

Stringenz und Kohärenz statt Autoritätsbeweis

Reflexionen über das Schreiben psychologischer Gutachten

Wolfgang Palm

Stringentes Argumentieren

Mehr als 12 Jahre Tätigkeit in Sachen psychologische Gutachten auf Basis von Tests und Fragebögen für Versicherungen, Sozialgerichte und Unfallkassen haben lehrreiche Erfahrungen beim Schreiben von mehr als 170, teils sehr ausführlichen Gutachten erbracht. Zum großen Teil handelt es sich um Gutachten zur Feststellung der Minderung von Berufs- oder Erwerbsunfähigkeit. Naiv und gutwillig begonnen, beruhen die Gutachten mittlerweile auf einer im Detail ausgefeilten Argumentationweise, die von der Analyse gewonnener Daten über die genaue Beobachtung der Mitarbeit der Probanden samt einer differenzierten Beschwerdvalidierung hin zu einer Einschätzung von Fähigkeiten und deren Einschränkungen führt. Begutachten in diesem Sinne heisst konsequent und konsistent *argumentieren*.

Befragung, Tests, Fragebogen, Beobachtungen und Vorberichte ergeben Daten, teils von sehr unterschiedlicher Qualität. Doch vor allem liefern sie zuerst einmal Daten, die meist auf subjektiven Angaben beruhen, wie es in der Psychologie und auch der Psychiatrie überwiegend der Fall ist. Doch Daten allein beweisen noch nichts, sie sind Stützen, die durch eine stringente Argumentation miteinander verbunden werden müssen. Eine solche Argumentation hat zu einer Herleitung der schlussendlichen Beurteilungen zu führen, derentwegen ein Gutachten in Auftrag gegeben worden ist. Das mag selbstverständlich klingen, doch das Durchforten von inzwischen tausenden von Aktenseiten hat gezeigt: Stringentes Argumentieren ist eher selten, Fähigkeitseinschränkungen werden vielfach nicht hergeleitet, sondern qua Fachautorität eingeschätzt. Am Ende von Klinikberichten und auch von seitenlangen Gutachten benötigen solche Einschätzungen eine Handvoll Zeilen. Daher kann es zu völlig gegensätzlichen 'Schlussfolgerungen' in zwei Gutachten über ein und dieselbe Person kommen, oft weil diese 'Schlussfolgerungen' eben keine sind, sondern bloße Einschätzungen durch die jeweiligen Gutachter.

Fähigkeiten beurteilen

Erfreulicherweise gibt es seit einigen Jahren Bemühungen, diesem Mangel abzuhelpfen. Dazu gehört zuerst einmal die verbindliche Festsetzung von *Fähigkeiten*, die einem menschlichen Wesen zuzuschreiben sind. Was sind sie und welche gibt es? Der erste Blick in das von der WHO herausgegebene Internationale Klassifikationssystem der Fähigkeiten (ICF) überschwemmt den Leser mit einem Meer von Einteilungen verschiedenster Arten und Ausdehnungen, so als ob ein Bildschirm aus Pixeln unterschiedlicher Größe bestünde, die sich zudem noch gegenseitig überlappen. Im psychischen Bereich gibt es hierzu eine handhabbare Reduktion auf 13 Fähigkeiten. Indes - diese dreizehn sind den Auftraggebern selten genug. Meist sind sie nur ein Unterpunkt in einer drei bis vierseitigen Liste von Anforderungen, deren Beantwortung durch ein Gutachten erfüllt werden sollte. Unmöglich zu erfüllende Anforderungen verführen aber zu bloßen Einschätzungen qua Fachautorität. Eingeschränkter, konsequenter und

zielgerichteter sollte daher eine vernünftige Anforderung an Gutachten lauten, nicht aber ausgedehnter und beliebiger. Nicht die Berufung auf die mehrheitliche Meinung von Fachautoritäten sollte in einer wissenschaftlichen Einzelfallanalyse, wie sie ein Gutachten darstellt, zur Begründung von Schlussfolgerungen dienen, sondern eine stringende, nachzuvollziehende Argumentationsweise.

Tests und Fragebögen

In den letzten Jahren werden in Klinikberichten und psychiatrischen Gutachten öfters auch die Ergebnisse psychologischer 'Tests' berichtet, doch es sind keine *Tests*, sondern überwiegend *Fragebögen*. Der Unterschied, der jedem universitär ausgebildeten Psychologen geläufig ist, wird in der Bericht- und Gutachtenpraxis aber leichthin übergangen. Auch Ausbilder für Neuropsychologie vergessen gelegentlich diesen Unterschied. Obwohl nach derselben, meist nach der klassischen Testtheorie konstruiert, ist er von erheblicher Tragweite: *Tests* messen - sofern sie überhaupt *messen* - Fähigkeitsausprägungen, wobei die Fähigkeit zugleich durch den Test, also das Gesamt seiner Aufgaben, definiert wird ¹. Bekanntlich lautet die zirkuläre Definition der Intelligenz: Intelligenz ist, was der Intelligenztest misst. Doch *misst* der Intelligenztest überhaupt? Die Summierung richtig beantworteter und mit je einem Punkt eingestufte Items ist keine Messung. Wer sein Kleingeld zusammenzählt, misst ebenfalls nicht, er oder sie zählt. Diejenige Größe, die bei der Testdurchführung gemessen werden könnte, weil sie mittels eines Instrumentes, nämlich einer Uhr, *direkt* messbar ist, ist die Zeit. Wie das Lehrbuch der Physik erklärt, gehört die Zeit zu den *direkt* messbaren Basisgrößen des physikalischen Messsystems. Dagegen ist im IQ-Test zwar die Zeitspanne für jede Aufgabengruppe (Skala) vorgegeben, aber es wird lediglich gezählt, nämlich die Anzahl der Aufgaben, die in der festgesetzten Zeitspanne richtig bearbeitet worden sind. Tatsächlich gemessen wird die Reaktionszeit in Aufmerksamkeitstests, die Fehler werden dort auch gezählt. Ein programmierter Aufmerksamkeitstest *misst* also. Nimmt man die Zeiteinheit als 'objektive' Messeinheit, so entsteht durch Aufmerksamkeitstests ein 'objektiver' Vergleich zwischen den Leistungen von Probanden, konstante und gleiche Umgebungsbedingungen vorausgesetzt. Zugleich erlaubt die *direkte* Messung individualspezifische Auswertungen, die jedoch von einer auf Stichproben ausgerichteten Testpsychologie nicht genutzt werden.

Fragebögen sind etwas anderes. Ihre Verwendung führt zur Selbstaussagen eines Probanden in einem Itempool, dessen Bedeutungen einem Lexikon entnommen worden sind. Doch was der jeweilige Proband unter dem Item versteht, welchen Sinn er/sie in der jeweiligen Behauptung im eigenen Lebensvollzug sieht, bleibt so eine große Unbekannte. Wer Fragebögen verwendet, will Zeit einsparen und wird deshalb nicht bei jedem bearbeiteten Item nachfragen, in welchem lebensgeschichtlichen Kontext es verstanden worden ist. Insofern ist in einem Gutachten eine Feststellung der Art "Wir haben unseren anamnestisch gewonnenen Eindruck durch einen Depressionstest untermauert" vielleicht eine Stütze für die Diagnose, darüber hinaus jedoch eine zu hinterfragende Behauptung, um der Gefahr eines falsch positiven Ergebnisses zu begegnen. Fragebogenergebnisse führen also über die Interpretationsbedürftigkeit anamnestisch gewonnener Daten nicht hinaus. Fraglich ist auch, was diese beiden Arten von Daten über die Minderung von Fähigkeiten besagen.

1 Kein Testergebnis, auch kein *neuropsychologisch* genanntes, bezieht sich in irgendeiner Weise direkt auf das erklärende Konstrukt *Hirnfunktion*. Denn jeder Test wird von einem *ganzen Menschen* bearbeitet und nicht von einem seiner Organe. Genau genommen erhebt der Test zunächst die *Leistungsperformance*, und die *Fähigkeit* erst unter der Voraussetzung, dass eine Testperson sich *willentlich so weit als ihr möglich anstrengt*, was jedoch von ihrer *Motivation* abhängt.

Glaubwürdigkeit und Wahrhaftigkeit

Glaubwürdigkeit und *Wahrhaftigkeit* sind nicht dasselbe, sie sitzen im Untersuchungsprozess einander gegenüber. Erstere seitens des Gutachters, letztere seitens des Probanden. Im Grunde ist der Eindruck, den selbst der erfahrendste Gutachter von der *Wahrhaftigkeit* der Aussagen eines Probanden hat, seine *Glaubwürdigkeit* also, schwer zu belegen. Denn der Sinn des Kriteriums *Wahrhaftigkeit* - also etwas zu sagen, was man für wahr hält, was aber nicht notwendig *wahr* sein muss – ist in der Untersuchungssituation oftmals unklar : Wie kann ein Proband an sich selbst unterscheiden, ob die Gefühle, die er/sie im Moment der Befragung hat oder erinnert, just jene Gefühle sind, die er/sie ansonsten überwiegend auch hat oder beim spezifisch befragten Vorfall gehabt hat? Wie könnte man sich darin überhaupt sicher sein, dass man sich 'gefühlsmäßig' richtig erinnert? Aufgrund solcher und anderer Unklarheiten scheidet das Kriterium der *Wahrhaftigkeit* zur Bewertung der gewonnenen Daten von vornherein aus. Was bleibt ist eine Analyse der Konsistenz bzw. der Kohärenz der gewonnenen Aussagen, z.B. geleitet von der Frage: Passen die Fragebogenergebnisse zu den mündlich gewonnenen Daten und den Testergebnissen? Nicht selten führen nämlich Fragebögen in der Begutachtung anlässlich eines Rentenverlangens zu extrem negativ auffälligen Ergebnissen, erzeugen also ein zusätzliches Problem.

Beschwerdendvalidierung

Leider ist *Wahrhaftigkeit* auch nicht durch die Verwendung von Verfahren zur *Beschwerdendvalidierung* zu klären. Denn mir ist keine Logik bekannt, mittels derer von ein, zwei oder drei Warnsignalen hinsichtlich Mitarbeit auf eine schlechte Mitarbeit während der gesamten Untersuchung zu schließen wäre. Bekanntlich gibt es keine Schlüsse von empirischen Einzeldaten auf eine allgemeine These, auch nicht auf eine beschränkt allgemeine. Schon gar nicht bei einem Menschen, der durch den einen Test oder Fragebogen gelangweilt, von einem anderen jedoch herausgefordert wird. Auch statistische Überlegungen helfen hier nur bedingt weiter, falls überhaupt, sobald es eben um die Beurteilung eines bestimmten Einzelfalles geht². Warnsignale aus der Beschwerdendvalidierung stellen eine Aufforderung zur eingehenden Konsistenz- und Kohärenzanalyse der Untersuchungsergebnisse dar, mehr jedoch nicht. Das auch offiziell bei Gericht zu verwendende Vokabular "Verdeutlichung, Aggravation, Simulation" hilft bei der Klärung der *Wahrhaftigkeit* ebenso wenig weiter. Diese Termini bezeichnen im Grunde drei Schweregrade der Anschuldigung, für die, genau genommen, immer nur einzelne Hinweise finden sind. Die Termini besagen etwas über die geringe *Glaubwürdigkeit*, die ein Gutachter den Aussagen und der Mitarbeit eines Probanden zubilligt.

2 Die Messfehler werden nach der klassischen *Testtheorie* mittels Reliabilitäten berechnet, die zwar anhand von Stichproben gewonnen werden, anschließend jedoch den Tests (oder Fragebögen) als *Eigenschaften* dieser 'Messinstrumente' zugeschlagen werden. Die damit berechneten Vertrauensintervalle für Ergebnisse berücksichtigen daher nicht die gewöhnlichen, alltäglichen Leistungsschwankungen des jeweils spezifischen Probanden. Das ist für die Fähigkeitsbeurteilung, in der die unbekannte Ausprägung durch einen Vergleich mit den bekannten ermittelt wird, eher unproblematisch. (Denn nur durch einen solchen Vergleich kann sie quantifiziert werden.) Problematisch ist aber ein solches Vergleichen mittels der Verfahren zur Beschwerdendvalidierung, weil darin meist die Fehlerzahlen verglichen werden, aber die personspezifische Arbeitsgenauigkeit (samt Schwankungen) der jeweiligen Testperson unbekannt ist. Dieses Problem wäre auch bei Kenntnis der Prävalenz und ihrer Verwendung zur Berechnung der Cut-Off-Werte (Bayes-Statistik) noch nicht beseitigt.

Es geht um's Geld

Nicht zuletzt geht es um's Geld. Begutachten bei Renten- und Entschädigungsfragen heisst Stellung beziehen und Entscheidungshilfe geben in einem Interessensgegensatz zwischen einer Person, die ein Anliegen hat und einer Organisation, die zahlen soll, aber bevor sie diese Verpflichtung übernehmen will, ein Gutachten in Auftrag gibt. Die Berufung auf Wissenschaftlichkeit ändert an der Funktion des Gutachten nichts, sie besagt nur etwas über die Mittel, die beanspruchen zu einem 'unparteiischen Urteil' zu führen - falls es so etwas gibt.

Autor

Dr. Wolfgang Palm, Dipl.-Psych, Dipl.-Phys, Psychotherapeut
Sachverständiger in Sozialrechtsfragen der Pt-Kammer Baden-Württemberg
www.wopalm.com

Stand dieses Papiers: Oktober 2019

Literatur:

M. Amelang, L. Schmidt-Atzert. Psychologische Diagnostik und Intervention. Springer Verlag, 2006

R. Dohrenbusch, Th. Merten, M. Kutzner (Hrsg). Psychologische Begutachtung in der Berufsunfähigkeitsversicherung. Referenz-Verlag, 2014

J. Francke, A. Gagel, D. Bieresborn (Hrsg). Der Sachverständigenbeweis im Sozialrecht. 2. Auflage, Nomos Verlagsgesellschaft, 2017

G. Gigerenzer. Risiko. Wie man die richtigen Entscheidungen trifft. Verlag C. Bertelsmann, 2013

H.P. Huber. Psychometrische Einzelfalldiagnostik. Beltz-Verlag, 1973

P. Janich. Was ist Wahrheit? Eine philosophische Einführung. 2. Aufl, Verlag C.H.Beck, 2000

M. Linden, S. Baron, B. Muschalla, M. Ostholt-Corsten. Fähigkeitsbeeinträchtigungen bei psychischen Erkrankungen. Verlag Hans Huber, 2015

T. Merten. Beschwerdevalidierung. Verlag Hogrefe, 2014

W. Palm. Genauigkeit in der Aufmerksamkeitstestung, 2018. www.w-palm.de/material/Genauigkeit.pdf

H. Vogel. Gerthsen Physik. 20. Auflage. Verlag Springer, 2000

W. Schneider, R. Dohrenbusch, H.J. Freyberger, P. Henningson, H. Irle, V. Köllner, B. Widder (Hrsg). Begutachtung bei psychischen und psychosomatischen Erkrankungen. 2. und erweiterte Auflage, Hogrefe-Verlag, 2016

B. Williams. Wahrheit und Wahrhaftigkeit. Suhrkamp Taschenbuch Wissenschaft, 2013

WHO. International Classification of Functioning, Disability and Health (ICF).
<https://www.who.int/classifications/icf/en/>