

ZUR ENTWICKLUNG DER UNTERSUCHUNGSMETHODE:

Grundlagen:

In **Leistungsdiagnostik** erwies es sich oft als nützlich **neben speziellen Leistungen** (z.B. Prüfung der Intelligenz, Prüfung des technischen Verständnisses, usw.) auch **gewisse allgemeine Voraussetzungen zur Erzielung von Leistungen zu berücksichtigen**, wie z.B. Konzentration, Aufmerksamkeit, Anstrengung, usw. -> Test-d2 ist ein solcher.

Spezielle Funktionen und allgemeine Voraussetzungen von Leistungen **können nicht unmittelbar gemessen werden**, sondern **Diagnose kann nur über das Verhalten** erfolgen, das bei bestimmten, paradigmatischen Tätigkeiten gezeigt wird.

Für allgemeine Leistungstests sind das **Tätigkeiten, die ein hohes Maß an Aufmerksamkeit und Konzentration, jedoch KEINE speziellen Fähigkeiten und Fertigkeiten erfordern**.

Derartige Tests verwenden:

- mehr oder weniger **einfache Rechenaufgaben** (z.B. Pauli-Test)
 [Nachteil: spezielle Rechenfähigkeit der TP oder auch ihre Aversion gegen Zahlen kann Ergebnis beeinflussen!] oder
- **Diskriminationsaufgaben** wie die von **Bourdon** eingeführten **Durchstreichtests**.
 [Vorteil: visuelle Aufmerksamkeit wird fokussiert, Buchstaben oder Zeichen haben bei TP ungefähr den gleichen Bekanntheitsgrad
 -> Leistungen werden so vergleichbar]

Definition von Aufmerksamkeit:

Aufmerksamkeit ist Selektion (RÜTZEL)

⇒ **Test-d2 ist somit also ein Aufmerksamkeitstest**

Definition von Konzentration:

Konzentration ist eine **leistungsbezogene, kontinuierlich und fokussierende Reizselektion**, die Fähigkeit eines Individuums, sich **bestimmten (aufgaben)relevanten internen oder externen Reizen selektiv**, d.h. unter Abschirmung gegenüber irrelevanten Stimuli, **ununterbrochen zuzuwenden** und diese **schnell und korrekt zu analysieren** (BRICKENKAMP)

⇒ **Test-d2** verlangt eine **auf externe visuelle Reize bezogene Konzentrationsleistung**.

Sie kommt zustande **durch individuelle Koordination von Antriebs- und Kontrollfunktionen**, die sich in drei Verhaltensweisen manifestieren:

- 1) im **Tempo** und in der **Quantität** (d.h. in der **Menge des in einer bestimmten Zeiteinheit bearbeiteten Materials**) -> **ANTRIEB**
 - 2) in der **Qualität** (d.h. in der **Güte, Sorgfalt und Genauigkeit der Bearbeitung**, die sich invers aus dem Fehleranteil erschließen läßt -> **KONTROLLE**
 - 3) im **zeitlichen Verlauf der Leistung**, der Rückschlüsse auf Besonderheiten des **Arbeitsverhaltens** (z.B. Anfangserregung, Konstanz oder Instabilität, frühzeitige Sättigung, Ermüdung, usw.) ermöglichen soll -> **KONTROLLE**
- **Konzentrationsleistung ergibt sich aus der individuellen Koordination dieser Verhaltenskomponenten.**
 - Theoretisch gesehen ist sie das **Resultat aus dem Zusammenwirken von Antriebs- und Steuerungsinstanzen.**

Konstruktionsgesichtspunkte:

- Test-d2 wurde entwickelt am Institut für Sicherheit in Bergbau, Industrie und Verkehr des Technischen Überwachungsvereins Essen.
- Sollte in erster Linie der **Beurteilung der Kraftfahreignung** dienen
- **Vorbild = Durchstreichtest von Lauer** hatte hohe Validität zur Eignungsauslese von Lenkern;
 ABER: **Reizselektion nur zwischen 2 ähnlichen Buchstaben [c und o]**
 -> zu niedrig im Vergleich zu komplexen Orientierungssituationen im modernen Straßenverkehr;
 Auswertung von Durchstreichtests war damals noch schwierig und oft fehlerhaft, weil ohne Schablonen; es gab noch keine verlässlichen Normen
- **anderer solcher Durchstreichtest war jener von MEILI**;
 ABER: Instruktionen für weniger intelligente Personen zu schwer verständlich; ging nur als Einzeltest.
- **Daher dann: Durchstreichtests mit unbekannten Zeichen:**
 Ziel = für alle TP gleiche Ausgangsbedingungen;
 ABER: längere Einübungszeit; Ziel wurde nicht erreicht wegen unterschiedlicher Lernfähigkeit der TP.

⇒ **DAHER TEST-d2:**

- mißt **attention to detail** auf einem **höheren Komplexitätsgrad** als bei Lauer
- Instruktionen sind **einfach und leicht verständlich** formuliert
- Instruktionsverständnis wird vor dem Test anhand einer **Übungszeile** kontrolliert
- **Testaufgaben** können **ohne lange Einübung** ausgeführt werden.
- Verfahren ist sowohl im **Einzel-als auch im Gruppenversuch durchführbar**
- Gesamttest **in Teilzeiten gegliedert** -> TP wird so Kontrolle über ihr Arbeitstempo ermöglicht, Auswerter erhält unmittelbaren Überblick über den zeitlichen Leistungsverlauf
- primäre **Testgütekriterien Objektivität, Reliabilität und Validität** sind in **optimaler Weise** gewährleistet
- Verfahren ist **ökonomisch** (geringer Zeit- und Materialaufwand in Anwendung und Auswertung)
- **Umfangreiche und differenzierte Normen** für die Auswertung vorhanden

Zur Bewährung des Verfahrens:

1) **Anwendungsbereiche:**

Einsatz des Test-d2 in vielen medizinisch-psychologischen Untersuchungsstellen **zur Beurteilung der Kraftfahreignung**. Darüber hinaus:

- in der **Klinischen Psychologie**: Überprüfung der differentialdiagnostischen Leistungsfähigkeit verschiedener Patientengruppen, zur Therapieverlaufskontrolle, für Rehabilitation; bei Neurotikern, Schizophrenen, Depressiven, Epileptikern, Hirnorganikern, Hirnverletzten, HIV-Infizierten, AIDS-Kranken, Alkoholikern, älteren Patienten mit präseniler und seniler Demenz vom Alzheimer-Typ, bei Kindern mit MCD.
- in der **Arbeitspsychologie**: darunter auch in der Verkehrs- und Sportpsychologie
- in der **Pädagogischen Psychologie**, Psychologischen Diagnostik, **Pharmakopsychologie**
- in der **Forensischen Psychologie** und der **Umweltpsychologie**

2) Kritik und Verbesserungsmöglichkeiten:

vor allem an **vier Bereichen:**

a) Frage der Intelligenzunabhängigkeit des Verfahrens:

→ Fehlannahme einer **Nullkorrelation zwischen Testd2 und allen Intelligenzvariablen wurde unterstellt;**

ABER: es gibt nur eine weitgehende Intelligenzunabhängigkeit (gemeint ist damit, daß Testinstruktionen keine besonderen intellektuellen Ansprüche stellen.

→ **Zusammenhang zwischen Intelligenz und Konzentrationsleistung** (Korrelation zwischen Testd2 und HAWIE) **ergibt sich** (laut Jäger) **aus der Speed-Komponente** beider Tests.

→ **auffällige Stichprobenabhängigkeit der Ergebnisse:**

- z.B. ***bei psychiatrischen Patienten mehr signifikante Korrelationen*** zwischen Intelligenz- und Konzentrationstestvariablen als bei „Normalen“;
 - bei ***psychiatrischen Patienten*** wies die Mehrheit ***signifikante Zusammenhänge zwischen d2- und Intelligenztestwerten*** auf.
 - ***Korrelationen innerhalb der psychiatrischen Patientengruppen höher*** (z.B. Korrel. von 0,71 zwischen GZ-F [= fehlerkorrigierte Mengenleistung des Testd2] und dem Symbol-Test des HAWIE)
 - ***bei „Normalen“ (Gymnasiasten) höchste Korrelation (0,55) zwischen Mengenleistung des Testd2 und der Merkfähigkeit des IST***
- ⇒ **Diskrepanz der Befunde erklärt Rauchfleisch so:**
Die mit dem Testd2 erfaßte Belastungsfähigkeit spielt bei psychiatrischen Patienten gegenüber Gesunden eine ungleich größere Rolle für die Erbringung von verschiedenen intellektuellen Leistungen, wie sie im HAWIE geprüft werden.

b) Testlänge bzw. Testdauer

→ kurze Bearbeitungszeit, daher kann **Testd2 angeblich nur kurzzeitige Konzentrationsleistungen messen** und vorhersagen (ist aber nicht erwiesen)

Antwort darauf:

- **Viele Tätigkeiten** erfordern oft **hohe Konzentration, aber meist nie durchgehend für längere Zeit** (Beispiel: Tennisturnier -> höchste Konzentration in einem Zweistunden-Match nur bei Aufschlag und Ballwechseln erforderlich, das ist aber nur ca. eine Viertelstunde insgesamt)
- **Aufmerksamkeitsschwankungen** wurden **bereits in kurzen Intervallen von 5-8 Sekunden festgestellt** (Rohracher) -> **äußerste Konzentration** kann also offensichtlich **nur in extrem kurzen Zeitspannen** aufgebracht werden; Befürchtung Testd2 wäre zu kurz ist daher unbegründet.
- **Reliabilität erwies sich als sehr hoch**, daher keine Testverlängerung zur Erhöhung der Reliabilität notwendig.
- Testverlängerung (bzw. Testwiederholung) könnte sinnvoll sein, wenn zuverlässigere qualitative Meßwerte (vor allem für F und F%) benötigt werden.
ABER: **Fehleranteil als Indikator für Sorgfalt und Genauigkeit des Arbeitsverhaltens** erreicht auch so eine **Reliabilität**, die der von diversen Persönlichkeitsfragebogen entspricht.

c) **Bedeutung der distributiven Aufmerksamkeit:**

→ **Testd2 mißt** zwar die **konzentrativen (fokussierende) Aufmerksamkeit**, **nicht** aber die **distributive Aufmerksamkeit**:

- **fokussierende Aufmerksamkeit**
= auf eine Handlung bzw. ein Objekt gerichtetes Aufmerksamkeitsverhalten
- **distributive Aufmerksamkeit**
= auf mehrere Handlungen bzw. Objekte verteiltes Aufmerksamkeitsverhalten

Abhilfe: Vorgabe einer Mehrfachaufgabe, z.B. Testd2 und Reaktionstest

Defizit der Aufmerksamkeitsdiagnostik laut Süllwold:

- ⇒ Nichtbeachtung des Aufmerksamkeitsumfanges, der im Alltag und bei vielen beruflichen Tätigkeiten aber eine Rolle spielt.
Aufmerksamkeitsumfang läßt sich nicht mit PP messen, daher schlägt Süllwold zusätzliche Anwendung seiner Zahlentafeln dafür vor.

d) Problematik des sogenannten Gesamtleistungswertes (GZ-F)

- dem einfach korrigierten **Mengenleistungswert GZ-F** wurde **mangelnde Validität** angelastet, die zu schweren Fehlurteilen führen könne; ist aber ein Holler.

Testwert GZ-F wird gebildet aus der **Gesamtzahl der bearbeiteten Zeichen minus der Fehlerzahl (F)**.

Letztes durchgestrichenes Zeichen in einer Zeile begrenzt jeweils den Bereich der als bearbeitet geltenden Zeichen. D.h. **GZ** ist die **Summe der Zeichen**, die in den 14 Zeilen (Teilzeiten) **nach relevanten Stimuli** (d mit 2 Strichen) **durchsucht** wurden.

Fehlerzahl setzt sich zusammen aus:

- **Auslassungsfehlern (F1)** = häufiger (d.h. relevante Zeichen werden übersehen und daher nicht durchgestrichen) und
- **Verwechslungsfehlern (F2)** = relativ selten (d.h. irrelevante Reize werden versehentlich durchgestrichen)

Diverse **Reliabilitätskontrollen** haben ergeben, daß **GZ-F ein äußerst zuverlässiger Meßwert** ist.

GZ-F mißt die fehlerkorrigierte Mengenleistung (zuverlässiger erfaßbare Quantität geht aber stärker in diesen Wert ein als weniger zuverlässiger erfaßbare Qualität). Testdurchführungszeit wird konstant gehalten, daher:

GF-Z entspricht dem Individuellen Tempo der visuellen Informationsverarbeitung. Ist vor allem geeignet, den **Grad einer eventuellen Verlangsamung valide zu diagnostizieren**, der für bestimmte psychiatrische Krankheitsformen charakteristisch ist (z.B. bei Neurotikern, Depressiven, Hirnorganikern). Ist auch ein guter Prädiktor für die Kraftfahreignung und für die Mathematiknote.

Fazit:

Forschungsergebnisse sprechen eindeutig gegen generelle Validitätsmängel;

ABER: **Fehler beeinflussen den GZ-F Wert nur geringfügig**, d.h. Fehlerkorrektur reicht nur bei einem instruktionskonformen Testverhalten aus.
Bei höheren Fehlerzahlen -> **Möglichkeit einer Überschätzung der tatsächlich erbrachten Konzentrationsleistung.** Grenzen des GZ-F liegen dort, wo einseitige Übergewichte in Antrieb oder Kontrolle das Testverhalten bestimmen.

Extrem hohe Fehlerzahlen vom Typ F1 (**Auslassungsfehler**) kommen zustande, wenn **TP instruktionswidrig Testteile überspringen**

⇒ extrem hohe F1-Werte bei hohen GZ-Werten sind daher ein Syndrom für instruktionswidrige, oberflächliche Testbearbeitung (= **Ü-Syndrom**)

Verfälschungstendenzen können **bei allen Leistungstest** auftreten

→ **vor allem das Multiple-Choice-Antwortformat** bietet Gelegenheit dazu...

Abhilfe: Auswertemodus muß eine wirksame Korrektur vorsehen. Auftreten von Verfälschungstendenzen kann selten verhindert werden, daher ist es wichtig, daß sie entdeckt werden

⇒ ***im Testd2 ist das Ü-Syndrom eine Möglichkeit dazu.***

Wichtig für die Interpretation der Ergebnisse des Testd2:

⇒ nicht nur auf den GZ-F schauen, sondern **Fehlerprozentsetz mitberücksichtigen**, wenn sich Diskrepanzen zwischen den PR dieser beiden Parameter ergeben.

e) **Alternativen zum bisherigen Gesamtleistungswert:**

- **GZ-F2:**

= **doppelt fehlerkorrigierte Mengenleistung**, beruht auf Anwendung einer Korrekturformel, die bei dichotomen und Mehrfachwahl-Items einsetzbar ist. Soll verhindern, daß sich TP durch instruktionswidriges Verhalten ungerechtfertigte Vorteile verschaffen kann.

Berechnung: von Gesamtzahl der bearbeiteten Zeichen wird die doppelte Fehlerzahl abgezogen.

Entscheidender Unterschied zum bisherigen GZ-F:

dem **qualitativen Leistungsaspekt** wird wesentlich **stärkeres Gewicht** beigemessen

- **KL:**

= **Konzentrationsleistungswert** (Gesamtzahl der richtig durchgestrichenen Zeichen minus der Verwechslungsfehler F2), ist ebenfalls wie alle anderen Werte normalverteilt.

Vorteil: bessere Verfälschungsresistenz als GZ-F2 und differentialdiagnostische Leistungsfähigkeit

-> **sollte daher bevorzugt werden vor dem GZ-F2**

Fazit:

- 1) Alternative Parameter GZ-F2 und KL unterscheiden sich in ihrer Reliabilität nicht vom bisherigen GZ-F -> daher besteht **keine Notwendigkeit einen alternativen Meßwert einzuführen**
- 2) Besonderheit von GZ-F, die als **Nachteil** angesehen werden kann, ist, daß **einfache Fehlerkorrektur die Mengenleistung nur geringfügig mindert**
 ⇒ kann vor allem **in Extremfällen die Einschätzung der Konzentrationsleistung beeinträchtigen:**
 - **Tempo geht stärker in GZ-F ein als Güte / Sorgfalt.**
 - **GZ-F ist vor allem dann valide, wenn es primär auf Tempo bzw. um Diagnose einer Verlangsamung der visuellen Informationsverarbeitung ankommt.**
- 3) **Meßwert KL gewichtet Fehler mehr**, verhindert sowohl Leistungsüberschätzung als auch Leistungsunterschätzung in Extremfällen, verdeckt aber auch wie jeder andere Gesamtleistungswert unterschiedliches Arbeitsverhalten, wenn er allein und nicht **vor dem Hintergrunde von GZ und F% interpretiert** wird.
- 4) Empfehlung: KL soll zusätzlich zu den anderen Parametern untersucht werden.

f) Testwiederholung und Lernfähigkeitsdiagnostik:

„Dynamisches Testen“

- = **mehrmaliges Testen mit demselben Verfahren**, wobei unter Umständen eine spezielle Übungsphase eingeschoben werden kann.
- Ziel = **Steigerung der zuverlässigkeit** bestimmter diagnostischer Aussagen

Durch **Wiederholung des Test-d2 kann Reliabilität der Meßwerte erhöht** werden. Das ist normalerweise nicht notwendig für GZ, GZ-F, GZ-F2 und KL, denn die sind auch so schon reliabel genug. **Für die weniger reliabler qualitativen Werte aber (F, F% und SB)** kann ein Retest zur absicherung von Befunden angezeigt sein. Testwiederholung **wirkt sich auf die Reliabilität aus wie eine Testverlängerung** -> z.B. Realibilitätskoeffizient von 0,75 erhöht sich nach Spearman-Brown-Formel dadurch auf 0,86.

- Wird z.B. so gemacht bei den Schweizern -> Vorgabe des Test-d2 am Anfang und am Ende der verkehrspsychologischen Untersuchung (damit läßt sich überprüfen, inwieweit eine TP die Anfangsspannung bis zum Ende der Untersuchung abgebaut hat)

- Differentialdiagnose zwischen Hirnorganikern, Neurotikern und Gesunden läßt sich durch dynamisches Testen verbessern (z.B. Test-d2 als Prä- und Post-Test bei Rechenaufgaben eingesetzt -> **bei Gesunden höhere Leistungssteigerung als bei Hirnorganikern**)

Fazit:

Methode des **wiederholten Testens mit dem Test-d2** hat sich in bestimmten Fällen als **zweckmäßig** erwiesen, und zwar:

- 1) zur **Erhöhung der Reliabilität** weniger zuverlässiger Testwerte und damit zur Absicherung der Befunde.
- 2) zur **Verbesserung von Differentialdiagnosen zwischen Gesunden, Neurotikern und Gehirngeschädigten**, wenn diese nach einmaliger Testung noch keine zufriedenstellenden Ergebnisse erbracht haben
- 3) zur Erstellung von **Lernerfolgsprognosen**

NORMIERUNG:

Eichstichprobe / Kinder und Jugendliche:

- N = 3132 Schüler und Schülerinnen aus Volksschule (9-14), Oberschule (11-20), Berufsschule (14-19) aus 7 deutschen Bundesländern.
- Getrennte Normen für Mädchen und Jungen (wegen **signifikanter geschlechtsspezifischer Unterschiede** -> im Gegensatz zu den Erwachsenen!).

Eichstichprobe / Erwachsene:

- N = 3000, aus allen Berufen, Bildungsgraden und sozialen Schichten (zahlenmäßiges Überwiegen der mittleren und unteren Schichten); vorwiegend männliche TP; ABER: bei Erwachsenen konnten keine signifikanten geschlechtsspezifischen Unterschiede gefunden werden -> KEINE getrennte Normierung für Männer und Frauen