

Glossar

Adaptives Testen. Vorgabe von Items in Leistungstests, deren Erstellung mittels Item-Response-Theorien oder dem Rasch-Modell erfolgt. Dem Diagnostikand wird zunächst eine Aufgabe mit mittlerer Schwierigkeit vorgegeben. Löst er sie, erhält er anschließend eine schwierigere Aufgabe; löst er sie nicht, folgt eine leichtere Aufgabe. Daraus ergibt sich ein funktionaler Zusammenhang zwischen der Wahrscheinlichkeit, ein Item richtig zu lösen und der wahren Merkmalsausprägung des Diagnostikanden. Adaptives Testen erfolgt meist computergestützt und hat den Vorteil der Itemreduktion bei zugleich hoher Messpräzision.

Additive Verknüpfung. bezieht sich auf die Frage, wie Einzelinformationen miteinander verknüpft werden. Bei der additiven Verknüpfung wird ein Grenzwert als Summe aller Prädiktoren festgelegt, der geeignete von ungeeigneten Probanden trennt. Eine wichtige Eigenschaft der additiven Verknüpfung ist die vollständige, wechselseitige Kompensierbarkeit der Variablen.
Siehe auch disjunktive und konjunktive Verknüpfung.

Akkommodationseffekt. Effekt, der den Kontexteffekten zugeordnet wird und bei dem ein diagnostisches Instrument dem diagnostischen Phänomen angepasst wird. Es kommt beispielsweise vor, dass Lehrer sich an die hohen Leistungen einer guten Klasse gewöhnen und daraufhin schwerere Klausuren vorgeben. Im Extremfall liegt dann der Notenspiegel einer leistungsschwachen Klasse über dem einer leistungsstarken Klasse.

Akquieszenz. Auch Ja-Sage-Tendenz genannt. Damit ist die Tendenz von Befragten gemeint, einer Frage oder Aussage unabhängig von ihrem Inhalt zuzustimmen. Beispielsweise bejahen Probanden zwei sich inhaltlich widersprechenden Items wie »Ich glaube an die große Liebe« und »Ich glaube nicht an die große Liebe«.
Akquieszenz zählt zu den Antworttendenzen. Es gibt motivationale und kognitive Ansätze zur Erklärung des Phänomens.

Akzeptanz. Akzeptanz ist eines der Nebengütekriterien und bezieht sich auf das Ausmaß, in dem ein diagnostisches Verfahren von Diagnostikanden akzeptiert wird. Die Akzeptanz eines diagnostischen Instruments hängt unter anderem von der Transparenz der Daten und der Nähe zur Fragestellung ab.

Anamnese. Die Erhebung der Krankheitsvorgeschichte zu Beginn einer psychotherapeutischen Behandlung mit dem Ziel der systematischen Sammlung von relevanten Daten für die Diagnosestellung.

Änderungssensitivität. Auch Messgelegenheitsspezifität genannt. Änderungssensitivität bezeichnet die Eigenschaft eines Messinstruments, wahre Veränderungen eines psychologischen Konstrukts abzubilden. Sie bezeichnet denjenigen Anteil, der durch systematische, aber veränderliche und messgelegenheitsspezifische Einflüsse bedingt wird. Es handelt sich um eine kontinuierliche Variable.

Ambulatory Assessment. Eine Methode der Einzelfalldiagnostik, bei der mittels eines mobilen Kleincomputers (Handheld-Computer), den die Diagnostikanden bei sich tragen, Daten im Alltag erfasst werden. Dies hat den Vorteil, dass die Angaben der Diagnostikanden nicht durch die künstliche Atmosphäre einer Laboruntersuchung beeinflusst werden.

Assessment Center. Assessment Center sind Gruppenauswahlverfahren, die überwiegend in der Personalauswahl eingesetzt werden. Es werden verschiedene Methoden zur Erfassung der Eignung der Bewerber eingesetzt, deren Beurteilung durch geschulte Beobachter erfolgt. Gängige Metho-

den, die zum Einsatz kommen, sind beispielsweise Einstellungsgespräche, Kurzvorträge, Leistungstests und Rollenspiele.

Auswertungsobjektivität. *Siehe Objektivität.*

Cronbachs Alpha. *Siehe Interne Konsistenz.*

Differential Item Functioning. Begriff, der den probabilistischen Testtheorien entstammt und den Sachverhalt beschreibt, dass sich die Item-Charakteristischen Kurven von verschiedenen Stichproben (wie z.B. Männer/Frauen, Kinder mit/ohne Migrationshintergrund) unterscheiden können und somit in diesen Gruppen verschiedene Merkmale messen (Gruppenabhängigkeit).

Dimensionales Modell. In dimensionalen Modellen werden Eigenschaften als kontinuierlich aufgefasst. Unterschiede zwischen Menschen können so differenzierter beschrieben werden. Eine Person ist beispielsweise mehr oder weniger extravertiert. Um Merkmale im Sinne quantitativer Modelle zu erfassen, können alle statistischen Verfahren eingesetzt werden, die mindestens Intervallskalenniveau voraussetzen.

Siehe auch Klassifikationsmodell.

Direkte Veränderungsmessung. Bei der direkten Veränderungsmessung findet die diagnostische Erhebung nur zu einem Zeitpunkt statt und Veränderungen werden von den Diagnostikenden rückblickend eingeschätzt. Direkte Veränderungsmessung wird beispielsweise dann eingesetzt, wenn wiederholte Messungen nicht möglich sind oder die subjektive Veränderungseinschätzung der Diagnostikenden von Interesse ist.

Siehe auch indirekte Veränderungsmessung.

Disjunktive Verknüpfung. Bezieht sich auf die Frage, wie Einzelinformationen miteinander verknüpft werden. Unterschieden wird zwischen stark und schwach disjunktiver Verknüpfung. Bei der stark disjunktiven Verknüpfung muss nur bei einem Prädiktor der Grenzwert überschritten werden. Die Ausprägung des anderen Prädiktors ist dann unerheblich. Bei der schwach disjunktiven Verknüpfung können sich die beiden Prädiktoren innerhalb eines Wertintervalls kompensieren.

Siehe auch additive und konjunktive Verknüpfung.

Diskriminante Validität. Auch divergente Validität genannt. Die diskriminante Validität ist ein Teilaspekt der Konstrukt-Validität und lässt sich mit der Multitrait-Multimethod-Methode erfassen (quantifizieren). Sie gilt als nachgewiesen, wenn die Ergebnisse von Instrumenten, die verschiedene Merkmale messen, nicht oder nur gering miteinander korrelieren.

Siehe auch Validität.

Divergente Validität. *Siehe diskriminante Validität.*

DSM-5. Das Diagnostic and Statistical Manual of Mental Disorders ist ein von der American Psychiatric Association (APA) vorgeschlagenes Klassifikationssystem, das vor allem im wissenschaftlichen Kontext eingesetzt wird. Die fünfte Version des DSM wurde im Mai 2013 veröffentlicht. Das DSM-5 ist hierarchisch aufgebaut (5 Achsen) und kombiniert sowohl kategoriale als auch dimensionale Beschreibungen.

Durchführungsobjektivität. *Siehe Objektivität.*

Eichstichprobe. Auch Normierungsstichprobe genannt. Bezeichnet eine Stichprobe, die zur Normierung eines Tests eingesetzt wird und die repräsentativ für das zu messende Merkmal sein muss. Dies ist nur durch die Ziehung einer hinreichend großen Zufallsstichprobe zu gewährleisten.

Eignungsquote. Die Eignungsquote beschreibt den Teil einer Population, der als geeignet eingestuft wird. Sie wird auch Basisrate oder Grundrate genannt.

Einstufige Diagnostik. Bei der einstufigen Diagnostik werden bei allen Diagnostikanden alle relevanten Merkmale gemessen. Die gewünschte Information wird i.d.R. zu einem Zeitpunkt erhoben und alle Diagnostikanden durchlaufen die gleiche Prozedur.

Externe Validität. Bezieht sich auf die Frage, ob die im Laborexperiment gefundenen Effekte sich in der Anwendungsrealität replizieren lassen (Generalisierbarkeit).

Siehe auch Validität.

Fairness. Fairness wird den Nebengütekriterien zugeordnet, ist aber streng genommen ein spezieller Aspekt von Validität. Ein diagnostisches Verfahren ist in dem Maße fair, in dem die ermittelten Messwerte durch die Gruppenzugehörigkeit von Diagnostikanden *nicht* in irrelevanter Weise beeinflusst werden, es also durch den Test nicht zu einer systematischen Benachteiligung der Diagnostikanden aufgrund soziokultureller oder geschlechterspezifischer Zugehörigkeit kommt.

Gitteranalyse. *Siehe Kontrastanalyse*

Grenzwert. Auch Cut-off-Wert genannt. Ein (meist subjektiv) festgelegter Wert, der Personen hinsichtlich eines bestimmten Merkmals in zwei Gruppen teilt.

Grundratenproblematik. Die Grundratenproblematik besteht darin, dass die verschiedenen Diagnosekoeffizienten auf die Veränderung von Eignungsquote und Selektionsquote unterschiedlich reagieren. Besonders relevant ist dieser Sachverhalt im Zusammenhang mit Krankheiten und der Bewertung von Risikofaktoren. Sie resultiert aus der Tatsache, dass die Prävalenz einer Krankheit (Erkrankungsquote) deutlich niedriger oder deutlich höher sein kann als die Prävalenz eines Risikofaktors (Risikoquote).

Halo-Effekt. Bezeichnet die Ausstrahlung eines Merkmals auf andere Merkmale, basierend auf einer subjektiven Zusammenhangsvermutung, die objektiv nicht gerechtfertigt ist (illusionäre Korrelation). Beispielsweise schließen Menschen oftmals von Attraktivität auf Intelligenz.

Homogenität. Homogenität von Items liegt vor, wenn die Items eines Tests alle dasselbe Konstrukt erfassen.

ICD-10. Die International Classification of Diseases ist ein hierarchisch aufgebautes Klassifikationssystem, das von der Weltgesundheitsorganisation (World Health Organization) vorgeschlagen wurde. Neben sämtlichen körperlichen Erkrankungen werden auch psychische Störungen aufgeführt (Kapitel V).

Indirekte Veränderungsmessung. Auch indirekte Prozessdiagnostik genannt. Bei der indirekten Veränderungsmessung werden Merkmale mehrmals gemessen. Anhand der Differenz der Messwerte zwischen erstem und zweitem Messzeitpunkt lässt sich auf die Richtung der Veränderung sowie deren Ausmaß schließen.

Siehe auch direkte Veränderungsmessung.

Inhaltliche Validität. Von manchen Autoren auch als logische oder Augenscheinvalidität bezeichnet. Ein diagnostisches Instrument ist inhaltlich valide, wenn direkt erkennbar ist, welches Merkmal erfasst wird, beispielsweise wird mit einer Puls-Uhr der Puls gemessen. In der Psychologie lässt man die inhaltliche Validität meist von Experten schätzen.

Siehe auch Validität.

Inkrementelle Validität. Bezeichnet denjenigen Anteil der Varianz des Kriteriums, der nur durch das zu validierende Instrument vorhersagbar ist und nicht durch bereits bestehende Verfahren.
Siehe auch Validität.

Interne Konsistenz. Eine Methode der Reliabilitätsschätzung, bei der jedes Item eines Tests als Maß für das zu messende diagnostische Merkmal aufgefasst wird. Die Korrelation der Items untereinander wird als Maß zur Bestimmung der Reliabilität verwendet.
Siehe auch Reliabilität.

Interpretationsobjektivität. *Siehe Objektivität.*

Item-Response-Theorie. *Siehe Probabilistische Testtheorie.*

Katamnese. Untersuchung in der Folge einer vorgenommenen Behandlung zur Evaluation des Behandlungserfolgs.

Klassifikationsmodell. Klassische psychologische Klassifikationsmodelle sind an die typischen Klassifikationsmodelle der Biologie angelehnt und sind vor allem in der Psychiatrie und der Klinischen Psychologie geläufig. Personen werden hierbei disjunkten Kategorien zugeordnet. Psychische Störungen werden dabei nicht als graduelle Abweichung normaler Gemütszustände angenommen, sondern als qualitative Andersartigkeit im Verhalten und Erleben. Eine Person ist also entweder krank oder gesund.
Siehe auch Dimensionales Modell.

Klassische Testtheorie. Grundannahme der Klassischen Testtheorie ist, dass sich jeder Messwert additiv aus dem wahren Wert (engl. *true score*) und einem Messfehlerwert (engl. *error score*) zusammensetzt. Aus den Implikationen der theoretischen Konzeption des wahren Werts als personbedingter Erwartungswert geht hervor, dass sich der Messfehler einer Person bei unendlich häufiger Messung ausmittelt und somit vom beobachteten Wert auf den wahren Wert geschlossen werden kann.

Klinische Urteilsbildung. Unter klinischer Urteilsbildung versteht man eine Vorgehensweise im klinisch-psychologischen Kontext, bei der die Auswahl, Gewichtung und Integration diagnostischer Information von Diagnostikern aufgrund theoretischer Überlegungen und Erfahrungen, aber ohne Rückgriff auf feststehende Regeln, vorgenommen wird. Diagnostiker greifen hierbei auf implizites Fakten- und Regelwissen zurück.
Siehe auch statistische Urteilsbildung.

Konjunktive Verknüpfung. Bezieht sich auf die Frage, wie Einzelinformationen miteinander verknüpft werden. Es wird zwischen stark und schwach konjunktiver Verknüpfung unterschieden. Bei der stark konjunktiven Verknüpfung muss eine Person bei beiden Prädiktoren den Grenzwert überschreiten, um als geeignet eingestuft zu werden. Bei der schwach konjunktiven Verknüpfung können sich die beiden Prädiktoren innerhalb eines Wertintervalls kompensieren. Damit ist die schwach konjunktive Verknüpfung eine Mischform der additiven und der stark konjunktiven Verknüpfung.
Siehe auch additive und disjunktive Verknüpfung.

Konkurrente Validität. Das Gütekriterium der konkurrenten Validität bezieht sich nicht auf die Schätzmethode, die zur Validierung herangezogen wird, sondern auf die zeitliche Abfolge der Erhebung der Daten. Bei der konkurrenten Validierung werden – im Gegensatz zur prädiktiven und postdiktiven Validität – Prädiktor und Kriterium gleichzeitig erhoben.
Siehe auch prädiktive und postdiktive Validität.

Konstruktvalidität. Ein Instrument zur Messung eines Merkmals (Konstrukts) ist konstruktvalide in dem Maße, in dem es Daten liefert, die von einer bewährten Theorie vorhergesagt werden, in

die das Merkmal (Konstrukt) eingebunden ist. Konstruktvalidität ist die überzeugendste und anspruchsvollste Art der Validierung.

Siehe auch Validität.

Kontexteffekte. Kontexteffekte sind unerwünschte Einflüsse auf den diagnostischen Prozess, die dadurch entstehen, dass die Erhebung diagnostischer Information in einem sozialen und psychischen Kontext stattfindet.

Kontrastanalyse. Auch Gitteranalyse genannt. Eine Methode zur Bestimmung von Soll-Werten. In einem ersten Schritt werden erfolgreiche und erfolglose Personen von einer ersten Gruppe von Experten nominiert. In einem zweiten Schritt werden erfolgreiche und erfolglose Personen von einer zweiten Gruppe von Experten systematisch miteinander verglichen. Merkmale oder Merkmalsausprägungen, die erfolgreiche Personen im Unterschied zu erfolglosen Personen aufweisen, werden zu Soll-Werten erhoben.

Kontrasteffekt. Eine Fehleinschätzung aufgrund eines Vergleichs zwischen Zuständen oder Personen. Die schlechten Schulnoten eines Schülers fallen in einer leistungsstarken Klasse beispielsweise stärker auf als in einer Klasse mit durchschnittlichem Leistungsniveau.

Konvergente Validität. Bezeichnet das Ausmaß, in dem ein zu validierendes Instrument mit anderen Instrumenten, deren Validität bereits bekannt ist und die das gleiche oder ein ähnliches Merkmal messen, zusammenhängt (korreliert). Die konvergente Validität ist ein Teilaspekt der Konstruktvalidität.

Siehe auch Validität.

Kriteriumsvalidität. Ein Maß für die Genauigkeit, mit der ein diagnostisches Instrument (z.B. ein Intelligenztest) ein relevantes Kriterium (z.B. Schulerfolg) vorhersagen kann. Sie wird über die Korrelation von Prädiktor und Kriterium geschätzt.

Siehe auch Validität.

Large Scale Assessment. Nationale und internationale Vergleichsuntersuchungen an großen und repräsentativen Stichproben, deren Ziel meistens darin besteht, den Lernstand von Schülern zu ermitteln (Schulleistungsuntersuchungen), beispielsweise die PISA-Studie zur Grundbildung von Schülern oder die TIMSS-Studie über mathematische Leistungen.

Latent-State-Trait-Theorie. Die Latent-State-Trait-Theorie (Steyer, 1987) ist eine Erweiterung der Klassischen Testtheorie. Ihr zentraler Unterschied zur klassischen Testtheorie ist, dass der wahre Wert als zeitlich und situativ bedingt und somit als veränderlicher Zustand aufgefasst wird.

Latente Variable. Nicht direkt beobachtbare Variable wie beispielsweise Intelligenz. Man schließt von der Ausprägung der manifesten Variablen auf die dahinterliegende latente Dimension.

Siehe auch Manifeste Variable.

Lokale stochastische Unabhängigkeit. Die lokale stochastische Unabhängigkeit ist eine Implikation der Eindimensionalität Rasch-skaliertter Items. Hält man die Fähigkeit der Person konstant, indem man nur Personen mit identischer Fähigkeit betrachtet, beträgt die Korrelation zwischen den Items 0. Dies bedeutet nichts anderes, als dass alle Items nur einen gemeinsamen Faktor teilen: die Fähigkeit der Person.

Manifeste Variable. Beobachtbare Variablen. Manifeste Variablen lassen Rückschlüsse auf dahinterliegende latente Dimensionen zu.

Siehe auch Latente Variable.

Messgelegenheitsspezifität. *Siehe Änderungssensitivität.*

Methodeneffekt. *Siehe Methodenspezifität.*

Methodenspezifität. Bezeichnet den Anteil beobachteter Varianz, der auf Effekte der verwendeten Messmethoden zurückzuführen ist. Die Ermittlung der Methodenspezifität mittels der Multitrait-Multimethod-Analyse (MTMM) ist wichtig, da sie die Validität eines Tests einschränkt.

Multimethodales Messen. Mithilfe des Einsatzes verschiedener diagnostischer Instrumente (Methoden) können die einzelnen Stärken diagnostischer Verfahren genutzt und ihre spezifischen Schwächen ausgeglichen werden.

Multitrait-Multimethod-Analyse. Die Multitrait-Multimethoden-Analyse (MTMM) ist eine von Campbell und Fiske (1959) entwickelte Methode zur Überprüfung der Konstruktvalidität diagnostischer Instrumente. Traits und Methoden werden hierbei systematisch kombiniert, um die Ergebnisse der Messung mehrerer Merkmale mit mehreren Methoden zu vergleichen.

Multivariate Diagnostik. Bei der multivariaten Diagnostik wird die Verteilung von zwei oder mehr Variablen gleichzeitig erhoben und untersucht, während bei der univariaten Diagnostik nur eine Variable erhoben wird.

Negativer Prädiktiver Wert. Der Anteil der als ungeeignet Diagnostizierten, die tatsächlich ungeeignet sind (d.h. Anteil der richtig Negativen).

Normierung. Bezeichnet das Herstellen eines Bezugsrahmens, der auf interindividuellen Vergleichen beruht und eine Verteilung der Messwerte liefert. Für die Normierung von diagnostischen Instrumenten wird eine repräsentative Eichstichprobe benötigt.

Nützlichkeit. Nützlich ist ein diagnostisches Verfahren, wenn es Informationen liefert, die bereits vorhandene Verfahren nicht liefern können oder wenn seine Verwendung hilft, Fehler zu verhindern, die bei mehrmaliger Verwendung eines bereits vorhandenen Instruments zu befürchten wären.

Objektivität. Ein diagnostisches Instrument ist objektiv, wenn die Durchführung, Auswertung und Interpretation der Testergebnisse unabhängig vom Diagnostiker ist. Die Objektivität ist neben der Reliabilität und der Validität das wichtigste Gütekriterium.

Ökonomie. Als ökonomisch gilt ein diagnostisches Instrument, wenn es diagnostische Daten gleicher Qualität zu einem günstigeren Preis liefert als alternative Instrumente.

Paralleltestreliabilität. Die Paralleltestreliabilität lässt sich ermitteln, wenn es außer demjenigen diagnostischen Verfahren, dessen Reliabilität geschätzt werden soll, noch mindestens ein zweites paralleles bzw. äquivalentes Verfahren zur Erfassung des gleichen Merkmals gibt. Parallel ist ein Test, wenn dieser auf ähnlichen Items, die die gleichen Messeigenschaften wie der erste Test aufweisen, basiert. Geschätzt wird die Paralleltestreliabilität über die Korrelation beider Verfahren. *Siehe auch Reliabilität.*

Positiver Prädiktiver Wert. Der Anteil der als geeignet Diagnostizierten, die tatsächlich geeignet sind (d.h. Anteil der richtig Positiven).

Postdiktive Validität. Bei der postdiktiven (= rückblickenden) Validierung ist das Kriterium bereits vor Einsatz des zu validierenden Instruments bekannt. *Siehe auch prädiktive und konkurrente Validität.*

Prädiktive Validität. Bei der prädiktiven (= vorhersagenden) Validierung wird das Kriterium in einem zeitlichen Abstand nach der Anwendung des zu validierenden Instruments erhoben.
Siehe auch konkurrente und postdiktive Validität.

Probabilistische Testtheorie. Oftmals auch als Item-Response-Theorie bezeichnet. Die probabilistische Testtheorie beschreibt den Zusammenhang zwischen dem wahren Wert einer Person und der Wahrscheinlichkeit, dass ein Item gelöst wird. Im Gegensatz zur klassischen Testtheorie kann mit der probabilistischen Testtheorie auch das Antwortverhalten von Probanden auf binäre bzw. kategoriale Variablen untersucht werden.
Siehe auch Rasch-Modell.

Projektive Verfahren. Projektive Verfahren beruhen auf dem Abwehrmechanismus der Projektion von Freud, bei dem der tabuisierte Handlungswunsch einer anderen Person unterstellt wird. Angenommen wird, dass Diagnostikanden unbewusste Wünsche und Ängste äußern, wenn ihnen mehrdeutige Bildvorlagen vorgelegt werden, die sie interpretieren sollen. Beispiele für Projektive Verfahren sind der Rorschach-Test, der Thematische Apperzeptionstest und die Picture Story Exercise. Da die meisten projektiven Verfahren nicht den heutigen Gütekriterien genügen, werden sie nur noch selten eingesetzt.

Prozessdiagnostik. Im Gegensatz zur Statusdiagnostik werden die interessierenden Merkmale bei der Prozessdiagnostik mehrmals erfasst. Dies ist beispielsweise notwendig, wenn das zu erhebende Merkmal natürlichen Veränderungen unterliegt oder wenn ein Merkmal durch eine Intervention verändert werden soll. Prozessdiagnostik liefert prinzipiell wertvollere Informationen als Statusdiagnostik, ist jedoch auch zeitlich aufwändiger und teurer.
Siehe auch Statusdiagnostik.

Pygmalioneffekt. Bezeichnet das Phänomen, dass Forscher von einer Theorie manchmal so überzeugt sind, dass sie empirische Überprüfungen unbewusst so gestalten, dass die Theorie bestätigt wird. Dieses Phänomen wird auch als sich selbsterfüllende Prophezeiung bezeichnet. Eine bekannte Untersuchung des Pygmalioneffekts ist die Studie von Rosenthal und Jacobson (1968).

Rasch-Homogenität. Items, deren Item-Charakteristische Kurven parallel verlaufen, werden auch Rasch-homogen genannt. Sie sind eindimensional, messen also nur eine einzige True-score-Variable oder Fähigkeit.

Rasch-Modell. Von Georg Rasch (1960) entwickeltes mathematisches Modell der probabilistischen Testtheorie. Im Rasch-Modell ist die Wahrscheinlichkeit, dass ein Item gelöst oder bejaht wird, als Funktion der Differenz von zwei Werten, nämlich dem wahren Wert oder der Fähigkeit der Person und der Schwierigkeit des Items, konzipiert. Die Schwierigkeiten von Items und die Fähigkeiten von Personen variieren kontinuierlich auf einer gemeinsamen Intervallskala.

Reaktivität. Ein diagnostisches Verfahren ist reaktiv, wenn die Diagnostikanden wissen, dass ihr Verhalten registriert, bewertet und zu einem diagnostischen Zweck genutzt wird. Die Reaktivität eines Verfahrens darf nicht mit der Transparenz eines Verfahrens verwechselt werden.

Regression zur Mitte. Ein von Sir Francis Galton (1886) beschriebenes Phänomen. Dieses Regressionsphänomen tritt bei Variablen auf, die nicht perfekt miteinander korrelieren, also bei wiederholter Messung eine Retestkorrelation < 1 aufweisen. Es handelt sich hier um ein symmetrisches Problem, das besonders bei der Evaluation von Interventionsstudien berücksichtigt werden muss.

Reihenfolgeeffekt. Wirkt sich die Reihenfolge, in der Informationen erhoben werden, auf das Ergebnis aus, entstehen Reihenfolgeeffekte. Sie entstehen beispielsweise durch Ermüdung, Langeweile, Lern- oder Erinnerungseffekte.

Reliabilität. Reliabilität ist eines der wichtigsten Gütekriterien diagnostischer Instrumente. Ein diagnostisches Instrument ist in dem Maße reliabel (zuverlässig), in dem es das genau misst, was es misst. Reliabilität ist gegeben, wenn ein Instrument bei wiederholter Messung der gleichen Objekte oder Merkmalsträger unter gleichen Bedingungen die gleichen Messergebnisse anzeigt. Reliabilität ist eine kontinuierliche Variable. Ein Messinstrument kann mehr oder weniger reliabel sein. Als Maßeinheit der Reliabilität werden Anteile (von 1) verwendet.

Retest-Reliabilität. Auch Testwiederholungsmethode oder Messwiederholungsmethode genannt. Bei mehrmaligem Messen des gleichen Objekts unter gleichen Bedingungen sollte ein reliables Instrument das gleiche Ergebnis anzeigen. Zur Schätzung der Retest-Reliabilität wird die Produkt-Moment-Korrelation der Ergebnisse zweier Messungen verwendet.

Siehe auch Reliabilität.

Sensitivität. Die Sensitivität oder auch Trefferquote ist der Anteil der richtig Positiven an den Geeigneten, also der Anteil der Geeigneten, die als geeignet erkannt werden.

Sequenzielle Diagnostik. Auch schrittweise Diagnostik genannt. Bei der sequenziellen Diagnostik werden die Daten in mehreren hintereinander geschalteten Schritten erhoben. Die Diagnostikanden müssen bei jedem einzelnen Schritt einen festgelegten Grenzwert überschreiten. Eingesetzt wird sequenzielle Diagnostik beispielsweise bei der Personalauswahl oder bei der Auswahl von Stipendiaten. Die Daten werden zwangsläufig konjunktiv verknüpft.

Siehe auch konjunktive Verknüpfung.

Soziale Erwünschtheit. Die Tendenz von Diagnostikanden, sich in psychologischen Untersuchungen möglichst positiv darzustellen und sozial anerkannte Merkmale in den Vordergrund zu rücken. Insbesondere tritt diese Tendenz zur Selbstbeschönigung auf, wenn die diagnostizierten Merkmale in hohem Maße erwünscht oder im Gegenteil sehr unerwünscht sind. Dadurch kann es zu systematischen Verfälschungen kommen.

Spearman-Brown-Formel. Wenn einzelne Testteile das gleiche Merkmal messen, lässt sich mit der Spearman-Brown-Formel der Testverlängerung bestimmen, wie die Reliabilität des Tests mit seiner Verlängerung zunimmt. Umgekehrt kann die Spearman-Brown-Formel auch verwendet werden, um zu ermitteln, um welchen Faktor ein Test mit einer unzureichenden Reliabilität verlängert werden muss, um eine geforderte Mindestreliabilität zu erreichen.

Siehe auch Reliabilität.

Spezifität. Bezeichnet den Anteil der richtig Negativen an den Ungeeigneten, also den Anteil der Ungeeigneten, die als ungeeignet erkannt werden.

Statistische Urteilsbildung. Bei der statistischen Urteilsbildung erfolgt die Auswahl, Gewichtung und Verknüpfung diagnostischer Information mittels eines Algorithmus, der eindeutig vorschreibt, wie verfügbare Information zusammengefasst wird.

Siehe auch Klinische Urteilsbildung.

State. Englische Bezeichnung für einen vorübergehenden psychischen Zustand.

Siehe auch Trait.

Statusdiagnostik. Statusdiagnostik findet im Gegensatz zur Prozessdiagnostik nur einmalig statt. Statusdiagnostik lässt sich z.B. einsetzen, wenn das zu erfassende Merkmal zeitlich stabil ist, die Merkmalsausprägung des zu diagnostizierenden Merkmalsträgers nur zu einem Zeitpunkt von Interesse ist oder wenn die zeitnahe Einleitung von Maßnahmen in Ausnahmesituationen, z.B. bei Suizidalität, notwendig ist.

Siehe auch Prozessdiagnostik.

Systematischer Messfehler. Systematische Messfehler entstehen beispielsweise, wenn ein diagnostisches Instrument neben dem Merkmal, das es erfassen soll, noch weitere, unerwünschte Merkmale misst. Systematische Messfehler beeinträchtigen daher die Validität eines diagnostischen Instruments, nicht aber dessen Reliabilität.

Siehe auch unsystematischer Messfehler.

Systematische Verhaltensbeobachtung. Bei der systematischen Verhaltensbeobachtung sind die Beobachtungsgegenstände und die Merkmale, die man beobachten möchte, zuvor festgelegt. Diese Festlegung setzt eine Hypothese oder konkrete Fragestellung voraus.

Taylor-Russel-Tafeln. In den Taylor-Russel-Tafeln lässt sich ablesen, wie sich die gleichzeitige Veränderung von Selektionsquote, Eignungsquote und Validität auf die Diagnosekoeffizienten auswirkt.

Trait. Englische Bezeichnung für eine Persönlichkeitseigenschaft, die zeitlich stabil ist.

Siehe auch State.

Transparenz. Eine diagnostische Untersuchung ist transparent, wenn die Person den Zweck der Untersuchung und die Bedeutung der Daten, die sie liefert, kennt.

Trennschärfe. Die Trennschärfe bezeichnet die Steigung der Item-Charakteristischen Kurven (ICC) an ihrem Wendepunkt. Je steiler die ICC eines Items, desto besser kann es zwischen Personen differenzieren, selbst wenn deren Fähigkeitsunterschiede gering sind.

Univariate Diagnostik. *Siehe Multivariate Diagnostik.*

Unsystematischer Messfehler. Auch zufälliger oder unregelmäßiger Messfehler genannt. Unsystematische Messfehler sind solche, die wir nicht vorhersagen können, da wir ihre Ursachen nicht kennen. Sie sind besonders problematisch, da sie sowohl die Validität als auch die Reliabilität eines diagnostischen Instruments beeinträchtigen.

Siehe auch systematischer Messfehler.

Validität. Valide ist ein diagnostisches Instrument in dem Maße, in dem es misst, was es messen soll. Maximal valide ist ein Instrument, wenn die mit ihm gewonnenen Messwerte exakt mit den wahren Merkmalswerten übereinstimmen. Validität ist eine kontinuierliche Variable, d.h. ein diagnostisches Instrument kann mehr oder weniger valide sein.

Verhaltensbeobachtung. Eine Methode zur Erfassung natürlicher Verhaltensströme, bei der beispielsweise die Interaktion zwischen zwei oder mehr Personen beobachtet und erfasst wird. Es gibt zahlreiche Varianten der Verhaltensbeobachtung, die sich danach unterscheiden lassen, ob die Beobachtung systematisch oder unsystematisch erfolgt, ob sie im Labor oder im Feld stattfindet, wie sehr der Beobachter in das Geschehen involviert ist und welche Protokollierungsmethode angewandt wird.

Vorhersagerichtigkeit. Bezeichnet den Anteil richtiger Entscheidungen an allen Entscheidungen.

Wachstumsanalyse. (engl. *latent growth curve analysis*) Die Wachstumsanalyse umfasst eine Familie latenter Strukturgleichungsmodelle, anhand derer ermittelt werden kann, ob Veränderungen über die Zeit einem bestimmten Trend oder Muster folgen. Wachstumsanalysen gehen von einem autoregressiven Prozess zwischen Zuständen aus; das bedeutet, dass ein früherer Zustand einen späteren Zustand beeinflusst bzw. ein späterer Zustand von einem früheren abhängt.

Zeitreihenanalyse. Eine Methode zur Entdeckung der Verlaufsform von psychischen Merkmalen wie z.B. Beschwerden. Hierbei wird das interessierende Merkmal wiederholt gemessen und die ermittelten Korrelationen werden in einer Matrix abgetragen.

Wichtig ist, dass das interessierende Merkmal zu allen Zeitpunkten erfasst wird, zwischen denen eine Änderung der Merkmalsausprägung erwartet wird.

Zufälliger Messfehler. *Siehe Unsystematischer Messfehler.*