

Quantitative und qualitative Bestimmung kognitiver Leistungsdefizite im Messmodell FAMOUS¹

1 Intelligenzforschung, Demenzforschung

Wenn wir im Alltag über „Intelligenz“ reden, wissen wir genau, was mit dem Wort gemeint ist. Ab einer bestimmten Stufe der Sprachentwicklung wird der Ausdruck, ohne unsere Vorstellungskraft besonders zu fordern, mit einem konkreten Sinn verbunden. Die Wortbedeutung realisiert sich automatisch im natürlichen Sprachempfinden. Ganz anders verhält es sich, wenn der Begriff „Intelligenz“ im Hinblick auf seinen Sinngehalt im System empirisch begründeter Psychologie untersucht werden soll. Beim Versuch, eine wissenschaftlich haltbare Definition des mit „Intelligenz“ intendierten Sachverhaltes zu geben drängt sich eine Fülle von schwierig fassbaren Bedeutungsaspekten auf, deren Berücksichtigung als unerlässlich erscheint. Noch mühseliger, ja geradezu waghalsig erscheint der Versuch, Intelligenz über ihre begriffliche Präzision hinaus auch *quantitativ* zu bestimmen. Zu den Inponderabilitäten, die sich durch die schwierige Begriffsbestimmung ergeben, kommen nun auch Unbestimmtheiten hinzu die auf die Modelleigenschaften des zur Anwendung gebrachten Messinstrumentes Zurückzuführen sind.

Dieser, in groben Zügen dargestellten Problematik, nämlich der Aufgabe, die intellektuelle Leistungsfähigkeit einem Messvorgang zu unterziehen, widmet sich ein Teilgebiet der Neuropsychologie, die Pathopsychometrie – seit Jahrzehnten, mit wechselndem Erfolg. Im Einführungsteil meines Referats unternehme ich in einem kurzen Überblick den Versuch, die Schwierigkeiten darzustellen, mit denen die neuropsychologische Forschung im hier zur Diskussion stehenden Problembereich der „Intelligenzmessung bzw. Demenzbeurteilung“ konfrontiert wird.

Viele von Ihnen werden den ironisch gemeinten Spruch kennen, Intelligenz sei eine Eigenschaft, die mit Hilfe von Intelligenztests gemessen wird. Hinter dieser Tautologie verbirgt sich die Frustration vieler Psychologen, die spätestens um die Mitte des alten Jahrhunderts erfahren mussten, dass die von ihnen entwickelten Testinstrumente sich im praktischen Gebrauch eher schlecht als recht bewähren konnten. In den Vereinigten Staaten wurden lange Zeit alle Rekruten wie auch Einwanderer einem Intelligenztest unterzogen. Bei der Analyse der im Lauf der Zeit erhobenen Daten stellte sich heraus, dass die Mehrzahl der mit Hilfe des Intelligenztests „Armee Alpha“ beurteilten jungen Amerikaner eine unterdurchschnittliche Intelligenz aufwies.

¹ Autoren: Viktor Szendrödi, Ernst-Jörg Friedl, Giselher Guttmann.

Ein solches Resultat weist aber auf die willkürliche Bestimmung des im Test definierten Intelligenzbegriffes hin. Es ergab sich auch, dass ein namhafter Nobelpreisträger, der vor dem Zweiten Weltkrieg in die USA einwanderte und im Sinne strenger Gesetze ebenfalls einer Testung unterzogen wurde, nur einen IQ von 80 aufwies und somit offiziell als grenzdebil gelten musste.

Mit ähnlichen, wenngleich weniger dramatisch erlebten Widersprüchen wurde unser Forschungsteam konfrontiert, als wir vor einiger Zeit Hirnleistungsstörungen mit Hilfe der gängigen Instrumenten zu quantifizieren versuchten: Zwischen den erhobenen Testergebnissen und dem klinischen Eindruck, den die Patienten boten, ergaben sich wesentliche, nicht selten eklatante Unterschiede. Wenn aber ein Messmodell die intuitiv gefassten Vorstellungen über das in ihm zur Darstellung gebrachte Phänomen nicht befriedigt, so ist der Vorwurf der Entfremdung oder zumindest verzerrten Wiedergabe des Untersuchungsgegenstandes berechtigt. Die Verlässlichkeit der von uns verwendeten Messinstrumente war also sichtlich unbefriedigend, und es war nahe liegend, nach den Ursachen der im praktischen Gebrauch sichtbar werdenden Fehler zu suchen. Um die typischen Probleme, mit denen sich die Intelligenzforschung seit jeher zu tragen hat, näher zu bringen, möchte ich an dieser Stelle einen kurzen Blick auf die Geschichte der Messung menschlicher Intelligenz werfen.

Einer der Ersten, der in der zweiten Hälfte des 19. Jahrhunderts den Versuch unternahm, wissenschaftlich fundierte Aussagen über mentale Fähigkeiten zu machen, war Francis Galton, ein Cousin von Charles Darwin. Galton hielt seinen Vetter Darwin, die gemeinsamen Vorfahren und nicht zuletzt auch sich selbst für außergewöhnlich intelligent, und es dürfte kein Zufall gewesen sein, dass genetisch bedingte Momente der Intelligenz einen zentralen Platz in seinem Forschungsprogramm einnahmen. In diesen, aus festen Überzeugungen resultierenden Ansatz flossen freilich unvermeidbar subjektiven Elemente ein, die sich wesentlich auf die Verlässlichkeit der gefundenen Resultate auswirkten.

Lebhaftes Vorstellungsvermögen hielt Galton für ein sicheres Zeichen guter Intelligenz. Im Sinne dieser eher einfachen, heute würden wir sagen *eindimensionalen* Modellvorstellung forderte er in Rahmen seiner Versuche die aus Familien- und Bekanntenkreis rekrutierten Versuchspersonen auf, sich ihren Frühstückstisch vorzustellen und zu beschreiben. Die Auswertung der Reaktionen erfolgte nach Maßgabe eines ausgeklügelten Punktesystems, in dem spontane Reaktionen ebenso eine Rolle spielten, wie Antworten auf geschlossene Fragekategorien bezüglich Deutlichkeit, Detailreichtum, Farbqualität und vergleichbare Merkmale der Vorstellungsbilder. Zur Überraschung Galtons warteten in seinem Experiment Frauen und Kinder mit besseren Ergebnissen als männliche Versuchspersonen auf, unter denen sich viele im abstrakten Denken geschulte Wissenschaftler befanden. Im Wortgebrauch unserer Gegenwart lässt sich nun festhalten, dass die Validität des von Galton entworfenen „Tests“ nicht besonders hoch war.

Das erste, auch nach heutiger Auffassung als Intelligenztest geltende Messinstrument wurde 1905 vorgelegt. Auch wenn dieses, von Binet konzipierte Verfahren bestimmte praktische Erwartungen recht gut zu erfüllen vermochte, die Frage nach sei-

ner Validität blieb offen. Die Bestimmung allgemein gültiger Kriterien der Validität von Intelligenztests – dass nämlich der Test tatsächlich das misst, was er vorgibt zu messen – stellt nach wie vor ein umstrittenes Problem dar. Nicht zuletzt deshalb, weil – wie dies einleitend angesprochen wurde – der Intelligenzbegriff von recht unterschiedlichen Vorstellungen geprägt wird.

Seit Galton und Binet widmete sich eine große Anzahl namhafter Forscher der Problemstellung, Intelligenz im Allgemeinen und in ihren individuellen Manifestationen zu begreifen. Guilford ermittelte in seinem 1956 entworfenen *Strukturmodell der Intelligenz* auf intuitivem Wege 120 elementare Leistungsbereiche. Unter den theoretisch erschlossenen Fällen, die im Raster seines dreidimensionalen Strukturmodells Platz hatten, konnte er 50 mit konkreten Beispielen illustrieren.

Bereits Guilfords intuitive Arbeitsweise führt zu der Überzeugung, dass die Auffassung, Demenz einfach als Gegensatz von Intelligenz anzusehen falsch sei. In einer viel zitierten Bemerkung bringt er die Ansicht zum Ausdruck, dass Intelligenz sich in nur 50 Verhaltensformen zu erkennen gäbe, die Zahl der Spielarten von Unvernunft sei aber Legion.

Die Messung intellektueller Leistungsfähigkeit basiert – auf ihre einfachsten Wesenszüge reduziert – bekanntlich auf einem standardisierten Verfahren, in Rahmen dessen die Testperson mit Aufgabenstellungen konfrontiert wird, deren Lösungsgüte – nach Auffassung der Testkonstrukteure – mit der Intelligenz der Testperson korreliert. Ich betone deshalb das Moment „individuell geprägter Ansichten“, weil bis heute keine Übereinstimmung darüber besteht, welche Leistungsmomente tatsächlich charakteristisch für das kognitive Leistungsvermögen sind. In den dreißiger Jahren des alten Jahrhunderts begann ein wahres Wettrennen, raffiniert erdachte Items zu kreieren, von denen jeder für sich den Anspruch erhob, das kognitive Leistungsvermögen zu reflektieren.

Der Verdienst, die unüberschaubare Vielfalt möglicher Leistungsformen auf wenige Grunddimensionen zurückzuführen, kommt Adolf Otto Jäger zu. Sein dimensional aufgespanntes Modell der Intelligenz entwarf Jäger 1965, indem er alle bis zum Zeitpunkt seiner Untersuchungen bekannte Itemtypen der Intelligenzmessung einer faktorenanalytischen Prüfung unterwarf.

Nicht unter dem Aspekt der differenziellen Psychologie betrachtet, sondern vom Blickwinkel der Kognitionspsychologie her gesehen überprüfte Friedhart Klix 1990 alle ihm bekannte Ergebnisse der Grundlagenforschung und konzipierte als Resultat eingehender Studien das „Modularsystem menschlicher Wissensrepräsentation“, in dem Intelligenz nicht als statisches Konstrukt, sondern als funktional dimensioniertes Gebilde kognitiver Vorgänge zur Darstellung gelang.

2 Kritik an gängigen Verfahren

In der kurz nachgezeichneten Entwicklung der Lösungsansätze ging es den einzelnen Autoren darum, ein wissenschaftlich gesichertes Bild der Intelligenz mit messbaren Eigenschaften zu entwerfen. Das Hauptanliegen pathopsychometrischer Mess-

vorgänge besteht allerdings nicht in einer einfachen Quantifizierung kognitiver Fähigkeiten, sondern in der Gewinnung *krankheitsrelevanter* Informationen. Der Wert einer klinisch-psychologischer Untersuchung wird stets durch die Differenziertheit diagnostisch und vor allem therapeutisch verwertbarer Angaben bestimmt.

Typische Abweichungen von der Norm mit unterschiedlicher Genese stellen differentialdiagnostische Hinweise dar. Kognitive Leistungsminderungen können unter anderem auf ungünstige Umwelteinflüsse, neurologisch bedingte Störungen, altersbedingte Änderungen, prozesshaft vollzogene Einschränkungen zurückgeführt werden aber auch als Folge situationsbedingter Leitungsschwankungen auftreten. Weit entfernt sind die derzeit gebräuchlichen Testinstrumente jedoch davon, die im klinischen Alltag gestellten Erwartungen zu erfüllen und Informationen über die Genese einer messbaren Intelligenzminderung zu liefern.

Guilfords Feststellung, wonach einer beschränkten Anzahl kognitiver Leistungskategorien eine ungeordnete Vielzahl dementieller Störungsbilder gegenübersteht, soll nun die These entgegenhalten werden, dass im Bezugsrahmen eines entsprechend aufgespannten *dynamischen* Modells der Intelligenz auch mögliche „Spielarten der Unvernunft“, das heißt diverse Formen der Demenz mit begrifflicher Klarheit definiert werden können. Diese Überzeugung diene als Grundsatz bei der Bestimmung der Modelleigenschaften des computerassistierten Verfahrens FAMOUS. Der Test wurde von den Autoren mit dem Vorhaben konstruiert, häufig beobachtbaren Unzulänglichkeiten gebräuchlicher Instrumente der Demenzbeurteilung durch eine neuartige Annäherung der Problematik beizukommen.

Die Kritik an rezenten Testinstrumenten, die zur theoretischen Fundierung von FAMOUS führte, lässt sich in den folgenden Punkten zusammenfassen:

- 1) *Kurrente Instrumente der kognitiven Leistungsbeurteilung decken nicht alle für die menschliche Problemlösefähigkeit relevante Funktionsbereiche ab.* Dieses Manko lässt sich anhand des Vergleichs mit dem Buchstabenstempel einer Schreibmaschine illustrieren, in dem bestimmte Symbole gar nicht, andere wieder mehrfach vorkommen, wodurch der Umfang korrekt darstellbarer Wörter wesentlich eingeschränkt wird.
- 2) *Weder das Tempo noch der motivationelle Hintergrund kognitiver Prozeduren, welche die Testergebnisse wesentlich beeinflussen, werden in gängigen Konstrukten berücksichtigt.* Ob ein schlechtes Resultat auf ein vorübergehend reduziertes Antriebsniveau, geistige Verwahrlosung oder ein nicht korrigierbares Leistungsdefizit zurückzuführen ist, lässt sich aus Resultaten bekannter Tests nicht ableiten
- 3) *Die Sensibilität kurrenter Verfahren erstreckt sich jeweils nur über ein relativ schmales Spektrum möglicher Leistungsbereiche, fein abgestufte Übergänge zwischen normalen und pathologisch erscheinenden Fällen können nicht verdeutlicht werden.* Ausgesprochen nachteilig wirkt sich dieser Mangel auf die Möglichkeit der Früherkennung dementieller Prozesse aus. Typische Demenz-

tests schlagen erst an, wenn der geistige Abbau bereits so weit fortgeschritten ist, dass er schon mit dem „klinischen Blick“ des Diagnostikers erkennbar ist.

- 4) *Dynamische Momente, die über Verlaufseigenheiten dementieller Zustandsbilder Auskunft erteilen und damit differentialdiagnostische Aussagen mit einer Messung ermöglichen, werden in Ergebnissen herkömmlicher Tests nicht ersichtlich.* Ein Testergebnis, das dem Probanden einen Messwert in der Größenordnung von 100 IQ zuweist, ist im Allgemeinen unauffällig – auch in dem Fall, in dem die betreffende Versuchsperson vor einiger Zeit noch einen Messwert von 140 IQ lieferte.
- 5) *Die Ergebnisse herkömmlicher Verfahren lassen kein dimensional aufgespanntes, krankheitsspezifisches Ausfallsmuster erkennen.* Dieser letzte Kritikpunkt führt zur einleitend erwähnten Forderung zurück, wonach der Bezugsrahmen eines entsprechend aufgespannten Modells intelligenten Verhaltens auch eine begrifflich klare Definition unterschiedlicher Formen kognitiver Einbußen ermöglichen soll.

3 Das Messmodell FAMOUS

Um den dargestellten Mängeln entgegenzusteuern, wurden in den konzeptuellen Grundlagen von FAMOUS drei bekannte Ansätze psychologischer Forschung zusammengeführt und integrativ verarbeitet:

- a) Faktorenmodell der Intelligenz nach Adolf Otto Jäger
- b) Modulare System menschlicher Wissensrepräsentation nach Friedhart Klix und
- c) Der Ergopsychometrische Ansatz von Giselher Guttman

Die wesentlichen – in den konzeptuellen Grundlagen von FAMOUS berücksichtigten – Eigenschaften der von Jäger und Klix erarbeiteten Systeme bestehen in theoretisch fundierten Vorstellungen über ein *erschöpfendes Repertoire intellektueller Leistungsformen*.

Primäre Leistungsdimensionen:

- Konzentrationskraft/Aufmerksamkeit
- Sprachgebundene Fertigkeiten
- Anschauungsgebundene Fertigkeiten
- Einfallsreichtum/Produktivität
- Selektion/Projektion
- Verkürzungen
- Vergleiche
- Verkettungen
- Verdichtung/Inversion
- Verarbeitungskapazität/Urteilsvermögen

Aus der Fusion der beiden Modellvorstellungen ergibt sich ein zehndimensionales Gefüge intelligenter Leistungsformen. Das Profil der in den diversen Bereichen erzielten Werte charakterisiert individuell ausgeprägte Begabungen ebenso wie die typische Gestalt pathologisch anmutender Leistungsausfälle.

Neben dem *Punktesummenscore* und den 10 *Funktionsbereichen*, deren standardisierte Skalenwerte zwischen 1 und 10 variieren können, enthält das automatisch erstellte Ergebnisprotokoll drei faktorenanalytisch ermittelte, die Einzelresultate verdichtende Größen:

- 1) Konvergentes Denken
- 2) Divergentes Denken
- 3) Integrative Denkvorgänge

In der Dimension „*Konvergentes Denken*“ sind alle Subtests, bei denen für die erfolgreiche Aufgabenbewältigung die Fähigkeit einer *konzentrierten und selektiven Aufmerksamkeitszuwendung* die Voraussetzungen bildet, mit einer hohen Ladungszahl repräsentiert.

Der Faktor „*Divergentes Denken*“ wird inhaltlich von Subtests definiert, die nach einem *ungestörten Denkfluss* und der begrifflich geordneten *Steuerung assoziativer Abläufe verlangen*.

Auf der Skala „*Integrative Denkvorgänge*“ erweisen sich jene Aufgabenstellungen als inhaltsbestimmend, welche den *Vollzug komplexer Denkvorgänge* bei gleichzeitiger Betrachtung mehrerer problemrelevanter Aspekte abverlangen.

Aus der Gegenüberstellung typischer Messwerte ergeben sich schließlich zwei, im Ergebnisprotokoll gesondert ausgewiesene, dynamische Parameter.

- a) Ergopsychometrischer Index
- b) Differenziertheitsgrad

Guttmanns Ergopsychometrisches Konzept

Erfassung persönlichkeitsrelevanter und motivationell bedingter Parameter, die auf signifikante Weise mit dem Erscheinungsbild dementieller Einbußen variieren.

Auf der Beobachtung, dass ein weites Spektrum menschlicher Leistungsdisposition in direkter Abhängigkeit der aktuellen Stressbelastung variiert, basiert die Grundidee des Ergopsychometrischen Ansatzes von Gisela Guttmann. In Anlehnung an das Ergopsychometrische Konzept werden in der Testanlage, durch kontrollierte Variation der Versuchsbedingungen, *dynamischen Momente* des Problemlöseverhaltens erfasst. In den Ergebnissen werden charakteristische Leistungsänderungen auf standardisierte Weise zur Darstellung gebracht.

Ergopsychometrischer Index

$E_{psyl} =$ Leistung unter *normalen*
Versuchsbedingungen -
Leistung bei *Belastung*

Der Index kann im Messbereich zwischen
- 7 und +7 variieren

Unauffällig gilt ein Wert in der Nähe von Null

Der im Test ermittelte *Ergopsychometrische Index* wird als Differenzbetrag zwischen Leistungen *ohne* und *mit* Stressbelastung gewonnen. Der Index gibt Auskunft über die Wirkung stabiler oder vorübergehend in Erscheinung tretender Persönlichkeitsmerkmale, die auf signifikante Weise mit dem Erscheinungsbild dementieller Einbußen variieren.

Ergopsychometrischer Index

*Differentialdiagnostische
Aussagemöglichkeiten:*

- Irritierbarkeit durch Stressbelastung (+)
- Ermüdbarkeit (+)
- Diffuse hirnorganische Beeinträchtigung (+)
- Psychomotorische Verlangsamung (-)
- Spätverwahrlosung (-)
- Neurotisch bedingte Denkhemmungen (-)

Der *Ergopsychometrischen Index* erbringt Hinweise auf Zustandsbilder, die aus Resultaten herkömmlicher Demenztests nicht ableitbar sind. Die Maßzahl lässt sich im Falle einer *positiven* Ausprägung als Kennwert verstehen, der über das Ausmaß der *Irritierbarkeit* unter Stressbelastung Auskunft erteilt und vor allem mit organisch bedingten *Denkstörungen* korreliert. Im *negativen* Skalenbereich kommt ihm die Bedeutung eines Wertes zu, der über die Wirksamkeit bestimmter, vor allem neurotisch bedingter *Denkhemmungen* oder *Auffassungsschwierigkeiten* informiert.

Mit dem Ausdruck „Spätverwahrlosung“, die ebenfalls in Form einer negativen Ausprägung des *Ergopsychometrischen Indizes* sichtbar wird, lässt sich ein Zustand „geistiger Trägheit“ beschreiben, der stets im Erwachsenenalter auftritt und bei Personen zu beobachten ist, die relativ gute Problemlösefähigkeiten besitzen, diese Anlage aber zunehmend verkümmern lassen. Ein typisches Beispiel für Störungen dieser Art liefern Alkoholiker im anfänglichen Abbaustadium.

Im Parameter „*Differenziertheitsgrad*“ werden Maßzahlen ontogenetisch früh in Erscheinung tretender Fertigkeiten mit Messwerten in Relation gebracht, die Leistungsformen charakterisieren, welche erst in einer späteren Entwicklungsphase zur vollen Ausbildung gelangen. Diese in der Abhängigkeit von Umwelteinflüssen heranreifenden Formen des kognitiven Leistungsvermögens besitzen eine gewisse Stabilität gegenüber altersbedingter Leistungsreduktion. Sie sind gegenüber diffus wirkenden Noxen, nicht aber gegenüber umschriebenen Hirnschädigungen stabil. Die früh in Erscheinung tretenden, „hypostasierenden“ Fähigkeiten reagieren vor allem auf altersbedingte Leistungseinbußen und diffus wirkende Noxen sensibel. Durch die Abweichung einzelner Werte von der zu erwartenden Gleichverteilung der Parameter *Konvergentes Denken*, *Divergentes Denken* und *Integrative Denkvorgänge* nimmt der Differenziertheitsgrad Ausprägung im positiven oder negativen Zahlenbereich an. Während aber die Ausschläge beim *Ergopsychometrischen Index* mit +7 bzw. -7 limitiert sind kann die für den Differenziertheitsgrad stehende Maßzahl in Extremfällen auch diese Grenzen überschreiten.

Bezeichnende Abweichungen werden beispielhaft auf der folgenden Folie dargestellt.

Differenziertheitsgrad

*Differentialdiagnostische
Aussagemöglichkeiten:*

- primäre Minderbegabung (0)
- manische Verstimmung (0 / -)
- depressives Bild (+)
- sekundäre Leistungseinbußen (++)
- neurologisch bedingte Ausfälle (+/-)
- intellektuelle Verwahrlosung (-)

Auf dem folgenden, konzentriert dargestellten Computerausdruck sind die Testwerte einer 71-jährigen Dame dargestellt, deren dementielle Beeinträchtigen im letzten, der Testabnahme voran gegangenen Jahr nach und nach ein Ausmaß annahm, dass sie in der familiären Umgebung deutlich registrierbar geworden ist.

Profil einer 71jährigen Frau mit corticalen Atrophiezeichen im CT

> Gesamtwert	39
> Klinische Einstufung	Leicht Beeinträchtigt
> Konvergentes Denken	13
> Divergentes Denken	15
> Integrative Denkvorgänge	11
> Ergopsychometrischer Index	1
> Differenziertheitsgrad	-3

Die mäßig fortgeschrittene, durch eine MR-Untersuchung verifizierte vaskuläre Demenz findet im Testprofil neben dem niederen Gesamtwert in der negativ ausgeprägten Maßzahl für den Differenziertheitsgrad und im relativ hohen Messwert entlang der Dimension „Divergentes Denken“ eine Entsprechung. Der Ergopsychometrische Index müsste, wie auch der Differenziertheitsgrad eine positive Ausprägung haben, wenn es sich nicht um eine Multiinfarkt-Demenz sondern um eine Demenz von Alzheimer-Typ handelte.

Beim klinischen Gebrauch des Instrumentes ließ sich feststellen, dass neben dem isoliert betrachteten Aussagewert der Parameter „Ergopsychometrischer Index“ und „Differenziertheitsgrad“, auch der kombinierte Informationsgehalt dieser Größen eigenständige, differentialdiagnostisch verwertbare Hinweise in sich verbirgt. Über einschlägige Forschungsergebnisse werden meine Kolleginnen Frau Mag. Auer-Steiner und Frau Mag. Gadinger-Häupl berichten.