

Anmerkungen zur Beschwerdvalidierung

Wolfgang Palm

1. Was ist eine Beschwerdvalidierung?

In der Klasse (Menge) von Proband*innen, die sich einer gutachterlichen Untersuchung unterziehen, äußert eine Teilklasse solche Beschwerden, die in Intensität und/oder Umfang über ihre *wirklichen* Beschwerden hinaus gehen. Dieser Eingangsthese ist der Anlass für das vorliegende Papiers. Der Grund für diese unwirklichen Beschwerden - er gehört nicht zu dieser Behauptung, sondern macht sie nur einleuchtend - ist ein geldwerter Vorteil, welcher durch ein Gutachten erwirkt werden kann. Die Behauptung selbst stellt eine zweifach beschränkte, allgemeine Aussage dar, einmal bezüglich der Umstände, zum anderen bezüglich der Personen. Die Frage ist nun: Wie lässt sich feststellen, ob eine Proband*in zur Teilklasse gehört oder nicht.

Beschwerden körperlicher oder psychischer Art werden von einer Person - Proband*in genannt - vorgetragen, um eine Entschädigung (meistens Geld) oder eine Vergünstigung zu erlangen, z.B. über einen Grad der Behinderung oder eine Minderung der Erwerbsfähigkeit. Die Adressaten dieses Verlangens sind Gerichte, Versicherungen, Unfallkassen und Berufsgenossenschaften, manchmal auch andere Personen. Führt dieses Verlangen auf Grund rechtlicher Einlassungen zu einem Verfahren, in dessen Verlauf die Proband*in sich einer Begutachtung unterzieht, so stellt sich insbesondere bei psychischen Beschwerden sehr oft die Frage: Sind die vorgetragenen Beschwerden auch valide? Will heißen: Sind die Beschwerden *wirklich* so, wie die Proband*in sie vorträgt? Das ist die Wahrheitsfrage.

Leider lässt sich die Frage, ob die Beschwerden der Proband*in *wirklich* so sind, wie sie sagt und vorzeigt, nicht auf die Weise beantworten wie die Frage, ob der Mond am Himmel steht oder nicht. Denn da die Beschwerden subjektiv sind, gibt es kein Objekt, das man gemeinsam beobachten könnte. Auch ist das Innere - mit welchen psychologischen Theorien man es auch zu erfassen versucht - nicht in Gestalt von Objekten zugänglich, die nur noch benannt werden müssten. Das Subjektive der Proband*in äußert sich - nicht nur während der Untersuchung - in sprachlichen Ausdrücken sowie in Mimik und Gesten und physiologischen Prozessen. Eine Gutachter*in mag versucht sein, dahinter bestimmte Vorstellungen, Erinnerungen, Empfindungen und Gefühle zu *erkennen*, doch deren notwendige *Benennung* im Gutachten führt wiederum zu sprachlichen Ausdrücken, wenn auch zu fachsprachlichen *Beobachtungsausdrücken* - mal abgesehen von der Frage, ob es ein *Erkennen* ohne *Benennung* überhaupt geben kann. Das Nachdenken der Gutachter*in über die Schwierigkeiten der Wahrheitsfindung verhilft dieser vielleicht zu 'theoretischen Einsichten', die zwar logisch fehlerfrei sein können, praktisch aber nicht weiter helfen werden. Denn das, was oft naiverweise von einem Gutachten erwartet wird, die *wirkliche* psychische Verfassung oder die *wirklichen* psychischen Einschränkungen einer Proband*in festzustellen, ist just jenes, das sich nicht feststellen lässt. So rückt an die Stelle der Wahrheitsfrage die Frage nach der Wahrhaftigkeit - seitens der Proband*in - und seitens der Gutachter*in die Frage nach der Glaubwürdigkeit der Aussagen der Proband*in. Sind Tränen und

Zittern ein Beweis für ihre Glaubwürdigkeit? Es ist auch sinnlos und nicht hilfreich, die Proband*in zu befragen, ob sie auch die Wahrheit über ihre Beschwerden gesagt habe, denn wie könnte diese wissen, was ihre *wirklichen* psychischen Beschwerden sind, zumal diese zwecks Feststellung *benannt* werden müssen? Zudem braucht, wer daraufhin 'nein' antwortet, erst gar nicht untersucht zu werden und wer 'ja' sagt, lügt vielleicht, d.h. antwortet nicht wahrhaftig. Wie auch immer - die Begutachtung des Psychischen verbleibt innerhalb der Grammatiken und Semantiken der Fach- und Alltagssprache.

In der Fachliteratur geht es daher meist nur noch darum, ob die geltend gemachten Beschwerden glaubwürdig oder *authentisch* sind. Oder ob es sich um eine *nicht-authentische, negativ verzerrte* Beschwerdendarstellung handelt. Zur Feststellung von *nicht authentischen* Beschwerden stellt die psychiatrische Fachsprache Begriffe wie Simulation, Aggravation und Verdeutlichung zu Verfügung: *Simulation* ist das bewusste Vortäuschen einer krankhaften Störung zu bestimmten Zwecken. *Aggravation* ist die bewusst intendierte, verschlimmernde bzw. gravierendere Darstellung einer vorhandenen Störung zum Zweck der Erlangung meist geldwerter Vorteile. Hingegen handelt es sich bei einer *Verdeutlichung* um den bewussten Versuch, eine Gutachter*in vom Vorhandensein der Beschwerden zu überzeugen. Zwar machen die gewählten Worte den Unterschied der drei Begriffe klar, doch ist die Feststellung von Inhalten und Differenzen zwischen den Begriffen alles andere als einfach. Die erste große Schwierigkeit liegt bereits in der Verwendung des Wortes *bewusst*, das im gegebenen Kontext als *absichtlich und vorsätzlich* zu verstehen ist. Wie aber stellt eine Gutachter*in zweifelsfrei fest, dass es *bewusst* geschieht?

Sieht man sich die von Fachleuten aufgelisteten Kriterien für das Zutreffen dieser Begriffe im einzelnen an, so finden sich darin Adjektive wie vage, unpräzise, ausweichend, vorher fehlend - nun vorhanden, gegensätzlich, zu verschieden u.a. Sodann vielerlei Diskrepanzen zwischen einem vorgetragenen Zustand A und einem fachlich zu erwartenden B, mangelnde Objektivierbarkeit¹ und mangelnde Erklärbarkeit durch die jeweiligen Fachtheoreme. Nicht zuletzt kommt dann die Subjektivität der Gutachter*in ins Spiel: Die Klagen lassen sie unbetroffen, erzeugen ein Gefühl des Unechten, sind nicht einfühlbar und vermitteln keine Erlebensqualität, klingen wie angelesen und angelernt, wie aus populären Darstellungen entnommen. Die Zuschreibung dieser Kriterien erfolgt durch die Gutachter*in, stellt also deren Einschätzung dar, und damit alles andere als eine Messung!

Soll die *Einzelfallanalyse* Gutachten eine wissenschaftliche Arbeit sein, müssten ihre Ergebnisse indes objektiv und reproduzierbar sein. Das sind die oben genannten Erlebensqualitäten der Gutachter*in aber nicht. Auch nicht in einer 'schwachen' Fassung der beiden Forderungen als intersubjektive Übereinstimmung von Gutachter*innenaussagen und wiederholbare Überprüfung. Denn der Moment des Erlebens ist in unwiederbringlicher Weise längst schon perdu. Gleiches gilt für das Kriterium der *Nachvollziehbarkeit* einer Beschwerde; auch dieses ist von der Subjektivität der Gutachter*in nicht abzulösen. Ein weiterer Grund, weshalb die an das situative Erleben der Gutachterin gebundenen Kriterien keine sind oder sein dürften, liegt im *be-rechtigten* Anliegen, genauer in der be-rechtigten Zwecksetzung der Proband*in, die unabhängig davon besteht, ob

1 Dabei besteht geradezu chronische Unklarheit in der Operationalisierung dieses Terminus

deren Beschwerden als authentisch oder nicht beurteilt werden. Drittens ist das Ziel von Gutachten, in denen eine Beschwerdvalidierung nötig oder verlangt ist, eher selten die Bestimmung der Krankheit, auf die die Beschwerden verweisen. Überwiegend geht es um die Bestimmung von Fähigkeits-, Leistungs- und Teilhabebeeinträchtigungen, für die es keine einfache, lineare Herleitung selbst aus authentischen Beschwerden gibt, schon gar nicht im psychischen Bereich. Letztlich handelt es sich bei deren Bestimmung um Einschätzungen, die, weil sie keine Messungen sind, unvermeidlich mit Vagheit und Unschärfe einher gehen. Dennoch sollte die Datengrundlage, auf der solche Einschätzungen erfolgen, eher stabilen Pfeilern ähneln als unscharfen und schwankenden Erlebensqualitäten einer Gutachter*in.

2. Fragebögen zur Beschwerdvalidierung.

Psychologen versuchen dem Problem der Objektivierung und Reproduzierung von *nicht authentischen* Beschwerden mit ihrer fachspezifischen wissenschaftlichen Methodik beizukommen². Die darauf beruhenden Instrumente sind Fragebögen und Tests. Fragebögen führen zu Selbstbeschreibungen anhand einer vorgegebenen Anzahl von anzukreuzenden Aussagen (Items), wobei die Angaben einer Proband*in mit denen einer Normierungstichprobe von Personen verglichen werden, die diese Items bereits bearbeitet haben. Das Ergebnis ist eine auf die Items eingeschränkte Selbstaussage mit einer Platzierung in Normwerten innerhalb der Stichprobe³. Die üblicherweise in Klinikberichten und Gutachten verwendeten klinischen Fragebögen und Persönlichkeitsinventare sind prinzipiell ebenso wenig gefeit vor *nicht authentischen* Angaben einer Proband*in wie mündliche Befragungen (Anamnese, Exploration). Während jahrelanger Anwendung solcher Fragebögen musste ich immer wieder feststellen, dass Fragebogenergebnisse für Begutachtungen noch eher zu Extremen tendieren als face to face Befragungen. Ja, manche Proband*innen neigen im Gespräch mit einem Psychologen dazu überwiegend körperliche Beschwerden zu benennen, während sie sodann in den klinischen Beschwerdebögen Höchstwerte für Angst, Depression, Somatisierung oder Zwangssymptome anführen, die sie zuvor mit keinem

2 Diese hat sicherlich ihre immanenten Schwächen, die indes in der Literatur bereits ausführlich abgehandelt worden sind (Palm 1991, Herzog 1984). Hier ist vor allem zu erwähnen, dass der Aufbau von statistisch behandelten Zahlen über semantischen Relationen keine Messhandlung ist, Fragebögen also nicht *messen* und insofern keine objektiven Daten liefern, als sie ihren Gegenstand bzw. dessen Dimensionen selbst erzeugen.

3 Dazu kommt die chronische Schwierigkeit des Verstehens: Da es keine feststehenden Wortbedeutungen gibt, diese sich für die meisten Leute eher pragmatisch aus dem gewohnten Lebensvollzug ergeben haben, führt die Anweisung, dass ein Item genau das bedeutet, was jemand darunter versteht, zu einer nicht einzuschätzenden Bedeutungsunschärfe, weil nicht anzunehmen ist dass zufällig alle Teilnehmer der Normstichprobe die Items gleich verstanden und nur quantitativ verschieden codiert haben, insbesondere da die Bedeutung der Items sich auf das je eigene Erleben bezieht. Doch durch das Konstruktionsverfahren des Bogens werden die Verstehensunterschiede in den Ausprägungsunterschieden aufgesaugt, Qualität derart in Quantität verwandelt. Denn individuelles *Verstehen* ist eben, wie immer auch seine Quantifizierung versucht worden ist, im Grunde keine messbare Größe. Der naive Behaviorismus versuchte das Erlernen von Wortbedeutungen durch hinweisende Gesten zu erklären: Vater deutet auf einen *Stuhl* und sagt wiederholt "Stuhl, Stuhl, Stuhl ..." So lerne das Kind die Bedeutung des Wortes Stuhl. Ich erinnere mich bei Wittgenstein den Einwurf gelesen zu haben, dass hierzu das Kind die Bedeutung der hinweisenden Geste des Vaters verstehe müsse, was heisst, dass diese Erklärung nichts taugt.

Wort erwähnt haben.

Die Brauchbarkeit klinischer Fragebögen zur Bestimmung der *wirklichen* Beschwerden einer Proband*in ist höchst zweifelhaft. Aufschlussreich aber sind sie in der kombinierten Verwendung mit einer anderen Art von Fragebögen, die der Feststellung *negativer Antwortverzerrungen* dienen, aus denen ein mehr oder weniger starker Hinweis auf *nicht authentische* Beschwerdenschilderung hervor geht (kurz: BSV-Bögen). Die Stärke dieses Hinweises erfolgt über sog. Cut-Off-Werte ⁴. Nehmen wir an, einen Proband*in sei darin stark auffällig geworden mit einer mehr als 90%igen Sensitivität und einer nur 5%igen Selektivität. Was folgt daraus? Man kann eine bedingte Wahrscheinlichkeit ⁵ dafür angeben, dass eine zufällig aus der Klasse aller Proband*innen heraus gegriffene Proband*in zur Teilklasse derjenigen gehört, die ihre Beschwerden negativ übertrieben darstellen oder sogar Beschwerden angeben, an denen sie nicht leiden oder nach derzeitigem Wissenstand nicht leiden können. Das allein ist aber kein Beweis, dass es so ist, ein solcher ist auch nur als eine Sammlung von Indizien möglich. Denn aus der Bejahung von 'Beschwerden', die in kein bekanntes Krankheits- bzw. Störungsbild passen, folgt logisch nicht, dass sie nicht existieren ⁶. Es gab auch keine schwarzen Schwäne, bevor man sie entdeckte; die meisten sind vermutlich nach wie vor weiß.

Unter- oder überschreiten also die Ergebnisse in einem solchen BSV-Bogen bestimmte Cut-Off-Werte und ergibt sich daraus entweder kein oder sogar ein starker Verdacht auf negative Antwortverzerrungen – und zwar nicht für den spezifischen BSV-Bogen, denn darin liegen sie ja faktisch vor – sondern bezüglich der übrigen schriftlich oder mündlich vorgetragene Beschwerden, so sind dennoch zwei Folgerungen unzulässig: Weder ergibt sich aus unauffälligen Ergebnissen im BSV-Bogen, dass in der übrigen Beschwerdendarstellung keine negativen Antwortverzerrungen enthalten sind, noch folgt aus auffälligen Ergebnissen im BSV-Bogen der 'Beweis' von Aggravation oder Simulation. Deren Zuschreibung ist nämlich eine Beurteilung durch die Gutachter*in, die sich auf möglichst viele Daten stützen sollte; es ist also ein Urteil, d.h. eine Behauptung, die sich Indizien stützt. Starke Hinweise auf Aggravation könnten beispielsweise sein: Überschreiten eines hohen Cut-Off-Werts im BSV-Bogen, keine Angaben über psychische Beschwerden in der mündlichen Befragung für ein psychologisches Gutachten, aber extreme Angaben im klinischen Beschwerdefragebogen. Dazu noch Angaben über den täglichen Lebensablauf, die mit den angegebenen Beschwerden kaum zu vereinbaren sind und die möglicherweise ihrerseits mit Angaben in Vorberichten nicht überein stimmen sind. Der wesentliche Gesichtspunkt bei der Beurteilung dieser 'Sachlage' ist aber nicht die Glaubwürdigkeit oder Unglaubwürdigkeit einzelner Angaben, sondern die *Inkohärenz* der Gesamtheit dieser Angaben: sie fügen sich nicht zu einem

4 Es versteht sich von selbst, dass zu dieser Art von Fragebögen und ihrer Anwendung hier keine ins Detail gehenden Angaben gemacht werden.

5 $P(\text{bed}) = p \cdot \text{sens} / [p \cdot \text{sens} + (1 - p) \cdot (1 - \text{spez})]$; p ist die Prävalenz

6 Zu den grundlegenden Einsichten der Wissenschaftstheorie gehört jene, dass aus einer Hypothese, d.i. eine Behauptung in Satzform, weder auf die Existenz noch auf die Nichtexistenz der behaupteten Gegenstände geschlossen werden kann. Existenznachweise haben so weit als möglich durch davon unabhängige Handlungen zur Feststellung (speziell also Messhandlungen) zu erfolgen. Es gibt keine logischen Schlüsse von einem Sollen auf ein Sein! Die oft zu lesende Formulierung "Im Umkehrschluss folgt ..." folgert meistens Unsinn! Existenznachweise sind nicht zu folgern.

zusammenhängenden Bild! Die *Inkohärenz*⁷ wird durch Angaben und Antworten herbei geführt, die nicht zusammen passen, Widersprüche oder Brüche offenbaren.

3. Tests zur Beschwerdvalidierung

Tests stellen die Probanden vor Aufgaben, die zu bearbeiten und zu lösen sind. Manchen erinnert das an Schulaufgaben – nicht zu Unrecht. Die wohl bekanntesten Testart heißen *Intelligenztest* (IQ-Test). Die Definitionen von Intelligenz schwanken zwischen dem Bonmot "Intelligenz ist, was der IQ-Test misst" und der weiten und damit vagen Formulierung "Intelligenz ist die generelle Fähigkeit Probleme effizient richtig zu lösen". Die IQ-Forschung hat eine lange Tradition und die daraus hervor gegangenen Tests führen zu den vergleichsweise stabilsten und am besten zu reproduzierenden Ergebnissen auf Basis recht großer Stichproben (1.000 bis mehr als 10.000 Personen). IQ-Werte korrelieren hoch mit beruflichem Erfolg, allerdings in Abhängigkeit von kovariierenden Persönlichkeitsfaktoren wie *Belastbarkeit* (Psychische Stabilität) und *Gewissenhaftigkeit* (Einsatzbereitschaft und systematisches Arbeiten).

Die in den letzten Jahren zunehmend von Versicherungen, Berufsgenossenschaften und Unfallkassen verlangten *neuropsychologischen* Testungen enthalten in erster Linie Tests zur Prüfung von *Aufmerksamkeitsprozessen*, *Gedächtniskapazitäten*, *Handlungsplanung* und *Handlungskoordination* (Exekutive Funktionen), dazu noch *BSV-Tests*, die oft Gedächtnistests sind oder Reaktionsgeschwindigkeiten messen und Fehler zählen⁸. Auch BSV-Tests haben die bereits beschriebenen Cut-Off-Werte, jenseits dessen ein Ergebnis negativ auffällig wird. Doch nicht wegen negativer Antwortverzerrungen und unglaublichen Beschwerden - diese Ausdrücke ergeben im Zusammenhang mit BSV-Tests keinen Sinn - sondern wegen *Anstrengungsvermeidung* bzw. *mangelnder Mitarbeit*. Genaugenommen ergibt das auffällige Ergebnis eines BSV-Tests einen Warnhinweis hinsichtlich schlechter Mitarbeit in anderen Tests, sofern für diese negativ auffällige Ergebnisse vorliegen. Aber auch für Ergebnisse im Normalbereich könnte eine Probandin sich zu wenig angestrengt haben, allerdings ist ein solcher Nachweis schwierig zu erbringen.

Bei mangelnder Anstrengung kann davon ausgegangen werden, dass das Testergebnis eine

7 Die Frage, ob trotz nachgewiesener negativer Antwortverzerrungen dennoch eine psychische Störung vorliegt, erscheint mir, obwohl in den meisten Anforderungen der Auftraggeber von hoher Priorität, als eher nachrangig. Denn aus der Diagnose einer psychischen Störung folgt unmittelbar nichts über die Fähigkeitseinschränkungen einer Proband*in im Alltags-, Arbeits- und Berufsleben. Kernanliegen vieler Guachten ist es aber, argumentativ einen stringenten Zusammenhang zwischen Beschwerden und Fähigkeitseinschränkungen herzustellen.

8 *Zählen* ist noch nicht *Messen*! Reaktionszeittests messen die physikalische Basisgröße 'Zeit'. Ein IQ-Test misst nicht. Der IQ einer Proband*in ergibt sich aus der Summierung von richtig gelösten Aufgaben, die mit einer Zahl gewichtet werden (meistens '1') und der Platzierung dieser Summe in einer Verteilung von Normwertem. In einem Messprozess hingegen wirkt ein nach menschlichen Zwecken zu einem Instrument umgearbeiteter Naturstoff auf einen nicht bearbeiteten Naturstoff, was je nach Präzision des Messvorgangs ein genaueres Ergebnis in realen Zahlen ergibt. Das Metaphysisch-Geheimnisvolle besteht darin, dass dieser Vorgang, der uns alltäglich so selbstverständlich vorkommt, überhaupt gelingt. Denn wodurch sollte dies garantiert sein? Die Gewichtung von richtig gelösten Rechenaufgaben oder angekreuzten Items mit einer Zahl '1' hat nichts Geheimnisvolles, sie ist eine pragmatische Setzung, mit der anschließend nach bekannten Rechenregeln verfahren wird, deren Ergebnis überhaupt nicht gemessen werden kann.

Leistung (eine Performance) unterhalb der Fähigkeit der getesteten Probandin darstellt. Diese Formulierung beinhaltet, dass das Fähigkeitsniveau durch einen Test festgestellt werden kann, sofern sich die Proband*in optimal anstrengt. Wozu sie auch motiviert sein muss ⁹. Eine Testkonstruktion beinhaltet also die Konzeption einer *Fähigkeit* oder beruht sogar darauf. Die Betonung liegt auf *Fähigkeit* und nicht etwa auf *Funktion*. Eine Funktion ist ein theoretisches Konstrukt und steht beispielsweise in einem Lehrbuch der Neuropsychologie oder der Medizin. Eine Funktion kann aber nicht unmittelbar durch einen Test festgestellt - oder wie man halt alltäglich so sagt 'gemessen' werden. Denn vor der Computertastatur sitzt ja nicht das Gehirn oder gar ein Teil desselben – was sollte es dort auch ohne Arme tun? - sondern der ganze Mensch mit Kopf, Leib, Armen und Beinen und seinen Sinnen sowie seinen Gedanken, Emotionen und Launen. Menschen aber schreiben sich gegenseitig *Fähigkeiten* (Vermögen) seit Jahrtausenden zu, Maschinen haben Funktionen, auch Mathematiker benutzen diesen Begriff eingehend. Der Vorteil bei der Verwendung psychologischer Tests ist also, dass die Ausprägung von Fähigkeiten festgestellt werden kann, optimale Mitarbeit vorausgesetzt. Das Bonmot über den IQ-Test lässt sich insofern erweitern.

Reaktionszeit-, Aufmerksamkeits-, Gedächtnis-, Handlungskordinations- und IQ-Tests haben zwar ihre je eigenen 'Theorien' oder sollen bestimmte Funktionen prüfen, doch werden sie – nicht nur in einer gutachterlichen Untersuchung - von einem Mensch durchgeführt und deshalb können die Testergebnisse voneinander nicht unabhängig sein. Von den Testkonstrukteuren wird diese Abhängigkeit als divergente und konvergente Validität der Tests bezeichnet und in Korrelationskoeffizienten quantitativ ausgedrückt. Indes sind in der Untersuchung einer Probandin noch stärkere und andere Zusammenhänge zu berücksichtigen: Reihenfolgeeffekte, Ermüdungerscheinungen, genaues Arbeiten, aber auch einzelne Schwächen. Banalerweise kann jemand, der unaufmerksam ist, unkonzentriert reagiert, keine guten Gedächtnisleistungen erbringen. Denn was nicht aufgenommen wird, kann nicht erinnert werden. Sollten also die Aufmerksamkeitsleistungen durchwegs schlecht, die geprüften Gedächtnisleistungen dennoch gut ausfallen, so ist das *inkonsistent*. Anstrengungsvermeidung führt ebenfalls zu inkonsistenten Ergebnissen; auffällige BSV-Tests verweisen auf eine notwendig gewordene *Konsistenzprüfung* von Testergebnissen.

Als Resultat der Überlegungen dieses Papiers ergeben sich *Kohärenz* und *Konsistenz* als die zentral wichtigen Kriterien für die Beschwerdenuvalidierung, bei deren Anwendung die aus der Untersuchung hervor gehende Datenlage als Basis zu dienen hat. Dadurch wird zwar der Einfluss der Subjektivität der Gutachter*in nicht ausgeschlossen, aber zumindest vom Ansatz her reduziert.

Autor und Copyright: Dr. Wolfgang Palm, www.psy-gutachten.de
Dipl.-Psych., Dipl.Phys., Psychotherapeut
Sachverständiger der Psychotherapeutenkammer BaWü

⁹ Das 'Gift', welches eine solche Motivation oft lähmt, ist die Hoffnung auf einen geldwerten Vorteil durch ein Gutachten auf Basis schlechter Testergebnisse.

Literatur

Dohrenbusch R, Schneider W, Merten T (2016). Zur Bedeutung der Testpsychologie bei ICF-orientierter Begutachtung. In W Schneider, R. Dohrenbusch et al (Hrsg), Begutachtung bei psychischen und psychosomatischen Erkrankungen, 2. überarb. u. erw. Aufl., Verlag Hogrefe

Janich P (2000). Was ist Wahrheit? Eine philosophische Einführung. Verlag C.H. Beck

Herzog W (1984). Modell und Theorie in der Psychologie. Verlag Hogrefe

Heubrock D, Scholl H, Petermann F (2013). Die differentielle Validität neuropsychologischer Testverfahren zum Nachweis nicht-authentischer Störungen. *Z. f. Neuropsychologie*, vol.24, 229-238

Merten M, Dettenborn H (2009). Diagnostik der Beschwerdenuvalidität. Deutscher Psychologen Verlag

Merten T, Dohrenbusch R. (2016). Psychologische Methoden der Beschwerdenuvalidierung. In W Schneider, R. Dohrenbusch et al (Hrsg), Begutachtung bei psychischen und psychosomatischen Erkrankungen, 2. überarb. u. erw. Aufl., Verlag Hogrefe

Merten T, Giger P, Merckelbach H, Stevens A (2019). Handbuch zum Self-Report Symptom Inventory – deutsche Version, Verlag Hogrefe

Miller M (2017). Begutachtung im Fachgebiet Psychiatrie/Psychotherapie. In J Francke, A Gagel, D Bieresborn (Hrsg), Der Sachverständigenbeweis im Sozialrecht, 2. Aufl. Verlag Nomos

Palm W (1991). Zur Validität physikalischer und psychologischer Messprozesse. Verlag Haag & Herchen

Palm W (2018). Genauigkeit in der Aufmerksamkeitstestung.
<https://www.wopalm.com/wp-content/uploads/2018/04/Genauigkeit.pdf>

Spektrum Kompakt (2020). Die Vermessung der Intelligenz, digitale Ausgabe 01/2020

Williams B (2013). Wahrheit und Wahrhaftigkeit. Suhrkamp Taschenbuch Wissenschaft 2050

Stand: Januar 2020