

Validierung eines klinischen Routine-Systems zur computerisierten Erstellung von MMPI-Befunden bei psychiatrischen Patienten*

Rolf R. Engel

Psychiatrische Klinik der Universität München,
Abteilung für Experimentelle und Klinische Psychologie,
Nußbaumstraße 7, D-8000 München 2, Bundesrepublik Deutschland

Validity of Computer-Generated MMPI Reports in Psychiatric Patients

Summary. Lachar's (1974) completely documented, automated MMPI interpretational system was translated into German and analyzed regarding validity and generalizability in 1190 German psychiatric inpatients. The 114 paragraphs of the text library, each interpreting independent aspects of the MMPI profile, were rated for their accuracy in each individual report by the psychiatrist in charge for the respective patient. As an average, 83 of every 100 printed paragraphs were rated as 'accurate.' Further analysis revealed that the accuracy of paragraphs describing elevated MMPI profiles was substantially higher than in those describing profiles within the normal range. Accuracy rates of individual paragraphs correlated significantly ($r=0.61$) with those reported by Lachar, thus confirming the specific accuracy of each paragraph. It is concluded that the present MMPI interpretational system offers descriptive clinical reports that have shown a high accuracy in culturally diverse psychiatric settings.

Key words: MMPI – Computerized interpretation – Psychological assessment – Transcultural comparison.

Zusammenfassung. Das MMPI-Interpretationssystem von Lachar (1974), das gegenüber den bisher verfügbaren Systemen eine Reihe von Vorteilen aufweist, wurde in einer deutschen Version an 1190 stationären psychiatrischen Patienten auf seine Gültigkeit hin untersucht. Von den 153 Texten der Inter-

* Eine frühere Fassung dieser Arbeit wurde auf dem 31. Kongreß der Deutschen Gesellschaft für Psychologie in Mannheim vorgetragen. Ich danke Frau Dipl. Psych. Kornelia Zangl und Frau Dr. Pola Engel-Sittenfeld für ihre Hilfe bei der Übersetzung der Interpretationstexte. Langjährige wertvolle Mitarbeit haben Frau Irmtraud Reinhold bei der Testdurchführung, Frau Gabi Kunze bei der Programmierung des Interpretationssystems und bei der Verarbeitung der Daten geleistet

pretationsbibliothek wurden diejenigen 114 Texte, die voneinander unabhängige Merkmale eines MMPI-Profiles interpretieren, bei jedem Vorkommen im Einzelfall von dem behandelnden Psychiater hinsichtlich der interpretativen Treffsicherheit beurteilt. Im Durchschnitt wurden von jeweils 100 Texten 83 als „zutreffend“ bezeichnet. Bei der Analyse der einzelnen Programmteile zeigte sich, daß die Treffsicherheit bei psychopathologisch auffälligen MMPI-Profilen wesentlich höher war als bei Profilen im Normbereich. Die in dieser Untersuchung gefundenen Trefferraten korrelierten hoch mit denen der amerikanischen Originaluntersuchung, was die spezifische und replizierbare Gültigkeit der einzelnen Texte bestätigt. Das vorliegende Interpretationssystem bietet damit auf der Grundlage des MMPI eine Persönlichkeitsbeschreibung, für die in verschiedenen psychiatrischen Einsatzbereichen und verschiedenen Sprachen eine gute interpretative Treffsicherheit im Einzelfall nachgewiesen ist.

Schlüsselwörter: MMPI – Computergenerierter Testbefund – Psychologische Tests – Interkulturelle Vergleichbarkeit.

Einleitung

In einer Besprechung im 7. Mental Measurements Yearbook bezeichnet Rodgers (1972) den MMPI (Minnesota Multiphasic Personality Inventory; Hathaway u. McKinley 1951) als „gegenwärtig wohl bestgeeignetes psychologisches Meßinstrument zur Erfassung von Art und Ausmaß psychischer Probleme bei erwachsenen Patienten, die professionelle Hilfe suchen für Symptome, die seelischen Ursprungs sind oder sein könnten.“

Im weiteren Zusammenhang der Besprechung wird deutlich, daß Rodgers diese recht prägnante Aussage auf zwei Informationsquellen stützt, die in der psychologischen Diagnostik gemeinhin selten übereinstimmende Aussagen anregen: zum einen führt er die „in der klinischen Praxis feststellbaren Vorzüge des Testverfahrens“ an, zum anderen beruft er sich auf die Unzahl empirischer Validierungsstudien zum MMPI (s. z. B. Dahlstrom et al. 1972). Die Doppel-funktion des MMPI — zum einen ein Test, der in der klinischen Praxis beliebt ist und gerne als Grundlage für eine Persönlichkeitsbeschreibung individueller Patienten benutzt wird und zum anderen ein Test, dessen statistische Diskriminationsfähigkeit für verschiedene diagnostische Gruppen nachgewiesen ist — dieser Doppelrolle des MMPI ist es wohl zu verdanken, daß in den 50er Jahren in den USA an Hand des MMPI maßgebliche und kontroverse Untersuchungen über klinische versus statistische Urteilsbildung durchgeführt wurden, in deren Gefolge Meehl 1956 seine bekannte Forderung „Wanted — A good cookbook“ aufstellte.

Die Meehl'sche Forderung geht zunächst von der expliziten Annahme aus, daß es in einem klinischen Kontext sinnvoll ist, Befundberichte psychologischer Testuntersuchungen in Form allgemeiner Persönlichkeitsbeschreibungen abzugeben (manchmal auch „prediction of free criteria“ genannt, s. Wiggins 1973,

p. 109). Die Forderung nach dem „Kochbuch“ bezieht sich auf die Art und Weise, wie ein solcher verbal formulierter Testbefund zustande kommt: statt immer wieder neue Generationen von Psychologiestudenten in der Testinterpretation aufgrund tradierter „klinischer“, meist nur impliziter Regeln zu trainieren, schlägt Meehl die Konstruktion eines expliziten Systems vor, bei dem jeweils ein Interpretationstext einem bestimmten Merkmal des MMPI-Profilzugeordnet ist.

Seit 1956 sind eine Reihe von Interpretationssystemen für den MMPI erschienen, deren wesentliches Merkmal die explizite und eindeutige Zuordnung einer bestimmten Interpretation zu einem gegebenen Testprofil ist und die dadurch — wenn sie einmal aufgestellt ist — in ihrer Anwendung stets reliabel und konsistent sind. Die meisten dieser Systeme sind auch computerisiert, was aber kein bestimmendes Merkmal, sondern eher eine ökonomische Begleiterscheinung ist.

Die vorhandenen Systeme unterscheiden sich allerdings sehr weitgehend hinsichtlich ihrer Konstruktionsmerkmale. Auf der einen Seite gibt es diverse kommerziell vertriebene Systeme, bei denen — wohl aufgrund finanzieller Überlegungen — weder die Regeln noch die Textbibliotheken dokumentiert sind (Ausnahme: das in der Mayo-Clinic entwickelte MMPI-System (Psychological Corporation MMPI Reporting Service, Pearson u. Swenson 1967), das eher für ein generelles diagnostisches Screening gedacht ist).

Obwohl gelegentlich eine Bewertung der Güte dieser Systeme versucht wurde (Eichman 1972; Kleinmuntz 1972; Dahlstrom et al. 1972; Klett u. Pumroy 1971; Manning 1972; Bürli 1972), wobei auch jeweils deutliche Unterschiede zwischen den verschiedenen Systemen dargelegt wurden, sind sie aufgrund der fehlenden Dokumentation von unabhängiger Seite letztlich nicht beurteilbar und schon gar nicht überprüfbar. Die gelegentlich von den Autoren kommerzieller Systeme durchgeföhrten Validierungsstudien (v. a. Fowler für das vom Roche Psychiatric Service Institute vertriebene System; Webb et al. 1970) geben stets nur pauschale und methodisch zweifelhafte (s. Eichman 1972) Information über die Güte des gesamten verbalen Berichtes, keine Information über die Güte der einzelnen Regeln und Texte.

Auf der anderen Seite gibt es drei Systeme, die im Rahmen von streng aktuarischen Studien entstanden sind (Gilberstadt 1970; Marks u. Seeman 1963; sowie Gynther et al. 1973). Bei diesen Systemen besteht eine volle Dokumentation von Klassifikationsregeln und Interpretationstexten. Die Texte basieren jeweils auf Informationen, die aus anderen Quellen (Q-Sorts, Check-Lists, u. ä.) stammen und sich für ein gegebenes Profil als spezifisch erwiesen haben. Der Hauptnachteil dieser Systeme besteht in ihrer begrenzten praktischen Anwendbarkeit: da nur für einige MMPI-Profilz Texte angeboten werden, lassen sich die Systeme von Gilberstadt (1970) und Marks u. Seeman (1963) nur bei ca. 15–30%, das System von Gynther et al. (1973) bei ca. 50–70% der MMPI-Profilz psychischer Patienten einsetzen (Gynther et al. 1973). Von einem klinischen Standpunkt aus ist die angebotene Information bei einigen der behandelten MMPI-Profilz dürftig.

Bei dieser unbefriedigenden Dichotomie — in einem Fall valide aber dürftige Information nur für einen Teil der Patienten, im anderen Fall umfassende und stets anwendbare Information, aber von zweifelhafter Gültigkeit — nimmt das

kürzlich vorgestellte MMPI-System von Lachar (1974 b) eine Sonderstellung ein. Es soll im folgenden ausführlicher beschrieben werden.

Das Interpretationssystem von Lachar

Von der Konstruktion her ist Lachar's System nicht-aktuarisch: es beruht also nicht auf einer empirischen Sammlung von Persönlichkeitsmerkmalen bestimmter MMPI-Profiles. Im wesentlichen handelt es sich um ein nach klinischen Gesichtspunkten konstruiertes System, das möglichst alle verfügbaren Informationen über Persönlichkeitskorrelate bestimmter MMPI-Konfigurationen miteinbezieht, „nur-klinische“ Informationen, wie z. B. das Manual von Carson (1969), die schon vorhandenen aktuarischen Informationen (Gilberstadt u. Duker 1965; Marks u. Seeman 1963), aber auch verstreute Detailpublikationen über spezifische Profilkonfigurationen. Alle Regeln und alle Interpretationstexte sind voll dokumentiert (Lachar 1974 a, 1974 b) und es liegt außerdem eine umfassende Validierungsstudie an 1410 psychiatrischen Patienten vor, in der für die einzelnen Textteile Gültigkeitsziffern berechnet wurden.

Abbildung 1 zeigt den strukturellen Aufbau des Interpretationssystems. Eine Interpretation für ein gegebenes MMPI-Profil umfaßt normalerweise zwischen 3 und 10 Textteile (im folgenden Paragraphen genannt), die aus einer Bibliothek von 153 Paragraphen ausgewählt werden. Die Länge der einzelnen Paragraphen beträgt im Durchschnitt etwa 6 bis 8 Zeilen. In Abhängigkeit von der generellen psychopathologischen Auffälligkeit des MMPI-Profs sind drei verschiedene Befundtypen möglich, die jeweils auf verschiedene Regeln (— und damit Texte —) zugreifen (im folgenden Unterprogramme genannt).

Das erste und das letzte Unterprogramm wird bei jedem Befundtyp aufgerufen: das erste („Gültigkeit“) interpretiert die Konfiguration der Validitäts-skalen, das letzte („Spezialkonfigurationen“) interpretiert spezielle Profilkonfigurationen, hauptsächlich solche der Skalen Mf (männlich-weibliche Interessenrichtung) und Si (soziale Introversion). Der erste Befundtyp interpretiert solche Profile, bei denen mindestens eine klinische Skala einen T-Wert größer 70 aufweist. Bei diesem Befundtyp wird zunächst das mit 51 Paragraphen umfangreichste Unterprogramm „Code“ aufgerufen, das auffallende Skalenerhöhung einer oder zweier klinischer Skalen interpretiert. Sind noch weitere Skalen auffällig (je nach Skala $T > 60$ oder $T > 65$) erhöht, schließt sich das Unterprogramm „Skalenerhöhungen“ an. Dieses Unterprogramm wird ebenfalls bei dem zweiten Befundtyp aufgerufen, der dann benutzt wird, wenn zwar keine Skala > 70 , aber mindestens 1 Skala größer als (je nach Skala) 60 oder 65 ist. MMPI-Profiles, die im Normalbereich liegen, stellen den dritten Befundtyp dar, bei dem das Unterprogramm „Skalerniedrigungen“ aufgerufen wird.

In der Validierungsstudie zu diesem System, deren Ergebnisse zusammengefaßt in Lachar (1974 a), detaillierter im eigentlichen Manual (Lachar 1974 b) dargestellt sind, wurde jeder Paragraph zunächst individuell evaluiert. Zusammen mit dem eigentlichen Computerbefund erhielt der jeweils behandelnde Psychiater einen Bewertungsbogen, auf dem er angeben konnte, ob der Inhalt eines Paragraphen bei diesem Patienten zutrifft oder nicht zutrifft.

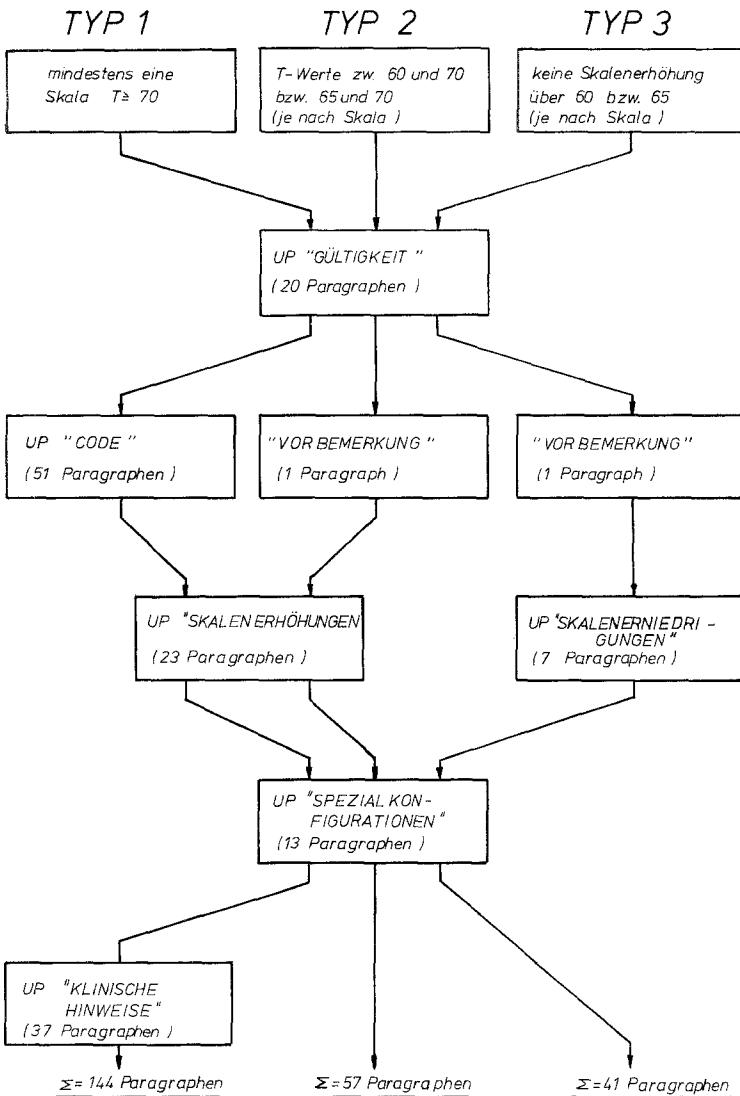


Abb. 1. Ablaufplan des MMPI-Interpretationssystems von Lachar (1974b)

Die Gesamt trefferrate des Systems betrug 90%. Bei den Einzelergebnissen zeigte sich, daß die Gültigkeit der einzelnen Paragraphen stark streute, wobei diejenigen Teile des Interpretationssystems, die sich auf psychopathologisch auffällige Profile bezogen, eine deutlich höhere Gültigkeit erreichten als die Teile, die „normale“ MMPI-Profile interpretieren. Andere Personenmerkmale wie Alter, Geschlecht, Schulbildung und stationäre vs ambulante Betreuung hatten nur einen geringen Einfluß auf die Gültigkeitsziffern (Lachar 1974 a).

Fragestellung

Aufgrund der nachgewiesenen Gültigkeit und Generalisierbarkeit des Interpretationssystems von Lachar bei psychopathologisch auffälligen Patienten wurde

Tabelle 1. Stichprobenmerkmale der vorliegenden und der amerikanischen Vergleichsstichprobe (nach Lachar 1974 b) (in Prozent)

Merkmal		vorliegende	amerikanische
		Stichprobe <i>N</i> = 1190	Stichprobe <i>N</i> = 1472
Geschlecht	männlich	52%	75%
	weiblich	48%	25%
Alter	bis 24	36%	54%
	25–34	24%	21%
	35–44	19%	17%
	45–54	13%	6%
	über 54	8%	2%
Quelle	stationär psychiatrisch	100%	39%
	ambulant psychiatrisch	—	50%
	allgemein medizinisch	—	11%

eine deutsche Version dieses Systems konzipiert und Mitte 1975 als Forschungsinstrument in die klinische Routinediagnostik der Psychologischen Abteilung der Psychiatrischen Klinik der Universität München eingeführt. Die vorliegende Arbeit berichtet die Ergebnisse der begleitenden Validitätsuntersuchung an 1577 Patienten (vorläufige Ergebnisse an kleineren Patientenzahlen wurden bereits mitgeteilt (Engel 1977)).

Neben der allgemeinen Frage nach der Validität der deutschen Version des Systems interessierte vor allem, ob die von Lachar (1974 a) gefundenen Gültigkeitsunterschiede der verschiedenen Systemteile sich bei dem interkulturellen Vergleich replizieren lassen.

Methode

Beschreibung der Patientenstichprobe

Die vorliegende Stichprobe umfaßt alle „testfähigen“ Patienten, die im Laufe von 20 Monaten stationär in der Psychiatrischen Klinik der Universität München neu aufgenommen wurden. Als „nicht testfähig“ galten dabei Patienten, die aufgrund von geringer Intelligenz, mangelnder deutscher Sprachkenntnisse oder Akuität ihrer Erkrankung nicht in der Lage waren, an der ca. einstündigen Untersuchung teilzunehmen. Infolge der Struktur einer Psychiatrischen Universitätsklinik handelt es sich um eine heterogene Stichprobe psychiatrischer Patienten, bei der akut Erkrankte den Hauptanteil darstellten.

Insgesamt 1577 Patienten wurden untersucht. Von 387 Patienten erfolgte keine oder nur ungenügende Information über die Gültigkeit des Befundes durch den behandelnden Arzt, so daß die Analysestichprobe letztlich aus 1190 Patienten bestand. In Tabelle 1 sind die wichtigsten Stichprobenmerkmale für die vorliegende und für die amerikanische Validierungsstichprobe (Lachar 1974 b) zusammengestellt.

Testdurchführung und Auswertung

Der MMPI-Saarbrücken (Spreen 1963) wird üblicherweise in der Gesamtlänge von 566 Items vorgegeben, obwohl für die Auswertung der Standardskalen nur 399 Items benötigt werden. Von diesen 399 Items kommen 16 Items doppelt vor, so daß sich die eigentliche klinische Standardinformation aus 383 Items errechnen läßt. Aus ökonomischen Gründen wurden in dieser Untersuchung lediglich diese 383 vorgegeben, und zwar wahlweise in zwei verschiedenen Formen:

(a) Bei der üblichen und ökonomischsten Version kreuzten die Patienten ihre Antworten direkt auf maschinell lesbaren OMR (Optical Mark Reader)-Karten an, die automatisch bearbeitet wurden.

(b) Als Alternative zu dieser Art der Testdarbietung stand bei weniger intelligenten oder bei sehr gestörten Patienten eine „leichtere“ Version zur Verfügung, bei der die Zuordnungsaufgabe (vom Item des Testheftes zur Antwortzeile der OMR-Karte) wegfällt. Bei dieser Kartenversion sortieren die Patienten normale Lochkarten, die auf der Vorderseite die Itemlochung und auf der Rückseite den Itemtext tragen, in zwei Stapel („TRIFFT ZU“; „TRIFFT NICHT ZU“). Auch bei diesen Kartenstapeln ist durch die Itemlochung eine automatische Auswertung möglich. Vor einer neuerlichen Vorgabe muß der Stapel allerdings wieder in die richtige Itemreihenfolge gebracht werden, was im Vergleich zu der OMR-Version recht unökonomisch ist.

Von den 1577 Patienten der Stichprobe bearbeiteten 136 Patienten die Kartenversion, alle anderen die OMR-Version.

Routinemäßig wurde die MMPI-Untersuchung vormittags zwischen 9.00 und 11.00 Uhr in einem speziellen Testraum durchgeführt, nur in Einzelfällen auch einmal nachmittags.

Testauswertung und -interpretation

Die apparative Ausstattung zur Verarbeitung des Tests umfaßte ein 8 K-Bytes-Tischrechner-System (Hewlett Packard 9830) mit Kartenleser, Schnelldrucker und zwei Digitalkassetten als Peripherie. Die Auswertung eines einzelnen MMPIs bestand aus Decodierung der OMR-Karte, Roh- und *T*-Wert-Berechnung, Itemweise Speicherung auf Digitalkassette, Interpretation sowie Ausdruck des gesamten Befundes mit Profilblatt und dauerte etwa 25 Minuten (ausführliche Beschreibung bei Engel u. Kunze 1979).

Bearbeitung des Interpretationssystems

Das MMPI-Interpretationssystem von Lachar (1974 b) besteht aus zwei Teilen: (a) eine Textbibliothek von insgesamt 153 Textabschnitten, die interpretative Aussagen zu bestimmten Charakteristika des gegebenen MMPI-Profils machen. Die Länge dieser Textabschnitte variiert zwischen einfachen Ein-Satz-Aussagen und ausführlichen Interpretationstexten, die bis zu 200 Worte lang sein können. (b) Eine Sammlung logischer Regeln, die jeden einzelnen Textabschnitt einem bestimmten Charakteristikum des MMPI-Profils zuordnen.

Bei der hier untersuchten deutschen Bearbeitung des Systems wurden die logischen Regeln exakt so übernommen, wie sie bei Lachar (1974 b) beschrieben sind, also einschließlich der zahlenmäßigen *T*-Wert-Kriterien zur Abgrenzung bestimmter Profiltypen. Die Textbibliothek wurde von drei klinischen Psychologen gemeinsam übersetzt. Die Übersetzung lehnt sich in der vorliegenden Fassung so eng wie möglich an das englische Original an.

Gültigkeitsbewertung

Zusammen mit dem MMPI-Bericht wurde dem behandelnden Arzt ein Bewertungsbogen zugeschickt, auf dem er den betreffenden MMPI-Bericht in zweierlei Art und Weise bewerten konnte: (a) jeder einzelne Textabschnitt sollte in einer Ja/Nein-Entscheidung bewertet werden (also: dieser Abschnitt ist im wesentlichen zutreffend versus dieser Abschnitt ist im wesentlichen nicht zutreffend). (b) Für den gesamten Bericht konnte eine Note zwischen 1 (im wesentlichen zutreffend) und 6 (völlig unzutreffend) vergeben werden. Die Richtlinien für die Bewertung

Tabelle 2. Vorkommenshäufigkeit von 114 Paragraphen, die auf unabhängigen Regeln beruhen

Häufigkeit des Vorkommens (in Klassen)	0	1-	20-	40-	60-	80-	100-	150-	200-	400-
Anzahl der Paragraphen	4	51	20	8	10	3	9	3	5	1

wurden von Lachar (1974b) übernommen und zur Erinnerung für den Arzt auf jedem Bewertungsbogen erneut ausgedruckt.

Die Ärzte, die den MMPI-Befund bewerteten, kannten die entsprechenden Patienten zum Zeitpunkt der Bewertung schon mindestens 5 Tage (meistens länger), so daß bei diesen stationären Patienten mit Sicherheit eine gute Beurteilungsgrundlage gegeben war. Die Rücklaufquote der Bewertungsbögen betrug 75%.

Ergebnisse

Eine Bewertung der Gültigkeit der einzelnen Paragraphen setzt voraus, daß der betreffende Paragraph auch häufig genug auftritt. In Tabelle 2 sind die Vorkommenshäufigkeiten für diejenigen Textabschnitte dargestellt, die aufgrund voneinander unabhängigen Regeln auftreten können ($N=114$ Paragraphen; die restlichen 39 Paragraphen sind kurze Hinweise, die jeweils vom Auftreten anderer Paragraphen abhängen).

Es zeigte sich, daß selbst bei der doch recht beachtlichen Zahl von 1190 Patienten immerhin 4 Paragraphen kein einziges Mal und 28 Paragraphen einmal bis neunmal vorkamen. Bei diesen Paragraphen, die alle zu extremen Profilkonfigurationen Stellung nehmen, ist die Fallanzahl so gering, daß eine individuelle Bewertung der Genauigkeit dieser Interpretationsteile nicht möglich ist. Bei den sehr häufig vorgekommenen Paragraphen handelt es sich entweder um solche Texte, die darauf hinweisen, daß in bestimmten Bereichen *keine* pathologischen Werte vorliegen oder um Texte, die zu solchen Profilcharakteristika Stellung nehmen, die häufig — sei es isoliert oder in Verbindung mit anderen Skalenerhöhungen — vorkommen, z. B. hohe D, Hy oder Si-Skala.

In Tabelle 3 sind die Trefferraten der einzelnen Unterprogramme in der untersuchten deutschen Version im Vergleich zu Lachar's (1974b) Daten dargestellt. Um eine Vergleichbarkeit zu gewährleisten, wurden in diese Tabelle nur diejenigen Paragraphen aufgenommen, für die Lachar (1974b) Gültigkeitsdaten angegeben hat, d. h. nur solche Paragraphen, die in der amerikanischen Patientenstichprobe ($N=1472$) mindestens 15mal vorgekommen waren.

Wie Spalte 1 von Tabelle 3 zeigt, reduzierte diese Restriktion die Anzahl verwendbarer Paragraphen von 114 auf 74. Spalte 2 gibt die absolute, Spalte 3 die relative Anzahl von Beurteilerratings für jedes einzelne Unterprogramm an. Eine χ^2 -Analyse für die Spalte 2 zeigte, daß die verschiedenen Programmteile in beiden Untersuchungen unterschiedlich oft benutzt wurden ($\chi^2 = 180,0$, df = 4, $P \leq 0.001$): im Vergleich zu den amerikanischen Daten wurden in unserer Stichprobe die Programmteile Code und Skalenerhöhungen seltener, die restlichen Teile häufiger.

Tabelle 3. Diagnostische Treffsicherheit des Interpretationssystems, aufgegliedert nach Programmteilen, bei 1190 psychiatrischen Patienten (in Klammern die Vergleichsdaten von Lachar 1974b)

Programmteil	Anzahl Paragraphen	Anzahl der abgegebenen ratings	Prozentuale Verteilung auf die Programmteile	Prozentualer Anteil zutreffender Paragraphen
Gültigkeit	10 von 20	1126 (1302)	22 (18)	93 (98)
Code	27 von 51	605 (1153)	12 (16)	89 (91)
Skalenerhöhungen	23 von 23	2165 (3328)	41 (45)	80 (89)
Spezialkonfigurationen	11 von 13	1224 (1472)	23 (20)	79 (87)
Skalenerniedrigungen	3 von 7	93 (51)	2 (1)	65 (86)
Total	74 von 114	5213 (7306)	100 (100)	83 (90)

ger verwendet. Dies bedeutet inhaltlich, daß in der amerikanischen Stichprobe relativ mehr Patienten psychopathologisch auffällige Profile hatten.

Spalte 4 von Tabelle 3 zeigt die Gültigkeitsraten der einzelnen Unterprogramme und des gesamten Systems. Die mittlere Gültigkeitsrate aller einbezogenen 74 Paragraphen betrug in der vorliegenden Version 83%, verglichen mit 90% im amerikanischen Original ($\chi^2 = 131,6$, $df = 1$, $P \leq 0,001$). Es wurden also rund 17 von jeweils 20 Abschnitten als zutreffend bezeichnet.

In den beiden Untersuchungen zeigte sich exakt die gleiche Rangreihe der Programmteile hinsichtlich ihrer Gültigkeit: diejenigen Programmteile, die sich vorwiegend auf „normale“ MMPI-Profile beziehen (Skalenerniedrigungen und Spezialkonfigurationen), wiesen in beiden Untersuchungen die niedrigsten Gültigkeitswerte auf und die Programmteile, die sich auf auffällige Profile beziehen, zeigten die höchsten Trefferraten. Die Güte der Interpretation war also umso genauer, je spezifischer die angebotene Information war: Profile, die auf deutliche psychopathologische Symptome hinweisen, wurden genauer interpretiert als Profile, die sich im wesentlichen im Normalbereich bewegen.

Um ein generelles Maß für die interkulturelle Übertragbarkeit des Interpretationssystems zu erhalten, wurden die bei Lachar (1974b) angegebenen Trefferraten aller einzelner Paragraphen mit den in der vorliegenden Untersuchung gefundenen korreliert. Da hier die Gültigkeit jedes einzelnen Paragraphen möglichst präzise erfaßt werden sollte, wurden nur die Texte berücksichtigt, die in *beiden* Stichproben mindestens 15mal vorgekommen waren ($N=55$). Die entsprechende Produkt-Moment-Korrelation beträgt 0,61 bei einem 95%-Vertrauensbereich von 0,41 bis 0,75. Die „guten“ Textteile der amerikanischen Version erbrachten also in der vorliegenden Untersuchung ebenfalls weitgehend gute Resultate, während die „schwachen“ Paragraphen der amerikanischen Version auch in der vorliegenden deutschen Fassung schwache bis unzureichende Trefferraten erzielten. Zumindest für die 55 Paragraphen, die oft genug vorkamen, ließ

sich damit eine gute interkulturelle Übertragbarkeit des Interpretationssystems nachweisen.

Diskussion

In der vorliegenden deutschen Version des MMPI-Interpretationssystems von Lachar (1974b) betrug die Gesamtgenauigkeit 83%. Da ein durchschnittlicher Testbefund ca. 7 Paragraphen umfaßt, kann man sagen, daß pro Testbefund im allgemeinen ein Paragraph als unzutreffend bezeichnet wurde. Die amerikanische Originalversion erreichte eine Treffergenauigkeit von 90%. Für die geringere Trefferrate der deutschen Version lassen sich mehrere Erklärungen finden.

Die einfachste Erklärung wäre wohl die, daß man Unzulänglichkeiten bei der Übersetzung der Textbibliothek dafür verantwortlich macht. Dies scheint allerdings nicht sehr überzeugend, da gerade diejenigen Paragraphen, in denen sehr detailliert, ausführlich (ca. 100–200 Worte) und in fachspezifischer psychologischer oder psychiatrischer Terminologie auf psychopathologisches Verhalten eingegangen wird (Unterprogramm Code), in der deutschen Version praktisch gleich gut sind wie in der amerikanischen. Eine deutliche Validitätseinbuße läßt sich vorwiegend bei den einfachen und kurzen Paragraphen des Unterprogramms Skalenerniedrigungen feststellen.

Als weitere Erklärung ließen sich die beobachteten Unterschiede in beiden Patientenstichproben anführen. Auch hier ergibt sich aber keine klare Situation: nach den Ergebnissen zur Vorkommenshäufigkeit der einzelnen Unterprogramme müßte man annehmen, daß Lachar's Patienten psychopathologisch auffälliger waren. Dies erscheint aber unglaublich, wenn man bedenkt, daß ca. 50% der amerikanischen Stichprobe aus ambulanten Patienten bestand, während die deutsche Stichprobe nur stationäre Patienten erfaßte.

Die plausibelste Erklärung für die beobachteten Diskrepanzen ist wohl in der direkten Übernahme der amerikanischen cut-off-scores zur Trennung der einzelnen Profiltypen zu sehen. Da diese cut-off-scores auf *T*-Werten beruhen, sind sie populationsabhängig. Butcher u. Pancheri (1976) haben bei einem Vergleich von MMPI-Standardisierungen in verschiedenen Ländern deutliche Unterschiede hinsichtlich der mittleren Skalenrohwerte feststellen können. Es ist anzunehmen, daß sich solche interkulturellen Unterschiede am ehesten bei „normalen“ Profilen auswirken. Je offensichtlicher eine Skalenerhöhung ist (bei einzelnen MMPI-Profilen können *T*-Werte von 100 vorkommen!), umso unwichtiger dürften die Effekte der Standardisierung werden.

Als weiteren Beleg dafür, daß ein gegebener *T*-Wert bei interkulturellen Vergleichen eine unterschiedliche Bedeutung hat, läßt sich anführen, daß in der deutschen Version einzelne Paragraphen sehr viel seltener, andere sehr viel häufiger als in der amerikanischen Version auftraten. Eine Verschiebung der cut-off-scores dürfte deshalb eine erfolgversprechende Maßnahme bei einer Revision des deutschen Systems sein.

Von 55 Paragraphen lagen in beiden Versionen mindestens 15 Beurteilungen vor, so daß sich die Trefferrate relativ genau erfassen ließ. Bei diesen 55 Paragraphen zeigte sich eine Korrelation von $r=0,61$ zwischen der amerikanischen

und der deutschen Trefferrate. Dieses Ergebnis läßt den Schluß zu, daß die gefundenen Trefferraten der einzelnen Paragraphen nur wenig Zufallseffekte beinhalten, sondern im wesentlichen echte „Stärken“ und „Schwächen“ des Systems widerspiegeln. Dies ist ein klarer Beleg dafür, daß es offensichtlich möglich ist, aufgrund von MMPI-Profilen verbale Beschreibungen psychopathologischer Merkmale abzugeben, die ein hohes Maß an Gültigkeit besitzen, selbst bei interkulturellen Vergleichen. Angesichts der starken interkulturellen Diskrepanzen hinsichtlich der psychiatrischen Diagnostik und Klassifikation (s. Kendell et al. 1974) bietet sich der MMPI (einschließlich der expliziten verbalen Befundinterpretation) als ein Instrument an, das auf einem niedrigeren, mehr deskriptiven Niveau der Beurteilung eine hohe interkulturelle Vergleichbarkeit aufweist.

Nach den Ergebnissen dieser Untersuchung läßt sich über den Anwendungsbereich des Systems zumindest eine klare Aussage machen: bei auffälligen Profilen werden wesentlich höhere Gültigkeitsraten erzielt als bei „normalen“ Profilen. Daraus läßt sich ableiten, daß das System bei psychopathologisch unauffälligen Personen keine Anwendung finden sollte. Inwieweit sich innerhalb der Gruppe psychiatrischer Patienten bestimmte Untergruppen hinsichtlich ihrer Beurteilungsgüte unterscheiden, kann aufgrund des vorliegenden Datenmaterials noch nicht entschieden werden. Lachar (1974a) fand geringfügige Validitätsunterschiede zwischen ambulanten und stationären Patienten, die in Wechselwirkung mit dem Typ des Befundberichts und dem Alter der Patienten standen. Adams u. Shore (1976) konnten — allerdings an einer sehr kleinen Stichprobe ($N=98$) — diesen Befund nicht replizieren. Lachar et al. (1976) zeigten, daß speziell bei jugendlichen Patienten (12 bis 17 Jahre) die Gültigkeit der Interpretation erhöht wird, wenn die neuen (amerikanischen) Normen für Jugendliche statt der üblichen Erwachsenennormen verwandt wurden. Da für den MMPI-Saarbrücken keine Normen für Jugendliche vorliegen, sollte man den Anwendungsbereich des deutschen Systems auf Erwachsene beschränken.

In der vorliegenden Untersuchung fand sich kein Anhaltspunkt dafür, daß der von Meehl (1956) so genannte „Barnum-Effekt“ die Gültigkeitsurteile der Psychiater beeinflußt hat.

Der Barnum-Effekt beschreibt eine Fehlerquelle bei der Interpretation von Urteilen: allgemeine Aussagen, die auf fast alle Leute zutreffen (— also lediglich das beschreiben, was Cohen (1965) das „gemeinsame Stereotyp“ der Zielpopulation genannt hat —), werden sehr oft als „richtig“ eingestuft, obwohl ihnen keine diagnostische Entscheidungskraft innerhalb der Zielpopulation zukommt. Die vorliegenden Ergebnisse wären mit einer solchen Interpretation nicht in Einklang zu bringen: das Unterprogramm Code, das durchweg aus prägnanten und spezifischen Aussagen besteht, weist eine höhere Gültigkeit auf als die anderen Unterprogramme (außer „Validität“), die weniger spezifische Aussagen machen und im allgemeinen auch höhere Vorkommenshäufigkeiten aufweisen.

Es sollte hier vielleicht nicht unerwähnt bleiben, daß die übliche Betrachtungsweise des Barnum-Effektes (s. z. B. Jackson 1978; O'Dell 1972) in einem handlungsbezogenen klinischen Kontext nicht ohne gewichtige Modifikationen übernommen werden kann. Nur bei einer Gültigkeitsprüfung, die den Aspekt der unterschiedlichen Handlungs- oder Entscheidungsrelevanz der einzelnen Aussagen (in unserem Fall: Wichtigkeit der einzelnen Paragraphen für das weitere

diagnostisch/therapeutische Vorgehen) unberücksichtigt läßt, kann es überhaupt einen Barnum-Effekt geben. Ein Beispiel soll dies verdeutlichen: bei psychiatrischen Patienten liegt der Kupfer-Serum-Spiegel in über 99% der Fälle innerhalb eines unauffälligen Normbereichs. Die diagnostische Aussage auf Grund der Populationsrate wäre also „Kupfer normal“, und der Fehler dabei betrüge maximal ein Prozent. Durch eine Bestimmung des Kupfer-Serum-Spiegels könnte sich eine maximale Steigerung der Aussagegenauigkeit von ein Prozent ergeben, was aufgrund rein statistischer Überlegungen — ohne den Handlungsbezug und die Berücksichtigung der Wichtigkeit dieser 1% — vielleicht als allzu geringe Steigerung der diagnostischen Effizienz angesehen werden könnte. Allerdings hat bei einem bestimmten psychiatrischen Syndrom ein abnormer Kupferspiegel eine pathognostische und damit unmittelbar therapieentscheidende Bedeutung, da er auf eine ätiologisch klar abgrenzbare Erkrankung hinweist. Unter gewissen Umständen kann also eine Steigerung der Trefferrate um 1% wichtiger sein als eine Steigerung um 50% bei einer weniger entscheidungsrelevanten Variablen. Es wäre zu überlegen, ob man bei zukünftigen Evaluationsstudien außer der Genauigkeit einer Aussage auch deren Wichtigkeit für den Handelnden von ihm beurteilen läßt.

Zum Abschluß sei noch auf einige ethische Implikationen einer „Computerdiagnostik“ hingewiesen. Maschinell aufbereitete Informationen können immer dann prinzipiell gefährlich sein, wenn sich aus ihnen auch für Nicht-Fachleute Entscheidungen oder Handlungen ableiten lassen. Rodgers (1972) hat speziell zu maschinellen MMPI-Interpretationssystemen Überlegungen angestellt, die der Autor der vorliegenden Arbeit teilt. Sie lassen sich etwa so zusammenfassen:

Da es zu den Eigenheiten des MMPI gehört, daß eine klinische Interpretation des Tests im Einzelfall — ungeachtet des umfangreichen empirischen Materials — beträchtliche klinische Erfahrung verlangt, muß man fordern, daß auch eine computererstellte MMPI-Interpretation nur von MMPI-erfahrenen Psychologen oder von solchen Psychiatern und klinischen Psychologen benutzt wird, die über erhebliche praktisch-klinische Erfahrung verfügen, die sie in die Lage versetzt, Testbefunde auch einmal zu ignorieren, wenn sie unangebracht erscheinen. Ärzte ohne psychiatrische Erfahrung und gelegentliche Testbenutzer zählen nicht zu diesem Personenkreis, da sie durch einen computerisierten MMPI-Befund ermutigt werden könnten, auf Informationen hin zu handeln, die außerhalb ihrer fachlichen Zuständigkeit liegen.

Die American Psychological Association hat 1966 Standardforderungen an kommerziell vertriebene Testinterpretationssysteme aufgestellt, die die obigen Gesichtspunkte berücksichtigen (Fowler 1969).

Literatur

- Adams KM, Shore DL (1976) The accuracy of an automated MMPI interpretation system in a psychiatric setting. *J Clin Psychol* 32:80-82
- Bürli A (1972) Automatische MMPI-Systeme. *Schweiz Z Psychol Anw* 31:281-291
- Butcher JN, Pancheri P (1976) A handbook of cross-national MMPI-research. University of Minnesota Press, Minneapolis
- Carson RC (1969) Interpretative manual to the MMPI. In: Butcher JN (ed) *MMPI: Research developments and clinical applications*. McGraw-Hill, New York, pp 127-180

- Cohen R (1965) Empirische Untersuchungen zur klinischen Urteilsbildung auf Grund psychologischer Tests. In: Hardesty F, Eyferth K (Hrsg) Forderungen an die Psychologie. Huber, Bern, S 139–152
- Dahlstrom WG, Welsh GS, Dahlstrom LE (1972) An MMPI handbook, Vol 1: Clinical interpretation. University of Minnesota Press, Minneapolis
- Dahlstrom WG, Welsh GS, Dahlstrom LE (1975) An MMPI handbook, Vol 2: Research applications. University of Minnesota Press, Minneapolis
- Eichman WJ (1972) Test review: MMPI computerized scoring and interpreting services. In: Buros K (ed) The seventh mental measurements yearbook. Gryphon Press, Highland Park (NJ), pp 250–262
- Engel RR (1977) Möglichkeiten und Grenzen des Einsatzes eines automatisierten MMPI-Befundsystems in einer psychiatrischen Klinik. In: Tack WH (Hrsg) Bericht über den 30. Kongreß der Deutschen Gesellschaft für Psychologie, Band 2. Hogrefe, Göttingen, S 51–53
- Engel RR, Kunze G (1979) Scoring and interpreting the MMPI with a desk-top calculator. *Behav Res Meth Instr* 11:317–320
- Fowler RD Jr (1969) Automated interpretation of personality test data. In: Butcher JN (ed) MMPI: Research developments and clinical applications. McGraw-Hill, New York, pp 127–180
- Gilberstadt H (1970) Comprehensive MMPI codebook for males. US Government Printing Office, Washington, DC
- Gilberstadt H, Duker J (1965) Handbook for clinical and actuarial MMPI interpretation. Saunders, Philadelphia
- Gynther MD, Altman H, Sletten IW (1973) Replicated correlates of MMPI two-point code types: the Missouri actuarial system. *J Clin Psychol Monograph Supplement* No 39, July 1973
- Hathaway SR, McKinley JC (1951) The MMPI manual. Psychological corporation, New York
- Jackson DE (1978) The effect of test taking on acceptance of bogus personality statements. *J Clin Psychol* 34:63–68
- Kendell RE, Pichot P, von Cranach M (1974) Diagnostic criteria of English, French, and German psychiatrists. *Psychol Med* 4:187–195
- Kleinmuntz B (1972) Test review: Roche MMPI computerized interpretation service. In: Buros OK (ed) The seventh mental measurements yearbook. Gryphon Press, Highland Park (NJ), pp 262–266
- Klett CJ, Pumroy DK (1971) Automated procedures in psychological assessment. In: McReynolds P (ed) Advances in psychological assessment, Vol 2. Science and Behavior Books, Palo Alto (CA), pp 14–39
- Lachar D (1974a) Accuracy and generalizability of an automated MMPI interpretation. *J Consult Clin Psychol* 42:267–273
- Lachar D (1974b) The MMPI: Clinical assessment and automated interpretation. Western Psychological Services, Los Angeles
- Lachar D, Klinge V, Grisell JL (1976) Relative accuracy of automated MMPI narratives generated from adult norm and adolescent norm profiles. *J Consult Clin Psychol* 44:20–24
- Manning HM (1971) Programmed interpretation of the MMPI. *J Pers Assess* 35:162–176
- Marks PA, Seeman W (1963) The actuarial description of abnormal personality: An atlas for use with the MMPI. Williams & Wilkins, Baltimore
- Meehl PE (1956) Wanted—A good cookbook. *Am Psychol* 11:263–272
- O'Dell JPT (1972) Barnum explores the computer. *J Consult Clin Psychol* 38:270–273
- Pearson JS, Swenson WM (1967) A user's guide to the Mayo clinic automated MMPI program. Psychological Corporation, New York
- Rodgers DA (1972) Test review: MMPI. In: Buros OK (ed) The seventh mental measurements yearbook. Gryphon Press, Highland Park (NJ), pp 243–250
- Spreen O (1963) MMPI Saarbrücken. Handbuch zur deutschen Ausgabe des MMPI von Hathaway SR & McKinley JC. Huber, Bern
- Webb JT, Miller ML, Fowler RD Jr (1970) Extending professional time: a computerized MMPI interpretation service. *J Clin Psychol* 26:210–214
- Wiggins JS (1973) Personality and prediction: Principles of personality assessment. Addison-Wesley, Reading (MA)