



Klassische Testtheorie – Eine kurze Einführung

(inklusive kurze Beschreibung der drei grossen
Kräfte der Psychologie)

Marc Schreiber, September 2019



Inhalt

Psychoanalyse, Behaviorismus und humanistische Psychologie als drei grosse Kräfte der Psychologie	3
Klassische Testtheorie	5
Objektivität	7
Reliabilität.....	7
Validität	8
Skalierung	9
Literatur	10

Psychoanalyse, Behaviorismus und humanistische Psychologie als drei grosse Kräfte der Psychologie

Die Psychologie wurde seit Mitte des 19. Jh. primär durch drei grosse Kräfte beeinflusst: Psychoanalyse, Behaviorismus und humanistische Psychologie. Im Sinne eines groben Überblickes und für das Verständnis der verschiedenen Beratungstheorien sowie der Paradigmen der Berufs-, Studien- und Laufbahnberatung (BSLB; siehe Schreiber, im Druck) werden die drei Ansätze in der Folge kurz beschrieben (siehe auch Schreiber, 2017).

In der *Psychoanalyse* mit dem einflussreichsten Vertreter Sigmund Freud (1856-1939) wird der Fokus auf unbewusste Prozesse gelegt, die unser Erleben und Verhalten bestimmen (siehe Rammsayer & Weber, 2016). Mit dem topologischen Modell der Persönlichkeit unterschied Freud zwischen den bewussten, vorbewussten und unbewussten Erlebnisinhalten. Zentrale Impulse für das Verhalten entspringen gemäss topologischem Modell dem Unbewussten, beispielsweise in Form von unbewussten Trieben, welche ihren zentralen Ursprung in den Kindheitserfahrungen haben und permanent befriedigt werden möchten. Dadurch hindern Sie uns daran, einen entspannten Alltag zu führen. Der grosse Fokus auf das Unbewusste sowie das Vorbewusste führte zu zahlreichen Beratungsmethoden, die darauf abzielen, unbewusste Erlebnisinhalte dem Bewusstsein zugänglich und dadurch auch therapierbar zu machen. Dazu gehören das freie Assoziieren oder die Rorschach-Tafeln (Rorschach, 1921), welche in der Persönlichkeitsdiagnostik lange Zeit eingesetzt wurden. In Therapie und Beratung werden dabei die unbewussten Inhalte ins Bewusstsein gerufen, um in der Folge einen Umgang damit finden, respektive sich von der Last der unbewussten Triebe befreien zu können. Der Fokus auf das Unbewusste bringt zwangsläufig mit sich, dass Personen keine Selbstauskunft über ihre Triebe und Eigenschaften geben können und dass Therapeutinnen das Unbewusste mithilfe spezifischer Methoden erfassen müssen. So ist es auch naheliegend, dass Therapeutinnen eine Expertenrolle einnehmen und die spontanen Assoziationen sowie die Antworten ihrer Klienten bei den Rorschach-Tafeln interpretieren und daraus die triebbedingten Persönlichkeitsdefizite ihrer Klienten ableiten.

Die beiden anderen zentralen Kräfte der Psychologie haben sich mindestens teilweise als Gegenentwürfe zur Psychoanalyse entwickelt.

Im *Behaviorismus* liegt der Fokus mit grosser Konsequenz ausschliesslich auf dem Beobachtbaren und der expliziten Abgrenzung gegenüber den unbewussten Prozessen der Psychoanalyse (siehe Rammsayer & Weber, 2016). Ein zentraler Vertreter des Behaviorismus war John B. Watson (1878-1938). Im Gegensatz zur Psychoanalyse gehen die Behavioristen davon aus, dass das menschliche Erleben und Empfinden einer Black-Box entspricht, die nicht zu entschlüsseln ist. Psychologie fokussierte fortan auf das konkrete und beobachtbare Verhalten und nicht auf innerpsychische Prozesse, die das Verhalten steuern sollen. Eine zentrale Annahme der Behavioristen ist, dass Personen als weisses Blatt auf die Welt kommen und dass die beobachtbaren Persönlichkeitsaspekte allesamt erlernt sind und dadurch auf Erfahrungen in ihrer Umwelt zurückzuführen sind. Der Einfluss der genetischen Prädisposition wird als unbedeutend angesehen und mit zahlreichen Experimenten wurde versucht herauszufinden, wie Menschen lernen oder in anderen Worten ausgedrückt, wie sich die Persönlichkeit eines Menschen entwickelt.

Im Zuge der *kognitiven Wende* in den 1960er und 1970er Jahren wird der Fokus in der wissenschaftlichen Psychologie dann wieder auf innerpsychische Prozesse gelegt (Neyer & Asendorpf, 2018). So hat Bandura (1977) beispielsweise das Konzept der Selbstwirksamkeitserwartung entwickelt. Mit dem Konzept wird die Wichtigkeit hervorgehoben, dass eine Person von sich selbst die Vorstellung oder Erwartung hat (oder diese in der Beratung entwickelt), dass sie eine konkrete Tätigkeit erfolgreich ausführen kann oder eine bestimmte Kompetenz oder Eigenschaft aufweist (z.B. die Kompetenz, sich im Bewerbungsgespräch authentisch präsentieren zu können). Dabei ist nicht die objektive Einschätzung der Kompetenz relevant, sondern die subjektive Erwartung der Person.

Als dritte grosse Kraft neben der Psychoanalyse und dem Behaviorismus gilt die *humanistische Psychologie*, welche sich im Gegensatz zur defizitorientierten Psychoanalyse auf die positive Selbstverwirklichung der Menschen konzentriert (Rammsayer & Weber, 2016). Ein zentraler Vertreter ist Abraham Maslow (1908-1970), welcher die Selbstverwirklichung als oberste Stufe einer Bedürfnishierarchie betrachtete. Die seit 15 Jahren sehr stark verbreitete *positive Psychologie* nimmt die Grundannahmen der humanistischen Psychologie auf und legt den Fokus primär darauf, Personen, denen es gut geht, darin zu unterstützen, dass es ihnen noch besser geht (siehe Rammsayer & Weber, 2016; oder auch Fredrickson, 2003; Gable & Haidt, 2005; Ruch & Proyer, 2015; Seligman, 2011; Seligman & Csikszentmihalyi, 2000). Abbildung 1 stellt einen Versuch dar, den Zeitgeist der Psychologie im Zeitraum zwischen 1850 und 2050 mit den soeben beschriebenen relevanten Strömungen abzubilden. Aktuell beschäftigen sich die *Neurowissenschaften* wieder damit, unbewusste Prozesse und damit verbunden auch die Architektur und Funktionsweise des menschlichen Gehirns zu beforschen (Asendorpf, 2015). Daraus konnten bereits zahlreiche Erkenntnisse für die Praxis extrahiert werden, wie beispielsweise die Plastizität und vor allem die damit verbundene Lernfähigkeit der äusseren Hirnareale oder eben die Wichtigkeit unbewusster Prozesse und teilweise sogar pränataler Erfahrungen für unser Erleben und Verhalten (Bauer, 2016; McGilchrist, 2009; Roth, 2016). Es ist davon auszugehen, dass in den kommenden Jahren weitere Erkenntnisse dazukommen werden.

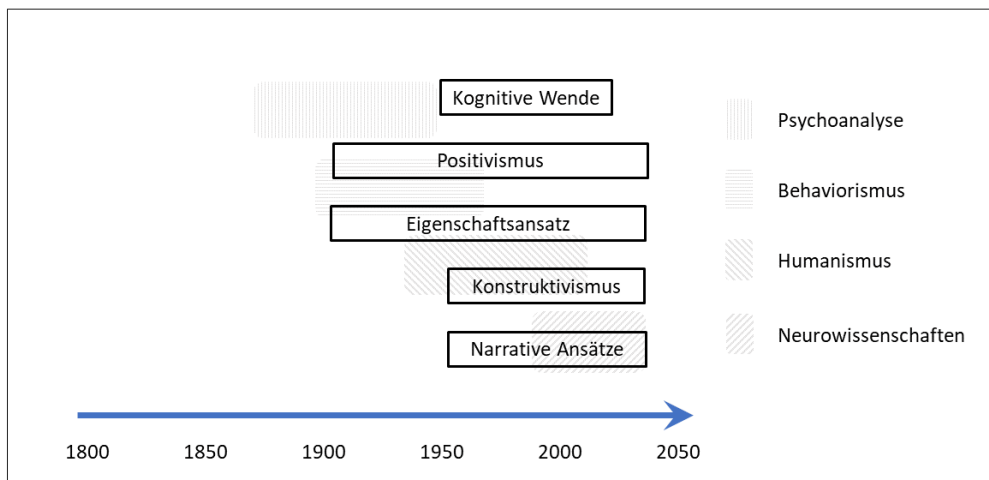


Abbildung 1: Veranschaulichung des Zeitgeists der Psychologie von 1850 bis 2050

Klassische Testtheorie

Das Wesen der klassischen Testtheorie (Bühner, 2011; Moosbrugger & Kelava, 2012) besteht darin, dass Messwerte aufgegliedert werden. Die Messwerte entsprechen den Antworten, die eine Person bei den einzelnen Fragen oder Items eines Fragebogens gegeben hat.

Abbildung 2 zeigt einen Ausschnitt eines ausgefüllten Persönlichkeitsfragebogens mit acht Items, die unterschiedliche Dimensionen der Persönlichkeit erfassen.

Fragebogen zur Erfassung der Persönlichkeit (IPIP-5F30F; 180 Fragen)

Bitte geben Sie an, inwiefern die Aussagen auf Sie zutreffen: *

	trifft nicht zu	trifft eher nicht zu	neutral	trifft eher zu	trifft zu
Ich spüre meist eine gewisse Anspannung in mir.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Es fällt mir leicht, Freundschaften zu schliessen.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich verliere mich oft in Tagträumen.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe das feste Vertrauen, dass mir andere Menschen wohlgesinnt sind.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Es gelingt mir schnell und einfach, Dinge zu verstehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Ich verspüre oft starke Wut.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin ungern alleine.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Schönheit der Natur berührt mich tief.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Abbildung 2: Beispielitems eines Persönlichkeitsfragebogens (aus: Plattform Laufbahndiagnostik, ZHAW/IAP, 2019)

Jeder Angabe wird eine Zahl zugeordnet: trifft nicht zu = 1; trifft eher nicht zu = 2; neutral = 3; trifft eher zu = 4; trifft zu = 5. In der Folge werden alle Items, welche dieselbe Dimension (z.B. Geselligkeit) erfassen, aufsummiert. Dieser Summenwert (oder auch der Mittelwert über alle Items) entspricht dann dem Wert, den eine Person bei dieser Dimension der Persönlichkeit erreicht.

Abbildung 3 zeigt ein Beispielprofil eines Persönlichkeitsfragebogens. Die Mittelwerte bei den Dimensionen Neurotizismus oder Extraversion entsprechen dem Mittelwert über alle Items, welche Neurotizismus, resp. Extraversion erfassen. Man spricht im Kontext der klassischen Testtheorie auch von *beobachteten Werten*, weil es sich um die Werte handelt, die sich aus den Antworten der Personen ergeben.

Jeder beobachtete Wert wird gemäss Definition aufgeteilt in einen hypothetischen *wahren Wert* und in einen sogenannten *Fehlerwert* (vgl. Bühner, 2011):

Definition in der klassische Testtheorie: $X_{vt} = \tau_v + \epsilon_{vt}$

Der *beobachtete Wert* (x) einer Person (v) zum Zeitpunkt (t) wird definiert als konstanter *wahrer Wert* (τ , Tau) der Person (v) plus *Messfehler* (ϵ , Epsilon) zum Zeitpunkt (t).

Der Messfehler repräsentiert alle unkontrollierten, unsystematischen Störeinflüsse, also beispielsweise Müdigkeit, Motivation sowie die Bedingungen beim Ausfüllen eines Fragebogens. Der beobachtete Wert (x_{vt}) entspricht in der Praxis wie erwähnt dem Summen- oder Mittelwert einer Person. Bühner (2011) erwähnt verschiedene Annahmen, die im Rahmen der klassischen Testtheorie erfüllt sein müssen (z.B. zufällige Verteilungen der Messwerte ohne systematische Messfehler).

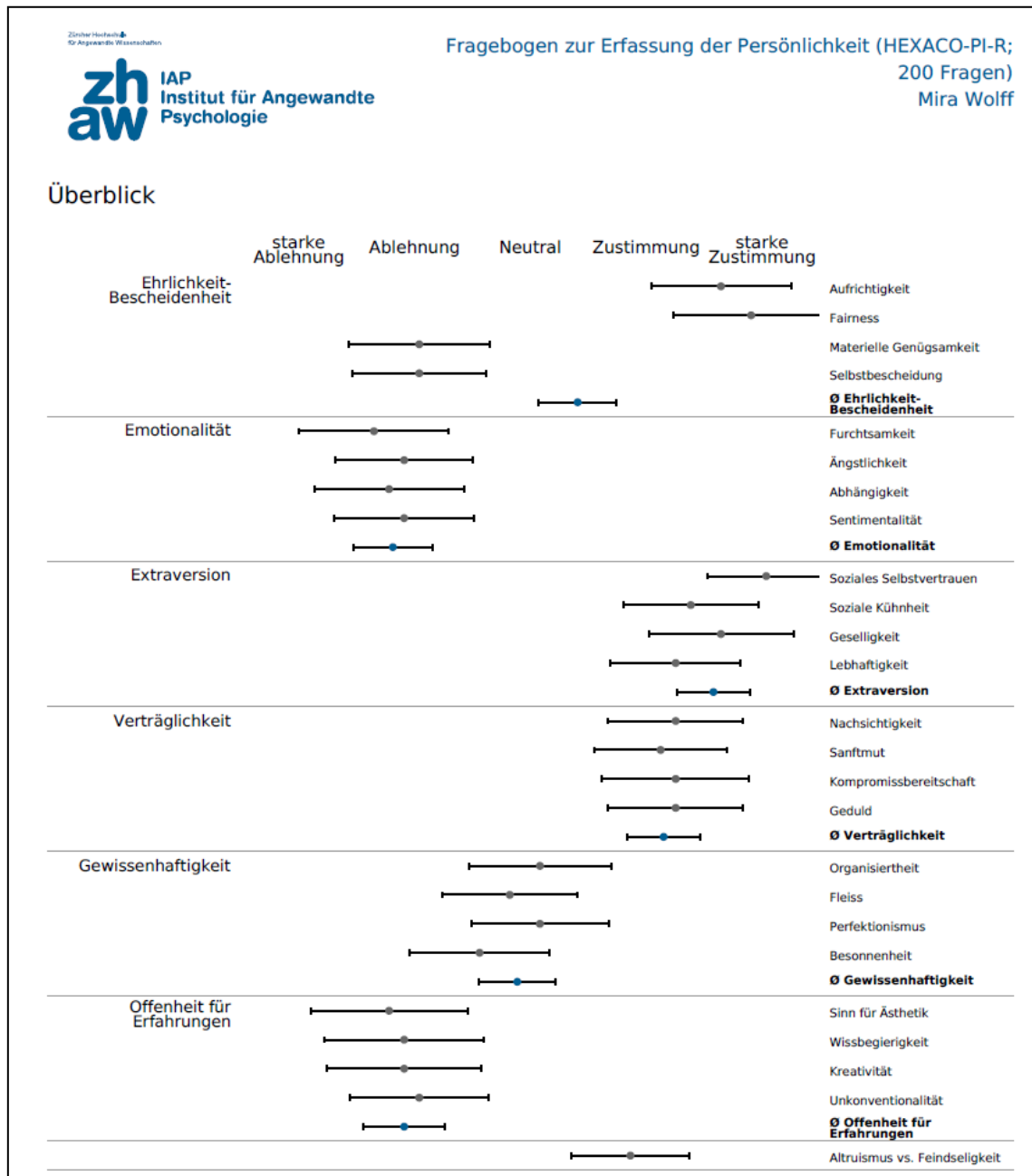


Abbildung 3: Beispielprofil HEXACO-PI-R inklusive Vertrauensintervall (schwarze Linie) (aus: Plattform Laufbahndiagnostik, ZHAW/IAP, 2019)

Die beobachteten Werte einer Person oder einer ganzen Gruppe sind also bekannt, sobald die Person den Fragebogen ausgefüllt hat. Gemäss Definition der klassischen Testtheorie wird der beobachtete Wert aufgliedert in einen wahren Wert (τ , Tau) und einen Messfehler (ε , Epsilon). Wie bereits erwähnt ist der wahre Wert hypothetisch, also nicht beobacht- oder messbar. Im Rahmen der klassischen Testtheorie wird eine Schätzung des Messfehlers vorgenommen. Diese Schätzung kann auf unterschiedliche Arten geschehen. Sie hängt linear mit der *Reliabilität* (Messgenauigkeit) eines Fragebogens zusammen und ist deshalb gleichzusetzen mit den verschiedenen Arten der Reliabilitätsberechnung, auf die später eingegangen wird. Vorweg gilt es festzuhalten, dass die Reliabilität *das* zentrale Konstrukt im Rahmen der klassischen Testtheorie darstellt. Anhand einer zuverlässigen Schätzung der Messgenauigkeit (oder anders ausgedrückt: des Messfehlers) eines Fragebogens kann gemäss der Definition der klassischen Testtheorie der Bereich festgelegt werden, innerhalb welchem der wahre Wert einer Person mit einer gewissen Wahrscheinlichkeit liegt.

Bereich, in welchem der wahre Wert liegt: $\tau_v = X_{vt} \pm \varepsilon_{vt}$

Der *wahre Wert* (τ , Tau) der Person (v) liegt um den *beobachteten Wert* (x) der Person (v) zum Zeitpunkt (t) herum, wobei der Bereich um den beobachteten Wert herum durch den *Messfehler* (ε , Epsilon) der Person (v) zum *Zeitpunkt* (t) bestimmt wird. Gesucht wird also eine zuverlässige Schätzung des Messfehlers einer Person zu einem bestimmten Zeitpunkt (ε_{vt}). Die Schätzung basiert wie soeben erwähnt auf verschiedenen Berechnungsarten der Reliabilität.

Bühner (2011) spricht von vier Hauptgütekriterien psychometrischer Fragebogen und Tests: Objektivität, Reliabilität, Validität und Skalierung. In der Folge wird auf diese für die Bewertung eines Fragebogens oder Test relevanten Kriterien eingegangen.

Objektivität

Bei der *Objektivität* geht es um eine Standardisierung von der Durchführung über die Auswertung bis hin zur Interpretation eines Fragebogens oder Tests. Es handelt sich nicht um die Objektivität im Sinne von *objektiv richtig*, sondern um die Unabhängigkeit des diagnostischen Prozesses von der Diagnostikerin. Damit ist die Forderung gemeint, dass es bei einer Fragebogendurchführung nicht darauf ankommen soll, wer den Fragebogen mit dem Klienten durchführt. Wenn die Instruktion eines Fragebogens akribisch befolgt wird, stellt die Objektivität bei der Durchführung keine grosse Herausforderung dar. Bei einer online Durchführung, bei der das Fragebogenprofil automatisch erstellt wird, sind Durchführung und Auswertung standardisiert und man kann davon ausgehen, dass die Objektivität erfüllt ist. Stehen dem Anwender zudem klare Interpretationshilfen zur Verfügung, so kann die Objektivität auch bei der Interpretation als gegeben betrachtet werden.

Reliabilität

Die *Reliabilität* ist ein Gütekriterium, welches sich wie bereits vorher erläutert auf die Messgenauigkeit des Instrumentes bezieht. Es zeigt an, wie genau ein Verfahren misst. Bei psychologischen Konstrukten, wie beispielsweise der Extraversion, kann die die Frage aufgeworfen werden, wie Messgenauigkeit überhaupt erreicht werden kann. Grundsätzlich steht die Forderung im Raum, dass bei einem Konstrukt, von dem man davon ausgeht, dass es über die Zeit hinweg konstant ist (z.B. eine Persönlichkeitseigenschaft), bei mehreren Messungen einer Person mit demselben Fragebogen derselbe Wert erreicht wird. So kann auch die Messgenauigkeit eines Fragebogens oder Tests überprüft werden. Wird dies gemacht, so spricht man von der sogenannten *Retest-Reliabilität*. Neben dieser Möglichkeit, die Messgenauigkeit eines Verfahrens zu überprüfen, gibt es aber noch zwei weitere Arten der Reliabilität.

Zum einen die *Paralleltest-Reliabilität*, bei welcher zwei identische Testversionen erstellt werden. Dies ist häufig bei Intelligenztests der Fall. Die Fragen der beiden Paralleltest-Versionen sollten denselben Schwierigkeitsgrad aufweisen. Dadurch können die Paralleltest-Versionen auch bei Gruppentestungen eingesetzt werden, in welchen die Personen direkt nebeneinandersitzen und die Gefahr des Abschreibens besteht. Die Paralleltest-Reliabilität entspricht dann der Korrelation der beiden parallelen Testversionen.¹

Zum anderen die *Testhalbierungs-Reliabilität*, bei welcher ein Fragebogen oder Test in zwei möglichst homogene Hälften unterteilt und dann aufgrund der Gleichheit der Ergebnisse in den beiden Hälften die Reliabilität berechnet wird. Als Grundlage für die Reliabilitätsberechnung wird analog zur Paralleltest-Reliabilität auch bei der Testhalbierungs-Reliabilität die Korrelation zwischen den beiden Testhälften verwendet. Der häufigste Reliabilitätswert ist ein Spezialfall der Testhalbierungs-Reliabilität. Es handelt sich um die sogenannte *innere* oder auch *interne Konsistenz*, bei welcher die Antworten der einzelnen Fragen eines Fragebogens betrachtet werden. Die interne Konsistenz wird als *Cronbach's α* ausgegeben. Der Wert gibt an, inwiefern die Personen die verschiedenen Items, mit denen ein Konstrukt wie beispielsweise die Extraversion gemessen wird, kongruent beantwortet haben. Er basiert auf der Annahme, dass Personen, die eine hohe Ausprägung in Extraversion aufweisen, auch bei allen Fragen zur Extraversion stark zustimmen. Nur wenn dies der Fall ist, kann davon ausgegangen werden, dass die Fragen zur Extraversion etwas Einheitliches erfassen – eben, weil die Personen mit hohen Gesamtwerten in Extraversion sich in den Antworten von den Personen mit tiefen Gesamtwerten unterscheiden. Wichtig zu erwähnen ist, dass bei der Reliabilität nicht geprüft wird, was gemessen wird. Es geht ausschliesslich darum, ob die Fragen etwas Einheitliches messen und die Messung dadurch genau ist.

Anhand der beschriebenen Möglichkeiten der Reliabilitätsschätzung (Retest, Paralleltest, Testhalbierung, interne Konsistenz) kann die Messgenauigkeit (der Messfehler) eines Verfahrens und in der Folge das Vertrauensintervall bestimmt und dadurch die zentrale Frage der klassischen Testtheorie beantwortet werden. Die Reliabilität liegt per Definition zwischen null und eins, wobei eine Reliabilität von eins einer perfekten Messung ohne Messfehler entsprechen würde. In der Realität der psychologischen Diagnostik gibt es keine perfekten Messungen. Für die Individualdiagnostik wird eine interne Konsistenz nach Cronbach's α von mindestens $\alpha = .70$ gefordert (Bühner, 2011). Ein Reliabilitätswert von $\alpha = .70$ kann wie folgt interpretiert werden: Die Items, die das Konstrukt erfassen (z.B. Extraversion), erfassen zu 70% etwas Gemeinsames. Die restlichen 30% sind aufgrund des Messfehlers zustande gekommen. Anders ausgedrückt kann man bei einer Reliabilität von $\alpha = .70$ festhalten, dass die Frage, ob jemand einen hohen oder einen tiefen Wert bei dem Fragebogen erreicht, zu 70% auf das interessierende Merkmal zurückzuführen ist. Nur gerade 30% der Varianz sind auf den zufälligen und deshalb auch nicht interpretierbaren Messfehler zurückzuführen.

Neben der Messgenauigkeit wird zur Beurteilung eines psychometrischen Fragebogens als weiteres Gütekriterium die Validität benötigt, um zu wissen, was der Fragebogen überhaupt misst.

Validität

Die *Validität* zeigt auf, ob ein Fragebogen das misst, was gemessen werden soll. Dabei unterscheidet man zwischen Inhalts-, Konstrukt- und Kriteriumsvalidität.

Zentral für die Praxis ist die *Inhaltsvalidität*, bei der es darum geht, dass die Fragen oder Items eines Fragebogens das zu erfassende Konstrukt (z.B. Extraversion) repräsentativ und gemäss Definition erfassen sollen. Hier zeigt sich, dass die Ausgangslage eines jeden Fragebogens die Definition des

¹ Bühner und Ziegler (2017) beschreiben in ihrem Buch Statistik für Psychologen und Sozialwissenschaftler die für den Umgang mit Fragebogen und Tests relevanten statistischen Kennwerte wie beispielsweise die Korrelation.

zu erfassenden Konstruktes darstellt. Darauf abgestimmt werden die Fragen formuliert. Zur Erfassung von Extraversion werden im HEXACO-PI-R (Lee & Ashton, 2004) beispielsweise die folgenden Beispielitems verwendet: «Wenn ich in einer Gruppe von Leuten bin, bin ich oft derjenige, der im Namen der Gruppe spricht.», «Ich geniesse es, viele Leute um mich herum zu haben, mit denen ich reden kann.», «Ich bin fast immer voller Energie.» (siehe Schreiber, Mueller, et al., 2018). Wenn die Items das definierte Konstrukt gleichmässig repräsentieren, ist die Inhaltsvalidität gegeben. Für die Inhaltsvalidität gibt es keine quantifizierbare Grösse. Sie wird ausschliesslich inhaltlich begründet. Mit Hilfe der Konstrukt- und Kriteriumsvalidität kann geprüft werden, inwiefern ein Fragebogen mit anderen relevanten Konzepten zusammenhängt. Beide Arten der Validität werden unter Angabe von Korrelationskoeffizienten deklariert.

Mit der *Konstruktvalidität* wird geprüft, ob bei zwei Fragebogen, die das gleiche Konstrukt erfassen – z.B. Extraversion beim HEXACO-PI-R (Lee & Ashton, 2004) verglichen mit der Extraversion aus einem Big Five Fragebogen – die Ergebnisse miteinander korrelieren. Je höher die Korrelation desto mehr Ähnlichkeit besteht zwischen den mit den beiden Fragebogen erfassten Konstrukten. Auch wenn mit unterschiedlichen Fragebogen dieselben Konstrukte gemessen werden (z.B. RIASEC-Dimensionen mit AIST – Version 3 (Bergmann & Eder, 2018) oder Explorix (Joerin Fux, Stoll, Bergmann & Eder, 2013)), ergeben sich bei solchen Überprüfungen meist keine kompletten Übereinstimmungen im Sinne von Korrelationen nahe bei 1. Die Items der Fragebogen sind zu unterschiedlich, auch wenn diese dasselbe Konstrukt erfassen. Für die Beratung bedeutet dies, dass man sich der Eigenheiten der unterschiedlichen Fragebogen bewusst sein sollte, um diese gezielt einsetzen zu können.

Die *Kriteriumsvalidität* bezieht sich auf ein Kriterium, welches in einem Zusammenhang mit dem gemessenen Konstrukt steht und in der Vergangenheit oder in der Zukunft liegen kann. So werden Intelligenztests oft mit dem Kriterium der Schulnoten oder des Berufserfolges validiert. Mit der Kenntnis solcher Zusammenhänge können in der Praxis Hinweise auf eine mögliche Eignung erschlossen werden. Solche Rückschlüsse dienen als Hinweise, sind aber aufgrund der stets vorhandenen Messfehler der Verfahren, wie auch der meist nicht allzu hohen Korrelationen mit den untersuchten Kriterien nur mit Vorsicht anzudenken.

Für die Praxis ist es aus verschiedenen Gründen wertvoll, sich mit den unterschiedlichen Validierungen der Verfahren auseinanderzusetzen. Die Informationen gehören in das zum Instrument zugehörige Handbuch zu. Nur so wissen Beraterinnen, dass das Instrument das misst, was diese gemessen haben wollen und zudem lernen sie die Zusammenhänge zwischen verschiedenen untersuchten Konstrukten besser kennen. Als Beispiel: Die Konstruktvalidierung des Explorix (Joerin Fux et al., 2013) zeigt eine positive Korrelation der RIASEC-Dimension A „kreativ“ mit der Dimension „Offenheit“ in einem Persönlichkeitsfragebogen. Zeigt sich nun in der Praxis bei einer Klientin im Interessensfragebogen eine hohe Ausprägung in der Dimension A sowie im Persönlichkeitsfragebogen eine tiefe Ausprägung in der Dimension Offenheit, so könnte dies ein Hinweis für ein mögliches Spannungsfeld sein. Solche Hypothesen können im Beratungsgespräch aufgenommen und mit der Klientin diskutiert werden.

Skalierung

Neben Reliabilität, Validität und Objektivität wird der *Skalierung* vielfach kein spezieller Fokus als Hauptgütekriterium beigemessen. Dies, obwohl Bühner (2011, S. 67/68) zu recht festhält, dass die Skalierung, mit welcher eine gültige Verrechnungsvorschrift und dadurch ein gültiges Messmodell gefordert und überprüft wird, essenziell ist für die valide Messung eines Konstruktes. Gültige Messmodelle werden im Rahmen der probabilistischen Testtheorie explizit überprüft. Bei Fragebogen oder Tests, die nach der klassischen Testtheorie konstruiert worden sind, geht man implizit davon aus, dass der Messvorgang und das Messinstrument für unterschiedliche Personen einer Gruppe dasselbe misst. Dies, obwohl diese Annahme, welche eine zentrale Bedingung für den Einbezug von

Normstichproben bei Fragebogen und Tests darstellt, von der klassischen Testtheorie nicht überprüft wird.

Literatur

- Asendorpf, J. B. (2015). *Persönlichkeitspsychologie für Bachelor*. Berlin, Heidelberg: Springer.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.
- Bauer, J. (2016). *Das Gedächtnis des Körpers. Wie Beziehungen und Lebensstile unsere Gene steuern*. München: Piper.
- Bergmann, C. & Eder, F. (2018). *Allgemeiner Interessen-Struktur-Test - Version 3 (AIST-3)*. Göttingen: Hogrefe.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bühner, M. & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler* (2. Auflage). Hallbergmoos: Pearson.
- Fredrickson, B. L. (2003). The value of positive emotions. *American Scientist*, 91(4), 330–335. <https://doi.org/10.1511/2003.4.330>
- Gable, S. L. & Haidt, J. (2005). What (and why) is positive psychology? *Review of General Psychology*, 9(2), 103–110. <https://doi.org/10.1037/1089-2680.9.2.103>
- Joerin Fux, S., Stoll, F., Bergmann, C. & Eder, F. (2013). *Explorix - Das Werkzeug zur Berufswahl und Laufbahnplanung*. Bern: Huber.
- Lee, K. & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39, 329–358. <https://doi.org/10.1207/s15327906mbr3902>
- McGilchrist, I. (2009). *The master and his emissary: The divided brain and the making of the western world*. New Haven and London: Yale University Press.
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2. Auflage.). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-20072-4_2
- Neyer, F. J. & Asendorpf, J. B. (2018). *Psychologie der Persönlichkeit*. Berlin: Springer.
- Rammsayer, T. & Weber, H. (2016). *Differentielle Psychologie - Persönlichkeitstheorien*. Göttingen: Hogrefe. https://doi.org/10.1007/978-3-7091-3424-5_4
- Rorschach, H. (1921). *Rorschach-Test. Psychodiagnostik Tafeln*. Bern: Hans Huber.
- Roth, G. (2016). *Persönlichkeit, Entscheidung und Verhalten: Warum es so schwierig ist, sich und andere zu ändern*. Stuttgart: Klett-Cotta.
- Ruch, W. & Proyer, R. T. (2015). Mapping strengths into virtues: the relation of the 24 VIA-strengths to six ubiquitous virtues. *Frontiers in Psychology*, 6(October). <https://doi.org/10.3389/fpsyg.2015.00460>
- Schreiber, M. [Marc Schreiber]. (29.11.2017). *IAP Kompakt «Persönlichkeitspsychologie von Freud bis zu den Neurowissenschaften»*. Verfügbar unter: <https://youtu.be/nvW21PG6Z2U>
- Schreiber, M. (im Druck). *Wegweiser im Lebenslauf*. Stuttgart: Kohlhammer.
- Schreiber, M., Mueller, I. M. & Morell, C. (2018). *Handbuch HEXACO Personality Inventory-Revised (HEXACO-PI-R)*. Zürich: IAP Institut für Angewandte Psychologie der ZHAW. Verfügbar unter: <https://laufbahndiagnostik.ch/de/download>
- Seligman, M. E. P. (2011). *Flourish: A visionary new understanding of happiness and well-being*. New York, NY: Free Press.
- Seligman, M. E. P. & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55(1), 5–14. https://doi.org/10.1007/978-94-017-9088-8_18
- ZHAW/IAP. (2019). Plattform Laufbahndiagnostik. Zugriff am 2.5.2019. Verfügbar unter: <https://laufbahndiagnostik.ch>