

Michael Berg

Experimentelles Denken in der Diagnostik kognitiver Funktionen

Als 1983 am Psychologischen Institut der Humboldt-Universität ein psychodiagnostisches Zentrum gegründet wurde, lag es nahe, der Tradition dieses Instituts folgend, experimentelles Denken in die Diagnostik kognitiver Leistungen einzubringen. Große Vorbilder wie Hermann Ebbinghaus, Wolfgang Köhler, Max Wertheimer und Kurt Lewin prägten seinerzeit die Berliner Schule, und auch die Zuordnung zur Naturwissenschaftlichen Fakultät war und ist eine begünstigende Bedingung für diesen Weg. Insbesondere war es Friedhart Klix, der einen richtungweisenden Einfluss auf Forschung und Entwicklung des Berliner Instituts hatte, mithin auch auf den Aufbau der Psychodiagnostik.

Die in den 1970er-Jahren aufkommende Kritik an der klassischen Testtheorie (Pawlik 1982) gehörte ebenfalls zu den Bedingungen, die experimentelles Denken auch in der Diagnostik förderte, wissend allerdings auch, dass das klassische Modell noch über lange Zeit fest etabliert sein wird: In den APA-Standards ist auch in der Ausgabe von 2014 keine Rede von experimentellen Validierungs-Methoden. Ein Blick in aktuelle Lehrbücher zeigt jedoch, dass experimentelles Herangehen an die Validierung diagnostischer Methoden durchaus schon als Möglichkeit gesehen wird (Rauch/Moosbrugger 2011; Schmidt-Atzert/Amelang 2012; Yousfi 2011).

Im Psychodiagnostischen Zentrum entstanden Arbeiten wie „Diagnostische Parameterfindung auf der Grundlage experimentell psychologischer Vorgehens“ (Schaarschmidt/Berg 1985) oder „Intelligence – its measurement on the basis of cognitive psychology“ (Berg/Schaarschmidt 1986). Einer der Bezugspunkte für Forschung und Entwicklung im Psychodiagnostischen Zentrum war das von Klix (1980) beschriebene Vereinfachungsprinzip:

„Die höhere Qualität einer Denkleistung stellt sich dar in der größeren Einfachheit und Effektivität des Lösungsgewinns.“ (Klix 1980, S. 264)

Einer der frühen Ansätze zur Entwicklung von Testverfahren (Berg/Schaarschmidt 1984) auf dieser Basis ging von dem so genannten Symbol-Distanz-Effekt aus (Potts 1972; Krause 1985). Nach dem Einlernen von paar-

weisen Relationen wie etwa: Paul ist größer als Fritz, Fritz ist größer als Emil, Emil ist größer als Klaus usw. und dem folgenden Abfragen aller möglichen Paarbeziehungen wie etwa: Ist Fritz größer als Klaus? – wurden nicht etwa die benachbarten Paarbeziehungen am schnellsten reproduziert, sondern die am weitesten entfernten. Für den entfernten Vergleich wurden also nicht alle jeweils benachbarten Relationen im Sinne einer Inferenz-Kette abgearbeitet, sondern die Elemente Paul usw. wurden – wahrscheinlich als visuelle Vorstellung – in eine Reihe gestellt, die man nur „anschauen“ muss, um die Fragen zu beantworten.

Für die Validierung von Test-Gegenständen zur Erfassung kognitiver Leistungen lassen sich klassische experimentelle Befunde heranziehen, die wir „Marker-Effekte“ nennen. Das sind Effekte, die wie eine Art Leit-Symptom typisch für einen Gegenstand der Erforschung kognitiver Leistung sind und an denen die Validierung entsprechender Leistungstests ansetzen kann. Der Symbol-Distanz-Effekt würde also die Fähigkeit kennzeichnen, räumliches Vorstellen zur Lösung von Inferenzaufgaben heranzuziehen. In weniger komplexen Anforderungssituationen sind Marker-Effekte im Allgemeinen noch klarer auf die jeweilige kognitive Leistung zu beziehen, um entsprechende Testverfahren daraus zu konstruieren. Das sind z.B. Interferenzeffekte als Marker für selektive Aufmerksamkeit oder Effekte mentaler Rotation für räumliches Vorstellen im engeren Sinne (spatial ability).

Das bedeutet zunächst auch:

Die Validierung eines Testverfahrens beginnt damit, es so zu konstruieren, dass das was erfasst werden soll, erfasst werden kann.

Für die Konstruktion von Testverfahren werden die Marker-Effekte dahingehend betrachtet, welche unabhängigen Variablen, den Effekt konstituierend bewirkt haben. Umgesetzt in Testanforderungen sind das jene Aufgabenvariablen, die die Schwierigkeit der Testaufgaben maßgeblich bestimmen.

Die Nutzung von Aufgabenvariablen als Ausgangspunkt für verschiedene Strategien der Testkonstruktion ist nicht ganz neu. Sie wurden und werden genutzt etwa von:

- Klauer (1987) zur Erzeugung eines „kontentvaliden“ Itempools im Rahmen kriteriumsorientierter Diagnostik,
- Hornke (1991), unter dem Begriff „rationale Itemkonstruktion“ zur Erzeugung von Items mit ‚bestimmbarer‘ Schwierigkeit,
- Embretson (1998), im Rahmen ihres „cognitive design systems approach“, um theoriegeleitet einen Itempool zu konstruieren. „... construct validity is given on the item level“ (S. 264).

Ganz allgemein können die IRT-(oder Rasch-)Modelle auch als verwandt zu den genannten Ansätzen gelten, weil darin die Aufgaben-Schwierigkeit auf derselben Skala liegt wie die angezielte Fähigkeit. Insbesondere expliziert Fischer (1974) in seinem LLTM-Modell (lineares logistisches Test-Modell) die Item-Schwierigkeit als gegeben durch kognitive Operationen.

In besonderer Weise steht der Ansatz von Borsboom, Mellenbergh und van Heerden (2004) für einen neuen Zusammenhang von Aufgabenvariablen („Attribute“) und Validität: „A test is valid for measuring an attribute if variation in the attribute causes variation in the test scores“ (S. 1067). Danach ist Validität keine quantitative, methodologisch bestimmte Größe, sondern eine qualitative, durch Hypothesenprüfung über die Wirkung von Aufgabenvariablen gegebene Eigenschaft von Testmethoden. „In conclusion, the present conception of validity is more powerful, simple, and effective than the consensus position in the validity literature.“ (S. 1070)

Berg (1991, 1993) beschrieb unter dem Begriff „Konstituentenansatz“ ein Vorgehen, bei dem Aufgabenvariablen zunächst den Zugang zu differentieller Validität von Testanforderungen bildeten, die mit unterschiedlichen Strategien (bzw. Leistungsvoraussetzungen) bearbeitet werden können. In der Folge entstehen unterschiedliche Schwierigkeitsprofile der Aufgabenvariablen. Später (Berg 1996) wurde der Ansatz auf interindividuell gleiche Schwierigkeitsprofile verallgemeinert.

Dieser Ansatz¹ stellt eine Möglichkeit der Konstruktion und Validierung psychologischer Testverfahren auf der Basis vorhandenen Wissens dar, mit dem Ziel, besonders ökonomische, erweiterbare Testsysteme mit unterschiedlichen Anforderungen zu konstruieren und zu validieren.

Ganz im Sinne von Borsboom et al. (2004) ist der Ansatz ein methodisch simpler und inhaltlich-substantiell bestimmter Zugang zur Validität, in dessen Zentrum auch, wie bei Hornke et al. (2000) und Embretson (2005) die Item-Schwierigkeit steht: Wenn auf der Grundlage kognitionspsychologischen Wissens erkannt ist, wodurch das Schwierigkeitsprofil einer Testanforderung entsteht, dann ist bekannt, was die Schwierigkeit einer Testaufgabe ausmacht und welche Art von Leistung die Aufgabe erfasst.

Der Konstituentenansatz

Den kognitionspsychologischen Ausgangspunkt der Testkonstruktion bilden, wie schon erwähnt, die aus dem jeweiligen Wissensgebiet bekannten, theo-

1 Der Ansatz ist somit kein Mess-Modell, kann aber eine inhaltliche Basis dafür hergeben.

retisch (nicht notwendigerweise „endgültig“) erklärten und empirisch nachgewiesenen Marker-Effekte. Auf diese bezogen werden Validitäts-Hypothesen formuliert, die die Wirkung der entsprechenden Aufgabenvariablen betreffen und mit dem Testmaterial ebenfalls empirisch prüfbar sind. Mit dem Konstituentenansatz wird Validität auch im Sinne von Borsboom et al. (2004) qualitativ definiert: Es wird gesagt, *was gemessen* wird und nicht (quantitativ) in welchem Maße das gemessen wurde, was gemessen werden sollte.

Beim Konstituentenansatz geschieht das durch Variation der (schwierigkeitsstiftenden) konstituierenden Aufgabenvariablen zur Verifikation von Hypothesen über deren Wirkung, abgeleitet aus den theoretisch erklärten Effekten der Validitäts-Marker. Alternativ-Hypothesen sind zu prüfen, wenn die Wirksamkeit anderer möglicher Testgegenstände oder von Strategien nicht ausgeschlossen werden kann (diskriminante bzw. differentielle Validierung).

Das Zueinander von systematischer Variation und bereichsweisem Konstant-Halten konstituierender Bedingungen wird durch Versuchspläne umgesetzt. Durch übergeordnete Variablen werden verschiedene Test-Anforderungen auseinander „verzweigt“. Das so entstandene „Testsystem“ erhält eine neue Validitäts-Komponente, die „Systemvalidität“ (Validität eines Tests oder Subsystems in Relation zu anderen). Validiert werden jene Versuchspläne, die Tests und Subtests in Bezug setzen und so die angezielte Leistung zu präzisieren bzw. zu untersetzen gestatten. So ist z.B. eine unzureichende Leistung in einem Gedächtnistest keineswegs von vornherein als Gedächtnisdefizit zu interpretieren. Die Testleistung kann z.B. durch ein Aufmerksamkeitsdefizit verursacht sein.

Der Begriff „Konstituentenansatz“ suggeriert eine ausschließlich auf die Wirkung und Variation der konstituierenden Bedingungen beschränkte Testkonstruktion. Um jedoch eine möglichst „reine“ Wirkung der konstituierenden Bedingungen zu ermöglichen, ist eine gründliche Kontrolle der modifizierenden Bedingungen zwingend erforderlich.

Das Vorgehen, wie kognitionspsychologisch begründete Annahmen mit dem Konstituentenansatz als Konstruktions- und Validierungsprinzip für die Entwicklung eines Testsystems (Berg 1996) genutzt werden können, soll am Beispiel des Testsystems Corporal demonstriert werden.

Das Konstruktionsziel war ein Test-System zur Erfassung kognitiver Funktionen im bildlich-räumlichen Bereich, deren verschiedene Anforderungen zum Zweck der Eingrenzung von diagnostischen Aussagen aufeinander beziehbar sein sollten.

Die Kontrolle der Randbedingungen als unverzichtbarer Bestandteil der Testkonstruktion

Damit die Schwierigkeitskonstituenten der Testaufgaben störungsfrei wirken können, müssen gleich am Beginn der Testkonstruktion modifizierende Variablen, die nicht grundlegend den Testgegenstand bedingen, durch die Anwendung experimenteller Techniken (Eliminierung, Konstant-Halten, Randomisierung oder Balancierung) kontrolliert werden.

Hier war der Einfluss der jeweils speziellen Testanforderung, der Wechsel von einer Anforderung zu einer anderen, möglichst störungsfrei zu halten. Die Reaktionen sollten per Tastendruck erfolgen, aber so, dass die Position der Tasten die Reaktion auf die Aufgabenmerkmale nicht stört. Die Idee dazu war, die Positionen der Testfiguren (übereinstimmend mit der Position der Tasten) zum aufgabenrelevanten Merkmal zu machen. Um komplexere Anforderungen zu konstruieren, die über einfache Wahlreaktionen hinausgehen, wurde ein zweites Tastatur-kompatibles Merkmal der Testfiguren definiert. Bei einer zweidimensionalen Anordnung der Tasten ergab sich als Antwortvektor für beide Merkmals-Dimensionen: oben, unten, links, rechts (gleiche Wortmarken für unterschiedliche Merkmals-Dimensionen). Dazu wurde eine Testfigur (ein Pfeil) konstruiert, die nicht nur eine jeweilige Position einnehmen, sondern auch in die entsprechende, mit gleicher Wortmarke benannte Richtung zeigen konnte. Um die Unabhängigkeit der Wirkung beider Merkmals-Dimensionen prüfbar zu machen, wurde eine zweite Testfigur konstruiert, die nur das Positions-Merkmal tragen sollte – ein Kreuz. Abbildung 1 zeigt die nach diesen Vorgaben konstruierten Testfiguren.

Um die Wahrnehmbarkeit beider Figuren konstant zu halten, wurde deren kognitive Komplexität berechnet. Abbildung 2 zeigt die Berechnung des Strukturellen Informationsgehalts (Leeuwenberg/Buffart 1984) der Testfiguren als Nachweis gleicher kognitiver Komplexität.

Als weitere Voraussetzung dafür, dass jene Aufgaben-Variablen (die Merkmale Lokation und Orientierung und deren Verknüpfungen), die die Schwierigkeit der Einzel-Aufgaben bestimmen sollten, möglichst störungsfrei wirken können, wurde die Abfolge der jeweils relevanten Merkmale ausbalanciert. Das heißt, in der Item-Abfolge sollten die vier Einzel-Merkmale (oben, unten, rechts und links) und deren paarweise Abfolgen gleich häufig vorkommen. Die konkrete Item-Abfolge wurde nach einem Überkreuzungs-(Berenblut) Design (Krauth 2000) konstruiert.

Damit war ein potentieller Itempool geschaffen, auf den verschiedene kognitive Anforderungen zugreifen konnten. Angezielt waren zunächst: Funktionen der Aufmerksamkeit, des Arbeitsgedächtnisses und der Raumorientierung (spatial ability) mit jeweils einem Subsystem.

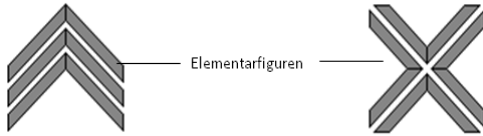


Abb. 1: Merkmalsträger des Testsystems Corporal Plus: Pfeil und Kreuz

Die Binnenstruktur der Grundfiguren

Die Binnenstruktur lässt sich nach Leeuwenberg & Buffart (1984) durch den so genannten "Strukturcode" darstellen. Dabei wird zunächst ein so genannter "Primärcode" über den Elementarfiguren (ein Parallelogramm bei der Grundfigur, ein Trapez bei der Kontrollfigur) gebildet, aus denen sich die Testfiguren zusammensetzen, der aus der Information tragenden Abfolge von Strecken und Winkeln besteht. Die gestrichelte Linie kennzeichnet das keine-Information-mehr-tragende Restelement.

Die Primärcodes für die linke und rechte Elementarfigur:
 $a\alpha b\beta a\alpha$ bzw. $a\alpha b\alpha\beta$

Der Strukturcode wird durch Operationen gebildet, die gleiche Strukturelemente "ausklammern":
 $2<\alpha\beta>/<ab>$ bzw. $<2\alpha\beta>/<ab>$

$<>/<>$ bedeutet, dass die eingeklammerten Elemente ineinander gesetzt werden, sich also abwechseln, die Zahl beschreibt die Anzahl erforderlicher Wiederholungen der folgenden Elemente oder Teilstrukturen.

Die Teilfiguren von Pfeil und Kreuz hätten dann den Strukturcode:
 SYM ($2<\alpha\beta>/<ab>$) bzw. SYM ($<2\alpha\beta>/<ab>$)

SYM bedeutet, dass eine zweite Teilstruktur symmetrisch zur ersten (Parallelogramm oder Trapez) gebildet wird. Dadurch entsteht aus beiden Elementarfiguren jeweils eine Art "Winkel". Als Strukturcode der jeweiligen Gesamtfigur wird der linke Winkel drei Mal nach oben oder unten versetzt, der rechte um $\gamma=90^\circ$ rotiert, bis eine Fortsetzung nicht mehr erforderlich ist:
 $3(\text{SYM } (2<\alpha\beta>/<ab>))$ bzw. $\gamma(\text{SYM } (<2\alpha\beta>/<ab>))$

Als struktureller Informationsgehalt (oder "I-Load") I gilt die Anzahl der verbleibenden Elemente oder Operationen. Für beide Strukturcodes wäre $I = 7$.

Abb. 2: Bestimmung der kognitiven Komplexität der Testfiguren

(nach Leeuwenberg/Buffart 1984)

Konstruktion von Testanforderungen als Subsysteme in einer Verzweigungsstruktur

Die Grundlage für Art und Anzahl der Tests innerhalb eines Subsystems bilden vorhandene Modelle und Taxonomien, jeweils erhalten anhand typischer experimenteller Paradigmen mit typischen empirisch unterscheidbaren Befunden, die wir dann als Marker-Effekte für die Validierung unserer Subsysteme und Tests verwenden.

Das Subsystem zur Erfassung von Funktionen der Aufmerksamkeit leitet sich im Wesentlichen aus den von Posner (1971, 1980) beschriebenen ab. Diese haben sich seit rund 40 Jahren zwar immer wieder in ihrer Bezeichnung geändert, kaum jedoch in ihrer angenommenen Wirkung. Funktionen des Arbeitsgedächtnisses knüpfen hauptsächlich an das Modell von Baddeley (2000) an, Test-Gegenstand ist der räumlich-visuelle Notizblock (sketch pad). Das Subsystem zur Erfassung der Orientierungsfähigkeit beruht auf Erkenntnissen aus dem so genannten MR- (mental rotation, Shepard/Metzler 1971) Paradigma, wonach die Schwierigkeit der Items vom Umfang der mentalen Bewegung abhängt, über die ein Vergleichs-Reiz mit einem Zielreiz zur Deckung gebracht wird. Abbildung 3 zeigt die Verzweigungsstruktur der aktuellen Subsysteme im Überblick.

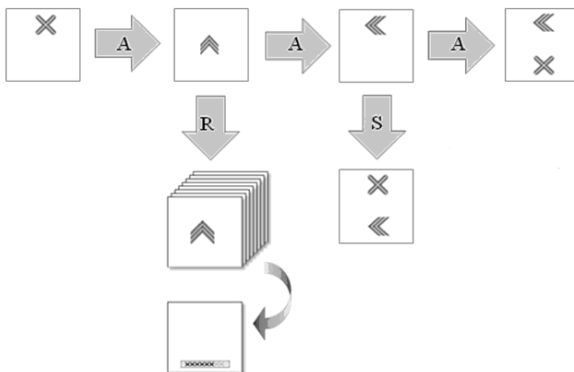


Abb. 3: Subsysteme des Testsystems Corporal

Subsystem A: Funktionen der Aufmerksamkeit (links Wahlreaktionen zur Erfassung diffuser und räumlich fokussierter Aufmerksamkeit, rechts zur Erfassung selektiver und verteilter Aufmerksamkeit)

Subsystem R: Funktionen des Arbeitsgedächtnisses (Recall, freies Erinnern von 8 Items),

Subsystem S: Orientierungsfähigkeit (Spatial Ability, Pfeilrichtung aus der Kreuz-Perspektive)

Experimentelle Validierung von Testanforderungen anhand von Marker-Effekten

Zur Demonstration der experimentellen Denkweise bei der Validierung von Tests innerhalb eines solchen Systems soll eine Test-Anforderung herausgegriffen werden, eine Variante verteilter Aufmerksamkeit², einer Doppelaufgabe (dual task).

Der Pfeil S1 und das Kreuz S2 erscheinen gleichzeitig, zu reagieren ist zuerst auf die Orientierung von S1 dann auf die Lokation von S2. Es gibt zwei Untervarianten: Dual A mit S1 in der Mitte, S2 am Rand und Dual C mit S1 am Rand, S2 gegenüber am Rand (siehe Abb. 4). Wenn auf S1 am Rand zu reagieren ist, dann ergibt sich eine Kombination von verteilter und selektiver Aufmerksamkeit, d.h. mit zusätzlich zu erwartendem Interferenz-Effekt. Als Ausgangs-Anforderung für die Validierungs-Effekte dient die jeweils vergleichbare einfache Wahlreaktions-Aufgabe: Reaktion nur auf die Orientierung des Pfeils, einmal in der Mitte, einmal am Rand.



Abb. 4: Varianten der Dual Task Anforderung

Links = Dual A ohne Interferenzeffekt

Rechts = Dual C mit Interferenzeffekt

Über das Zusammenwirken von Subfunktionen im Sinne der verteilten Aufmerksamkeit wird für Dual-Tests ganz allgemein angenommen, dass hier eine Subfunktion des Reagierens auf Reize wirksam wird, die über einfache Wahlreaktionen hinausgeht, nämlich der Reaktionsentwurf bzw. die Handlungsplanung (Neumann 1992). Der relevante Marker-Effekt für Doppelaufgaben ist als „concurrency costs“ bekannt. Bereits Noble et al. (1981) konnten zeigen, dass allein das Vorbereitet-Sein auf eine zweite Handlung die Reaktionszeit für die erste Handlung verlängert.

2 Zu verstehen als Unterschied zur sogenannten Aufmerksamkeits-Teilung, wo auf zwei simultan dargebotene Reize meist unterschiedlicher Modalität getrennt simultan zu reagieren ist.

Wenn durch die Aufgaben-Variable „zweite Reaktion“ unterschiedliche Schwierigkeitsgrade innerhalb dieses Aufgabentyps entstehen, dann sollte das entstandene Schwierigkeitsprofil als empirische Bestätigung der hypothesengeleitet vorhergesagten Schwierigkeitsprofile gelten und somit dafür, dass der Aufgabentyp „verteilte Aufmerksamkeit“ erfasst, mit und ohne Einschluss selektiver Aufmerksamkeit.

Validierungs-Hypothesen

Entsprechend dem Vorwissen wird für den Marker-Effekt „concurrency costs“ angenommen, dass die Reaktionen auf die Figuren S1 und S2 im Sinne des Reaktionsentwurfs als *eine* Handlungskette ausgeführt werden.

- H1: Die Reaktion auf S1 sollte in der Dual-Task-Situation schwieriger sein als eine vergleichbare Reaktion in einfachen Wahlreaktionen.
- H2: Die Zeit für die Reaktion auf S2 sollte wesentlich kürzer sein als für S1.

Ergebnisse zu den Hypothesen

Zu H1 zeigt Abbildung 5 die Reaktionszeit auf die Orientierung des Pfeils S1 in einfachen Wahlreaktion und der Dual-Task-Situation.

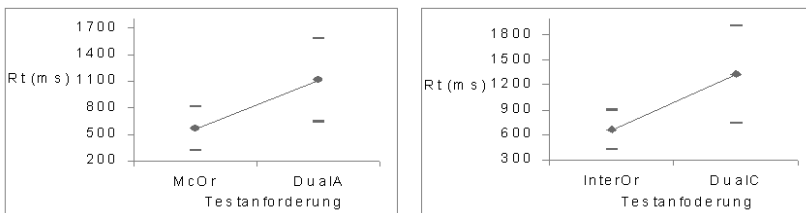


Abb. 5: Reaktion auf S1 in einfachen Wahlreaktionen und der Dual-Task-Situation

Mc: Multiple Choice Aufgabe, Inter: Mc-Aufgabe mit Interferenzeffekt
 Reaktionszeit (links, mittlere pro Item) $t = 17,82$; sig. (2-seitig) $< ,001$
 Reaktionszeit (rechts, mittlere pro Item) $t = 14,37$; sig. (2-seitig) $< ,001$

Zu H2 zeigt Abbildung 6 den Vergleich der Reaktionszeiten für die erste (S1) und die zweite (S2) Reaktionszeit auf S1 in der Dual-Task-Situation.

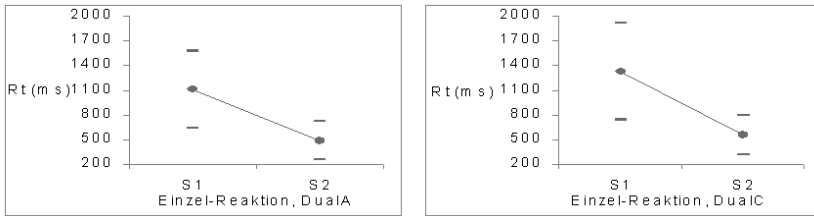


Abb. 6: Reaktion auf S1 im Vergleich zur Reaktion auf S2 in Dual A und Dual C

Reaktionszeit (links, mittlere pro Item) $t = 26,14$; sig. (2-seitig) $< ,001$

Reaktionszeit (rechts, mittlere pro Item) $t = 24,75$; sig. (2-seitig) $< ,001$

Es ist zu vermerken, wie mit H1 angenommen, dass die Reaktion auf S1 in Dual A und Dual C schwieriger ist als die vergleichbare Reaktion in einfachen Wahlreaktionsaufgaben (Abb. 5). Und es zeigt sich auch, dass die Reaktion auf S1 deutlich langsamer ist als auf S2 (Abb. 6).

Somit bestätigt sich die Annahme, dass die Reaktionen auf die Figuren S1 und S2 als eine Handlungskette ausgeführt werden, ganz im Sinne des Reaktionsentwurfs bzw. der Handlungsplanung (Neumann 1992), d.h. die zweite Reaktion wird zusammen mit der ersten in einen Handlungsplan integriert, der dann insgesamt ausgeführt wird. Abbildung 7 zeigt den Vergleich beider Anforderungen an verteilte Aufmerksamkeit mit und ohne zusätzlichen Interferenzeffekt. Der Unterschied von etwa 300 ms ist größer als der Interferenzeffekt bei einfacher Wahlreaktion, dort beträgt er etwa 100 ms. Es ist anzunehmen, dass es sich um eine zweifache Interferenzwirkung handelt: Sowohl die Lokation der Zielfigur als auch die der irrelevanten Figur (gegenüber) stören die Reaktion auf die Orientierung der Zielfi-

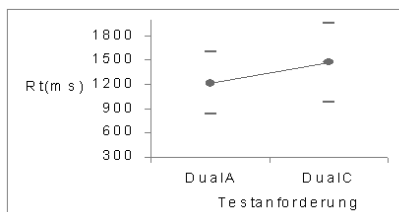


Abb. 7: Reaktion auf S1 ohne (links) und mit Interferenz (rechts)

Reaktionszeit (mittlere pro Item) $t = -4,29$; sig. (2-seitig) $< ,001$

gur, gerade weil auch die zweite Reaktion innerhalb einer Handlungskette anzulegen ist.

Als Validitäts-Aussage lässt sich insgesamt schließen, dass mit diesem Aufgabentyp eine Variante verteilter Aufmerksamkeit erfasst wird, die typische Effekte aus dem vorhandenen Vorwissen zu reproduzieren vermag.

Kontrolle der interpersonellen Variablen

Mit einem experimentell aufgebauten Testsystem, bedingt durch die Eingrenzbarkeit diagnostischer Aussagen, lässt sich nicht nur eine hohe inhaltliche Präzision erreichen, es weist konstruktionsbedingt auch eine bereichsweise hohe Homogenität und Reliabilität auf.

Um diese auch formal hohe diagnostische Präzision nicht am Ende zu verschenken, sollte auch die Normierung nicht durch Mengeneffekte, sondern nach kontrollierten Variablen erfolgen. Für die Erfassung elementarerer kognitiver Funktionen bieten sich dabei Variablen wie z.B. Alter, Geschlecht und Bildungsgrad an. Durch Nutzung schon vorhandener Informationen zur Verteilung dieser Variablen in der Gesamtbevölkerung oder spezifischen Teilen der Bevölkerung (z.B. Statistiken des Statistischen Bundesamtes oder der Bundesanstalt für Straßenwesen (BAS) lassen sich Quoten-Annäherungsmodelle als Grundlage für die Zusammensetzung der Normstichprobe bilden.

Vorteile für die Testanwendung

Dass solche am derzeitigen Wissen der Kognitionswissenschaften validierten Testverfahren Vorteile für den Anwender mit sich bringen, soll kurz am Beispiel Verkehrspsychologie demonstriert werden, wo das hier behandelte Testverfahren überwiegend eingesetzt wird.

Hier wird immer wieder die Ansicht vertreten (z.B. Vetter et al. 2015), und ihr widersprochen (Berg/Schubert 2015) „verkehrspsychologische Testverfahren“ hätten eine Vorhersage des Fahrverhaltens zu leisten. Man stelle sich vor, eine Testperson erreicht in einem Test mit abstrakten geometrischen Figuren einen Prozentrang von 16. Wie bitte lautet die Vorhersage ihres Fahrverhaltens? Und nun die Umkehrung: Ein alternder PKW-Fahrer steht an einem Vorfahrtsschild, von links nähert sich ein Fahrzeug, rechts ist alles frei und er fährt los. Hat er a) das linke Fahrzeug nicht gesehen?, b) es gesehen aber nicht wahrgenommen?, oder c) es gesehen und wahrgenommen, aber beim Kopfwenden vergessen, dass da noch etwas war? Zwar

kann (vgl. Fußnote 1) nicht im Sinne eines Messmodells (quantitativ) gesagt werden, wieviel Aufmerksamkeit in einer Gedächtnisleistung enthalten ist. Aber kognitionspsychologische Defizite wie in diesem Beispiel angedeutet, sind durch die konstruktionsbedingt mögliche Bezugnahme innerhalb experimentell validierter Testsysteme qualitativ eingrenzbar und können somit Hinweise für Interventionsmaßnahmen hergeben.

Zusammenfassende Betrachtung

Der Tradition des Berliner Instituts und dem Vermächtnis von Friedhart Klix geschuldet, kann experimentelles Denken auch in der Psychodiagnostik angewendet werden, um Testverfahren zur Erfassung kognitiver Funktionen zu konstruieren und zu validieren. Sowohl die Konstruktion von Testsystemen als auch deren Validierung ist anhand von Aufgabenvariablen möglich: innerhalb der Testanforderungen zur Reproduktion von Marker-Effekten als Hypothesenlieferanten für die Validierung, zwischen den Testanforderungen zur Abzweigung jeweils neuer, unterschiedlicher Anforderungen. Auf der Basis eines einheitlichen Testmaterials ergeben sich durch den Vergleich unterschiedlicher Testanforderungen Möglichkeiten der Untersetzung und Eingrenzung diagnostischer Aussagen (vgl. das soeben beschriebene Beispiel aus der Verkehrspsychologie).

Um die Aussagekraft der Validierungsbefunde zu optimieren, die Aufgabenvariablen als Schwierigkeits-Konstituenten mit Bezug auf den jeweiligen Testgegenstand störungsfrei zu variieren und deren Wirkung präzise messen zu können, ist ein ernst zu nehmender Aufwand erforderlich, um eine gründliche Kontrolle der modifizierenden Variablen zu erreichen: Wahrnehmbarkeit (hier: Komplexität) der Testfiguren, Kompatibilität zwischen Reizumgebung und Reaktionsmuster, Variablen der Item-Abfolge, interpersonelle Variablen.

Als Fazit sei festgestellt, dass es möglich und machbar ist, Gegenstände von Testverfahren zur Erfassung kognitiver Funktionen am vorhandenen Wissen aus den Kognitionswissenschaften zu validieren. Die experimentelle Methodik erlaubt es, Testverfahren so zu konstruieren, dass das was erfasst werden soll, erfasst werden kann.

Literatur

- Berg, M.; Schaarschmidt, U. (1984): Überlegungen zu neuen Wegen in der Intelligenzdiagnostik. In: *Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Reihe*, 8, S. 565–573
- Berg, M.; Schaarschmidt, U. (1986): Intelligence – its measurement on the basis of cognitive psychology. In: Klix, F.; Hagendorf, H. (eds.): *Human Memory and Cognitive Capabilities – Mechanisms and Performances*. North Holland: Elsevier Publishers B.V., pp. 499–512
- Berg, M. (1986): Experimentelle Diagnostik – Eine Herausforderung für Theorie und Praxis. In: Schaarschmidt, U.; Berg, M.; Hänsgen, K. (Hg.): *Diagnostik geistiger Leistungen*. Berlin: Psychodiagnostisches Zentrum, S. 24–36
- Baddeley, A. D. (2000): The episodic buffer: A new component of working memory? In: *Trends in Cognitive Sciences*, 4, pp. 417–423
- Berg, M. (1993): Der Konstituentenansatz – Ein Weg zu höherer Ergiebigkeit leistungsdagnostischer Methoden. In: Trost, G.; Ingenkamp, K. H.; Jäger, R. S. (Hg.): *Tests und Trends 10, Jahrbuch der pädagogischen Diagnostik*. Weinheim und Basel: Beltz, S. 40–82
- Berg, M. (1996): „Corporal“: Ein thematisches Testsystem zur Erfassung von Funktionen der Aufmerksamkeit. In: *Verhaltensmodifikation und Verhaltensmedizin*, Jg. 17/Heft 4, S. 295–310
- Berg, M.; Nädtke, J. (2015): Psychometrisches Testsystem Corporal Plus – Testsystem zur Erfassung kognitiver Funktionen im bildlich-räumlichen Bereich. Olching: Vistec
- Berg, M.; Schubert, W. (2015): Zum Begriff der theoriegeleiteten Validierung von Fahrübungen und Leistungstests – Eine Stellungnahme zu Vetter et al. (2015) ZVS 4. *Zeitschrift für Verkehrssicherheit*, 4, 33–37
- Borsboom, D.; Mellenbergh, G. J.; van Heerden, J. (2004): The Concept of Validity. In: *Psychological Review*, Vol. 111/No. 4, pp. 1061–1071
- Embretson, S. E. (1998): A cognitive design system approach to generating valid tests: Application to abstract reasoning. In: *Psychological methods*, 3, pp. 380–396
- Embretson, S. E. (1999): Generating Items during testing: Psychometric issues and models. In: *Psychometrika*, 64, pp. 407–433
- Embretson, S. E. (2005): Measuring Intelligence with Artificial Intelligence: Adaptive Itemgeneration. In: Sternberg, R. J.; Pretz, J. E.: *Cognition & Intelligence: Identifying the Mechanisms of the Mind*. Cambridge University Press: New York, pp. 251–267
- Fischer, G. H. (1974): *Einführung in die Theorie psychologischer Tests*. Bern: Huber
- Hornke, L. F. (1991): Neue Itemformen für computergestütztes Testen. In: Schuler, H.; Funke, U. (Hg.): *Eignungsdiagnostik in Forschung und Praxis. Beiträge zur Organisationspsychologie*, 10. Stuttgart: Verlag für Angewandte Psychologie, S. 67–70
- Hornke, L. F.; Küppers, A.; Etzel, S. (2000): Konstruktion und Evaluation eines adaptiven Matrizentests. In: *Diagnostica*, Vol. 46/No. 4. Göttingen: Hogrefe, S. 182–188
- Klauer, K. J. (1987): *Kriteriumsorientierte Tests*. Göttingen: Hogrefe
- Klix, F. (1980): *Erwachendes Denken*. Berlin: VEB Deutscher Verlag der Wissenschaften
- Krause, W. (1985): Komponentenanalyse des Symbol-Distanz-Effektes mit Hilfe von Augenbewegungsmessungen. In: *Zeitschrift für Psychologie*, 193, S. 259–272

- Krauth, J. (2000): *Experimental Design – A handbook and dictionary for medical and behavioral research*. Vol. 14 of Huston, J. P. (Series editor) *Techniques in the Behavioral and Neural Sciences*. Amsterdam: Elsevier
- Leeuwenberg, E.; Buffart, H. (1984): The perception of foreground and background as derived from Structural Information Theory. In: *Acta Psychologica*, 55, pp. 249–272
- Lienert, G. A. (1961): *Testaufbau und Testanalyse*. Weinheim: Beltz
- Neumann, O. (1992): Theorien der Aufmerksamkeit: von Metaphern zu Mechanismen. In: *Psychologische Rundschau*, 43, S. 83–101
- Noble, M. E.; Sanders, A. F.; Trumbo, D. A. (1981): Concurrence costs in double stimulation tasks. In: *Acta Psychologica*, 49, pp. 141–158
- Pawlik, K. (Hg.) (1982): *Diagnose der Diagnostik*. Stuttgart: Klett-Cotta
- Posner, M. I. (1980): Orienting of attention. In: *Quarterly Journal of Experimental Psychology, Section A*, 32, pp. 3–25
- Posner, M. I.; Boies, S. J. (1971): Components of attention. In: *Psychological Review*, 78, pp. 391–408
- Potts, G. R. (1972): Information processing strategies used in the encoding of linear ordering. In: *Journal of verbal learning and verbal behaviour*, 11, pp. 727–740
- Rauch, W. A.; Moosbrugger, H. (2011): Klassische Testtheorie. Grundlagen und Erweiterungen für heterogene Tests und Mehrfacettenmodelle. In: Hornke, L.; Amelang, M.; Kersting, M. (Hg.): *Methoden der psychologischen Diagnostik. Enzyklopädie der Psychologie, Themenbereich B, Serie II, Band 2*. Göttingen: Hogrefe, S. 1–86
- Schaarschmidt, U.; Berg, M. (1985): Diagnostische Parameterfindung auf der Grundlage experimentell psychologischen Vorgehens. In: *Zeitschrift für Psychologie*, 193, S. 117–131
- Schmidt-Atzert, L.; Amelang, M. (2012): *Psychologische Diagnostik* (5., vollständig überarbeitete Auflage, unter Mitarbeit von Thomas Fydrich und Helfried Moosbrugger, mitbegründet von Prof. Dr. Werner Zielinski). Berlin, Heidelberg: Springer
- Shepard, R. N.; Metzler, J. (1971): Mental rotation of three-dimensional objects. In: *Science*, 171, pp. 701–703
- Vetter, M.; Schünemann, L.; Debelak, R.; Gatscha, M.; Herle, M.; Mandler, G.; Ortner, T. M. (2015): Vorhersage von sicherheitsrelevantem Fahrverhalten bei Berufskraftfahrern: eine theoriegeleitete Validierung von Leistungs- und Persönlichkeitstests. In: *Zeitschrift für Verkehrssicherheit*, 4, S. 222–234
- Yousfi, S. (2011): Methoden der Item- und Skalenkonstruktion. In: Hornke, L.; Amelang, M.; Kersting, M. (Hg.): *Methoden der psychologischen Diagnostik. Enzyklopädie der Psychologie, Themenbereich B, Serie II, Band 2*. Göttingen: Hogrefe, S. 151–213