmt.

Methodeneffekte 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.

Methodenspezifitätskoeffizient Der Methodenspezifitätskoeffizient gibt den Anteil an beobachteter Varianz wieder, der auf den Einfluss eines Methodeneffektes zurückzuführen ist. Je höher der Methodenspezifitätskoeffizient ausfällt, desto stärker ist der Einfluss der Messmethode auf die Messung.

Mischverteilungs-Rasch-Modelle (Mixed-Rasch Models) Kombination aus Rasch-Modell und Latent-Class-Analyse. Innerhalb jeder Klasse wird versucht, jeweils ein eigenes Rasch-Modell anzupassen. Zwischen den latenten Klassen unterscheiden sich die Parameter des Rasch-Modells.

Mixed-Rasch-Models ► Mischverteilungs-Rasch-Modelle.

Modelldifferenztest Werden mit der CFA hierarchisch geschachtelte Modelle spezifiziert und gegeneinander getestet, so kann der Unterschied im Modellfit statistisch über die Differenz der χ^2 -Werte beider Modelle überprüft werden, die wiederum χ^2 -verteilt ist.

Modellfit Der bezeichnet in der Statistik ganz allgemein die Güte der Anpassung des Modells an die Daten. Je schlechter die Passung von Modell und Daten, desto »schlechter« ist der zur Beurteilung der Passung gewählte Index (z.B. χ^2 -Wert, BIC etc.).

Multidimensionales adaptives Testen 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.

Multiple Regression Mittels einer multiplen Regression wird eine Kriteriumsvariable anhand mehrerer Prädiktoren vorhergesagt.

Multitrait-Multimethod-Analyse Verfahren zum Nachweis der Konstruktvalidität unter Berücksichtigung einer systematischen Kombination von mehreren Traits und mehreren Messmethoden.

N

Nested Models Unter »nested models« versteht man ► geschachtelte Modelle, die durch Parameterrestriktionen ineinander überführbar sind.

Niveautests ► Powertests

Nomologisches Netz Ein nomologisches Netz stellt ein Beziehungsgeflecht zwischen (latenten) Konstrukten und beobachtbaren Testvariablen dar. Die beiden Ebenen werden mit Axiomen bzw. empirischen Gesetzen beschrieben und durch Korrespondenzregeln miteinander verbunden.

Normaktualisierung Unter Normaktualisierung versteht man eine erneute Testeichung, sobald die empirische Überprüfung der Gültigkeit von Normen ergeben hat, dass sich die Merkmalsverteilung in der Bezugsgruppe seit der vorherigen Testeichung bedeutsam verändert hat.

Normalisierung Bei der Normalisierung wird eine nicht-normalverteilte Testwertvariable zur besseren Interpretierbarkeit so transformiert, dass die Variab-

le danach normalverteilt ist. Die Normalisierung ist von der Normierung zu unterscheiden, die bei der Testeichung vorgenommen wird.

Normdifferenzierung Unter Normdifferenzierung versteht man die Bildung von separaten Normen für die einzelnen Subpopulationen aus der Eichstichprobe hinsichtlich eines mit dem Untersuchungsmerkmal korrelierten Hintergrundfaktors (z.B. separate Normen für Männer und Frauen).

Normierung, auch Testeichung (Gütekriterium) Die Normierung dient dazu, zur normorientierten Testwertinterpretation Vergleichswerte zu gewinnen. Dazu werden Testergebnisse von Personen einer »Eichstichprobe« in Normierungstabellen zusammengestellt.

Normorientierte Testwertinterpretation Die normorientierte Testwertinterpretation besteht darin, dass zu einem individuellen Testwert ein Normwert bestimmt wird, anhand dessen die Testperson bezüglich ihrer individuellen Merkmalsausprägung hinsichtlich ihrer relativen Position innerhalb der Bezugsgruppe eingeordnet wird.

Normwert Ein Normwert (z.B. Prozentrang, z_v -Wert) ermöglicht es, den Testwert x_v einer Testperson hinsichtlich seiner Position in der Testwertverteilung einer bestimmten Bezugsgruppe zu interpretieren.

Nützlichkeit (Gütekriterium) Ein Test ist dann nützlich, wenn die auf seiner Grundlage getroffenen Entscheidungen (Maßnahmen) mehr Nutzen als Schaden erwarten lassen.

O

Objektivität eines Tests (Gütekriterium) Ein Test ist dann objektiv, wenn er dasjenige Merkmal, das er misst, unabhängig von Testleiter, Testauswerter und von der Ergebnisinterpretation misst.

Odd-Even-Methode 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

ungeradzahligen (»odd«) Items, der andere die geradzahligen (»even«) Items des Gesamttests.

Ordnungsaufgabe Aufgabentyp, bei dem die einzelnen Bestandteile der Aufgabe so umgeordnet oder einander zugeordnet werden, dass idealerweise eine logisch passende Ordnung entsteht.

P

Parallele Tests Zwei Tests heißen parallel, wenn sie gleiche wahre Werte und gleiche Fehlervarianzen aufweisen.

Paralleltest-Reliabilität Methode der Reliabilitätschätzung. Die Reliabilität eines Tests, von dem zwei parallele Formen existieren, wird über die Korrelation der Testwerte der beiden parallelen Testformen geschätzt.

Personenparameter Der Personenparameter kennzeichnet in der IRT die Merkmalsausprägung ξ_v einer Person v auf der latenten Variable ξ .

Persönlichkeitsmerkmale Persönlichkeitsmerkmale sind mehr oder weniger zeitlich stabile psychische und physische Eigenschaften von Probanden (z.B. Extraversion, Körpergröße).

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

Perzentil Das Perzentil bezeichnet jenen Testwert x_v , der einem bestimmten Prozentrang in der Normierungsstichprobe entspricht. Beispielsweise wird derjenige Testwert, welcher von 30% der Testwerte unterschritten bzw. höchstens erreicht wird, als 30. Perzentil bezeichnet.

Powertests, auch Niveautests Powertests sind Leistungstests mit eher schwierigen Aufgaben, wobei erhoben wird, welches Schwierigkeitsniveau der Aufgaben der Proband ohne Zeitbegrenzung bewältigen kann.

Probabilistische Modelle 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 IC-Funktion entspricht einer monoton steigenden Funktion.

Projektive Tests Bei projektiven 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.

Prozentrang Ein Prozentrang gibt an, 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 .

Q

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 (vgl. Perzentil).

R

Rasch-Modelle Rasch-Modelle stellen eine Klasse von spezifisch objektiven Modellen in der IRT (Item-Response-Theorie) dar.

Ratekorrektur Die Ratekorrektur zieht bei der Testwertbestimmung jene Anzahl an »richtigen« Lösungen ab, die nur durch Erraten der richtigen Antworten entstanden ist.

Reliabilität (Gütekriterium) Reliabilität bezeichnet die Messgenauigkeit eines Tests. 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.

Reliabilitätskoeffizient (KTT) Konkrete Bezeichnung für die Messgenauigkeit eines Tests (► Reliabilität Gütekriterium). In der KTT wird Reliabilitätskoeffizient als das Verhältnis zwischen true

score-Varianz $\text{Var}(\tau)$ und Testwertvarianz $\text{Var}(x)$ definiert.

Repräsentative Aufgabenstichprobe Eine repräsentative Aufgabenstichprobe stimmt hinsichtlich der Schwierigkeitsverteilung mit der Grundgesamtheit aller merkmalsrelevanten Aufgaben überein und erlaubt somit eine kriteriumsorientierte Testwertinterpretation in Bezug auf die Aufgabeninhalte.

Repräsentativität Eine Stichprobe ist dann repräsentativ, wenn sie hinsichtlich ihrer Zusammensetzung die jeweilige Zielpopulation möglichst genau abbildet.

Retest-Reliabilität 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.

Retrospektive Befragung In der Testentwicklungsphase wird der Proband »zurückblickend« über Schwierigkeiten bei der Beantwortung der einzelnen Items befragt.

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

S

Schwellenwert Im Rahmen kriteriumsorientierter Testwertinterpretation bezeichnet ein Schwellenwert jenen Testwert, ab dem das Kriterium als zutreffend angenommen wird. Schwellenwerte können z.B. mittels ROC-Analyse empirisch bestimmt werden.

Schwierigkeitsparameter/IRT σ_i ist ein Itemparameter, der durch jene Merkmalsausprägung ξ definiert ist, bei der die Lösungswahrscheinlichkeit des Items 50% beträgt.

Sicherung Unter Sicherung versteht man die Pflicht zur Regelung der Verfügbarkeit, Aufbewahrungsdauer und Verwendung von Testdaten (incl. des Testprotokolls und aller schriftlichen Belege) und Schutz der Identität von Probanden.

Skalierung (Gütekriterium) Ein Test erfüllt das Gütekriterium Skalierung, wenn die laut Verrechnungsregel resultierenden Testwerte die empirischen Merkmalsrelationen adäquat abbilden.

Soziale Erwünschtheit, auch soziale Desirabilität Die Soziale Erwünschtheit beinhaltet die Antworttendenz eines Probanden, sich selbst so darzustellen, wie es soziale Normen seiner Wahrnehmung nach erfordern.

Spearman-Brown-Korrektur 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.

Speedtests, auch Geschwindigkeitstests Speedtests sind Leistungstests mit meist einfachen Aufgaben, wobei erhoben wird, wie viele Aufgaben unter Zeitdruck gelöst werden können.

Spezifische Objektivität/FRT Spezifische Objektivität liegt vor, wenn alle IC-Funktionen die gleiche Form aufweisen, d.h. lediglich entlang der ξ -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 ξ untersucht wurden. In Umkehrung kann auch der Fähigkeitsunterschied zweier Personen ($\xi_w - \xi_v$) unabhängig von den verwendeten Items festgestellt werden.

Spezifität Die Spezifität einer Messvariablen beschreibt in der Latent-State-Trait-Theorie das Ausmaß der durch die Situation und die Person-Situation-Interaktion erklärten Varianz relativiert an der Gesamtvarianz der Messvariablen.

Splithalf-(Testhalbierungs-) Reliabilität Methode der Reliabilitätsschätzung. Aus den Items eines Tests

werden zwei möglichst parallele Testhälften gebildet (s. h.: Itemzwillinge, Odd-Even-Methode, Zeitpartitionierungsmethode). Aus der Korrelation der Testwerte der Halbttests wird mittels Spearman-Brown-Korrektur die Reliabilität des Gesamttests geschätzt.

Standardabweichung $SD(x)$ Die Standardabweichung gibt die Streuung der Testwertvariable x um den Mittelwert \bar{x} an. Ist die Testwertvariable normalverteilt, so befinden sich im Bereich $\bar{x} \pm 1 SD(x)$ ca. 68% der Testwerte, im Bereich $\bar{x} \pm 2 SD(x)$ ca. 95% der Testwerte.

Standardmessfehler $SD(\epsilon)$ Der Standardmessfehler $SD(\epsilon)$ eines Tests resultiert aus der Unreliabilität des Tests und errechnet sich als Wurzel aus der Fehlervarianz eines Tests. Dabei gilt: $SD(\epsilon) = SD(x) \cdot \sqrt{1 - Rel}$. Der Standardmessfehler ist bei höherer Reliabilität kleiner und bei niedrigerer Reliabilität größer.

Standardnormen Als Standardnormen werden die z-Norm sowie weitere durch Lineartransformationen gewonnene Normen (z.B. IQ- oder T-Norm) bezeichnet.

State Ein State ist ein zeitlich begrenzter biologischer, emotionaler und kognitiver Zustand, in dem sich eine Person befinden kann. Er kennzeichnet sich durch personenbedingte (d. h. trait-bedingte), situativ bedingte und durch die Interaktion zwischen Person und Situation bedingte Einflüsse.

Stichprobenunabhängigkeit Stichprobenunabhängigkeit bedeutet, dass in Rasch-Modellen die Itemparameter unabhängig von den Personen und die Personenparameter unabhängig von den Items geschätzt werden können.

Strukturell unterschiedliche Methoden Strukturell unterschiedliche Methoden sind solche Methoden, die nicht austauschbar sind, weil sie sich qualitativ von anderen Methoden unterscheiden und kein Zufallsauswahl darstellen. Strukturell unterschiedliche Methoden sind z.B. Selbst- und Fremdbeurteilungen.

T

τ -Äquivalenz Zwei Tests p und q heißen τ -äquivalent, wenn beide den gleichen wahren Wert τ messen.

TBS-TK Das TBS-TK ist ein veröffentlichtes Testbeurteilungssystem des ► Testkuratoriums zur standardisierten Erstellung und Publikation von Testrezensionen anhand eines vorgegebenen Kriterienkataloges.

Teaching to the test Ist ein Phänomen, das die Validität bewertender Interpretationen im Bildungssystem dadurch gefährdet, dass gezielt spezielle Aufgaben geübt werden, um ein besseres Abschneiden der Schülerinnen und Schüler bei den Tests zu gewährleisten.

Tendenz zur Mitte Als Tendenz zur Mitte wird eine Antworttendenz bezeichnet, die extreme Antworten eher vermeidet und mittlere Antwortkategorien eher bevorzugt.

Testadaptation Testadaptation 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 Die Testeichung 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/Gütekriterien Testgütekriterien stellen ein System zur Qualitätsbeurteilung psychologischer Tests dar. Üblicherweise werden folgende 10 Kriterien unterschieden: Objektivität, Reliabilität, Validität, Skalierung, Normierung (Eichung), Testökonomie, Nützlichkeit, Zumutbarkeit, Unverfälschbarkeit und Fairness.

Testitem Zu beantwortende Aufgabe (Frage, Statement etc.) eines Tests.

Testkuratorium Das Testkuratorium ist ein Gremium der Föderation Deutscher Psychologievereinigungen (Deutsche Gesellschaft für Psychologie e.V. und Berufsverband Deutscher Psychologinnen und Psychologen e.V. (DGPs und BDP)), dessen Aufgabe es ist, die Öffentlichkeit vor unzureichenden diagnostischen Verfahren und vor unqualifizierter Anwendung diagnostischer Verfahren zu schützen.

Testökonomie (Gütekriterium) Ein Test erfüllt das Gütekriterium Ökonomie, wenn er, gemessen am diagnostischen Erkenntnisgewinn, relativ wenig Ressourcen wie Zeit, Geld oder andere Formen beansprucht.

Teststandards Teststandards sind vereinheitlichte Leitlinien, in denen sich allgemein anerkannte Zielsetzungen zur Entwicklung, Adaptation, Anwendung und Qualitätsbeurteilung psychologischer Tests widerspiegeln.

Testwert Der Testwert (= Rohwert) x_v ist das individuelle numerische Testresultat und wird aus den registrierten Antworten einer Testperson durch Anwendung definierter Regeln unmittelbar gebildet (vgl. aber ► Personenparameter).

Testwertestreuung $SD(x)$ Die Testwertestreuung 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 $Var(x)$ Die Testwertestreuung $Var(x)$ ist die Varianz der beobachteten Testwerte. Sie setzt sich aus der wahren Varianz $Var(\tau)$ und der Fehlervarianz $Var(\varepsilon)$ zusammen.

Trait Ein Trait ist ein mehr oder weniger zeitlich stabiles Merkmal (Disposition), das personeninhärent und transsituativ überdauernd ist.

Trait-Methoden-Einheit In der Multitrait-Multimethod Analyse wird angenommen, dass in jeder Messung Einflüsse des zu messenden Konstrukts und der verwendeten Messmethode zu finden sind.

Messungen eines Traits repräsentieren somit eine Trait-Methoden-Einheit.

Treffsicherheit Index zur Beurteilung der Güte eines LCA-Modells. Definiert als die durchschnittliche Höhe der maximalen bedingten Klassenzuordnungswahrscheinlichkeit $P^{\max}(g|a_v)$ über alle in der Stichprobe vorkommenden Antwortmuster (N_a) hinweg.

Trennschärfe ▶ Itemtrennschärfe

True score τ_v Der »true score« bzw. wahre Wert τ_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 τ_v nicht völlig überein. Ein Konfidenzintervall für τ_v kann mit Hilfe des ▶ Standardmessfehlers bestimmt werden.

U

Unbedingte Antwortmusterwahrscheinlichkeit $P(a_v)$ Bei der dichotomen LCA: Wahrscheinlichkeit eines Antwortmusters a_v in der Stichprobe.

Unbedingte Itembejahungswahrscheinlichkeit $P(x_{vi}=1)$ Bei der dichotomen LCA: Wahrscheinlichkeit, mit der ein Item i bejaht wird.

Unbedingte Kategorienwahrscheinlichkeit $P(x_{vi}=k)$ Bei der polytomen LCA: Wahrscheinlichkeit, mit der ein Item i mit der Antwortkategorie k beantwortet wird.

Unbedingte Klassenzuordnungswahrscheinlichkeit $P(g)$ Bei der dichotomen LCA: Wahrscheinlichkeit, mit der eine beliebige Person v zur Klasse g gehört (auch: relative Klassengröße π_g).

Unverfälschbarkeit (Gütekriterium) Unverfälschbarkeit eines Tests liegt vor, wenn das Verfahren derart konstruiert ist, dass die zu testende Person durch vorgetäushtes Verhalten (»Faking«) die konkreten Ausprägungen ihrer Testwerte nicht steuern bzw. verzerren kann.

V

Validität (Gütekriterium) Ein Test gilt dann als valide (»gültig«), wenn er das Merkmal, das er messen soll, auch wirklich misst und nicht irgendein anderes. Validität bezeichnet darüber hinaus die Menge der zutreffenden Schlussfolgerungen, die aus einem Testergebnis gezogen werden können.

W

Wahre Varianz Die wahre Varianz $\text{Var}(\tau)$ ist die Varianz der wahren Werte τ_v in einem Test. Sie ist meistens niedriger als die Testwertvarianz $\text{Var}(x)$. Aus dem Verhältnis beider Varianzanteile resultiert in der KTT die Reliabilität.

Z

Zeitpartitionierungsmethode (auch: Zeitfraktionierungsmethode) 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 Die im Rahmen der Testeichung zu definierende Zielpopulation ist diejenige Bezugsgruppe, für welche die zu erstellenden Testnormen gelten sollen und aus der entsprechend die ▶ Normierungsstichprobe zu ziehen ist.

Zumutbarkeit (Gütekriterium) Zumutbarkeit 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.

Zweiparameter-Logistisches-Modell (2PL-Modell) Im Unterschied zum 1PL-Modell wird beim 2PL-Modell ein zusätzlicher Itemparameter λ_i ins Modell aufgenommen, der die Diskriminierungsfähigkeit des Items (ähnlich der Trennschärfe in der KTT) repräsentiert.

z_v -Normwert Der z_v -Normwert gibt an, wie stark der Testwert x_v einer Testperson v vom Mittelwert \bar{x} der Verteilung der Bezugsgruppe in Einheiten der Standardabweichung $SD(x)$ abweicht.

Sachverzeichnis

A

Abbruchkriterien 312
 Adaptation 200
 ad-hoc-Stichprobe 188
 AIC (Akaike's Information Criterion) 293
 AID, Adaptives Intelligenzdiagnostikum 20
 AID 2, Adaptives Intelligenzdiagnostikum 2 20, 231, 232, 237
 Akaike's Information Criterion (AIC) 293
 Akquieszenz 61, 296
 Aktualisierungseffekte 67
 Algorithmus, adaptiver 262
 α -Fehler 150
 Alternativhypothese 150
 American Educational Research Association 194
 American Psychological Association 137, 194
 AMT, Adaptiver Matrizen Test 71
 Analogskala 51
 Anker-Items 240
 Ansatz
 – experimenteller 148
 – korrelativer 150
 Antwortformat 39, 140
 – atypisches 55
 – freies 40
 – gebundenes 43
 Antwortmodus, dichotomer 221
 Antwortmuster 233, 281
 Antwortskala, Polarität 51
 Antworttendenzen 60, 158
 APM, Advanced Progressive Matrices 25
 Attenuationskorrektur 151
 ARES, Action Regulating Emotion Systems 152
 Aufgaben, dichotome 48
 Aufgabenmerkmale 247

Aufgabentyp 38
 Auftraggeber 196
 Auftragnehmer 196
 Augenscheinvalidität 14
 Auswahlaufgaben 45
 Auswertung
 – deskriptive 150
 – inferenzstatistische 150
 Auswertungsobjektivität 9
 Autokorrelation 117
 Autokorrelationseffekt 365
 Axiome 100, 145

B

Basisparameter 253
 Bayesian Information Criterion (BIC) 293
 Bayes-Theorem 286
 Befragung, retrospektive 70
 Bereich
 – der Beobachtung 145
 – der Theorie 145
 β -Fehler 150
 Beurteilungsaufgaben 50
 Bias 327
 BIC (Bayesian Information Criterion) 293
 Birnbaum-Modell (2PL-Modell) 237, 281
 Blinder Empirismus 147
 Bootstrap-Verfahren 292

C

CAIC (Corrected Information Criterion) 293
 CAT-ASVAB, Computerized Adaptive Testing Version of the Armed Services Vocational Aptitude Battery 263

CDI, Child Depression Inventory 379
 CFA (confirmatory factor analysis), Konfirmatorische Faktorenanalyse 316
 CFT 3, Grundintelligenztest Skala 3 24
 Cognitive lab 153
 Conditional Maximum-Likelihood-Methode (CML-Methode) 230
 Consequential validity 158
 Corrected Information Criterion (CAIC) 293
 Cronbachs α 124, 125, 127
 CTCM-Modell 336
 CTCU-Modell 337
 CTUM-Modell 336
 Culture fair tests 24, 160
 Curriculare Validität 142
 Cutoff-Punkt 280

D

Daten, psychometrische Aufbereitung 74
 Datenmatrix 74
 Debriefing 70
 Determinationskoeffizient 107
 Deterministische Modelle 222
 Differenzierungsfähigkeit 79
 Dimensionalität 28, 154
 Dimensionen, latente 216
 DIN 33430 186, 195
 Disjunktheit 46
 Diskriminationsindex 294
 Diskriminationsparameter λ_i 237
 Distraktoren 45
 Dreiparameter logistisches Modell (3PL-Modell, Ratemodell von Birnbaum) 223, 238
 Durchführungsfairness 24
 Durchführungsobjektivität 9

E

EFA 308
 Effekte, situationsspezifische 118
 Eichstichprobe, Normierungsstichprobe 33, 168ff
 Eichung 19
 Eigenwert 311
 Eigenwerteverlauf 312
 Eignungsbeurteilungen 196
 Eindimensionalität 127
 Einfachstruktur 314
 Einparameter logistisches Modell (1PL-Modell, Rasch-Modell) 223, 224
 Eisbrecheritems 266
 EPQ, Eysenck Personality Questionnaire 143, 152
 Ereignisse, unabhängige 217
 Ergänzungsaufgaben 42
 Erinnerungseffekte 118
 Erinnerungseinflüsse 102
 Erprobung, Testversion 69
 Erwartungswert 103
 Erwünschtheit, soziale 23, 59
 Evaluation, Items 73
 Evidence-centered assessment design 143
 Exhaustivität 46
 Exposure control 269
 Exzess 91

F

FAIR, Frankfurter Aufmerksamkeitsinventar 9, 30, 48, 189
 Fairness 23
 Faking bad 305
 Faking good 305
 FAKT, Frankfurter Adaptiver Konzentrationsleistungs-Test 271
 FAKT-II, Frankfurter Adaptiver Konzentrationsleistungs-Test II 17, 32, 169, 237, 271, 274, 275

Faktorenanalyse 282
 – exploratorische 16, 154, 308
 – konfirmatorische 17, 121, 154, 316, 335
 Faktorenrotation 314
 Faktorladung 310
 Faktorwert 310
 Fehlervariablen, paarweise unabhängige 101
 Fehlervarianz 104, 115
 – systematische 159
 – Verdoppelung der 108
 Fehlervarianz $\text{Var}(\varepsilon)$ 102
 Fehlervarianzanteil 104
 Finger-Labyrinth-Test 32
 Fit-Maße 319
 Fixierungsrestriktionen 297
 Flächentransformation 93
 Flynn-Effekt 190
 Forced choice 49
 Forschungsansatz, hypothetisch-deduktiver 145
 FPI, Freiburger Persönlichkeitsinventar 38, 60, 220
 Funktion, logistische 223
 Funktionen, monoton steigende 223

G

Geltungsbereich 32
 Genauigkeitsmaß 132
 Geschlechtsrollenelbstkonzept 281, 282
 Geschwindigkeitstests (Speedtests) 30
 Gleichheitsrestriktionen 297
 Grenzwerte 150
 Guttman-Modell 222

H

Halbttestkorrelation 122
 Hauptachsenanalyse 310

Hauptkomponentenanalyse 309
 HAWIE-R, Hamburg-Wechsler-Intelligenztest für Erwachsene – Revision 10, 17, 66
 HAWIK, Hamburg-Wechsler-Intelligenztest für Kinder 37, 44, 67
 Heterogenität 33
 – der Items 126
 Hitrate 293
 Homogenität 33, 111, 353

I

Indikatoren 220, 314ff, 330, 336ff, 357, 364ff
 Individualdiagnostik 129
 Induktion, experimentelle 149
 Informationsfunktion 234
 Informationskriterien 292, 295
 Inhaltsvalidität 14, 137, 140, 197
 Instruktionen 68
 International Test Commission 195
 Internationale Richtlinien für die Testanwendung 202
 Interne Konsistenz 13, 124
 Interpretation
 – extrapolierende 144, 154
 – Testwerte (IRT) 240
 – verallgemeinernde, von Testergebnissen 141
 Interpretationsobjektivität 10
 Interquartilabstand 91
 IRT, Item-Response-Theorie 183, 216
 IRT-Modell 17, 199, 220
 I-S-T 70, Intelligenz-Struktur-Test 70 254
 I-S-T 2000 R, Intelligenz-Struktur-Test 2000 R 38, 46, 129
 ISTA, Instrument zur Stressbezogenen Tätigkeitsanalyse 127
 Itemauswahl, adaptive 266
 Itembejahungswahrscheinlichkeit 284
 Itemcharakteristiken, klassenspezifische 253

Itemcharakteristische Funktion (IC-Funktion) 198, 221
 Itemcharakteristische Kurve (ICC, item characteristic curve) 269
 Item-fit-indices 233
 Itemformulierung 62
 Itemgenerierung 198
 Itemhomogenität 83, 217, 303
 Iteminformationsfunktion I_i 235, 236, 269
 Iteminhalt 139
 – Post-Hoc-Analyse 246
 Itemlösungswahrscheinlichkeit 81
 Itempolung 61
 Itempool 264
 Item-Response-Theorie (IRT) 112, 154, 183, 216, 281
 Items
 – dichotome 83
 – Heterogenität 126
 – invertierte 78
 – kalibrierte 231, 236
 – unabhängige 101
 Itemschwierigkeit 74, 154, 219
 Itemselektion 85
 Item-Testwert-Korrelation r_{it} 83
 Itemtrennschärfe 111
 Itemtyp 38
 Itemvarianz 79
 Itemzwillinge 123

J

joint scale 225, 265

K

Kaiser-Kriterium 312
 Kategoriencharakteristiken 252
 Kategorienwahrscheinlichkeit, bedingte 301
 Kategorisierung von Itemarten 62
 Klassenzuordnung, Genauigkeit 293

Klassenzuordnungswahrscheinlichkeit, bedingte 281, 283, 286
 Klassische Testtheorie (KTT) 100, 151, 216
 – Erweiterung der 346
 Kognitives Vortesten 70
 Kommunalität 311
 Kompetenzniveaus 245
 Konfidenzintervall 109
 Konkordanzkoeffizient 10
 Konsistenz 344, 349
 – interne 107, 124, 136, 198
 Konsistenzanalyse 13, 124
 Konsistenzeffekte 68
 Konsistenzkoeffizient 349
 Konstrukt
 – latentes 148
 – theoretisches 137, 139
 Konstruktäquivalenz 200
 Konstruktionsstrategien 36
 Konstruktvalidierung 308
 – schwacher Ansatz 147
 – starker Ansatz 147
 Konstruktvalidität 14, 137, 145, 197, 326
 Konvergenz 289
 Koordinatensystem 152
 Korrelation 83, 101, 104ff, 115ff, 131, 150
 – punktbiseriale 83
 Korrelationskoeffizient 107, 150
 Korrespondenzregeln 145
 Kriterien, externe 155, 174
 Kriterium 137
 Kriteriumsvalidität 14, 18, 137, 197
 KTT, Klassische Testtheorie 100, 151, 216
 Kurzaufsatzaufgaben 40
 Kurztest 151

L

Latent-Class-Analyse (LCA) 220, 281
 – exploratorische 295
 – konfirmatorische 297

– polytome Antwortformate 300
 Latent-Class-Modelle 220
 Latent-Curve-Modell 386
 Latent-State-Trait-Theorie (LST-Theorie) 343ff
 Latent-Trait-Modelle
 – dichotome 221
 – polytome 251
 Layout 68
 Leistungstests 29, 75, 136
 Lernziele 181
 Likelihood (L) 289, 293
 – standardisierte 290
 Likelihood-Funktion 227, 288
 Likelihood-Quotienten-Test 232
 Likelihood-Ratio (LR) 290
 Linear logistic model with relaxed assumptions (LLRA) 254
 Lizenzprüfung nach DIN 33430 205
 Logarithmierung 93
 Lokale stochastische Unabhängigkeit 217, 264, 284
 LST-Modell (Latent-State-Trait-Modell), hierarchisches 372
 LST-Theorie (Latent-State-Trait-Theorie) 343ff
 Lückentext 42

M

MAS, Manifest Anxiety Scale 37
 Matrix-Sampling 240
 Maximum Likelihood 289
 Maximum-Likelihood-Methode 319
 Median 170
 Mediansplit 280
 Mehrfachwahlaufgaben 49
 Merkmal, theoretisch begründete Definition 139
 Merkmalsarten 28
 Messeffizienz 272
 Messfehler 115, 150
 Messgenauigkeit 114, 355
 – eines Tests 106

Messmodell 264
 Messpräzision 272
 Methode des lauten Denkens 153
 Methodeneffekte 327, 363
 Methodenspezifitätskoeffizient 375
 Methodenvarianz 127
 Minderungskorrektur, doppelte 151
 Mindestkorrelation 151
 Mischverteilungsmodelle 253, 300
 Mischverteilungs-Rasch-Modelle 302
 Mitwirkende 196
 Mixed-Rasch-Modelle 252, 304
 MMPI, Minnesota Multiphasic Personality Inventory 23, 38
 MMPI 2, Minnesota Multiphasic Personality Inventory 2 31, 38, 60
 Modelldifferenztest 319, 357
 Modelle
 – deterministische 222
 – hierarchisch geschachtelte 316
 – linear-logistische 253
 – probabilistische 223
 Modellgüte 282, 288, 295
 Modellidentifikation 318
 Modellkonformität 231
 Modellkontrollen, empirische 231
 Modelltest, graphischer 232
 Modellvergleich 299
 Motivation zur Testbearbeitung 274
 Multiconstruct-LST-Modell 368, 371, 372
 Multimethod-LST-Modell 368
 Multioccasion-Correlated States-Correlated Method minus 1-Modell 369
 Multioccasion-MTMM-Modell 368, 369
 Multiple-Choice-Aufgabe 49
 Multiplikationstheorem für unabhängige Ereignisse 217
 Multistate-Modell 350

Multitrait-Multiinformant-Analyse 328
 Multitrait-Multimethod-Analyse 17, 326, 328
 Multitrait-Multioccasion-Analyse 328
 Multitrait-Multistate-Modell 353

N

National Council on Measurement in Education 195
 NEO-FFI, NEO-Fünf-Faktoren-Inventar 31, 129, 304
 NEO-PI, NEO-Personality Inventory 23, 71
 NEO-PI-R, NEO-Personality Inventory – Revision 31
 Nested models 299
 Netz, nomologisches 145
 Netzwerk, nomologisches 16
 Niveautests (Powertests) 77
 Normaktualisierung 190
 Normalisierung 93, 173
 Normdifferenzierung 184
 Normen 199
 – Überprüfung der Gültigkeit 190
 – verteilungsunabhängige 170
 Normierung 19
 Normierungsstichprobe, Eichstichprobe 33, 186
 – Umfang 189
 Normierungstechniken 19
 Normtabellen 33, 169
 Normwert 167
 Nullhypothese 150
 Nützlichkeit 22

O

OA-TB 75, Objektive Testbatterie 31
 Objektivität 8, 31, 130, 136
 – spezifische 234

Odd-Even-Methode 123
 OLMT, Objektiver Leistungs-motivations-Test 31
 Ökonomie 20
 Optimizing-Satisficing-Modell 57
 Ordnungsaufgaben 43
 Ordnungsrelation 298
 Organismusfaktor 148
 Overadjustment 185

P

Parallelanalyse 313
 Parallelförmigkeit 119
 Parallelität 120
 Paralleltest-Reliabilität 12, 106, 119
 Parameter, Separierbarkeit 230
 Parameterschätzung 226, 319
 Parsimonitätsprinzip 292
 Partial-Credit-Modell 252
 PERLE, Persönlichkeits- und Lernentwicklung an sächsischen Grundschulen Subtest »Zahlen Lesen« 244
 Person in einer Situation 347
 Personenparameter ξ_v 240
 Personenparameterschätzung 270
 Personenselektion 233
 Person-fit-indices 233
 Persönlichkeitstests 30
 Person-Situation-Interaktion 367
 Perzentil 170
 PFT-K, Rosenzweig Picture Frustration Test für Kinder 32, 41
 Pick any out of n 50
 Polarität der Antwortskala 51
 Populationsabhängigkeit 131
 Post-Hoc-Analyse von Iteminhalten 246
 Powertests (Niveautests) 30
 Prozentrang 168
 Prozentrangnorm 168

Q

Qualitätsanforderungen 7
 Qualitätsbeurteilung 194
 Quartile 170
 Quotenstichproben 188

R

Range 91
 Rasch-Homogenität 232
 Rasch-Modell (1PL-Modell) 224, 281, 300
 – dichotomes 224
 – polytomes 251
 Ratekorrektur 77, 88
 Rate-Modell von Birnbaum (3PL-Modell) 238
 Rate-Parameter ρ_i 238
 Ratingskala 51
 Ratingskalenmodell 251
 RCMAS, Revised Children's Manifest Anxiety Scale 379
 Receiver-Operating-Characteristics-Analyse (ROC-Analyse) 175
 Regression, multiple 158
 Reihenfolge, Items 67
 Reliabilität 11, 102, 105, 113, 114, 136, 197, 349
 – Bestimmung 113
 – Herleitung 106
 Reliabilitätskoeffizient 11, 349
 – Höhe 129
 Repräsentationsschluss 141
 Repräsentativität 187
 – globale 187
 – spezifische 187
 Residuum 346
 Response sets 60, 303
 Restriktionen 294, 297
 Retest-Intervall 119
 Retest-Reliabilität 12, 106, 116
 Revision, Itemselektion 85
 Rohrschachtest 32

ROC-Analyse (Receiver-Operating-Characteristics-Analyse) 175
 ROC-Kurve (Receiver-Operating-Characteristics-Kurve) 178
 Rohwert 103, 166
 Rotationsverfahren 314

S

Schiefe 91
 Schlüsselwort 42
 Schlüsselwortergänzungsaufgaben 42
 Schwellen 251
 Schwellenwerte 174
 Schwierigkeitsanalyse 74
 Schwierigkeitsindex 75
 Schwierigkeitsparameter σ_i 221
 Screening 130
 Screeningverfahren 35
 Screeplot 312
 Scree-Test 312
 Separierbarkeit der Parameter 230
 Shadow Tests 270
 Signifikanztest 151
 Singletrait-Multistate-Modell 351
 Situational judgement 62
 Situationseffekt 367
 Skalendignität 217
 Skalenniveau 111
 – der latenten Persönlichkeitsmerkmale 217
 Skalierung 18, 111
 Skalogramm-Modell 222
 Social desirability 59
 Sokrates-Effekt 386
 SON, Non-verbaler Intelligenztest 2½-7 30
 Soziale Erwünschtheit 23, 59
 Spaltensummen 228
 Spearman-Brown-Formel 108
 Spearman-Brown-Korrektur 122
 Speedtests (Geschwindigkeitstests) 30
 Spezifische Objektivität 234
 Spezifität 176, 349
 Spezifitätskoeffizient 349
 Split-Half-Reliabilität 107, 108, 122
 Stabilität 344
 STAI, State Trait Anxiety Inventory/State-Trait-Angst-Inventar 345, 355
 Standard
 – kriteriumsorientierter 184
 – normorientierter 184
 Standardabweichung 109
 Standardisierte Likelihood 290
 Standardmessfehler $SD(\epsilon)$ 104, 109
 Standardmethode 363
 Standardnorm 172
 Standardnormalverteilung 109
 Standards for Educational and Psychological Testing 195
 Standards
 – Itemgenerierung 198
 – Normen 199
 – psychologisches Testen 194
 – Qualitätsbeurteilung 209
 – Reliabilität 197
 – Testauswertung 207
 – Testdurchführung 2043
 – Übersetzung 200
 – Validität 197
 Standardwert 172
 Stanine-Norm 172
 Startwerte 289
 State-Residuum-Variable, latente 348
 States 29, 344
 State-Variable, latente 348
 Statistiken, erschöpfende (suffiziente) 230
 STAXI, State-Trait-Ärgerausdrucksinventars 29, 305
 Stichprobe
 – anfallende 188
 – geschichtete 188
 Stichprobenunabhängigkeit 217, 230
 Strategien
 – maßgeschneiderte 268
 – mehrstufige adaptive 267
 – zweistufige adaptive 267

Strukturen, autoregressive 386
 Strukturgleichungsmodelle 150
 Subtest 38
 Summennormierung 230
 Simpson-Hetter-Methode 270

T

TAI-G, Test Anxiety Inventory/
 Prüfungsängstlichkeitsinventar
 355
 τ -Äquivalenz 105
 – essentielle 125
 Taylor-Russell-Tafeln 130
 TBS-TK (Testbeurteilungssystem
 des Testkuratoriums) 209
 Teacher Report Index of Anxiety
 379
 Teacher Report Index of Depression
 379
 Tendenz zur Mitte 60
 Test Adaptation Guidelines 200
 Test für mathematisches Fach-
 wissen von Mathematiklehrern
 240
 Test zur Messung des räumlichen
 Vorstellungsvermögens 254
 Testadaptation 194
 Testadministration 35
 – computerbasierte 271
 Testanwendung 194, 204
 Testarten 29
 Testaufbau 34
 Testauswertung 207
 Testbatterien 34
 Testbearbeitung, Motivation 274
 Testbeurteilung 209
 Testbeurteilungssystem des Test-
 kuratoriums (TBS-TK) 209
 Testeichung 186
 Testen, adaptives 236, 262
 – Auswirkungen 271
 – multidimensionales 275
 Testentwicklung 198
 Testergebnisse
 – erklärende Interpretation 142

– Validität von Interpretationen
 136
 – verallgemeinernde Interpretation
 141
 Testfairness 159, 207
 Testgütekriterien 8
 Testhalbierung 123
 Testhalbierungs-Reliabilität 12, 122
 Testhälften 124
 Testinformation I_v 236
 Testinhalte 140
 Testitem 100
 Testkonstruktion 194
 Testkuratorium der Föderation
 Deutscher Psychologiever-
 einigungen 195
 Testlänge 34
 – Zusammenhang mit Reliabilität
 108
 Testlets 38
 Testmanual 9, 199
 Testmodell, linear-logistisches
 (LLTM) 254
 Testökonomie 20
 Testplanung 28
 Testrezensionen 209
 Testroutine 24
 Tests
 – adaptive 32
 – apparative 32
 – Messgenauigkeit 106
 – parallele 106, 108
 – Unreliabilität 109
 Teststandards 194
 Test-Test-Korrelation r_{tt} 106
 Testversion, vorläufige 67
 Testwert 82, 103, 136, 166
 – Uneindeutigkeit 167
 – wahrer 102
 Testwertermittlung, Testwert-
 bildung 86, 166
 – einfache 87
 – Leistungstests 86
 – Persönlichkeitstests 89
 Testwertinterpretation 167
 – kriteriumsorientierte 167, 183
 – normorientierte 167, 183
 – Validität 144

Testwerttransformation 93
 – lineare 171
 – nicht-lineare 168
 Testwertvarianz 91, 104
 Testwertverteilung 90
 Testzeit 35
 TIMSS/III, Teiluntersuchung
 der Dritten Internationalen
 Mathematik- und Naturwissen-
 schaftsstudie 246, 247, 248
 TMS, Test für medizinische Studien-
 gänge 22
 Traits 29, 344
 Trait-Methoden-Einheit 369
 Trait-Variable 363
 – latente 348
 Treffsicherheit 293
 Trennschärfe 82, 294
 – Berechnung 82
 – bei dichotomen Items 83
 – Interpretation 83
 Trennschärfeanalyse 82
 True score 100, 102
 Typenkonzepte 220

U

Übereinstimmungsvalidität 18,
 157
 Überprüfung der Gültigkeit von
 Normen 190
 Übungseffekte 117
 Umordnungsaufgaben 44
 Unabhängigkeit, lokale stochas-
 tische 217, 264, 284
 Unverfälschbarkeit 23
 UTCM-Modell 336
 UTCU-Modell 337
 UTUM-Modell 336

V

Validierung 136, 155
 Validierungsmethode, Wahl 160

Validität 13, 136, 197, 273, 327, 362
 – curriculare 142
 – diskriminante 150, 273, 326, 362
 – diskriminante bzw. divergente 17
 – faktorielle 154
 – inkrementelle 158
 – von Interpretationen von Testergebnissen 136
 – konvergente 17, 150, 273, 326, 362
 – konvergente und diskriminante, längsschnittliche Analyse 367
 – prognostische 157
 Variable
 – abhängige 149
 – kontinuierliche 149
 – manifeste 150, 216
 – unabhängige 148
 Varianz
 – erklärte 158
 – Vervierfachung 108
 – wahre 104, 115
 Varianzanteil, wahrer 104
 Varianzeinschränkung 131
 Varianzkomponenten

– situativ bedingte 372
 – stabile 372
 Variation, experimentelle 148
 Verfahren, projektive 32
 Vergleiche, spezifische Objektivität 234
 Vergleichsmaßstab 167
 Verhaltenskodierung 70
 Versuchsplan, experimenteller 149
 Verteilungseigenschaften 189
 Vorhersagevalidität 18

W

Wert, wahrer 115, 151
 WIE-III, Wechsler-Intelligenztest für Erwachsene 30, 35, 36, 37

Y

Youden-Index 178
 YSR/11-18, Youth Self Report 201

Z

Zeilensummen 228
 Zeitpartitionierungsmethode, Zeitfraktionierungsmethode 123
 Zielgruppe 33
 Zielpopulation 187
 Zufallsstichprobe 187
 Zumutbarkeit 22
 Zuordnungsaufgaben 43
 Zusammenhangsstruktur 138
 Zustimmungstendenz 61
 z_v -Normwert 171
 z_v -Transformation von Testwerten, lineare 171
 ZVT, Zahlen-Verbindungs-Test 55, 56
 Zweiparameter logistisches Modell (2PL-Modell, Birnbaum-Modell) 223, 237