Genauigkeiten in Aufmerksamkeitstestungen

Wolfgang Palm

Dieses Papier zeigt, wie aus den 'Schwankungen der Reaktionszeiten' und den 'Reaktionsfehlern' in Aufmerksamkeitstestungen gemeinsame Größen errechnet werden können, und zwar mit Hilfe einfacher mathematischer Funktionen. Die Größen erhalten den Namen 'Genauigkeiten', die Funktion wird 'Genauigkeit' genannt. Hinweise werden gegeben, wie diese 'Genauigkeiten' interpretiert werden können, beispielsweise zur Prüfung, ob schlechte Leistungsergebnisse glaubwürdig sind.

1. Normierte und nicht normierte Größen

Aufmerksamkeitstestungen sind Reaktionszeitestungen: Einfache Zeichen erscheinen auf einem Bildschirm, schnellstmoglich ist die eine oder die andere Taste zu drücken. Aufmerksamkeitstestung geschieht am besten programmiert. Die durch das Programm zu messenden oder zu zählenden Größen sind: Der zeitliche Abstand zwischen Reiz und Reaktion, die Zahl der falschen Reaktionen, die Zahl der ausgelassenen Reaktionen und die Zahl der voreiligen Reaktionen. Da die Reaktionszeiten schwanken, werden von den Programmen als Kennzahlen der Mittelwert oder der Median ausgegeben, sowie die Streuung der Werte um die mittlere Größe, die als Standardabweichung berechnet wird. Bei Mittelwertberechnungen konnen zudem noch die 'Ausreißer' bestimmt werden, also Extremwerte, die die Größe des Mittelwerts beeinflussen.

Üblicherweise werden die Ergebnisse eines Probanden, der einem psychologischen Test berarbeitet hat, mit denen einer Stichprobe verglichen, deren Ergebnisse moglichst normalverteilt vorliegen sollen, um in Normwerten dargestellt zu werden. Letztere unterscheiden sich lediglich in Mittelwert und Standardabweichung und haben dementsprechend verschiedene Benennungen: IQ-, T-, S- oder z-Werte konnen ineinander umgerechnet werden. Sind die Stichprobenergebnisse nicht adäquat normalverteilt, werden oft Prozentränge angegeben. Ein bestimmter Prozentrang besagt, dass dieser %-Satz von Personen einer Stichprobe einen geringeren oder bestenfalls einen gleich hohen Anteil an entsprechenden Antworten (Reaktionen) erreicht haben wie eine Testperson (ein Proband). Normwerte liegen vor - je nach Test - für Median, Standardabweichung, und Fehler. Als Fehler gelten Auslassungen, falsche Reaktionen und vorschnelle Reaktionen (Antizipationen).

Die Vorgehensweise, Testergebnisse in Normwerten auszudrücken, ist gängige Routine. Dabei wird kaum je erwogen den basalen Mess- oder Zahlergebnissen eine eigene Bedeutung zuzuweisen, obwohl diese der elementare Ausdruck der individuellen Leistung sind. Doch genau das kann in gutachterlichen Fragestellungen zweckmäßig sein, sobald zu erörten ist, ob die vorliegende Leistungsperformance eines Probanden seine bestmogliche ist, also seiner Fähigkeit entspricht. Hierbei ist hilfreich, dass diese Ergebnisse auf direkten Messungen und Zählungen beruhen, die zu ihrer Ausführung keiner psychologischen Theorie bedürfen, weil sie nur Zeitdauer

und Anzahl enthalten [1]. Meist wird die Größe 'Zeit' durch die Kennwerte Mittelwert oder Median und Standardabweichung angegeben. Zusammen mit den fehlerhaften Reaktionen sind diese Größen durchaus Kandidaten für den intraindividuellen Vergleich.

Indes weisen verschiedene Reaktionszeitaufgaben auch unterschiedliche Schwierigkeitsgrade auf, welche die Reaktionszeit eines Probanden beeinflussen. Schwierigere Aufgaben werden eine längere Reaktionszeit erfordern, insbesondere, um fehlerfrei ausgeführt zu werden, ein Vorgang, der allerdings von Proband zu Proband unterschiedlich ausfallen wird. Daher ist der Grad der Schwierigkeit ist nicht unabhangig von der Stichprobe zu bestimmen und geht folglich implizit in die Verteilung der Reaktionszeiten und deren Normierung ein. Anhand dieser Normierung wird die Leistung eines Probanden in der Bewältigung einer Schwierigkeit relativ zu anderen Leistungen plazierbar. Grundgröße hierfür ist die mittlere Reaktionszeit, entweder Mittelwert oder Median der Reaktionszeit. Wegen dieser impliziten Abhängigkeit von der jeweiligen Aufgabenschwierigkeit sind die mittleren Werte der Reaktionszeiten eines Probanden für verschiedene Aufgaben nur über den Umweg einer Normierung miteinander zu vergleichen. Folglich scheidet genau diese Größe für einen nicht normierten Vergleich von Ergebnissen aus. Selbstverständlich ist sie als Vergleichsgröße wichtig.

2. Fähigkeits-Testungen unter Begehrensvoraussetzungen

Die Ergebnisse von Reaktionszeitmessungen sind nicht bereits als solche der Ausdruck der Fähigkeit, des 'eigentlichen' Könnens eines Probanden, weil diese Ergebnisse immer mitbestimmt sind von emotional-motivationalen und willentlichen Prozessen, umgangssprachlich formuliert, von Anstrengung und Mitarbeit. Sie beinflussen die mittlere Reaktionszeit, die Reaktionszeit-schwankungen und die Fehlerzahlen, gehen also mit in eben jene Größen ein, die als Ergebnisse der Testdurchführungen vorliegen. Deshalb sind die Ergebnisse – weder in Rohwerten noch in Normwerten – unbefragt als Ausdruck der Fähigkeiten eines Probanden anzuerkennen. Doch es gibt keine Methode, mit der sich der Einfluss der Motivation aus den vorliegenden Tester-gebnissen eindeutig herausrechnen ließe. Dieses Problem ist besonders brisant in Testunter-suchungen, die mit Probanden mit einem Renten- oder Entschädigungsbegehren durchgeführt werden. Solche Untersuchungen sollen klären, inwieweit Fähigkeitsminderungen vorliegen, die die Ausübung einer Berufs- oder Erwerbstätigkeit einschränken, oder – wie es juristisch heißt – sich "kausal" auf ein Ereignis (z.B. Unfall, Gewalttat) zurückführen lassen [2].

In solchen 'Fällen' gibt es eher selten neurologische Befunde über eine Schädigung des Gehirns und somit keine Hinweise darauf, welche Art von Funktionsstörungen und damit verbunden, welche Fähigkeitseinschränkungen zu erwarten sind. Zudem unterliegt eine 'Breitband'-Testauswahl zur Überprüfung von Fähigkeiten gewissen situativen Beschränkungen: Wegen der geforderten Anstrengung sind einem Probanden in der Regel an einem Tag nicht mehr als vier bis

1 Palm W. Zur Valdität psychologischer und physikalischer Messprozesse. Frankfurt 1991 (Haag + Herchen)

2 Franke J, Gagel A, Bieresborn D (Hrsg). Der Sachverständigenbeweis im Sozialrecht, 2. Auflage, Baden-Baden 2017
Schneider W, Dohrenbusch R, Freyberger HJ, Henningsen P, Irle H, Köllner V, Widder B (Hrsg). Begutachtung bei psychischen und psychosomatischen Erkrankungen, 2., überarbeitete und erweiterte Auflage, Bern 2016

höchstens fünf Stunden Testuntersuchung zuzumuten, in denen zusätzlich zur Aufmerksamkeit Gedächtnis, Intelligenz, exekutive Funktionen u.a. zu prüfen sind. Ein zweiter Termin lässt sich aus anderen Gründen (Entfernung, Kosten) nur selten realisieren. Dadurch werden der Testauswahl enge Grenzen gesetzt. Langjährige Erfahrungen mit Reaktionszeittestungen unter Begehrens-voraussetzungen zeigen zudem, dass deren Ergebnisse von Tests mit längeren Bearbeitungszeiten (etwa länger als fünf Minuten) durchwegs zu schlechteren Ergebnissen tendieren [3].

Spezifische Tests zur Beschwerdenvalidierung [4] haben bei der Aufklärung von Verfälschungs-tendenzen im obigen Kontext einen beschränkten Nutzen. Einmal treten die bekanntensten B-Tests als 'Gedächtnistests' auf, zum anderen muss zur Logik ihrer Interpretation klar gesagt werden: Die Auffälligkeit in *jedem einzelnen* B-Test *beweist* nicht, dass ein Proband in allen anderen Tests schlecht mitgearbeitet hat. Erstens sind induktive Schlüsse aus einzelnen empirischen Ergebnissen auf allgemeine Aussagen logisch unmöglich. Zweitens wird man auch aus einem auffälligen Ergebnis im *Alterness*-Test nicht den Schluss ziehen, dass man die übrigen Aufmerksamkeitsprüfungen nicht durchzuführen braucht [5]. Jedoch kann man auffälligen B-Tests ein Warnsignal entnehmen, um die Mitarbeit eines Probanden bei anderen negativ auffälligen Testergebnissen genauer unter die Lupe zu nehmen.

Hierfür soll im folgenden eine mathematische Funktion entwickelt werden, die sich graphisch darstellen lässt, und die es gestattet, den Einfluss von Anstrengung und Mitarbeit auf die Testergebnisse besser einzuschätzen. In diese Formel werden die nicht normierten, auf direkter Messung beruhenden Größen 'Reaktionszeitschwankungen' und 'Fehlerzahlen' eingehen.

Zunächst erfolgt zuerst eine knappe Darlegung der Definitionen und der daraus entwickelten Formeln, danach werden Hinweise zur Interpretation gegeben und schließlich folgt eine Erläuterung und eine Begründung (Rechtfertigung) dieser Vorgehensweise.

3. Formeln für die 'Genauigkeit'

In den Handbüchern zu den einschlägigen Testverfahren dienen die Fehlerzahlen (falsche Reaktionen und/oder Auslassungen) oft als Maß für die *Sorgfalt* (So) im Ausführen der

3 Die in den nachfolgenden Diagrammen stehenden Abkürzungen von Testnamen stehen für Tests aus den Testbatterien bzw. Testsystemen:
P. Zimmermann, B. Finn. Testbatterie zur Aufmerksamkeitsprüfung (TAP); www.psytest.net
Der Test SZT ist im Hogrefe-Testsystem HTS 4 enthalten; www.testzentrale.de
BSV_A, BSV_G sind Abkürzungen für Tests aus einer Testbatterie zur Beschwerdevalidierung, die hier nicht namentlich genannt wird, um Probanden keinen direkten Hinweis auf die benutzte Software zu geben.

4 Merten T, Dettenborn H (Hrsg). Diagnostik der Beschwerdenvalidierung, Berlin 2009 (Deutscher Psychologen Verlag GmbH)
Merten T, Dohrenbusch R. Psychologische Methoden der Beschwerdenvalidierung, in Schneider W. et al, aaO, Kapitel 7

5 Handbuch zur TAP 2.3, Teil 1: "Unter Alertness ist zunächst der allgemeine Wachzustand zu verstehen, der es einer Person erlaubt schnell und angemessen auf konkrete Anforderungen zu reagieren. Es ist die Voraussetzung für ein adäquates Handeln und stellt insofern die Basis jeder Aufmerksamkeitsleistung dar." (S.10) Wer wird aus einer solchen allgemeinen Behauptung folgern wollen, dass auf Grund einer schlechten oder einer sehr guten Leistung in der *Alertness* sich weitere Aufmerksamkeitsprüfungen erübrigen?

Anforderungen. Sei fz die Zahl der Fehler und Iz die Zahl der kritischen Items (auf die zu reagieren ist), so lässt sich die *Sorgfalt* quantitativ als eine Größe definieren, die proportional zu dem Quotienten (Iz - fz)/(fz+1) ist [6].

Analog lässt sich durch die Verwendung der Standardabweichung der Reaktionszeiten eine Größe *Konstanz* (Ko) definieren, die die Gleichmäßigkeit des Reagierens beschreibt: Sei m das Mittel (Mittelwert oder Median) und s die Standardabweichung, so sei Ko proportional zum Quotienten (m – s)/(s) [7].

Aus Gründen der Anschaulichkeit - die gesuchte Funktion soll ihre Interpretation anhand von Diagrammen erhalten - sollen sich die Werte der Funktion im Intervall]0 ; 1] bewegen. Die Werte sollen also die obere Grenze +1 nicht überschreiten und nach unten möglichst nicht unter 0 absinken. Die gesuchte Funktion erhält den Namen *Genauigkeit* (Gn).

Verwendet man Ko und So wie oben definiert, so werden die Quotienten mit kleinerem fz bzw. kleinerem s zwar größer, doch es entstehen durch Multiplikation oder durch Addition Werte größer als 1. Die Erfahrung zeigt zudem, dass s meist in der Größenordnung von 20% bis 50% von m liegt, der %-Satz von fz jedoch geringer ausfällt. Dadurch wird nach obiger Definition von So der Einfluss der Fehler in Gn zu gering gewichtet.

Um zu einer brauchbaren Formel im Wertebereich]0 ; 1] zu kommen, müssen die Formeln für So und Ko 'umgestülpt' werden zu So: = (Iz + fz) / (Iz – fz) und Ko: = m / (m -s). So enthält zudem eine höhere Gewichtung der Fehleranteile. Soweit – so einfach; doch wie ergibt sich ein Zusammenhang zwischen den beiden Größen So und Ko, die beide auf von einander unabhängigen Messvorgängen beruhen, auf der Messung von Zeiteinheiten [ms] und der Zählung von Fehlern?

Die Antwort lautet: Der Zusammenhang wird durch eine Formel angegeben, die als solche durch direkte Messungen von Zeit und Anzahl leider nicht bestätigt werden kann [8]. Sofern So und Ko voneinander unabhängig sind (geringe Korrelation), böte es sich an, die beiden als Komponenten eines Vektors zu betrachten, dessen skalare Größe aus der Summe der quadrierten Komponenten hervorgeht. Doch dieses Verfahren führt zu keinen brauchbaren Ergebnissen innerhalb des Intervalls]0 ; 1]. Besser sind einfache Mittelungsverfahren, beispielsweise das geometrische und das harmonische Mittel, und zwar mit den folgenden Formeln [9]:

Formel (1): Gn: = 2 - Wurzel(Ko * So) = 2 – Wurzel[m / (m -s) * (Iz + fz) / (Iz – fz)]

Gn kann negativ werden und nähert sich nach oben der ganzen Zahl 1 an. Die Werte aus dieser Formel existieren nur punktweise und lassen sich grafisch darstellen. Die absoluten Höhen sind weniger bedeutsam als deren Verlauf, der u.a. mit dem Verlauf der normierten Ergebnisse für die

6 Die 1 im Nenner verhindert, dass der Bruch für fz = 0 undefiniert wird.

7 Falls die Standardabweichung = Null ist, verliert die Rede vom Mittelwert ihren Sinn. Korrekterweise müsste an die genannten Größen der Index i angehängt werden (i = 1 ... bis ... n; n ist die Zahl der verwendeten Tests oder Testabschnitte). Der Übersichtlichkeit halber wird er jedoch weggelassen.

8 Zur Rechtfertigung siehe Punkt 5.

9 Damit in den folgenden Formeln kein Nenner mit dem Wert 0 entsteht, gilt für für diese und alle weiteren Formeln: fz < Iz; s ist positiv; und s < m. Falls die Daten dem nicht entsprechen, kann man sich fragen, ob eine sinnvolle Messung vorliegt.

mittleren Reaktionszeiten (Mediane) verglichen werden kann, sofern diese sich im Intervall] 0 ; 1] bewegen (z.B. mit %-Rängen/100).

Verwendet man statt des geometrischen das harmonische Mittel, so fallen die Differenzen in den Ergebnispunkten meist marginal aus:

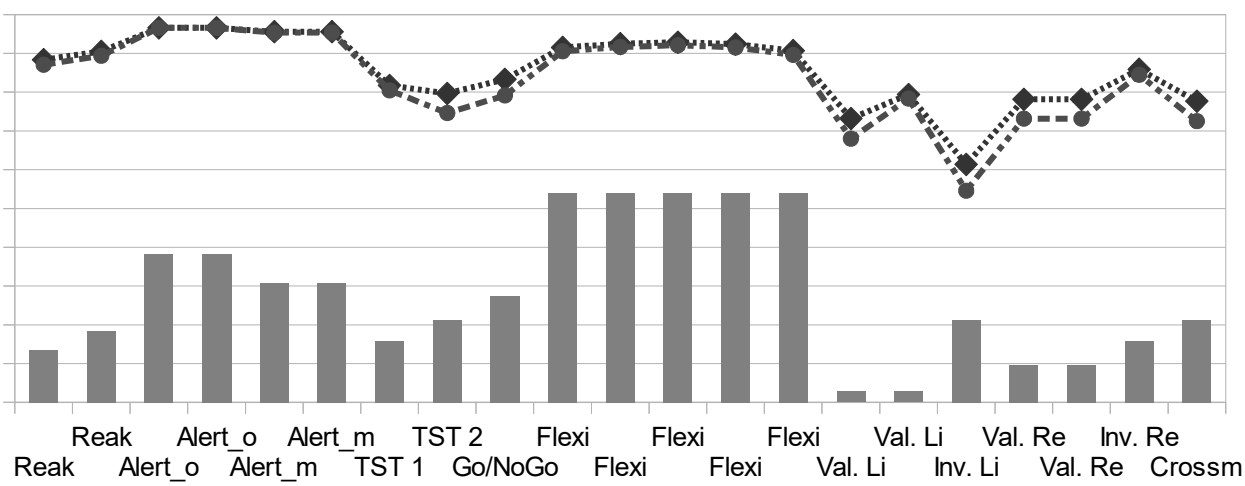
Formel (2): Gn: = 2 – 2/(1/Ko + 1/So) = 2-2/[(m – s) / m + (Iz – fz) / (Iz + fz)]

Etwas betonter fallen die Unterschiede aus, wenn man statt Formel (1) eine modifizierte Formel für das harmonische Mittel verwendet.

Formel (3): Gn: = 2 – 2*1/Wurzel(1/Ko² + r*Ko*So + 1/So²); r: = r(s/m; fz/Iz) = 0,24

Die folgende Diagramm 1 zeigt die Verläufe der Genauigkeit berechnet nach den Formeln (1) und (3). Die grauen Balken bilden die Mediane der Reaktionszeiten in %-Rängen/100 ab.

Die eher geringen Unterschiede sind für die Interpretation bedeutungslos. Detailgenaueres Rechnen führt also nicht notwendig zu besserer Erkenntnis. Man kann sich also auf Formel (1) oder Formel (2) beschränken. Doch sind drei Eigenschaften der Genauigkeitsfunktion diskussionswürdig.



- Diagramm 1: Normwerte der Mediane der Reaktionszeiten [%-Ränge/100] und Genauigkeiten. Die Benennungen unter der waagrechten Achse sind Abkürzungen für die verwendeten Tests. Jeder Balken steht für eine Itemzahl Iz = 20.

a) Das obige Diagramm 1 mit 'echten Daten' könnte evtl. einen Zusammenhang vermuten lassen, der nicht besteht. Infolge der Definition von So und Ko bzw. So und Ko hängt die Höhe der Genauigkeiten nicht direkt von der Höhe der mittleren Reaktionszeiten ab, weil es für einen bestimmten Quotienten eine unendliche Zahl von möglichen Eingangsgrößen für Zähler und Nenner gibt. Die Höhe der grauen Balken hat also keinen direkten Zusammenhang mit den Genauigkeiten. Vielmehr zeigt sich, dass im schwierigen Test *Flexibilität* eine gut durchschnittliche Reaktionszeit bei gleichzeitig hohen Genauigkeiten erreicht worden ist, während bei weniger

anspruchsvollen Aufgaben die Genauigkeiten teilweise etwas geringer ausfallen, wobei die Leistungen in den Reaktionszeiten teils deutlich geringer ausfallen.

b) Der Verlauf der beiden Genauigkeitsfunktionen verdeckt ein Problem: Setzt man nämlich dieselben Eingangswerte in die drei Formeln ein [10], so fallen die Rechenergebnisse nach Formel (3) etwas, aber immerhin sichtbar niedriger aus als die nach Formel (1) und (2). Da jedoch die Eingangswerte (mengentheoretisch: die Definitionsmenge) für die Größen So und Ko bzw. So und Ko wohl definierte Messergebnisse sind, sollten solche Unterschiede eben nicht vorkommen! Der Grund hierfür liegt eben darin, dass die Ergebnisse aus den Formeln (1), (2) und (3) nur rechnerisch existieren, aber messtechnisch nicht nachzuprüfen sind. Das Problem hat bei dem hier gewählten Ansatz vermutlich keine Lösung, ist aber 'hebbar'. D.h. man kann es grafisch fast verschwinden lassen, indem der Maximalwert nach Formel (3) auf den Maximalwert nach Formel (1) oder (2) 'angehoben' wird und der Differenzwert zu all jenen Ergebnissen addiert wird, die die Formel (3) für eine bestimmte Definitionsmenge erzeugt. Das mag unbefriedigend sein, funktioniert indes hinreichend gut, weil die danach noch bestehenden, minimalen Unterschiede in der Interpretation der Genauigkeiten keine Rolle spielen.

c) Eine weitere Schwierigkeit, die jedoch im Grunde 'behebbar' ist, also eine Lösung hat, steckt in den Formeln für Ko und Ko. Während die Größen für Iz und fz direkt durch Abzählen gewonnen werden und dadurch eine Größe wie fz/Iz auf direkten Messhandlungen beruht, fehlen diese für eine Größe wie s/m; Zähler und Nenner sind statistische Größen. Die darin liegende Schwierigkeit besteht nicht nur in den hier entwickelten Formeln, sie ist im Grunde typisch für wichtige, in der psychotherapeutisch-psychiatrischen Forschung verwendete Formeln [11]. Für die Zwecke, die in diesem Papier dargelegt werden, ließe sich zwar eine entsprechende Formel unter Bezug auf die direkten Messhandlungen entwickeln, doch setzte sie voraus, dass alle einzelnen Zeitmessergebnisse einer Testung bekannt sind. Praktisch gibt aber eine Software wie die TAP lediglich Mittelwerte, Mediane und Standardabweichungen aus. Also ist mit diesen als Eingangsgrößen zu rechnen.

4. Hinweise zur Interpretation

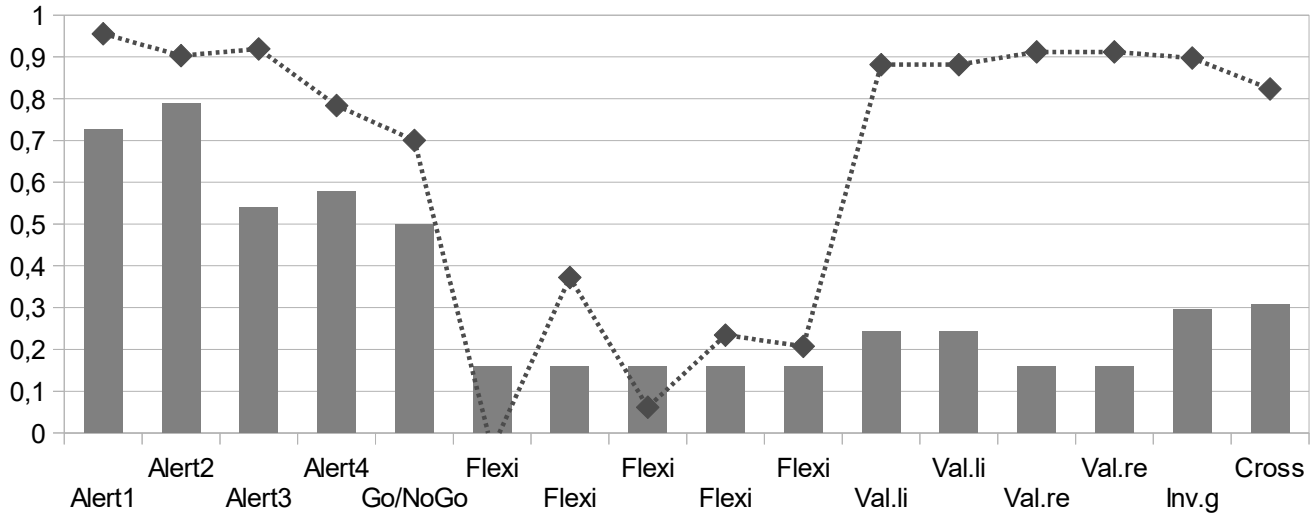
Eine einfache Regel, nach der die Genauigkeitslinie zu interpretieren sei, gibt es nicht! Freilich tritt in der Begutachtung hin und wieder auch der 'einfache Fall' auf, in dem die Ergebnisse der Aufmerksamkeitstestung fast alle im Normalbereich liegen und die Genauigkeiten sehr hoch verlaufen. Selbst in Zusammenhang damit durchgeführte und auffällig gewordene Test zu *Beschwerdevalidierung* würden dann einen Hinweis ohne weitere Bedeutung geben.

Noch recht einfach zu interpretieren ist das folgende, 'echte' Beispiel im Diagramm 2. Die niedrigsten Normwerte für die Reaktionszeiten liegen noch auf dem unteren Rand des

10 Die Berechnungen lassen sich rasch mittels Tabellenkalkulation durchzuführen.

11 Formeln, wie solche für die Effektstärke, enthalten Mittelwertdifferenzen und Varianzen. Eine Größe wie Δm/σ enthält die Rohwerte von beispielsweise mit Fragebögen gewonnenen Ergebnissen, deren messtheoretische Grundannahmen in der Praxis der Forschung und der Einzefallanalyse selten noch thematisiert werden. Meist beruhen weder in Δm noch in σ auf direkten Messhandlungen. Zur Effektstärke siehe z.B. https://de.wikipedia.org/wiki/Effektstärke (Stand 09/18).

Normalbereichs. Auffällig ist der Absturz der Genauigkeiten über dem Test *Flexibilität*. Obwohl die Zeiten noch 'normal' ausfallen, nehmen die Standardabweichungen und vor allem die Fehler



- Diagramm 2: Reaktionszeiten (Mediane) in Normwerten (graue Balken; %-Ränge/100) und Genauigkeiten (Rhomben mit gepunkteter Linie)

drastisch zu. Der Proband hat offenbar erhebliche Schwierigkeiten bei der Bewältigung dieser Aufgaben. Die Auffälligkeiten bilden eine Ausnahme in einer Reihe sonst unauffälliger Ergebnisse, auch denen vor und nach der Aufmerksamkeitstestung.

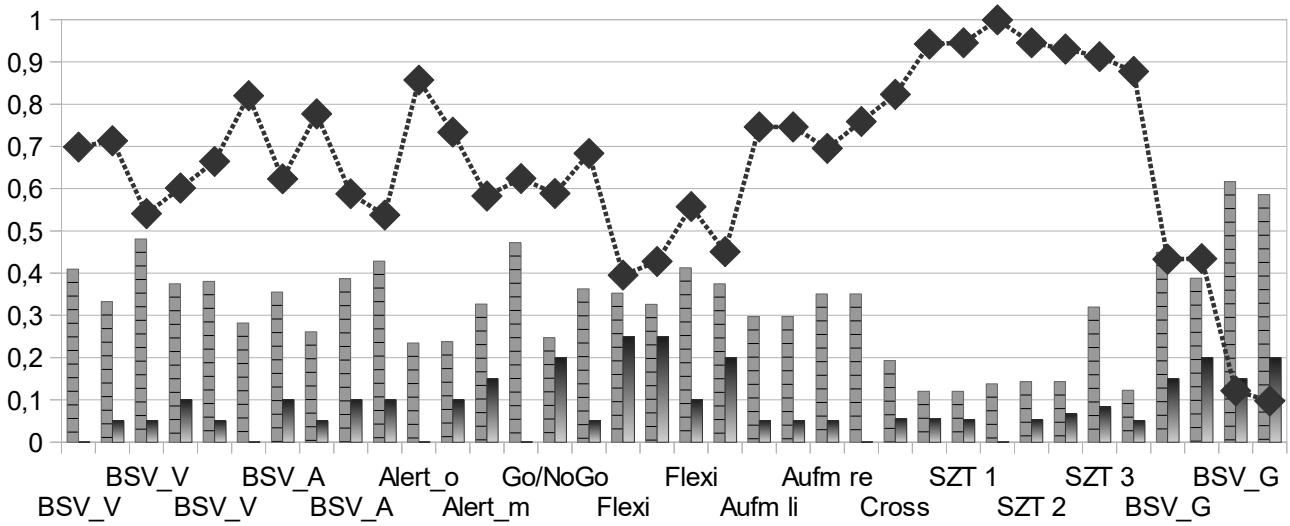
Viele Fälle in der Begutachtung sind jedoch nicht so offensichtlich; deshalb sind die Zusammenhänge mit den Ergebnissen in Normwerten und in Zeiteinheiten zu beachten, wobei es auch auf die Schwierigkeiten der jeweiligen Aufgaben und den Zweck des jeweiligen Tests ankommt. Auch deren Reihenfolge kann von Bedeutung sein [12]. Darüber hinaus sind die Ergebnisse und die Zwecke der übrigen Tests heran zu ziehen, sofern die Aufmerksamkeitstestung nur ein Teil einer umfangreicheren Testbatterie ist. Auch hierbei ist der zeitliche Ablauf der Testung und die Plazierung der einzelnen Tests eventuell von Bedeutung. Selbstverständlich ist auch die Beobachtung der Arbeitsweise eines Probanden in der Untersuchungssituation sehr wichtig. Schließlich kommt es auf den gesamten Kontext von Daten und auf Vorwissen (aus Berichten) an, die zur Einschätzung der Motivationslage eines Probanden in einer Testsituation heran zu ziehen sind.

Das folgende Diagramm 3 enthält die Daten eines weiteren 'echten' Beispiels. Aus der Vorgeschichte dieses Falles liegen keine Befunde über eine Schädigung von Hirnarealen vor, die Diagnose einer mittelschweren depressiven Episode wurde indes gestellt. Die vorhandenen Normwerte für die Reaktionszeiten und die Schwankungsbreiten fallen mit Ausnahme derer für *Go/NoGo* und *SZT_3* nach unten aus dem Normalbereich heraus. Zu fragen ist nun: Sind anhand der vorliegenden, durchwegs schlechten Ergebnisse, diverse Aufmerksamkeitsstörungen zu

12 Man könnte beispielsweise aus dem Winkel einer Regressionsgeraden einen Hinweis auf eine mögliche Ermüdung entnehmen. Andererseits können Aufgaben auch langweilig oder herausfordernd sein, was wiederum Einfluss auf die Mitarbeit haben kann. Der Winkel einer Regressionsgeraden würde in einem solchen Fall von der Position der Testergebnisse abhängen.

diagnostizieren oder sind die schlechten Ergebnisse auf mangelhafte Mitarbeit (ja sogar auf Aggravation) zurück zu führen? Der Verlauf der Genauigkeiten verweist auf Letzteres. Denn deutliche Funktionseinbußen haben meistens klare Einbrüche ähnlich wie in Diagramm 2.

In Diagramm 3 erkennt man zunächst, dass die Genauigkeiten gegenläufig zu den Größen s/m und fz/Iz sind. Im ersten Drittel schwanken die Genauigkeiten etwas, sinken dann über dem Test *Flexibilität* ab, klettern danach über dem Test *SZT* zur höchsten Ausprägung, um schließlich gegen Null abzustürzen. Die Tests *BSV_V*, *BSV_A,* und *BSV_G* sind *Beschwerdenvalidierungstests* [13]. Sie enthalten höchst simple Aufgaben, die als schwierig angekündigt werden. Ihre Ausführung ähnelt



- Diagramm 3: Genauigkeiten, relative Schwankungen s/m (quer gestreifte Balken) und relative Fehler fz/Iz (hell-dunkle Balken). Unter der waagrechten Achse stehen die Abkürzungen für die verwendeten Tests.

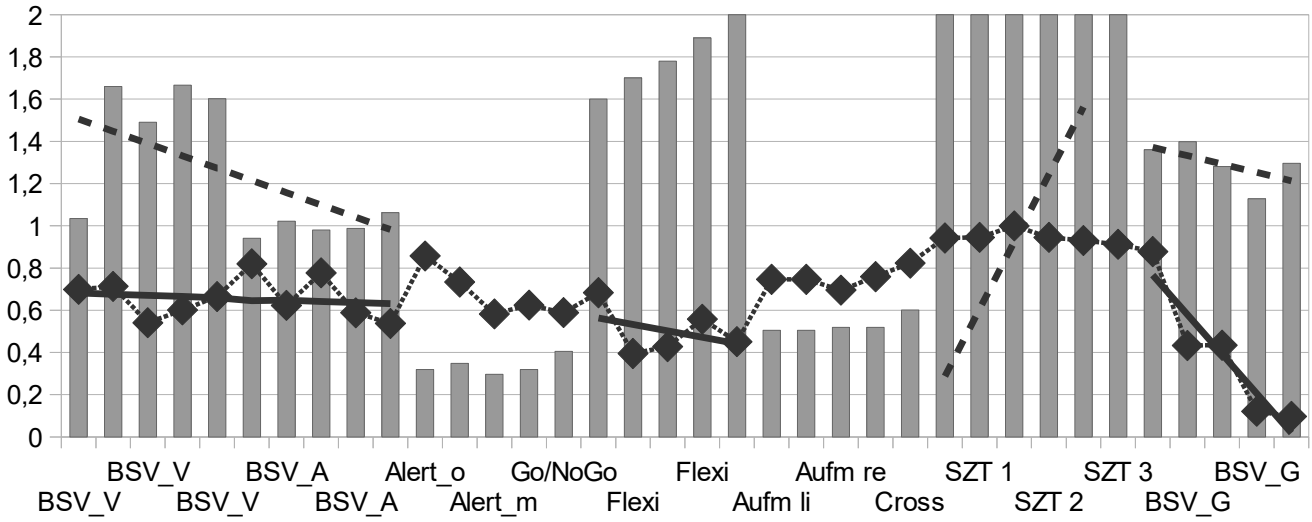
jedoch stark der von *Alertness* und *Go/NoGo*. Zu reagieren ist allerdings mit zwei Tasten (links und rechts auf der Tastatur). Die Zahl der Fehler in den drei BSV-Tests ist auffällig, sie überschreitet jeweils den Cut-Off-Wert und so ergeben sich drei Warnsignale hinsichtlich der Mitarbeit des Probanden. Da aus einer mangelnden Mitarbeit in einzelnen Tests nicht auf eine mangelnde Mitarbeit in allen anderen Tests geschlossen werden darf, muss der Verlauf der Genauigkeiten eingehender analysiert werden.

Eine grobe Abschätzung des Gesamtniveaus der Genauigkeit unter Verwendung von Formel (1) führt zu einem niedrigen Wert knapp über 0,5. Augenfällig weichen davon die hohen Werte über dem Test SZT ab, der drei Teile hat, wovon der letzte der schwierigste ist. Während die Aufmerksamkeitstests aus der TAP eine Erwartungsspannung erzeugen, verhindert sie der *SZT,* der allerdings Vergleichs- und Selektionsaufgaben enthält [14]: Nach einen Tastendruck erscheint direkt

13 Für die Tests der Beschwerdenvalidierung (BSV) existieren keine Normwerte für die Reaktionszeiten.

14 Alltäglich wird von Probanden über Konzentrationsstörungen geklagt, nicht jedoch über Aufmerksamkeitsdefizite. Herkömmliche Konzentrationstests verlangen einen weitaus höheren Anteil an Unterscheiden und Vergleichen von Merkmalen als die Tests aus der TAP. Sie stellen daher höhere Anforderungen an das Arbeitsgedächtnis als computerisierte Reaktionszeitmessungen. Papiervorlagen haben meist auch einen erhöhten motorischen Aufwand, dessen Einfluss schwer zu kontrollieren ist. In Abwandlung einer Konzeption

auf dem Bildschirm die nächste Aufgabe. Obwohl zwischen neun Zifferntasten jeweils zu wählen und daraufhin ein längerer Weg mit der Hand zurück zu legen ist, steigen die Genauigkeiten beachtlich an! Das zeigt zuerst, dass es keine motorische Hemmung gibt. Sodann lässt der SZT für den Probanden keine Pause aufkommen, in der Ablenkungen oder Motivationsschwund aufkommen könnten, was die Schwankungsbreiten und die Fehlerzahlen reduziert und zu Genauigkeiten knapp unterhalb des Maximalwerts 1 führt. Da die Reaktionszeiten bereits aufgrund der Testkonstruktion länger ausfallen, führt die generell zu beobachtende langsame Reaktionsweise des Probanden zu Zeiten, die zwei bis drei Sekunden über den oberen Rand des folgenden Diagramms 4 hinaus ragen. Der Proband reagiert zwar langsam, jedoch in der schwierigsten Aufgabe in Normwerten sogar am besten, und zwar mit einem Ergebnis im Normalbereich, dem zweiten neben *Go/NoGo!* Er kann also genau arbeiten und folglich sind die



- Diagramm 4: Reaktionszeiten in Sek. (graue Balken), Genauigkeiten (Rhomben mit gepunkteter Linie) und einige Regressionsgeraden der Genauigkeiten (dunkelgraue Geraden) und der Reaktionszeiten (gestrichelte Geraden; für SZT nach unten verschoben).

erheblich schlechteren Genauigkeiten in den anderen Tests in erster Linie auf motivationale Einflüsse zurückzuführen.

Aus Diagramm 4 ist zu entnehmen, dass der Proband nach dem Einüben in *BSV_V* in *BSV_A* an Tempo zulegt ohne merklich an Genauigkeit einzubüßen. Dieser Lernprozeß stoppt in der *Alertness*, die Genauigkeiten sinken ab, während das Tempo einigermaßen konstant bleibt. (Die Zeiten sind kürzer als in der *BSV*, weil nur mit einer Taste zu reagieren ist.) Im Test *Flexibilität,* der wieder mit zwei Tasten auszuführen ist und höhere Anforderungen stellt, ist ein leichtes Absinken der Genauigkeiten zwar zu erwarten, doch die Daten gehen hier deutlich in die gegenläufige Richtung: Die Zeiten nehmen steiler zu als die Genauigkeiten abnehmen. Insbesondere springen die Fehlerzahlen in die Höhe (siehe Diagramm 3) und führen zu den niedrigsten Genauig-

von Bredenkamp und Häsgen lässt sich Konzentration als ein Vektor mit den beiden Komponenten 'Reaktionszeit' (in Normwerten) und 'Genauigkeit' auffassen, der je nach Ergebnissen in einen von vier Quadranten zeigt: 'konzentriert', 'refelexiv', 'konzentrationsgestört' und 'impulsiv'. Im vorliegenden Fall würde eine hohe Konzentration nur im Test SZT_3 erreicht.

keitswerten. Danach kommt eine eher gleichmäßigere Ausführung in der *Verdeckten Aufmerksamkeitsverschiebung,* gefolgt vom bereits beschriebenen *SZT.* Am Schluss stürzen die Genauigkeiten in der *BSV_G* ab, während die Zeiten nur etwas absinken. Der Schwund der Mitarbeit ist hier sofort zu sehen. Da sich für plötzlich und kurzfristig eintretende Ermüdungszustände aus dem weiteren Verlauf der Testuntersuchung keine Hinweise ergeben, ist davon auszugehen, dass der Proband in den Aufmerksamkeitstests seine Erwartungsspannung nicht aufrecht erhält, weil er zur Mitarbeit nur unzureichend motiviert ist und/oder ablenkende Vorstellungen hat. Folglich verweisen die insgesamt schlechten Ergebnisse unterhalb des Normalbereichs nicht auf Funktionseinbußen.

5. Erläuterungen und Begründungen (Rechtfertigungen)

Direkte Messvorgänge, die auf dem jeweiligen historisch-technischen Stand des Alltagshandelns ohne Rückgriff auf fachspezifische Theorien durchgeführt werden können, sind in der Psychologie eher selten; in der Experimentalphysik bilden sie im MKSA-System eine unentbehrliche Grundlage aller Messvorgänge [15]. Die Genauigkeitsfunktion beruht auf direkten Messvorgängen: Jeder 'normale' Erwachsene kann die Zeit messen und Fehler abzählen. Dennoch bedarf die Konstruktion dieser Funktion einer Begründung, ja einer Rechtfertigung, weil zur Ergebnisinterpretation verwendet wird. In der Testtheorie ist man an abgeleitete Messvorgänge gewöhnt, für die vorgängig bereits Annahmen über Zusammenhänge mathematisch ausformuliert sind, im einfachsten Fall durch Additionen. Ein bekanntes Beispiel hierfür ist ein Intelligenztest. Darin werden Wortschatzübungen mit Logikaufgaben durch Abzählen gelöster Aufgaben gleich gesetzt und eine gewisse Anzahl solcher Gleichsetzungen zu einem IQ-Koeffizienten addiert. Die Lösung der Aufgaben setzt allerdings ein Wissen und Können voraus, das über bloßes Reagieren für Zeitmessungen und und das Zählen von Fehlern hinaus geht. Die Rechtfertigung für die Gleichsetzung der unterschiedlichen Aufgaben erfolgt indes erst im Nachhinien durch die Berechnung von bestimmten Korrelationskoeffizienten (Itemschwierigkeit, Reliabilität, Validität), die an großen Stichproben gewonnen werden. Man hat sich an dieses Vorgehen gewöhnt, bei dem apriori unklar ist, was gemessen wird, der übergreifende Zweck jedoch die Mittel heiligt [16].

Daher sei folgende Rechtfertigung gewagt: Für die Formeln der Genauigkeit besteht die einfachste und trivialerweise vermutlich 'immer wahre' Annahme darin, den Zusammenhang in der Testperson (dem Probanden) anzusiedeln, weil es ein- und dieselbe Person ist, die sowohl die Reaktionsschwankungen als auch die Fehlerzahlen produziert. Die hierfür nötige 'intuitive Idee' für den Zusammenhang lautet, dass der Mensch ein schlecht funktionierender Computer oder ein solcher ist, auf dem ein fehlerhaftes Programm läuft ist. Denn ein gut funktionierender Computer reagiert im konstanter Verzögerung und fehlerlos auf einen eingehenden Reiz, wodurch s ~ 0 und fz ~ 0 werden und die höchste *Genauigkeit* erzielt wird. Im Grunde folgt die computerisierte Aufmerksamkeitstestung dieser Idee, allerdings wird der 'menschliche Normalbereich' über die Verteilung einer Stichprobe definiert. So gesehen werden Reaktionsschwankungen ebenfalls zu Fehlern und Fehler dürfen mit anderen Fehlern in einen Zusammenhang gebracht und in

15 Vogel H. Gerthsen Physik, 20. Aufl., Berlin-Heidelberg (Springer Verlag)
16 Janich P. Was ist Wahrheit? Eine philosophische Einführung. München 2000

Zahlenverhältnissen ausgedrückt werden. Kurz, was bei ersten Lesen der Formeln für die Genauigkeit befremdlich oder irritierend sein mag, ist nur das in der Psychologie Übliche – jedoch hier noch Ungewohnte. Folglich gibt es zwei Arten von Fehlern: Reaktionsschwankungen einerseits und Auslassungen sowie falsche Reaktionen andererseits. Damit verschwinden die qualitativen Unterschiede und es bleiben nur noch solche des Skalenniveaus. Denn es gibt eine halbe Sekunde, aber keinen halben Fehler. Dennoch erlaubt die übliche Rechenpraxis eine halbe Sekunde mit ganzzahligem Fehler zu multiplizieren, zu dividieren und zu addieren. Bereits durch die Formeln für So und So werden die ganzzahligen Messergebnisse in rationale Zahlen transformiert, die nur rechnerisch existieren, aber nicht gemessen werden. Da sich die Art des Zusammenhangs von Ko und So nicht aus den Messvorgängen selbst ergibt, wird dieser durch die mit den Ergebnissen verfolgten Zwecke bestimmt, nämlich auf möglichst einfache Weise eine gut aufgefächerte Verteilung der Werte im Intervall]0 ; 1] zu erhalten. Also wird der Zusammenhang axiomatisch formuliert und rechtfertigt sich im Nachhinien durch seine Brauchbarkeit, seinen erfüllten Zweck.

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

Stand des Papiers: Dezember 2018