

Was ist aus der Kritik der „Testrezensionen: 25 einschlägige Verfahren“ (*Zeitschrift für Differentielle und Diagnostische Psychologie*, 18, Heft 1-2) geworden?

Joachim F. Punter & Klaus D. Kubinger

Zusammenfassung:

5 Jahre nach dem Erscheinen eines Themenhefts der *Zeitschrift für Differentielle und Diagnostische Psychologie* mit 25 Testrezensionen zu einschlägigen psychologisch-diagnostischen Verfahren wird im vorliegenden Beitrag zusammengetragen, welche Konsequenzen verlagsseits daraus gezogen wurden. Als Rückblick wird jeweils das wörtlich zitierte Fazit aus den Rezensionen angeführt, das eine ziemlich eindeutige Stellungnahme pro oder contra abgibt. Ergebnis der Recherchen ist, daß es in der Zwischenzeit hinsichtlich eines Drittels der Verfahren zu Produktverbesserungen gekommen ist, sei es durch Testrevison und/oder durch die Entwicklung neuer Tests. Zwei Drittel der Verfahren erfüllen (weiterhin) die Qualitätsstandards nicht.

1. Einleitung

Das Themenheft „*Testrezensionen: 25 einschlägige Verfahren*“ (Kubinger, 1997) diene „primär dem Zweck, diagnostisch praktisch tätigen Psychologinnen und Psychologen eine evaluierende Übersicht der (vermutlich) häufigst eingesetzten Verfahren zu geben. Dabei geht es um die *aktuelle Sicht international beachteter Fachleute*, in einer ziemlich vereinheitlichenden Darstellungsweise“ (S. 1). Und bereits im Editorial konnte „resümierend gesagt werden: Zumindest mit den in der Praxis üblichen Ansprüchen sind die meisten der (hier behandelten) 25 *einschlägigen Verfahren* nicht tauglich; reduziert man aber diese Ansprüche auf das, was das jeweilige Verfahren tatsächlich – wenn überhaupt – zu leisten imstande ist, verlieren etliche davon entscheidend an Attraktivität für die Praktiker“ (S. 3).

Nun sind 5 Jahre seit dem Erscheinen dieser Testrezensionen vergangen, die grundsätzlich dazu verlagsseits genutzt hätten werden können, der statuierten „Mängelrüge“ durch Testrevison und/oder neue Tests zumindest punktuell zu begegnen. Intention des vorliegenden Beitrags ist es daher, für Praktiker(innen) den im Themenheft kritisierten Status „*up-zu-daten*“. Dies unter dem Motto: „So schlecht war ‚unser‘ Testinventar; wie gut ist es jetzt?“

2. Rückblick und *up-date*

Im folgenden wird das jeweils im Themenheft gegebene Fazit zitiert, um darauf – sofern gegeben – über seither erfolgte Produktveränderungen zu informieren¹.

Auf der linken Seite werden jeweils die bibliographischen Angaben der Verfahren aus dem angesprochenen Themenheft angeführt. Auf der rechten Seite werden jeweils die heute (Stand: Jänner 2002) erhältlichen Verfahren aufgelistet². Nachdrucke sowie unveränderte Auflagen werden nicht extra angeführt.

Rezensiert: Spreen, O. (1963). <i>MMPI Saarbrücken</i> . Bern: Huber.	Neu: Engel, R.R. (Hrsg.). (2000). <i>Minnesota Multiphasic Personality Inventory-2 (MMPI-2)</i> . Bern: Huber.
---	--

Fazit: „In Anbetracht der Qualität des MMPI verwundert das jüngste Umfrageergebnis, wonach ca. 20% der befragten Psychologen den MMPI zu den 5 am meisten von ihnen benutzten

¹ Die Reihenfolge der Tests im genannten Themenheft erfolgte nach dem Namen des Rezensenten alphabetisch; sie wurde hier so belassen. Bei allen Fazits wird im folgenden zwar der Rezensent namentlich angeführt, die Zitate entfallen aber im Literaturverzeichnis, weil die Quelle mit dem bewußten Themenheft eindeutig ist.

² Die vollständigen Literaturangaben zu den Verfahren werden im Literaturverzeichnis nicht wiederholt.

Tests zählen (Schorr, 1991)! So bleibt aus Gründen der Sachlichkeit nur zu hoffen, daß das MMPI als ‚psychometric nightmare‘ (Rodgers, 1975; S. 467) zunehmend weniger Anwendung findet“ (Angleitner, 1997, S. 9).

Mit dem MMPI-2 gibt es nunmehr eine neue Fassung und damit, im neuen Manual angeführt, folgende „wesentliche“ Änderungen (S. vii): Die „Formulierung der Itemtexte wurde überarbeitet, viele Items ... wurden gestrichen und dafür neue Items hinzugefügt, die im alten MMPI nicht enthaltene Inhaltsbereiche erfassen. Der alte ‚klassische‘ MMPI läßt sich aus den ersten 370 Items des MMPI-2 berechnen. ... Mit dem gesamten Itempool wurden neue Validitätsskalen, Inhaltsskalen und Zusatzskalen konstruiert Der MMPI-2 wurde mit hohem Aufwand an einer repräsentativen Stichprobe der Bundesrepublik Deutschland professionell standardisiert.“

Sieht man von der tatsächlich beispielhaften Neunormierung aus 1996 ab (über alle Bundesländer geschichtete Zufallsstichprobe mit 1244 untersuchten von 1466 kontaktierten Personen), ist damit alles im Anschluß an die ursprüngliche Kritik (Fazit) gesagt: Die testtheoretische Fundierung ist unverändert mangelhaft bis nicht gegeben, die Iteminhalte muten teilweise mehr denn je „unzumutbar“ (laut Testkuratorium, 1986, S. 359, „energetisch-motivational und emotional“ ungemäß beanspruchend) an: „Manchmal bin ich von bösen Geistern besessen“, „Ich lese nicht jeden Tag alle Leitartikel in der Zeitung“, „Ich liebe Gedichte“, „Meine Seele verläßt manchmal meinen Körper“, *Meine Tischmanieren sind zu Hause nicht ganz so gut wie auswärts in Gesellschaft*, „Ich gehe gerne ins Theater“, „Wenn ich ohne Bezahlung in ein Kino hineinkommen könnte und sicher wäre, nicht gesehen zu werden, würde ich es wahrscheinlich tun“, „Ich glaube an ein Leben nach dem Tode“, „Ich wäre gerne Soldat(in)“, „Manchmal habe ich Lust, eine Schlägerei mit jemanden anzufangen“, „Ich habe keine besondere Angst vor Schlangen“, „Ich mag nicht jeden leiden, den ich kenne“, „Ich gehe sehr gerne auf die Jagd“, „Manchmal klatsche ich ein wenig“, „Ich bin in der Liebe enttäuscht worden“, *Manchmal stimme ich bei Wahlen für Leute, von denen ich sehr wenig weiß*, „Ich glaube, ich bin ein verdammter Mensch“, „Alles schmeckt gleich“, „Ab und zu lache ich über einen unanständigen Witz“, „Ich habe nie in meinem Leben gern mit Puppen gespielt“, „Man hat mir gesagt, daß ich schlafwandle“, „Manchmal kommen mir sonderbare Gerüche“ – alles schon z.B. 1976 in einem „Testknacker“ belächelte Items (Paczensky, 1976). Davon abgesehen finden sich unter den neuen Items solche mit Formulierungen, wie sie jeder Übung zur Fragebogenkonstruktion als Negativbeispiel dienen würden; z.B.: „Man sagt mir, daß ich ein Alkoholproblem hätte, aber ich bin nicht der Meinung“.

Rezensiert:

Beckmann, D., Brähler, E. & Richter, H.E. (1991). *Der Gießen-Test (GT)* (4. überarbeitete Aufl. mit Neustandardisierung 1990). Bern: Huber.

Fazit: „Der GT ist ein an einer klinischen Stichprobe nach der klassischen Testtheorie konstruierter und 1990 an einer großen Stichprobe neu normierter Persönlichkeitsfragebogen mit differenzierten Normen. Er erfaßt auf sehr ökonomische und auf objektive Weise sechs psychosozial relevante bipolare Persönlichkeitseigenschaften, nicht jedoch klinische Symptome. Eine Besonderheit besteht im psychoanalytischen Hintergrund des Verfahrens. Als vorteilhaft erweist sich die Möglichkeit zur Diagnostik von *Selbst-* und *Fremdbildern*, die Einblicke in Zweierbeziehungen (z.B. Paar- oder Klient-Therapeut-Beziehungen) eröffnet. Der Fragebogen eignet sich bedingt zur Therapieevaluation (Erfassung von Veränderungen im Selbst- bzw. Fremdbild, nicht jedoch in Beschwerden und Symptomen). Die bei Fragebogen in der Regel gegebene leichte Durchschaubarkeit verbietet den Einsatz des GT in der Eignungsdiagnostik. Alles in allem ein *Breitbandverfahren*, das vor allem klinischen Psychologen bzw. Untersuchern mit psychoanalytischer Orientierung nützliche Informationen liefert, wohingegen zur Beschreibung der Persönlichkeit unausgelesener Personen andere Instrumente, mit homogeneren Skalen sowie einer theoretischen Verankerung in der empirischen Persönlichkeitsforschung, geeigneter erscheinen“ (Becker, 1997, S. 18).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert:

Melchers, P. & Preuß, U. (1994). *Kaufman-Assessment Battery for Children (K-ABC)* (2. korrigierte Aufl.). Lisse: Swets & Zeitlinger.

Fazit: „Die K-ABC kann als eine gut standardisierte Intelligenztestbatterie für Kinder von 2;6 bis 12;5 gelten; sie hat aktuelle Normen. Sie zielt darauf ab, die Art und Weise, wie ein Individuum Probleme löst, zu diagnostizieren. Dies gelingt nicht. Stattdessen wird auf klassische Art durch das Anbieten unterschiedlicher Aufgabentypen die Affinität zu unterschiedlichen Problemtypen erfaßt. Die praktische Handhabbarkeit kann als vorbildlich angesehen werden“ (Berg, 1997, S. 22).

Seither ist die Testbatterie K-ABC in 5. Auflage erschienen, allerdings ohne wesentliche Änderungen; insbesondere erfolgte keine Neunormierung (s. zum K-ABC auch nähere Ausführungen bei Schmidt & Leiss, 2002).

Rezensiert:

Kubinger, K.D. & Wurst, E. (1991). *Adaptives Intelligenz Diagnostikum (AID)* (3. ergänzte Aufl.). Weinheim: Beltz.

Neu:

Kubinger, K.D. & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum 2 (AID 2)*. Göttingen: Beltz.

Fazit: „Obwohl das AID – trotz der erweiterten Fähigkeitsmaße – an den Modellvorstellungen Wechslers ausgerichtet ist und somit neuere Entwicklungen der differentiellen Intelligenz- und Kognitionsforschung nicht diagnostisch zu nutzen vermag, hat sich der Test als außerordentlich hilfreiches Instrument zur diagnostischen Befunderhebung sowie zur Planung und Kontrolle psychologischer Interventionen bei Kindern bewährt. Nach den Ergebnissen einer informellen Befragung durch die Rezenten ist festzustellen, daß zurückhaltende Kritik eher von Berufspraktikern geäußert wurde, die mit dem AID weniger vertraut waren. Hingegen fielen die Bewertungen im Trend umso positiver aus, je mehr die Anwender über Erfahrungen und Routine sowohl mit dem AID als auch mit entsprechenden konventionellen Verfahren der Intelligenzdiagnostik verfügten“ (Conrad & Hatzipouleidis, 1997, S. 27).

Wie bereits in der im bewußten Themenheft gegebenen Replik angekündigt, ist zwischenzeitlich die „2. Generation“, der AID 2 erschienen, mit zahlreichen inhaltlichen Überarbeitungen (Aktualisierung von Iteminhalten; Ergänzung um Zusatztests, die besondere Teilleistungsfähigkeiten abprüfen sollen), einer Neunormierung sowie einem automatisch mitgelieferten Computer-Auswertungsprogramm (Genauerer s. bei Schmidt & Leiss, 2002).

Rezensiert:

Revers, W.J. (1979). *Der thematische Apperzeptionstest (TAT)*. Bern: Huber.

Fazit: „Der TAT ist ein – allerdings zweifelhaft ökonomisches – Verfahren zur Gewinnung von Hypothesen als Erklärung von Bedingungsbeziehungen; mehr (oder weniger) ist er nicht“ (Deegener, 1997, S. 30).

Schon in seiner Rezension verwies Deegener (1997, S. 30) darauf, daß man den TAT besser „in seiner ursprünglichen Form aufgrund der ‚veralteten‘ Bilder z.B. durch den Thematischen Gestaltungstest – (Salzburg) (TGT-(S); Revers & Allesch, 1985)“ ersetzen sollte. In der Praxis findet sich dieses Verfahren aber praktisch überhaupt nicht.

Rezensiert:

Krampen, G. (1981). *IPC-Fragebogen zu Kontrollüberzeugungen (IPC)*. Göttingen: Hogrefe.

Neu:

Krampen, G. (1991). *Fragebogen zu Kompetenz- und Kontrollüberzeugungen (FKK)*. Göttingen: Hogrefe.

Fazit: „Die Attraktivität des Konzepts, verbunden mit der hohen Ökonomie des IPC-Fragebogens sollte den möglichen Anwender nicht über die zweifelhafte Validität hinwegtäuschen. In jedem Fall hinzuweisen ist darauf, daß seit 1991 ein als FKK ‚überarbeiteter‘, neuer Fragebogen zum fraglichen Konzept vom selben Autor vorliegt“ (Deusinger, 1997, S. 34).

Bereits dieses Fazit verweist also auf die revidierte Form, den *Fragebogen zu Kompetenz- und Kontrollüberzeugungen* (FKK) bzw. plädiert auch schon Krampen (1997) in seiner Replik dafür, den IPC nicht mehr anzuwenden (ausgenommen für ältere Längsschnittstudien). Der FKK erfaßt gegenüber seinem Vorgänger zusätzlich Kompetenzerwartungen. Und zwar beziehen sich diese auf die generalisierte Erwartung einer Person, in Lebenssituationen (kraft eigener Fähigkeit) auch über eine Handlungsalternative zu verfügen. Das Verfahren beinhaltet demnach die Skalen *Selbstkonzept eigener Fähigkeiten*, *Internalität*, *Soziale Externalität* und *Fatalistische Externalität*. Diese theoretisch hergeleiteten Skalen konnten durch Faktorenanalysen bestätigt werden. Allerdings läßt die Meßgenauigkeit des Verfahrens zu wünschen übrig: die Reliabilitäten liegen zwischen .64 und .71, was jedoch bei jeweils nur acht Items pro Skala nicht überrascht. Die Normierung erfolgte 1989 an einer repräsentativen Stichprobe westdeutscher Erwachsener.

Rezensiert: Horn, W. (1969). <i>Prüfsystem für Schul- und Bildungsberatung (PSB)</i> . Göttingen: Hogrefe.	
--	--

Fazit: „Es stimmt nachdenklich, daß das PSB auf der Schorr’schen ‚Hit-Liste‘ (Schorr, 1995) immerhin den 15. Rang einnimmt. Aus allen Kritikpunkten, insbesondere was die Überalterung der Normierung betrifft, folgt, daß von einer Anwendung des Verfahrens strikt abzuraten ist“ (Fay, 1997, S. 38).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Tewes, U. (1991). <i>Hamburg-Wechsler Intelligenztest für Erwachsene – Revision 1991 (HAWIE-R)</i> . Bern: Huber.	
---	--

Fazit: „Der HAWIE-R stellt gegenüber dem HAWIE eine bereits längst fällige Verbesserung eines praktisch bewährten Tests dar. Allerdings gestattet er lediglich eine globale Abschätzung des Intelligenzstatus‘ durch Vorgabe möglichst unterschiedlicher Aufgaben, die nach der Theorie von Wechsler eine Vorhersage intelligenten Alltagsverhaltens zulassen soll. Der überzeugende Beweis für diese Behauptung ist aber noch immer nicht erbracht. Davon abgesehen sollten Profilinterpretationen nur bei sehr massiven Untertestdifferenzen gegeben werden. Trotz der aufgezeigten Mängel bleibt der HAWIE-R für die klinische Einzelfalluntersuchung Erwachsener und wegen seiner internationalen Verbreitung für die klinisch-psychologische Forschung gegenwärtig konkurrenzlos“ (Guthke & Herzberg, 1997, S. 41f.).

Das Handbuch der 2., korrigierten Auflage gibt keine näheren Hinweise auf die Art der Korrekturen; daher ist dieses Fazit aus 1997 infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Stumpf, H., Angleitner, A., Wieck, T., Jackson, D.N. & Beloch-Till, H. (1985). <i>Deutsche Personality Research Form (PRF)</i> . Göttingen: Hogrefe.	
--	--

Fazit: „Die Ergebnisse zur konvergenten und diskriminanten Validierung sind auf den ersten Blick eine Bestätigung dafür, daß die PRF innerhalb des Dimensionsraumes von bereits bekannten Persönlichkeitsfragebogen liegt. Auf den zweiten Blick taucht das Problem auf, ob es

eine Varianzaufklärung über das bisher erreichte Maß gibt. Dem Anwender ist zu raten, bei der Interpretation der Dimensionen vorsichtig zu sein und nicht zu meinen, mit der Anwendung der PRF seien nun alle Probleme der psychometrischen Persönlichkeitsdiagnose gelöst“ (Häcker, 1997, S. 47).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Oswald W.D. & Roth, E. (1987). <i>Der Zahlen-Verbindungs-Test (ZVT)</i> (2. überarbeitete und erweiterte Aufl.). Göttingen: Hogrefe.	
--	--

Fazit: „Der ZVT kann als brauchbares Verfahren zur Abschätzung eines besonderen Aspekts der intellektuellen Leistungsfähigkeit im Sinne eines *Screenings* empfohlen werden. Er ist relativ fair gegenüber kulturellen Einflüssen, theoretisch begründet und bezüglich aller Gütekriterien ausreichend. Die Normen bedürfen einer Überprüfung, obgleich der Verdacht einer Normveränderung weniger begründet ist als etwa bei wissens- oder problemlösebezogenen Leistungstests“ (Hänsgen, 1997, S. 52).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Weiß, R.H. (1987). <i>Grundintelligenztest Skala 2 (CFT 20) mit Wortschatztest (WS) und Zahlenfolgentest (ZF)</i> (3. verbesserte und erweiterte Aufl.). Göttingen: Hogrefe.	Neu: Weiß, R.H. (1998). <i>Grundintelligenztest Skala 2 (CFT 20) mit Wortschatztest (WS) und Zahlenfolgentest (ZF)</i> (4. überarbeitete Aufl.). Göttingen: Hogrefe.
--	--

Fazit: „Der CFT 20 samt seiner Ergänzungstests besitzt Attraktivität hinsichtlich der Möglichkeit, *fluid* und *crystallized* Intelligenz einer Tp gegenüberzustellen; erstere kann allerdings, u. U. schneller, auch mit anderen Tests erfaßt werden, letztere genauer nur durch andere. Bleibt der Einsatz für Fragestellungen, die Einflüsse der Sprache weitgehend ausschalten wollen – dafür ist aber die Normierung schon bedenklich alt“ (Heller, 1997, S. 55).

Die Überarbeitung zur 4. Auflage bezieht sich nur auf das Manual. Das Testmaterial und der Antwortbogen sowie die Handanweisung zu den beiden Ergänzungstests bleibt unverändert. Das Manual wurde um einige Abschnitte (z.B. Diagnose von Hochbegabungen) ergänzt. Die Stabilität der Normen des CFT 20 aus dem Jahr 1977 wurde anhand einer Stichprobe von Hamburger Schülern ($N = 12300$) aus dem Jahre 1996 überprüft. Aufgrund der nur geringen Unterschiede der durchschnittlichen Rohwerte zu beiden Erhebungszeitpunkten zieht der Autor für die Kurzform (1. Testteil des CFT 20) den Schluß, daß es keine Notwendigkeit für eine Neunormierung gebe.

Rezensiert: Jäger, A.O. & Althoff, K. (1994). <i>Der WILDE-Intelligenz-Test (WIT)</i> (2. revidierte Aufl.). Göttingen: Hogrefe.	
--	--

Fazit: „Ohne daß den geäußerten Kritikpunkten mit entsprechenden Studien, insbesondere mit einer Neunormierung begegnet wird, ist die Anwendung des WIT nicht gerechtfertigt“ (Jäger, 1997, S. 63).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Benton Sivan, A. & Spreen, O. (1996). <i>Der Benton-Test</i> (7. vollständig überarbeitete Aufl.). Bern: Huber.	
---	--

Fazit: „Der Benton-Test nach Instruktion A ist als ein Teil, als Untertest einer neuropsychologischen Testbatterie *neben* anderen Verfahren, wie dem *Nonverbalen Lerntest* von Sturm und Willmes (1994), Tests aus der *Wechsler Memory Scale* oder der *Rey-Osterrieth Figur*, einsetzbar. Die zahlreichen Varianten sind weniger empfehlenswert“ (Kessler & Kaschel, 1997, S. 68).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Tewes, U. (1985). <i>Hamburg-Wechsler Intelligenztest für Kinder – Revision 1983 (HAWIK-R)</i> (3. korrigierte Aufl.). Bern: Huber.	Neu: Tewes, U., Rossmann, P. & Schallberger, U. (Hrsg.). (2000). <i>HAWIK-III. Hamburg-Wechsler-Intelligenztest für Kinder – Dritte Auflage</i> . Bern: Huber.
---	--

Fazit: „Beim HAWIK-R handelt es sich um ein traditionsreiches, informatives, aber zeitaufwendiges Intelligenzdiagnostikum, dessen Validität insbesondere für die Profilinterpretation nur mäßig belegt ist. Die Normierung stammt aus 1983 und sollte aktualisiert werden“ (Moosbrugger, 1997, S. 70).

Der HAWIK-III ist aus der Übersetzung und Adaptierung der amerikanischen WISC-III (Wechsler Intelligence Scale for Children – Third Edition) hervorgegangen und stellt keine direkte Weiterentwicklung des deutschen HAWIK-R dar (Genaueres dazu s. bei Schmidt & Leiss, 2002).

Rezensiert: Laux, L., Glanzmann, P., Schaffner, P. & Spielberger C.D. (1981). <i>Das State-Trait-Angstinventar (STAI)</i> . Weinheim: Beltz.	
--	--

Fazit: „Angesichts der hohen Relevanz des Konstrukts Angst in der psychologischen Theoriebildung, in den verschiedenen Psychotherapie-Richtungen und im medizinischen Kontext erscheint der Angstfragebogen STAI unter den obigen Einschränkungen grundsätzlich weiter interessant für Forscher wie Anwender“ (Muthny, 1997, S. 73).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Janke, W., Erdmann, G. & Kallus, W. (1985). <i>Stressverarbeitungsfragebogen (SVF)</i> . Göttingen: Hogrefe.	Neu: Janke, W., Erdmann, G. & Kallus, W. (1997). <i>Stressverarbeitungsfragebogen (SVF) mit SVF 120</i> . Göttingen: Hogrefe.
--	---

Fazit: „Die Verwendbarkeit des SVF für die individuelle Diagnostik, beispielsweise im klinischen Bereich, ist wegen der eingeschränkten Vorhersagemöglichkeiten des Bewältigungsverhaltens in konkreten Situationen gering zu bewerten“ (Neumer & Margraf, 1997, S. 76).

Das Manual des SVF wird durch eine Kurzbeschreibung des SVF 120 ergänzt. Gegenüber dem SVF ist der SVF 120 um die Skala „Entspannung“ (zusätzliche 6 Items) erweitert, darüber hinaus wurden drei Items ersetzt. Neu eingeführt werden die Testkennwerte *Positiv-Strategien* und *Negativ-Strategien*. Natürlich lassen sich auch mit dem SVF 120 weiterhin nur Coping-Strategien erfassen; das Verfahren soll eine Aussage ermöglichen, zu welchen Strategien eine Person in Belastungssituationen habituell neigt. Nicht erfaßt werden können also die Ressourcen zur Stressbewältigung und die Effizienz der verwendeten Strategien. Die Normierungsstichprobe aus dem Jahr 1995 setzt sich aus lediglich 512 Personen im Alter von 20-79 Jahren zusammen. Die nach Alter, Geschlecht und Schulbildung geschichtete Stichprobe wurde vor allem in Berlin und Würzburg erhoben. Im Manual des SVF 120 finden sich nur dürftige Angaben zur Validität, das Literaturverzeichnis fehlt zur Gänze. Der Anwender wird u.a. zur Beschreibung des theoretischen Hintergrunds auf das Handbuch des SVF zurückgreifen müssen.

Rezensiert: Fahrenberg, J., Hampel, R. & Selg, H. (1994). <i>Das Freiburger Persönlichkeitsinventar – Revidierte Fassung (FPI-R)</i> (6. ergänzte Aufl.). Göttingen: Hogrefe.	Neu: Fahrenberg, J., Hampel, R. & Selg, H. (2001). <i>Das Freiburger Persönlichkeitsinventar (FPI-R)</i> (7. überarbeitete und neu normierte Aufl.). Göttingen: Hogrefe.
---	--

Fazit: „Beim FPI-R handelt es sich um ein sorgfältig konstruiertes und normiertes Persönlichkeitsinventar; allerdings erfolgte diese Normierung bereits vor 15 Jahren, und zwar vor der Wiedervereinigung Deutschlands. Die Reliabilität der Skalen ist befriedigend und für *gruppenstatistische* Zwecke ausreichend. Zur Validität, insbesondere zur Konstrukt- und Kriteriumsvalidität der fünf gegenüber dem FPI neuen FPI-R-Skalen liegen auch in der aktuellen, 6. Auflage der Handanweisung noch zu wenige Angaben vor. Das Inventar dürfte für Psychologen interessant sein, die sich dem ‚FPI-Menschenbild‘ anschließen können und die speziell den Merkmalsbereich Emotionalität bzw. Neurotizismus detailliert zu erfassen suchen. Im Vergleich zum gemeinsamen Merkmalsraum mehrerer anderer Persönlichkeitsinventare besitzt hingegen auch das erweiterte FPI-R nur einen Meßbereich mittlerer Bandbreite. Von einer Verwendung des FPI(-A1) ist aufgrund der dort nach wie vor enthaltenen, von den Autoren ja selbst kritisierten Skala *M* sowie der z. T. unzureichenden Güte einiger Skalen abzuraten“ (Ostendorf, 1997, S. 81).

Die 7. Auflage des FPI-R besticht durch die 1999 in den alten und neuen deutschen Bundesländern durchgeführte Repräsentativerhebung zur Normierung – der Fragebogen mit seinen 138 Fragen ist unverändert geblieben. Die Normierungsstichprobe umfaßt 3740 Personen (4097 Personen wurden kontaktiert) im Altersbereich von 16 bis 70 Jahren und älter. Das Manual umfaßt nun auch neuere Beiträge zur Validierung. Gewonnen allerdings anhand von nur kleinen Stichproben, werden mäßige Zusammenhänge der FPI-R-Skalen mit jenen des SVF und des Gießen-Tests berichtet. Leider fehlen die genauen Angaben über Korrelationen zum NEO-FFI (*NEO-Fünf-Faktoren Inventar*; Borkenau & Ostendorf, 1993), die den Bezug zur Theorie der „Big-Five-Persönlichkeitstheorie“ herstellen könnten, obwohl sich im Manual diesbezüglich ein Verweis auf das Manual des NEO-FFI findet. An der Normierungsstichprobe wurden zahlreiche Ergebnisse u.a. zur Faktorenstruktur gewonnen, wobei die Autoren eine 10-Faktorenlösung präferieren, die im wesentlichen den Ergebnissen von 1982 (anlässlich der Revision des Verfahrens) entspricht.

Rezensiert: Brickenkamp, R. (1994). <i>Test d2. Aufmerksamkeits-Belastungs-Test</i> (8. erweiterte und neugestaltete Aufl.). Göttingen: Hogrefe.	Neu: Brickenkamp, R. (2002). <i>Test d2. Aufmerksamkeits-Belastungs-Test</i> (9. überarbeitete und neu normierte Aufl.). Göttingen: Hogrefe.
--	--

Fazit: „Beim Test d2 handelt es sich um ein bewährtes und wissenschaftlich gesichertes Verfahren zur Erfassung *kurzfristiger* Aufmerksamkeit und Konzentration im Zusammenhang unterschiedlichster Fragestellungen. Der Test wird seitens der Tpn gut angenommen, ist ökonomisch durchzuführen und dank des Protokollblattes mit integrierten Normwerten gut zu interpretieren – allerdings sind die Normen überaltert“ (Oswald & Hagen, 1997, S. 88).

Die 9. Auflage zeichnet sich in erster Linie durch aktuelle Normen aus. Außerdem sind im Handbuch erweiterte und aktualisierte Angaben zu den Gütekriterien Reliabilität und Validität zu finden. Auf dem Testbogen kann nun die Händigkeit und die Benutzung von Sehhilfen eingetragen werden; diese Angaben mögen bei der Interpretation der Ergebnisse von Bedeutung sein. Damit hat sich aber grundsätzlich nichts daran geändert, daß der Test aufgrund der Kleinheit seines Materials Personen mit Sehschwächen besser nicht vorgegeben wird; auch werden weiterhin Personen mit gewissen Teilleistungsschwächen bei der Testbearbeitung Probleme haben. Die Daten ($N = 3251$) für die Normen wurden in den Jahren 1999/2000 in den Bundesländern Berlin, Brandenburg, Sachsen und Bayern erhoben. Die Normen umfassen den Altersbereich von 9 bis 60 Jahren. Erstmals sind auch Normen zu den Testkennwerten *Gesamtzahl der Fehler* und *Konzentrationsleistungswert* tabelliert.

Rezensiert: Brem-Gräser, L. (1995). <i>Familie in Tieren. Die Familiensituation im Spiegel der Kinderzeichnung</i> (7. neugestaltete Aufl.). München: Ernst Reinhardt.	
--	--

Fazit: „Berücksichtigt man die üblichen Vorgaben der Psychologischen Diagnostik, so erscheint der Einsatz der Familie in Tieren als *Test* nicht verantwortbar und als Explorationshilfe zu spekulativ und nicht interpretierbar“ (Petermann, 1997, S. 92).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Rorschach , H. (1992). <i>Psychodiagnostik</i> (11. unveränderte Aufl.). Bern: Huber.	
--	--

Fazit: „Trotz aller kritischer Einwände, die sich gegen das Rorschach-Form-Deute-Verfahren mit dem Anspruch, ein psychologischer Test zu sein, vorbringen lassen, stellt es für routinierte und mit kritischer Umsicht interpretierende Testleiterinnen und Testleiter ein hilfreiches, ergänzendes Instrument zur *systematischen Verhaltensbeobachtung* dar“ (Rauchfleisch, 1997, S. 96).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Irle, M. & Allehoff, W. (1984). <i>Berufs-Interessen-Test II (B-I-T II)</i> . Göttingen: Hogrefe.	
---	--

Fazit: „Der BIT II ist an inhaltsanalytischen wie auch testtheoretischen Kriterien gemessen dringend revisionsbedürftig. Das betrifft sowohl die Neukonstruktion von (für die aktuelle Berufsstruktur repräsentativen) Items wie auch die Mitteilung von Reliabilitäts- und Validitätsbefunden. Ohne eine entsprechende Überarbeitung entspricht der BIT II nicht mehr heutigen wissenschaftlichen Forderungen an ein psychologisch-diagnostisches Verfahren“ (Richter, 1997, S. 101).

Dieses Fazit aus 1997 ist infolge keinerlei Produktveränderung aktuell und relevant.

Rezensiert: Staabs, G. von (1995). <i>Der Scenotest</i> (8. unveränderte Aufl.). Bern: Huber.	Neu: Ermert, C. (1997). <i>Scenotest-Handbuch</i> . Bern: Huber. Neu: Zusatzmaterial zum Scenotest (7 Teile)
---	---

Fazit: „Je schwerwiegender die Diagnose in ihren Konsequenzen für die Tp ist, umso weniger kann man den Einsatz des Verfahrens als einzige diagnostische Informationsquelle empfehlen. Als Ergänzung und zur Gewinnung von Hypothesen, die durch besser abgesicherte diagnostische Verfahren überprüft werden müssen, hat der Sceno-, Test‘ jedoch durchaus seinen Platz – so wie er auch als therapeutisches Hilfsmittel gute Dienste leisten kann“ (Rollett, 1997, S. 104).

Das ergänzend zum erschienenen Scenotest-Textband (v. Staabs, 1995) neu erschienene Scenotest-Handbuch (Ermert, 1997) bietet einen Überblick über die historische Entwicklung des Scenotests und stellt eine große Anzahl von Untersuchungen zum Verfahren vor. Wesentliches Ziel des Handbuchs ist es, in die Durchführung und in die verschiedenen Auswertungsmodi einzuführen. Ein Kapitel widmet sich den Testgütekriterien und den Funktionen des Scenotests; die Autorin stellt diesbezüglich zusammenfassend fest: „Es liegen kaum explizite Untersuchungen zur Objektivität, Reliabilität und Validität des Scenotests vor. ... Die Validität des Scenotestverfahrens ist noch nicht hinreichend gesichert“ (S. 145). Gleichzeitig mit dem Sce-

notest-Handbuch ist Zusatzmaterial zum Scenotest erschienen; dieses besteht aus folgenden Teilen: Einhorn, Elefant, Karfunkelstein, Fernsehgerät, Löwe, Prinz und Schrubber.

Rezensiert: Amthauer, R. (1973). <i>Intelligenz-Struktur-Test (I-S-T 70)</i> (4. unveränderte Aufl.). Göttingen: Hogrefe.	Neu: Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). <i>Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)</i> . Göttingen: Hogrefe. Neu: Institut für Test- und Begabungsforschung & Gittler, G. (2001). <i>Intelligenz Struktur Analyse (ISA)</i> (2. erweiterte Aufl.). Frankfurt: Swets.
---	---

Fazit: „Bezogen auf die Gesamtestleistung weist der IST 70 hinreichende Reliabilität und Validität auf. Tieferegehende Analysen seiner Gütekriterien, insbesondere auf Untertest- und Itemebene, lassen allerdings zahlreiche Probleme erkennen. Gravierend sind sie in bezug auf eine differenzierte Strukturbetrachtung (Profilaussage), die ja einen wesentlichen Vorteil des Verfahrens ausmachen sollte. Die dafür erforderlichen Voraussetzungen sind in mehrfacher Hinsicht nicht erfüllt, die gegebenen Interpretationsempfehlungen bewegen sich auf spekulativem Niveau. Schließlich ist entscheidend, daß trotz der außerordentlich umfangreichen Normierungsdaten die Gültigkeit der Normen keinesfalls mehr vorausgesetzt werden kann; sie müssen als stark überaltert gelten. So bleibt aus heutiger Sicht zu resümieren, daß – bei allem erforderlichen Respekt gegenüber diesem ‚ehrwürdigen‘ Test – eine zunehmende Diskrepanz zwischen der noch immer großen Popularität des Verfahrens und seiner real nachweisbaren diagnostischen Leistungsfähigkeit konstatiert werden muß“ (Schaarschmidt, 1997, S. 108).

Die bereits in der Vorgängerversion I-S-T 2000 angebotene Möglichkeit zur Diagnose kristalliner und fluider Intelligenz (*sensu Cattell*) stellt im I-S-T 2000 R eine für den Anwender interessante Erweiterung gegenüber dem I-S-T 70 dar – der Nachweis der entsprechenden Struktur gelingt im I-S-T 2000 R eher überzeugend. Der „Sprachlastigkeit“ des I-S-T 70 wurde insofern begegnet, als das Übergewicht verbaler Untertests (5 von 9 Aufgabengruppen im I-S-T 70) reduziert wurde. Einige Untertests wurden neu konstruiert und Items überarbeitet. Kritisch ist zu bewerten, daß die Eindimensionalität (im Sinne des *Rasch*-Modells) für keinen Untertest untersucht oder gar bestätigt wurde, obwohl zumindest für die *Würfelaufgaben* des I-S-T 70 seit langem bekannt ist, daß die Aufgaben verschiedene Fähigkeiten gleichzeitig erfassen (vgl. Gittler, 1984). Mit dem I-S-T 2000 R sollen fünf der sieben Primärfaktoren von *Thurstone* erfasst werden: Verbale, numerische und figurale Intelligenz sowie Merkfähigkeit und Schlußfolgerndes Denken (*reasoning*). Der letzte „Faktor“ wird als Gesamtwert, nämlich als Summenwert aus verbaler, numerischer und figuraler Intelligenz bestimmt. Die Testkennwerte der übrigen Faktoren bestimmen sich aus jeweils drei bzw. zwei Untertests – mit Nachdruck wird von den Testautoren darauf hingewiesen, daß die Interpretation einzelner Untertests nicht möglich ist. Der Gesamtwert (im Protokollbogen: IQ!) erscheint befremdlich, zumal die dreifaktorielle Struktur von verbaler, numerischer und figuraler Intelligenz in der Faktorenanalyse bestätigt wurde. Wie immer ist es für einen solchen Gesamtwert zweifelhaft, daß verschiedene Fähigkeitsbereiche in der Lebensbewältigung eine kompensatorische Wirkung haben! Die neuen, in den Bundesländern Berlin, Hamburg, Brandenburg, Sachsen, Bayern und Baden-Württemberg erhobenen Normen (ohne Angabe des Erhebungszeitraums) basieren gegenüber jenen im I-S-T 2000 auf einer wesentlich größeren Stichprobe ($N = 3484$). Inwieweit die Testbatterie als solche und insbesondere ihre Normierung Gültigkeit für alle deutschsprachigen Länder besitzt, wird im Manual weder angesprochen noch garantiert.

Die ISA baut ebenfalls auf *Thurstone* auf. Vier Faktoren, die sich in Faktorenanalysen annähernd bestätigt finden, sollen erfasst werden: Verbale, numerische und figural-räumliche Intelligenz sowie Gedächtnis. Von insgesamt neun Untertests erfassen vier den verbalen Fähigkeitsbereich (Faktor); dies ist deswegen problematisch, weil die sprachlichen Fähigkeiten dadurch übermäßig in den Gesamtwert (d.i. die Summe der Rohwerte in allen Untertests) einfließen.

Auf diesen Gesamtttestwert trifft sinngemäß dieselbe Kritik wie auf den Gesamtwert des I-S-T 2000 R zu. Wenigstens die Aufgaben des Untertests *Würfel erkennen* wurden aus einem im Sinne des *Rasch*-Modells homogenen Itempool entnommen. Die beiden erst in der 2. Auflage enthaltenen Spezialformen ISA-L(eicht) und ISA-S(chwer) erfüllen das Gütekriterium Zumutbarkeit natürlich besser. Die Normen beziehen sich auf (teils anfallende) Stichproben in der Größe von $N = 3813$, die zwischen 1997 und 1999 erhoben wurden; nicht erwähnt ist allerdings die regionale Verteilung der Stichproben. Einige der Altersnormen beziehen sich aber auf eher kleine Stichproben (z.B. $N = 94$ für die Gruppe der 25- bis 29jährigen). Die Computer-Version der ISA ist im *Wiener Testsystem* (Dr. G. Schuhfried GmbH) integriert. Die Äquivalenz zwischen Computer- und Papier-Bleistift-Version ist empirisch nachgewiesen.

Rezensiert: Schneewind, K.A., Schröder, G. & Cattell, R.B. (1994). <i>Der 16-Persönlichkeits-Faktoren-Test (16 PF)</i> (3. korrigierte Aufl.). Bern: Huber.	Neu: Schneewind, K.A. & Graf, J. (1998). <i>Der 16-Persönlichkeits-Faktoren-Test, Revidierte Fassung (16 PF-R)</i> . Bern: Huber.
---	---

Fazit: „Für die 1994 in dritter Auflage erschienene deutschsprachige Form des 16 PF fehlen nach wie vor umfangreiche Validierungsuntersuchungen, so daß der praktische Wert des Verfahrens für die zahlreichen im Manual genannten Anwendungsbereiche nicht beurteilt werden kann. Einer Anwendung in der Praxis stehen auch fehlende Untersuchungen zum Einfluß von Antworttendenzen und zur Testfairneß entgegen sowie die nun schon lange zurückliegende Erhebung der Normierungsstichprobe“ (Schwenkmezger, 1997, S. 115).

Im zwischenzeitlich erschienenen 16 PF-R heißt es einleitend: „Gegenüber der ersten deutschen Fassung ist die vorliegende Revision durch folgende Neuerungen gekennzeichnet: zeitgemäße und geschlechtssensitive Itemformulierungen, offene Gestaltung der Antwortalternative ‚b‘ durch Fragezeichen, separate Erfassung des logischen Schlußfolgerns am Ende des Tests, Einbeziehen von Skalen zur Überprüfung von Antworttendenzen, vereinfachte Handhabbarkeit der Primärdimensionen durch leicht verständliche Bezeichnungen (z.B. ‚Wärme‘), Normierung an einer aktuellen repräsentativen Eichstichprobe (1998), Validierung anhand bekannter psychometrischer Persönlichkeitsinventare sowie kriteriumsorientierte Analysen.“ (S. 7). Dazu ist mehreres zu sagen. Die Normierung umfaßt $N = 1209$ Personen aus Deutschland (alte und neue Bundesländer), ohne genauere Angabe, wie diese Personen rekrutiert wurden, insbesondere wieviele angesprochene Personen die Testung verweigerten. Der Beleg, daß gerade 16 Faktoren methodisch gerechtfertigt sind, erfolgt nicht überzeugender als er schon nicht in der ersten deutschsprachigen Version erfolgte: Der gewiefte Anwender kann aus den angegebenen Korrelationen die nunmehr „verheimlichten“ Eigenwerte berechnen und stellt fest, daß nur 4 (!), nicht einmal wie noch früher 5 Eigenwerte größer als 1 sind, was eben nur 4 Faktoren, nicht 16 zu extrahieren inhaltlich sinnvoll macht! Traditionell kommt man dann ohnehin zu 5 (noch immer korrelierenden (!)) sog. „Sekundärfaktoren“. Daß diese der heute – wissenschaftlicherseits – präferierten „Big-Five-Persönlichkeitstheorie“ (vgl. Amelang & Bartussek, 1997) entsprechen, zeigt sich bei der „Validierung anhand bekannter psychometrischer Persönlichkeitsinventare“ gerade nicht: Die höchste Korrelation zwischen 16 PF-R und NEO-FFI (*NEO-Fünf-Faktoren Inventar*; Borkenau & Ostendorf, 1993) beträgt .67, 16 der 25 Korrelationen sind $< .20$! Bemerkenswert ist, daß die Beziehungen mit NEO-FFI, FPI-R, PRF und TPF (*Trierer Persönlichkeitsfragebogen*; Becker, 1989) als Beleg der Konstruktvalidität des 16 PF-R interpretiert wird, was als Konsequenz die Nützlichkeit des 16 PF-R stark in Frage stellt! Als „kriteriumsorientierte Analysen“ sind wieder nur Korrelationen mit Fragebogenverfahren, speziellen Persönlichkeitsfragebogen gemeint. Die Empfehlung zum Einsatz zum Beispiel in „Anwendungsfeldern wie Arbeits-, Betriebs- ... Psychologie“ (S. 7) ignoriert – wie Persönlichkeitsfragebogen häufig – das empirische Faktum, daß diese grundsätzlich leicht verfälschbar sind und auch tatsächlich, wenn mit dem Testergebnis persönliche Konsequenzen verbunden sind, verfälscht werden (vgl. z.B. die Übersicht dazu von Benesch & Raab-Steiner, 2002).

<p>Rezensioniert: Kratzmeier, H. & Horn, R. (1988). <i>RAVEN-Matrizen-Test. Standard Progressive Matrices (SPM)</i> (2. erweiterte und überarbeitete Aufl.). Weinheim: Beltz.</p>	<p>Neu: Heller, K.A., Kratzmeier, H. & Lengfelder, A. (1988). <i>Matrizen-Test-Manual, Band 1. Ein Handbuch mit deutschen Normen zu den Standard Progressive Matrices von J. C. Raven</i>. Göttingen: Beltz.</p> <p>Neu: Bulheller, S. & Häcker, H. (Hrsg.). (1999). <i>Standard Progressive Matrices. Ausgabe 1999 mit der Parallelform und der SPM-Plus-Version</i>. Frankfurt: Swets.</p>
--	--

Fazit: „Der Aufgabentyp von Matrizentests ist zweifellos ein geeignetes Mittel, die Fähigkeit zum induktiven Denken zu erfassen. Die SPM allerdings sind auf Grund problematischer Konstruktionseigenschaften und vielleicht ihrer Länge, jedenfalls aufgrund ihrer überalterten und altersmäßig stark eingeschränkten Normierung sehr kritisch zu sehen“ (Willmes, 1997, S. 119).

Mit dem neuen Manual zu den SPM (Heller et al., 1998) stehen nun aktuelle deutsche Normen zur Verfügung. Die Daten für die Neunormierung wurden 1996/97 in den Bundesländern Baden-Württemberg, Bayern, Nordrhein-Westfalen, Berlin und Sachsen erhoben. Die Autoren sprechen von einer repräsentativen Normierungsstichprobe ($N = 2134$; 6- bis 18jährige Schüler). Zudem enthält das Handbuch Normtabellen für hörgeschädigte und lernbehinderte Schüler (14- bis 16jährige) und für Studierende (19- bis 35jährige) sowie Nomogramme für Personen über 60 Jahre. Das Testmaterial – die bekannten Matrizen von J. C. Raven – ist gleichgeblieben, obwohl Kubinger, Formann und Farkas (1991) anhand von Analysen nach dem *Rasch*-Modell feststellten, daß von insgesamt 60 Aufgaben 43 nicht eindimensional zu den anderen skaliert sind. Neue Untersuchungen zur Reliabilität und Validität anhand deutscher Stichproben werden mitgeteilt.

Die noch neuere Ausgabe eines Manuals zu den SPM eines anderen Verlags (Bulheller & Häcker, 1999) führt verschiedene Formen ein. Einerseits die SPM in der klassischen Version samt erstmaliger Parallelform; andererseits wird mit einer Plus-Version eine Form eingeführt, die „verbesserte Differenzierungen zwischen den begabteren Jugendlichen und jungen Erwachsenen der heutigen Zeit“ (S. 3) erlaubt. Sowohl die Parallelform als auch die Plus-Version enthält neue Aufgaben, wobei in der Plus-Version einige Aufgaben der klassischen SPM mit mittlerer Schwierigkeit durch Aufgaben mit hohem Schwierigkeitsgrad ersetzt wurden. Der Testkonstruktion dieser Formen liegen Analysen nach dem *Rasch*-Modell zugrunde. Die Normierung der SPM-Plus erfolgte 1998/99 in Deutschland an 1796 Personen mit einer Alterstreuung zwischen 6 und 66 Jahren. Primär setzt sich die Normierungsstichprobe aus Schülern zusammen. Für die deutsche Stichprobe wurde ein Reliabilitätskoeffizient von .89 ermittelt. Die Validität wird überwiegend durch Studien im angloamerikanischen Raum belegt.

3. Schluß

Die eingangs gestellte Frage: „So schlecht war ‚unser‘ Testinventar; wie gut ist es jetzt?“ kann zusammengefaßt so beantwortet werden: „Weniges ist besser geworden!“ Nimmt man die vier Projektiven Verfahren außer Betracht, bezüglich derer wohl kaum Qualitätssteigerungen zu erwarten waren, so finden sich zwischenzeitlich nur relativ wenige, sieben von 21, also ein Drittel, die nennenswerte Produktverbesserungen mit sich brachten.

Nun kann man erstens argumentieren, innerhalb von 5 Jahren ist diese Quote im Vergleich zu früheren „Epochen“ schlicht sensationell hoch; man kann auch argumentieren, daß hier diejenigen 25 psychologisch-diagnostischen Verfahren zu Diskussion standen, welche vor fünf Jahren als „einschlägig“ galten, das heute verfügbare und in der Praxis relevante Inventar aber durch wichtige andere und vor allem neue „Gattungen“ von Verfahren „getragen“ wird: Zu letzteren zählen Computertests, die eine breite Palette von Eigenschaften erfassen, Lerntests und Objektive Persönlichkeits*tests*.

Das darf jedoch nicht davon ablenken, daß Zweidrittel der (ehemals?) einschlägigen Verfahren sicher nicht die Qualitätsstandards erfüllen. Als Konsequenz aus der definitiv beschlossenen Norm DIN 33430 *Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen* (DIN Deutsches Institut für Normung e.V., 2002) ist zu erwarten, daß diese in allen Anwendungsbereichen der *Psychologischen Diagnostik* in juristischen Auseinandersetzungen als „nächste Rechtsnorm“ herangezogen wird, natürlich auch in Österreich; und dabei greift schon allein ein einziger Passus gewichtig gegen diese Zweidrittel, nämlich die Festlegung: „Normen sind spätestens alle 8 Jahre zu überprüfen. Die Darstellung von Normierungsuntersuchungen muss das Jahr der Datenerhebung, ... das Design der Stichprobenziehung und die Teilnehmerquoten ausführlich enthalten“!

Literatur

- Amelang, M. & Bartussek, D. (1997). *Differentielle Psychologie und Persönlichkeitsforschung* (4. überarbeitete und erweiterte Aufl.). Stuttgart: Kohlhammer.
- Becker, P. (1989). *Der Trierer Persönlichkeitsfragebogen (TPF)*. Göttingen: Hogrefe.
- Benesch, M. & Raab-Steiner, E. (2002). Zur Verfälschbarkeit von Persönlichkeitsfragebogen: Strategien zur Verhinderung und deren Grenzen. *Psychologie in Österreich*, 22, 39-41.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI)*. Göttingen: Hogrefe.
- Deegener, G. (1997). Thematischer Apperzeptionstest (TAT). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 18, 29-30.
- DIN Deutsches Institut für Normung e.V. (2002). *Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. DIN 33430*. Berlin: Beuth.
- Gittler, G. (1984). Entwicklung und Erprobung eines neuen Testinstruments zur Messung des räumlichen Vorstellungsvermögens. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 5, 141-165.
- Guthke, J., Kubinger, K.D., Westhoff, K. & Wottawa, H. (2002). Der Fortschritt der Psychologischen Diagnostik im letzten Jahrzehnt. *Psychologie in Österreich*, 22, 3-10.
- Krampen, G. (1997). Replik zur Rezension des IPC-Fragebogens zu Kontrollüberzeugungen (IPC). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 18, 34-35.
- Kubinger, K.D. (Hrsg.). (1997). Themenheft „Testrezensionen: 25 einschlägige Verfahren“. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 18 (1/2).
- Kubinger, K.D., Formann, A.K. & Farkas, M.G. (1991). Psychometric shortcomings of Raven's Standard Progressive Matrices, in particular for computerized testing. *European Review of Applied Psychology*, 41, 295-300.
- Paczensky, S. von (1976). *Der Testknacker*. Reinbek/Hamburg: Rowohlt.
- Schmidt, I. & Leiss, U. (2002). Gegenüberstellung von Intelligenztests für Kinder und Jugendliche: HAWIK-III, AID 2 und K-ABC. *Psychologie in Österreich*, 22, 16-21.
- Testkuratorium (der Föderation deutscher Psychologenverbände). (1986). Mitteilung. *Diagnostica*, 32, 358-360.

Adresse der Autoren:

Joachim F. Punter
Arbeitsbereich Psychologische Diagnostik
Institut für Psychologie
Universität Wien
Liebiggasse 5
A-1010 Wien
eMail: joachim.punter@univie.ac.at

Univ.-Prof. Dr.Mag. Klaus D. Kubinger
Arbeitsbereich Psychologische Diagnostik
Institut für Psychologie
Universität Wien
Liebiggasse 5
A-1010 Wien
eMail: klaus.kubinger@univie.ac.at