

ILICA. EIN SIMULATIONSTEST ZUR ERFASSUNG DES ENTSCHEIDUNGSVERHALTENS¹

Joachim Funke²

1 TESTART

Computergestützter Persönlichkeitstest in Einzelanwendung

2 TESTMATERIAL

Manual zum Test (34 Seiten A4); 3,5-Zoll-MS-DOS-Diskette mit 100 Anwendungen. Zur Installation wird ein handelsüblicher PC benötigt (Windows 3.1 und höher). Genauere Angaben über die benötigte Hardware (Art und Größe des Bildschirms, Grafikauflösung, benötigter Plattenplatz etc.) werden im Manual nicht gemacht. Im Verlags-Prospekt wird ein IBM-kompatibler PC ab AT-286 mit einer VGA-Grafikkarte (640x480), Microsoft-kompatibler Maus sowie 1,3 MB freiem Plattenspeicher gefordert.

3 TESTGLIEDERUNG

Der computerpräsentierte Test besteht aus einem Instruktionsteil (Bildschirmseiten: Personaldata, Instruktion 1, Instruktion 2, Instruktion 3) und dem

¹ Doris Möseneder & Johannes Ebenhöf (Programmautor: Johannes Ebenhöf) (1996). *ILICA. Ein Simulationstest zur Erfassung des Entscheidungsverhaltens*. Frankfurt: Swets Test Services.

² Für hilfreiche Anmerkungen zu einer Erstfassung danke ich Frau Dr. Marlene Endepohls, Handschuhshelm, sowie Dr. Ernst Fay, Bad Godesberg.

eigentlichen Testablauf, der sich an die Instruktion anschließt. Dieser Ablauf ist von den schrittweise getroffenen Entscheidungen der Testperson (Tp) abhängig, "die Anzahl der Geschweiß[!]verzweigungen ist außerordentlich hoch" (S. 12).¹ Die anfänglichen Instruktionen werden nachfolgend etwas genauer dargestellt, da sie den Rahmen für den Testablauf setzen.

Auf der ersten Instruktionsseite wird der Tp gesagt, es ginge darum, "einen Tag – gerafft – durchzuspielen und sich mit der Planung des Urlaubs zu beschäftigen" (S. 8). Dabei wird explizit auf die Vielfalt möglicher Vorgehensweisen hingewiesen: "Es gibt keinen einzigen, optimalen Weg zur Planung eines Urlaubs. Jeder muss die Aufgaben auf seine Art und Weise erledigen, um am Ende des Tages zufrieden zu sein." (S. 8). Für die Bearbeitung wird explizit keine Eile angeraten, da ausreichend Zeit zum Durcharbeiten zur Verfügung stehe. Der zu verplanende Zeitraum beträgt 630 Minuten (von 7:30 bis 18:00 Uhr), die aktuelle Uhrzeit dieses Tages kann durch Anklicken einer Uhr eingeblendet werden.

Auf der zweiten Instruktionsseite werden Rahmenbedingungen definiert: Vier Wochen Urlaub auf ILICA ("Es bleibt Ihnen überlassen, was Sie sich darunter vorstellen: Beachten Sie aber bitte, dass das Ziel NUR mit dem Flugzeug zu erreichen ist und einige IMPFUNGEN sowie einen neuen PASS erfordert!", S. 9) sind vorzubereiten. Sieben mögliche Urlaubsziele werden dann zur Auswahl gestellt (romantisches kleines Dorf am Meer; Ort von großer historischer Bedeutung; Feriendorf zum Austoben; einsame Bauernsiedlung im Gebirge; Blockhütte in menschenleerer Wildnis; Grossstadt in weit entferntem Land; Brieffreund, der sein Land zeigen und Kontakte herstellen will), von denen eines oder mehrere von der Tp ausgewählt und von ihr ggf. noch mit weiteren Angaben angereichert werden soll. Wofür diese Phase dient, bleibt unklar – im weiteren Verlauf wird darauf genauso wenig Bezug genommen wie in der späteren Auswertung.

Auf der dritten Instruktionsseite werden die beiden notwendigen Erledigungen (Impfung, Pass) sowie drei bereits im Kalender festgelegte Termine (12:00 Mittagessen bei Hans und Gabi; 15:00 Zahnarzt; im Laufe des Tages Mutter im Spital besuchen) erwähnt, und zugleich wird davor gewarnt, sich nicht zu übernehmen, man habe eine anstrengende Zeit hinter sich und noch zwei aufreibende Wochen vor sich.

Am Ende dieser dritten Seite kommt das erste Item, das auch in der Auswertung wichtig wird und in dem angegeben werden soll, ob man sich im Alltags-

¹ Alle Zitate ohne weitere Angabe sind dem Manual entnommen.

leben eher notiere, was im Laufe des Tages zu machen sei, oder ob man spontan entscheide.

Wählt man hier die Option "notieren", erhält man eine Liste von 40 Aktivitäten (im Text steht irreführend "40 Termine"), von denen maximal 15 durch Anklicken ausgewählt werden sollen, und erneut die etwas schwierige Aufforderung, das Hauptziel ("Verwirklichung des Urlaubes" [?], S. 11) nicht zu vernachlässigen, sich aber trotzdem auch Zeit für anderes zu nehmen. Die Randbedingungen sind hierbei: "Stellen Sie sich im folgenden vor, dass Sie sich in Ihrer gewohnten Umwelt befinden und handeln Sie entsprechend! Eine wichtige AUSNAHME gibt es jedoch: Stellen Sie sich vor, dass die Wegzeiten immer zwischen 10 und 20 Minuten betragen." (S. 11).

Hätte man sich auf die eben erwähnte Frage hin als Spontan-Entscheider (anstatt als "Notierer") bezeichnet, hätte man die gerade beschriebenen Informationen zu den 40 Aktivitäten nicht gesehen, sondern wäre direkt zu der für alle Tpn verbindlichen Abschlusss instruktion gekommen. Diese letzte Instruktionsseite beginnt mit dem Wecken um 7:30 Uhr ("Heute sind Sie besonders müde, nachdem Sie den gestrigen Abend ausgiebig mit Freunden gefeiert haben.", S. 12) und der Frage, ob man sofort aufstehen würde (obwohl man noch weiterschlafen möchte), ob man weiterschlafen würde oder ob man aufstehen würde, weil man ein Morgenmensch sei. Damit beginnt der eigentliche Test.

4 GRUNDKONZEPT

Der Test ILICA ist Teil einer neuen Testreihe "Hi-Psi" bei "Swets Test Services", die mit diesem Label den "hoch gesetzten Anspruch an künftige psychologische Tests" (S. 4) dokumentieren will. ILICA (so der Name einer Phantasieinsel) soll nach Angaben des Reihen-Herausgebers Klaus D. Kubinger "das Konstrukt der 'Selbstverwaltungs-fähigkeit' in generalisierbarer Weise erfassen" (S. 4). Gemeint sei damit "der persönliche 'Stil' einer Testperson, mit Ablenkungen während einer angestrebten Zielerreichung sowie mit aktualisierten Problemen und Frustrationen umzugehen".

Diese einführenden Konstruktbeschreibungen des Reihen-Herausgebers werden von den Testautoren nicht wesentlich vertieft: Das Grundkonzept ihres Tests stellen sie auf gerade einmal anderthalb Seiten vor. Die Tp erhält die Aufgabe, vor einem bevorstehenden vierwöchigen Urlaub die individuelle Gestaltung eines "freien Tages" vorzunehmen. Neben einer Reihe von "freien" Gestaltungsmöglichkeiten (in Form verschiedener Aktivitäten) sind auch einige

Vorgaben zu erledigen (z.B. die kranke Mutter im Spital besuchen) sowie überraschende Ereignisse zu bewältigen (z.B. die Bitte der Nachbarin um Beaufsichtigung ihres Kindes).

5 DURCHFÜHRUNG

Der Test kann nur als Einzelversuch durchgeführt werden. Die Durchführungsdauer wird im Verlagsprospekt mit "zwischen 40 und 50 Minuten" angegeben, im Manual (S. 17) wird eine Durchführungszeit von 30 Minuten (exklusive Instruktion) genannt und das Verfahren daher als ökonomisch eingestuft. In bezug auf die Testleiter-Zeit wird es sogar als "höchst ökonomisch" bewertet. Seitens der Tp werden keine Computererfahrungen vorausgesetzt.

6 AUSWERTUNG UND INTERPRETATION

Auswertung. Mit einem eigenständigen Programmteil (ILICA ERGEBNISSE; mit diesem Programm werden auch die noch zur Verfügung stehenden Testabnahmen aus dem Testkonto⁴ angezeigt) werden die Ergebnisse der Testbearbeitung in Form von fünf "Ergebnisfenstern" angezeigt. Neben (a) einer Angabe darüber, ob die gestellten Aufgaben vollständig erfüllt sind oder nicht, werden weiterhin Angaben (b) zur Strategie, (c) zur Orientierung mittels der Hilfsmittel (Kalender, Uhr, Telefon), (d) zum Anteil spezifischer Aktivitäten (Urlaub, Freizeit, Alltag, Soziales) an den Gesamtaktivitäten sowie (e) zum resultierenden Typ ("vermittelnd", "leicht abzulenkend" bzw. "zielorientiert"; auf S. 6 des Manuals wurden diese drei Typen "reflexiv", "impulsiv" und "flexibel" benannt) gemacht. Für (a) erfolgt eine dichotome Angabe, für (b) bis (e) erfolgt die Auswertung in Form von Prozenträngen. Absolute Werte werden nicht angegeben. Die auf der Festplatte gespeicherten Daten können mit diesem Programmteil auch ausgedruckt oder in eine ASCII-Datei für die Weiterverarbeitung in einem Statistikprogramm konvertiert werden.

⁴ Inwiefern die mit dem Programm ILICA DEMO mögliche Testdarbietung ohne Datenerhebung dem Testkonto (=Zahl der noch möglichen Anwendungen des Testverfahrens, für die bezahlt wurde) angerechnet wird, bleibt im Text unerwähnt. Wofür diese Funktionalität dem Anwender dienen soll, bleibt unklar. Ebenso unklar bleibt, ob eine bereits begonnene Anwendung, die jedoch abgebrochen wurde, dem Testkonto belastet wird.

Insgesamt 17 Testkennwerte werden bestimmt, von denen 2 die Antwort auf je 1 Frage wiedergeben, 5 Kennwerte geben die Anzahl bestimmter Aktivitäten an, 7 Kennwerte sind relative Anteile (relativer Anteil von Freizeit-, Urlaubs-, Alltags- und Sozialaktivitäten an den Gesamtaktivitäten sowie relativer Anteil der Antworten in Richtung flexibel, leicht abzulenkend bzw. zielorientiert) und 3 Kennwerte sind Benutzungshäufigkeiten (von Kalender, Uhr und Telefon). Drei der 17 Kennwerte sind nur bestimmbar, wenn die Tp eine "Tagesplanung" vorgenommen hat (siehe Abschnitt 3, dritte Instruktionseite). *Interpretation.* Obwohl es im Manual ein Kapitel "Interpretation" gibt (S. 18-19), werden dort lediglich Screenshots der Ausgabefenster abgedruckt und geringfügig kommentiert. Interpretationshilfen werden dort nicht gegeben, denn "die angeführten Prozentränge ermöglichen die unmittelbare Interpretation der Testwerte" (S. 19). Wie der Anwender allerdings mit einer Situation umgehen soll, wie sie das auf S. 18 abgedruckte Beispiel darstellt, bleibt unklar: dort ist die Typ-Einstufung "leicht abzulenkend" mit PR=94.8 und die für "zielorientiert" mit PR=78.4 angegeben - für die beiden eher konträren Ausprägungen also jeweils erhöhte Wahrscheinlichkeit!

7 GÜTEKRITERIEN

7.1 Objektivität

Durchführungs- und Auswertungsobjektivität sind gegeben, da sowohl das eine wie das andere durch den Rechner übernommen wird. Interpretationsobjektivität ist angesichts fehlender Interpretationshinweise nicht empirisch bestimmbar.

Aber Vorsicht: die objektive Durchführung garantiert keineswegs fehlerfreie Darbietung des Testmaterials! So wird etwa eine Terminabsage beim Zahnarzt vom Programm bestätigt, erscheint aber nach wie vor im Terminkalender. Wenn man um 12:40 Uhr das gemeinsame Essen mit Hans und Gabi absagen möchte, bestätigen diese den Termin um 12:00 Uhr. Wenn man um 16:30 Uhr Reiseprospekte einsammeln will, erhält man den Hinweis, daß man auf dem Weg zum Mittagessen sowieso am Reisebüro vorbeikomme. Eine ganze Reihe anderer unlogischer, weil automatisch erzeugter und damit nicht situationsangepasster Hinweise zeigt, dass nicht alle Tpn gleichermaßen sinnvolles Testmaterial präsentiert bekommen. Würde man Durchführungsobjektivität in diesem Fall um ein Sinnhaftigkeitskriterium erweitern, müssten klare Mängel konstatiert werden.

7.2 Reliabilität

Reliabilitätsangaben werden nicht gemacht. Mit dem Hinweis darauf, dass selbst für einfache Scores wie die Anzahl der erfüllten Aufgaben die Annahme lokaler stochastischer Unabhängigkeit verletzt und damit eine Rasch-Skalierung dieser Kennwerte nicht möglich sei, sowie mit dem Hinweis darauf, dass wenn schon die Annahmen der probabilistischen Testtheorie verletzt seien, dann die der klassischen Testtheorie "erst recht" scheiterten, wird dieser Punkt abgehandelt. Weder seien interne Konsistenzen bestimmbar noch Verfahren der Retest-Reliabilität angemessen, da im letzteren Fall bei wiederholt abverlangter Bearbeitung ein alternatives Vorgehen seitens der Tp geradezu provoziert werden würde.

7.3 Validität

Validitätsbelege für ILICA werden auf drei Ebenen vorgelegt: (a) divergente und (b) konvergente Validitätshinweise durch Vergleich mit anderen Skalen sowie (c) Hinweise aus den Interkorrelationen der ILICA-Skalen untereinander. Zwei Validierungsstichproben wurden rekrutiert, auf deren Daten nachfolgend Bezug genommen wird, und zwar Stichprobe 1: 224 Wiener Psychologiestudierende ohne weitere Angabe; Stichprobe 2.1: 162 Wiener Psychologiestudierende ohne weitere Angabe; Stichprobe 2.2: 131 Personen im Alter bis 55 Jahre, und zwar Bankangestellte, Versicherungsangestellte, Angestellte diverser Unternehmungen sowie Fliegeranwärter des österreichischen Bundesheeres.

Ad (a) divergente Validitätshinweise. In Stichprobe 1 fanden sich keinerlei signifikante Korrelationen von ILICA-Kennwerten zum Konstrukt der "Informationsverarbeitungsgeschwindigkeit", gemessen mit dem "Zahlen-Verbindungs-Test" (ZVT; Oswald & Roth, 1987), ebenso in Stichprobe 2 keinerlei Korrelationen zu "Reasoning", gemessen mit dem Computertest "Rechnen in Symbolen" (Schmatzer, Kubinger & Maryschka, 1994).

Ad (b) konvergente Validitätshinweise. Es werden aus Stichprobe 1 einige signifikante Korrelationen zwischen ILICA-Kennwerten und den "Frankfurter Selbstkonzeptskalen" (FSKN; Deusinger, 1986), dem "Fragebogen zur Erfassung von Kontrollüberzeugungen" (IPC; Krampen, 1981) sowie dem "Stressverarbeitungsfragebogen" (SVF; Janke, Erdmann & Kallus, 1985) berichtet. In Stichprobe 2 bestehen einige signifikante (aber niedrige) Korrelationen zwischen ILICA und dem "Myer-Briggs-Typenindikator" (MBTI; Bents & Blank,

1991), die Testautoren sehen die konvergente Validität hier als "niedrig bzw. nicht gegeben" an (S. 16).

Herzstück des knapp zweiseitigen Validitätskapitels ist die Darstellung von zwei Diskriminanzanalysen. Die erste diskriminiert erfolgreich anhand von 5 ILICA-Kennwerten zwischen 32 erfolgreichen und 60 weniger erfolgreichen Psychologie-Studierenden. Die zweite, als "Kreuzvalidierung" bezeichnet, nimmt mit 2 der im ersten Lauf erfolgreichen ILICA-Kennwerten (Anzahl geplanter und ausgeführter Aufgaben; zielorientierter Typ) eine Diskrimination von 247 Personen vor (ohne 46 mittel erfolgreiche Psychologiestudierende), die zu immerhin 54.3% die Einschätzung "erfolgreich/weniger erfolgreich" des jeweiligen Vorgesetzten trifft; nimmt man 3 weitere ILICA-Kennwerte hinzu, steigt die Trefferquote auf immerhin 58.0%.

Die Darstellung beider Diskriminanzanalysen erfolgt so sparsam, dass eine genauere Beurteilung des Vorgehens und der Ergebnisse nicht erfolgen kann.

Die aus der ersten Diskriminanzanalyse resultierende Charakteristik erfolgreicher Studierender lautet (S. 16): zielorientiertes Verhalten, häufiges Telefonieren, hohes Mass an Planerfüllung, geringe Bereitschaft zur Erfüllung von Vorgaben, seltene Zeitkontrolle. Aus der zweiten Analyse resultiert folgendes Bild erfolgreicher Personen: Erhöhte Anzahl geplanter Aktivitäten sowie durchgeführter Aktivitäten, viele Freizeitaktivitäten, häufiges Telefonieren.

Ad (c) Validitätshinweise aus Interkorrelationen der ILICA-Skalen untereinander. Hinsichtlich der Korrelationen von ILICA-Kennwerten untereinander werden ganze 5 genannt, besonders freuen sich die Testautoren über die in beiden Stichproben replizierte Korrelation von 0.57 bzw. 0.54 zwischen Kalender- und Uhr-Nutzung.⁵ Die restlichen Korrelationen werden als trivial bezeichnet und sind daher nicht aufgeführt.

7.4 Normen

Bereits im Vorwort des Herausgebers wird der Leser gewarnt: "Dem gesetzten hohen Standard der vorliegenden Computertest-Reihe zuwider läuft regelmäßig die Normierung. Notgedrungen sind Normierungen bei Computertests viel schwerer als bei herkömmlichen Gruppentests in befriedigender Weise zu

⁵ Natürlich könnte diese hohe Korrelation ein Artefakt darstellen, da beide Variablen – so die Normtabellen – linkssteile Verteilung aufweisen und daher bereits einige wenige Ausreißer starke Effekte erzielen könnten. Da keinerlei Verteilungsangaben gemacht sind, ist diese Vermutung weder zu belegen noch zu entkräften.

vollbringen, handelt es sich doch bei ihnen, realiter, um Einzeltestverfahren. Dem Anwender müssen daher (vorläufig) die gegebenen Normierungsmängel zugemutet werden" (S. 4). Dies ist zwar nachzuvollziehen, entschuldigt aber den Mangel nur wenig: Gerade für den Einsatz in individualdiagnostischem Kontext sind Normdaten von großer Bedeutung, wird doch an ihnen das erreichte Niveau der Tp festgemacht. Auf den fünf Textzeilen im Manual (S. 17) führen die Testautoren diese Klage und konstatieren, es mangle dem Test an der "Voraussetzung einer Normierung", weswegen nur vorläufige Normwerte auf der Basis der beiden Standardisierungs-Stichproben (N=517) sowie einer dritten, ebenfalls nicht-repräsentativen Stichprobe von 156 Psychologie-Studierenden mitgeteilt werden. Auf vier Seiten werden dann für insgesamt 15 Testkennwerte Prozentränge berichtet.

7.5 Fairness, Zumutbarkeit und Benutzerfreundlichkeit

Diese Kriterien sind Zusatzkriterien, die den von der APA (1986) bzw. vom Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen (1986) vorgeschlagenen Standards psychologischer Diagnostik beim EDV-Einsatz entsprechen.

Was die *Testfairness* betrifft, geht es gerade im Bereich computergestützter Diagnostik darum, Personen ohne einschlägige Vorerfahrung nicht von vornherein zu benachteiligen. Da keine Untersuchung im Manual erwähnt wird, in der Computernovizen mit hinsichtlich soziodemografischer Variablen parallelisierten Experten beim Bearbeiten von ILICA verglichen werden, scheinen die Testautoren sich diesbezüglich keine Sorgen zu machen.

Die Testautoren diskutieren Fairness auch unter dem Aspekt der Kennwerte. So schreiben sie auf S. 14 etwa, ein Testkennwert "Häufigkeit der Benutzung der Notizblock-Funktion" sei "ganz klar unfair", da er mit der Anzahl geplanter Aktivitäten wie auch mit der Gedächtniskapazität der Tp konfundiert sei. Generell müsse für jeden Kennwert die Möglichkeit bestehen, seine Fairness zu überprüfen. Gerade für interdependente Aktivitäten sei dies aber nicht möglich. Nach diesen Bemerkungen überrascht es den Rezensenten dann doch, wenn gleich im nächsten Satz drei Testkennwerte (Häufigkeiten der Benutzung von Kalender-, Uhr- und Telefon-Funktion) ohne jede weitere Kontrolle als fair mit Bezug auf das Poisson-Modell von Rasch deklariert werden. Über die Fairness wird in Abschnitt 8.1 unter dem Blickwinkel des Instruktionsverständnisses nochmals gesprochen.

Hinsichtlich des Kriteriums "Zumutbarkeit" wird ILICA von den Testautoren als "uneingeschränkt zumutbar" bezeichnet, da nach Kubinger (1995, S. 236) "das Ambiente des Computers die Bereitschaft einer Person zur psychologischen Testung" erhöhe.

Benutzerfreundlichkeit ist in zweierlei Hinsicht zu beurteilen: (a) aus Sicht des Testleitenden und (b) aus der Sicht der Tp. Was den Testleitenden betrifft, ist zunächst die Installation zu betrachten. Diese ist alles andere als trivial (vgl. die Ausführungen dazu in Abschnitt 8.5). Die spätere Testanwendung dagegen ist einfach zu starten durch Anklicken des entsprechenden Icons. Wie der Testleitende bei entleertem Testkonto zu verfahren hat, bleibt ihm überlassen.

Was die Bedienung des Programms durch die Tp betrifft, ist zunächst die ausschließliche Verwendung mausgesteuerter Bedienung zu konstatieren. Alternative Eingabeformen (Lichtgriffel, Touch-Screen) stehen nicht zur Verfügung.

8 KRITISCHE ANMERKUNGEN

8.1 Anmerkungen zur Instruktion

Eine nicht unwichtige Frage dürfte sein, wie die Tp die Aufgabenstellung auffasst. Wie bereits bei der (aus diesem Grund ausführlich vorgenommenen) Beschreibung der Testdurchführung hervorgeht, ist dies keinesfalls eindeutig: Wenn ich den Urlaub planen soll, warum sind dann bereits wichtige Entscheidungen über Ziel und Dauer getroffen? Und wie genau soll sich die Tp verhalten: Sie soll sich nicht übernehmen, sie soll das Hauptziel (welches ist denn nun genau das Hauptziel? Doch nicht die Verwirklichung des Urlaubs, wie es instruiert wird!) nicht vernachlässigen, sie soll so handeln, wie es ihrer Person entspricht, sie soll am Ende des verplanten Tages zufrieden sein – schade nur, dass ausgerechnet diese Zufriedenheit nicht erfragt wird.

Nach meinem Eindruck hinterlässt die Instruktion die Tp in einem verwirren Zustand. Ich würde mir wünschen, dass in weiteren Untersuchungen die von den Tpn selbst verfolgten Ziele explizit erhoben werden, um das möglicherweise sehr heterogene Instruktionsverständnis abzuklären. Sollten Tpn tatsächlich unterschiedliche Aufgaben darin erkennen, halte ich eine normative Bewertung für außerordentlich problematisch. Fairer wäre es, Klarheit über die zu erbringende Testleistung zu schaffen.

Ebenfalls ein Problem der Instruktion: die den einzelnen Aktivitäten zugeordneten Zeiten. Diese werden keinesfalls offen gelegt, sondern müssen durch

Aufrufen der Uhr nachgeprüft werden. "Hund füttern" dauert z.B. 15 Minuten, "Staubsaugen" gleich 25 Minuten. Die meisten Aktivitäten kann man beliebig oft wiederholen, was schon beim Füttern selbst hungriger Hunde sehr zweifelhaft sein dürfte! Die Leistung der Tp besteht im wesentlichen darin nachzuschauen, welche Zeiten die Testkonstrukteure für einzelne Aktivitäten vorgegeben haben. Eine wirklich freie Tagesplanung, in der ich z.B. jeweils nur 5 Minuten füttere bzw. staubsauge, ist nicht möglich. Der angekündigten Gestaltung eines freien Tages werden damit erhebliche Grenzen gesetzt.

8.2 Anmerkungen zu den Leistungsparametern

Ich räume gerne ein, dass die Ableitung von brauchbaren Leistungsparametern aus komplexen Problemstellungen eine ausserordentlich schwierige Angelegenheit ist. In der Geschichte der Forschung zum sog. "Komplexen Problemlösen" (Übersicht bei Frensch & Funke, 1995) hat sich dieses Problem von Anfang an als hartnäckig erwiesen. Bereits die frühen Arbeiten zur "Schneiderwerkstatt" (z.B. Putz-Osterloh, 1981) mussten aufgrund einfacher Überlegungen zum Gütemaß (z.B. Funke, 1983) ihre Aussagen einschränken. Aber selbst 10 Jahre nach der Erstpublikation zu diesem komplexen System konnten Süß, Kersting und Oberauer (1991) darlegen, dass ein differenziertes und brauchbares Gütemaß ("Gewinnspanne") bis zu diesem Zeitpunkt übersehen wurde. Umso schwieriger ist das Ansinnen der ILICA-Autoren, "typische Planungsstrategien einer Person in organisatorisch minder anspruchsvollen Aufgabenstellungen" (S. 5) zu erfassen. Gerade die Vielfalt möglicher planerischer Aktivitäten entzieht sich einer einfachen Betrachtung und erst recht einer einfachen Diagnostik (vgl. Fritz & Funke, 1995). Entsprechend widersprüchlich ist das Selbstverständnis der Testautoren: "Obwohl mit dem Begriff der 'Selbstverwaltungsfähigkeit' unmittelbar ein leistungsbezogenes Verhalten angesprochen wird, geht es doch primär um eine Persönlichkeitsvariable ..." (S. 5, Kursivierung von mir, JF) – ist es nun Leistung oder Persönlichkeit, die da erfasst werden soll? Dass im übrigen bereits die Antwort auf die eine Frage, ob man sich Tagesaktivitäten notiere oder eher spontan entscheide, als "Strategie-Kennwert" bezeichnet wird, zeigt in meinen Augen die Leichtfertigkeit der Testkonstrukteure beim Umgang mit diesem Problem.

Völlig unklar bleibt, auf welcher Grundlage die für die Typologie entscheidenden Zuordnungen der Antworten zu den drei Stilen vorgenommen wird. Damit ist der zentrale Standard 4.1 der von Häcker, Leutner und Amelang 1998 herausgegebenen deutschen Version der amerikanischen Teststandards verletzt.

Welche Bedeutung es hat, wenn eine Tp einen spezifischen Anteil von Freizeit-, Urlaubs-, Alltags- und Sozialaktivitäten an den Gesamtaktivitäten aufweist, bleibt unklar. Aber schon in der Unterscheidung dieser vier Aktivitätsbereiche beginnt die Unklarheit. Neben der Tatsache, dass aufgrund von ungleichen Ausgangsmengen (9 Freizeit-, 17 Urlaubs-, 8 Alltags- und 6 Sozialaktivitäten) die Wahlwahrscheinlichkeiten nicht konstant sind, bleibt die entsprechende Kategorisierung auf einer willkürlichen Ebene: Ist "Briefe schreiben" tatsächlich eine soziale, "mit Susi sprechen (war im Vorjahr in ILICA)" dagegen eine Urlaubsaktivität? Ich könnte es mir genau anders herum ebenfalls vorstellen. Die Unklarheit der diesbezüglichen Leistungsparameter offenbart eine Unklarheit der theoretischen Konzeption.

8.3 Anmerkungen zur Normierung

Da bereits der Reihenherausgeber auf S. 4 die gegebenen Normierungsmängel konstatiert, ist dem nichts hinzuzufügen. Da dort wie auch später bei den Testautoren das Wort "vorläufig" verwendet wird, könnte daraus abgeleitet werden, dass Normierungen im Gange sind und demnächst mitgeteilt werden. In einer neueren Veröffentlichung des Reihenherausgebers (Kubinger, 1998) finden sich zwar lobende Worte (sehr ähnlich zu den im Manual befindlichen) über ILICA, aber keine Hinweise auf neue Normierungen. Nach Angaben des Verlegers (Horn, 1999) sind allerdings rund 1000 weitere Datensätze von Wiener Studierenden erhoben worden, die demnächst ausgewertet werden sollen.

8.4 Anmerkungen zur Reliabilität und Validität

Reliabilität. Dass es für komplexe Szenarien schwieriger mit der Bestimmung der Reliabilität wird, ist nicht neu. Überrascht hat mich dennoch, mit welcher Leichtigkeit auf nur 6 Zeilen die Nicht-Existenz von Reliabilitätsangaben abgehandelt wird. Gerade hier hätte ich mir ausführlichere Überlegungen gewünscht, welche Alternativen die Autoren sehen, um die Mess-Zuverlässigkeit ihres Verfahrens wenigstens ansatzweise abzuschätzen. Warum wird beispielsweise nicht das Konzept eines Paralleltests diskutiert? Die Autoren behaupten ja, dass der ausgewählte Bereich "Urlaubsplanung" eine generalisierte Eigenschaft erfasse – warum also nicht einen Parallelbereich auswählen, in dem etwas anderes zu planen ist, aber strukturell gleiche Kennwerte abgeleitet werden könnten? Damit könnte zum einen die (durchaus zweifelhafte) An-

nahme einer bereichsübergreifenden Planungskompetenz geprüft als auch das Reliabilitätsproblem wenigstens ansatzweise in Angriff genommen werden.

Validität. Da der Test im wesentlichen auf das Persönlichkeitsmerkmal der "Impulsivität/Reflexivität" (I/R) abhebt, verwundert das Fehlen einschlägiger Aussenkriterien: Das geradezu klassische Verfahren zur I/R-Messung ("Matching Familiar Figures"; vgl. etwa Egeland & Weinberg, 1976) taucht ebenso wenig auf wie Hinweise auf Sternbergs (1994) Vorstellungen zu "thinking styles", in denen anstelle eines kognitionszentrierten gerade ein aktivitätszentriertes Vorgehen in den Fokus gerückt wird.

Die als besondere Stärke von ILICA angesehene Diskriminationsleistung zwischen erfolgreichen und weniger erfolgreichen Personen in der Größenordnung von 80% (Stichprobe 1, homogene Gruppe von N=92 Psychologie-Studierenden) bzw. 58% (Stichprobe 2, gemischte Gruppe von N=247 Berufstätigen) sollte nicht überbewertet werden: Zum einen überschätzen stichprobenbedingte Trefferraten generell die wahren Verhältnisse in der Population, zum anderen sind Hitraten ohne Einbezug der Priorwahrscheinlichkeiten kaum aussagekräftig (vgl. zu beiden Punkten Bortz, 1993, p. 573f.). Die wenigen im Manual dazu mitgeteilten Kennwerte erlauben keine adäquate Beurteilung, inwiefern hier tatsächlich überzufällige Klassifikationen erbracht wurden.

Methodisch gesehen ist das gewählte Vorgehen äußerst kritisch zu betrachten: Gerade die schrittweise Diskriminanzanalyse ist anfällig für Zufallseffekte! Generell sollten pro Variable mindestens 20 Probanden vorliegen (vgl. z.B. Stevens, 1992, p. 273f.), wenn nicht, sollte die Diskriminanzfunktion (für zwei Gruppen gibt es genau eine) durch eine unabhängige Replikation überprüft werden. Leider sind jedoch weder die Eigenwerte noch die Diskriminanzfunktion mitgeteilt.

Die aus den Diskriminanzanalysen resultierenden Charakteristiken erfolgreicher Personen ist im Falle der Studierenden nicht ohne gewisse Widersprüchlichkeit (gleichzeitiges Vorliegen von geringer Vorgabenerfüllung und hoher Planerfüllung), im Falle der Berufstätigen nicht ohne gewisse Trivialität (viele geplante und durchgeführte Aktivitäten, häufiges Telefonieren). Lässt sich daraus wirklich eine verlässliche Aussage über das Entscheidungsverhalten ableiten?

In der Bewertung dieser Validitätsbelege schließe ich mich der Einschätzung von Wahl (1997) an, wonach für die angezielten Anwendungen im Bereich des Personalwesens sowie der Berufs- und Bildungsberatung die vorliegenden Ergebnisse keineswegs ausreichen.

8.5 Abschließende Anmerkungen

Standard 3.9 der deutschen Version der amerikanischen Teststandards verlangt bei adaptiven Tests, dass die "zugrundeliegenden Überlegungen und empirischen Befunde für die Verfahren, die bei der Itemselektion, beim Abbruch und bei der Auswertung des Tests verwendet werden, im Handbuch beschrieben werden" (Häcker et al., 1998, S. 32). Genau dies unterbleibt im vorliegenden Fall (auch wenn ILICA nicht im eigentlichen Sinn des Wortes ein adaptiver Test ist). Der trockene Hinweis auf die außerordentlich hohe Anzahl an Verzweigungen verdeckt damit einen zentralen Punkt rationaler Testkonstruktion, nämlich die Offenlegung der Vorgehensweise. Zwar könnte diese – prinzipiell – durch einen geduldsigen Rezensenten, der sukzessive und systematisch die verschiedenen Varianten durchspielt, aufgedeckt werden; angesichts des Aufwands und auch angesichts einer begrenzten Anzahl von Anwendungen scheidet dieses Verfahren jedoch aus.

Die Qualität des sehr knapp gehaltenen Manuals ist durch zahlreiche Rechtschreib- und Setzfehler, aber auch durch sinnentstellende Fehler gemindert (z.B. wird auf S. 27 Aktivität Nr. 21 bezeichnet als "Geschirr vom Vortag verkaufen", obwohl sich das Item im Test auf das Spülen des Geschirrs bezieht). Ein extrem großzügiger Umbruch versucht wohl zu verdecken, wie sparsam hier Informationen bereitgestellt werden.

Angesichts der Schwierigkeiten, die der Rezensent bei der Installation auf einem älteren 486er PC erfuhr, muss eine Warnung ausgesprochen werden – Auszug aus den Fehlermeldungen:

- "Data access objects require VBDB300.DLL" oder
- "Can't load Custom Control DLL: , C:\ILICA\THREED.VBX".

Beide Fehler sind nicht im Handbuch erwähnt, in dem ganze 3 andere Fehler⁶ beschrieben sind, aber der doppelte Platz für eine Beschreibung von Sicherheitsmaßnahmen eingeräumt wird, mit denen unerlaubte Änderungen verhindert werden sollen. Also: Nichts für Anfänger!

⁶ Einer der Fehler tritt auf (vgl. S. 24), wenn im \SYSTEM-Verzeichnis von WINDOWS die Datei VER.DLL ein Datum größer/gleich 1.11.93 hat und die Bytegröße größer/gleich 9856 Byte beträgt – von einem komfortablen Installationsprogramm erwarte ich, dass dieser Fehler selbsttätig entdeckt und behoben wird und nicht vom Benutzer in einer komplizierten Schrittfolge identifiziert und beseitigt werden muss.

Abschließende und zugleich zusammenfassende Bewertung: Ein interessanter Ansatz, der allerdings derzeit nicht die hohen Ansprüche erfüllt, von denen die Rede ist. Eine theoretische Fundierung der an die Tp gerichteten Aufgaben (etwa durch Bezug auf ein Handlungsregulationsmodell) fehlt derzeit. Möglicherweise kann das Produkt nach mehreren "Reifezyklen" (sowohl seitens der Testautoren als auch des Verlags) eine bessere Bewertung erhalten – im derzeitigen Entwicklungsstadium lautet das Urteil "für eignungsdiagnostische Zwecke im Bereich des Personalwesens noch nicht brauchbar".

9 LITERATUR

- American Psychological Association (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: APA.
- Bents, R. & Blank, R. (1991). *Myers-Briggs-Typenindikator*. Weinheim: Beltz.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler (4., vollständig überarbeitete Auflage)*. Heidelberg: Springer.
- Deusinger, I.M. (1986). *Die Frankfurter Selbstkonzeptskalen (FSKN)*. Göttingen: Hogrefe.
- Egeland, B. & Weinberg, R.A. (1976). The Matching-Familiar-Figures-Test: A look at its psychometric credibility. *Child Development*, 47, 483-491.
- Frensch, P.A. & Funke, J. (Eds.). (1995). *Complex problem solving: The European perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fritz, A. & Funke, J. (1995). Übersicht über vorliegende Verfahren zur Planungsdiagnostik. In J. Funke & A. Fritz (Eds.), *Neue Konzepte und Instrumente zur Planungsdiagnostik* (pp. 47-78). Bonn: Deutscher Psychologen Verlag.
- Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? *Diagnostica*, 29, 283-302.
- Häcker, H., Leutner, D. & Amelang, M. (1998). Standards für pädagogisches und psychologisches Testen. *Diagnostica, Supplementum 1*, 1-116.
- Horn, R. (1999). Persönliche Mitteilung vom 4.1.1999.
- Janke, W., Erdmann, G. & Kallus, W. (1985). *Stressverarbeitungsfragebogen (SVF)*. Göttingen: Hogrefe.
- Krampen, G. (1981). IPC-Fragebogen zu Kontrollüberzeugungen ("Locus of control"). Deutsche Bearbeitung der IPC-Scales von Hanna Levenson. Handanweisung. Göttingen: Hogrefe.
- Kubinger, K.D. (1995). *Einführung in die Psychologische Diagnostik*. Weinheim: Psychologie Verlags Union.
- Kubinger, K.D. (1998, May 17). *Ein neuer Ansatz psychologischer Persönlichkeitsdiagnostik: Die Erfassung persönlicher Stil-Merkmale mit Hilfe "Objektiver Persönlichkeitstests"* [WWW-Dokument]. URL <http://www.human-factor.de/volltext/9804-01/op-tests.htm>
- Oswald, W.D. & Roth, E. (1987). *Der Zahlen-Verbindungs-Test (ZVT)*. Göttingen: Hogrefe.
- Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg. *Zeitschrift für Psychologie*, 189, 79-100.

- Schmotzer, C., Kubinger, K.D. & Maryschka, C. (1994). *Rechnen in Symbolen. Software und Manual*. Frankfurt am Main: Swets Test Services.
- Sternberg, R.J. (1994). Thinking styles: Theory and assessment at the interface between intelligence and personality. In R.J. Sternberg & P. Ruzgis (Eds.), *Intelligence and personality* (pp. 169-187). New York: Cambridge University Press.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Second edition. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Süß, H.-M., Kersting, M. & Oberauer, K. (1991). Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen. *Diagnostica*, 37, 334-352.
- Testkuratorium der Föderation Deutscher Psychologenverbände (1986). Richtlinien für den Einsatz elektronischer Datenverarbeitung in der psychologischen Diagnostik. *Psychologische Rundschau*, 37, 163-165.
- Wahl, S. (1997). *Testbeschreibung "ILICA" in der Datenbank PSYTKOM* (Dok. Nr. 3402). Trier: Zentralstelle für Psychologische Information und Dokumentation (abgedruckt 1998 in ABOaktuell, 5, Heft 4, 47-51).