

## Zur Validität der Eppendorfer Stimmungs-Antriebs-Skala (ESTA III)\*

P. Richter und W. Schmitt

Psychiatrische Klinik (Direktor: Prof. Dr. W. Janzarik) der Universität Heidelberg, Voßstraße 4, D-6900 Heidelberg 1, Bundesrepublik Deutschland

### On the Validity of the Eppendorfer Mood and Drive Scale (ESTA III)

**Summary.** The paper examines the validity of the Eppendorfer Mood and Drive Scale (ESTA III). ESTA III is a self-rating instrument for evaluating the effect of therapy and for studying the course of affective psychoses. Empirical research on the validity of the scale has been carried out solely by the author of the test. Subsequent to a critical review of hitherto existing research on the validity of the instrument, our own validity studies are discussed. Our studies cast doubt on the concurrent validity of the instrument (comparison with the adjective check list / EWL-K). Analysis of our own data and Reanalysis of the data published by the author of ESTA III do not support the construct validity of the instrument.

**Key words:** Drive – Mood – Measurement – Validity – ESTA

**Zusammenfassung:** Die vorliegende Arbeit setzt sich mit Fragen der Validität der Eppendorfer Stimmungs-Antriebs-Skala (ESTA III) auseinander. Die Eppendorfer Stimmungs-Antriebs-Skala ist ein Selbstbeurteilungsbogen zur Therapiekontrolle und Verlaufsdarstellung affektiver Psychosen. Empirische Arbeiten zur Validität der Skala wurden bisher nur vom Testautor durchgeführt. Nach einer kritischen Durchsicht bisheriger Arbeiten zur Gültigkeit des Instruments werden eigene Validierungsstudien vorgestellt. Eigene Untersuchungen lassen Zweifel an der konkurrierenden Validität (Vergleich mit Skalen der Eigenschaftswörterliste / EWL-K) des Instrumentes aufkommen. Die Analyse der eigenen Daten und eine Reanalyse der Daten des Testautors sprechen gegen die Konstruktvalidität von ESTA III.

**Schlüsselwörter:** Antrieb – Stimmung – Messung – Validität – ESTA

### Testbeschreibung

Die Eppendorfer Stimmungs-Antriebs-Skala (abgekürzt: ESTA) ist ein Selbstbeschreibungsinstrument zur Verlaufserfassung manisch-depressiver und schizoaffectiver Psychosen. Sie wurde von Supprian im Rahmen von Verlaufsstudien bei

(bipolaren) Zykllothymien entwickelt. Die Genese des Instrumentes sowie die Vorläufer (ESTA I und ESTA II) der aktuellen Version (ESTA III) sind Supprian (1975a) zu entnehmen. Die folgenden Ausführungen beziehen sich auf die aktuelle Version des Tests, die Version ESTA III.

Der Test besteht aus insgesamt 20 Items, die in der Form von Gegensatzpaaren (z.B. Bedrücktheit – Heiterkeit) formuliert sind. Der Proband wird aufgefordert, sich auf einer 17stufigen Skala für einen der Gegensatzbegriffe zu entscheiden. Die Abstufungen der Skala sind spiegelbildlich wie folgt bezeichnet: äußerst stark, sehr stark, stark, deutlich, etwas mehr, wenig, sehr wenig, spurenweise, weder/noch, spurenweise, etc. Die gegensätzlich formulierten Begriffe (wie z.B. Arbeitsunfähigkeit - Betätigungslust) werden dem depressiven (im Beispiel Arbeitsunfähigkeit) bzw. dem manischen Pol (Betätigungslust) zugeordnet. Von den insgesamt 20 Items des Selbstbeschreibungsbogens dienen 10 Items der Erfassung von Antrieb, die restlichen 10 Items messen Stimmung.

Mögliche Anwendungsgebiete des Tests sind nach Supprian (1975a) der „Aufbau sehr langer Zeitreihen“ sowie die „Durchführung von Zeitgestaltanalysen des Verlaufs manisch-depressiver Psychosen... zur Verfolgung der Verhältnisse bei den sog. manisch-depressiven Mischbildern (Stimmungs-Antriebs-Entmischung)“. Der Hinweis auf die Möglichkeiten einer „Durchführung von Zeitgestaltanalysen des Verlaufs manisch-depressiver Psychosen“ wird verständlicher bei einer Durchsicht der empirischen Arbeiten des Testautors. Supprian (1975b, 1977, 1978) verwendet den Selbstbeschreibungsbogen zur Überprüfung seines chronopathologischen Modells „für den Gesamtablauf manisch-depressiver Psychosen“ (Supprian 1975c). Dieses Modell, das sich auf das „Zeitverhalten des Stimmungsantriebssystems bei phasischen Psychosen“ (Supprian 1977) bezieht, trifft Aussagen zu den Bereichen „Struktur der Tagesschwankung“ (a.a.o.), „Aufbau der Phase“ (a.a.o.), das heißt der „Datierung des Phasenansangs“ (a.a.o.) und der „Periodik von Stimmung und Antrieb“ (a.a.o.).

### Bisherige Studien zur Validität des Instrumentes

Eine Überprüfung teststatistischer Gütekriterien (wie Konsistenz und Reliabilität) des ESTA III liegt bisher nicht vor. Der Informationswert derartiger Analysen für die Beurteilung von Skalen zur subjektiven Befindlichkeit wird eindrucksvoll von v. Zerssen (1976) bzw. von Janke und Debus (1978) belegt. Da nach Lienert (1961) ein Test zur Erfassung der augenblicklichen Stimmungslage zeitlich nicht sehr stabil sein darf, ist ein Vergleich von Konsistenzschätzungen und Koeffizien-

\* Für die freundliche Unterstützung bei der Durchführung der Untersuchungen bedanken wir uns bei Dr. Dr. Tress und Dr. Holm-Hadulla.  
Offprint requests to: P. Richter at the above address

ten der Retest-Reliabilität auch im Hinblick auf die Validität eines Meßinstrumentes zur subjektiven Befindlichkeit bedeutsam.

Testanalysen des ESTA III wurden bisher ausschließlich vom Testautor durchgeführt und beziehen sich lediglich auf den Aspekt der Validität des Instrumentes. Die Validierungsbemühungen des Testautors basieren in erster Linie auf Faktorenanalysen von ESTA III. Supprian (1977, 1978) stützt sich dabei auf Ergebnisse von Einzelfallanalysen an lediglich zwei bipolar-depressiven Patientinnen, deren Daten mithilfe von *p*-Faktorenanalysen ausgewertet wurden. Hierbei ist er bemüht, eine Zwei-Faktorenstruktur mit den Faktoren Stimmung und Antrieb nachzuweisen. Die Faktoreninterpretation der Untersuchung von 1977 (Supprian 1977) wurde andernorts kritisiert (Richter 1984). Die Faktorendeutung der Zwei-Faktorenlösung der Einzelfallanalyse von 1978 (Supprian 1978) ist insofern fraglich, als das Ankeritem des Faktors Antrieb (Arbeitseifer) und das Schlüsselitem von Stimmung (Stimmung) sowohl auf dem Antriebs- als auch dem Stimmungsfaktor annähernd gleich hoch laden. Betrachtet man die extrahierte Varianz bei einer Extraktion von 5 Faktoren (Supprian 1978), so wird aufgrund der Varianzaufklärung der Faktoren eine 3-Faktorenlösung nahegelegt (der erste klärt 33.06%, der zweite 20.92% und der vierte Faktor 21.42% der Varianz auf). Die Faktorenanalyse der Daten von 1977 hingegen sprechen eindeutig für eine 2-Faktorenlösung (der erste klärt 56.1% der zweite Faktor 26.3% der Varianz auf), wobei jedoch abweichend von Supprian (1977) die Restfaktoren aufgrund der rechtwinkligen Rotationsmethode nicht im Sinne „eines zweiten Antriebsfaktors“ (Supprian 1977) bzw. „zweiten Stimmungsfaktors“ interpretierbar sind.

Eine kritische Durchsicht aktueller Arbeiten zur Zeitreihenanalyse (z.B. McCleary und Hay (1981), Box und Jenkins (1976)) lassen Zweifel aufkommen, ob das Verfahren der *p*-Faktorenanalyse zur Validierung eines Instrumentes geeignet ist, das vorgibt, Prozesse zu erfassen. Ausgangspunkt der Faktorenanalyse ist nach Bortz (1977) eine Korrelationsmatrix, im Falle serieller Messungen die Matrix der Kreuzkorrelationen der Items. Bei der Analyse von Kreuzkorrelationen ist nach McCleary und Hay (1981), Jenkins und Watts (1968) die Möglichkeit einer Überlagerung der Kreuzkorrelationen durch Autokorrelationen in Rechnung zu stellen. Nach Jenkins und Watts (1968) führt diese Überlagerung in der Regel zu einer künstlichen Erhöhung der Kreuzkorrelationen, und somit im Falle einer Faktorenanalyse zu einer artifiziellen Reduktion der Anzahl der Faktoren. Jenkins und Watts (1968) empfehlen daher, Kreuzkorrelationen lediglich anhand gefilterter Daten zu berechnen. Wie die Reanalyse der Daten Supprian (1977) zeigt, wurden entsprechende Filterungsmaßnahmen bei Supprian (1977) nicht durchgeführt (Richter 1982).

### Eigene Validierungsversuche

Bereits die von Supprian (1977, 1978) durchgeführten Faktorenanalysen lassen Zweifel an der Replizierbarkeit einer 2-Faktorenlösung mit den Faktoren Stimmung und Antrieb aufkommen. Diese Zweifel wurden durch die Faktorenanalyse eigener Daten bei einer Patientin mit schizoaffectiver Psychose erhärtet, bei der sich eindeutig eine 1-Faktorenlösung ergab.

Der Versuch, ESTA III mithilfe von *p*-Faktorenanalysen zu validieren, wurde von uns aus den oben genannten Grün-

den nicht weiter verfolgt. Lohnender erschien es uns, folgende Zugänge zur Validierung des Instruments zu wählen:

- Vergleich des ESTA III mit einem Instrument mit vergleichbarem Validitätsanspruch (konkurrierende Validität) und
- Genauigkeit in der Abbildung der im chronopathologischen Modell explizierten Prozesse (Konstruktvalidität).

### Untersuchungsvorgehen

In Übereinstimmung mit Supprian (1975a–d, 1977, 1978) wurde ein einzelfallanalytischer Untersuchungsplan verwendet. Zur Überprüfung der Validität der Eppendorfer Stimmungs-Antriebs-Skala wurden Einzelfallstudien an drei depressiven Patienten durchgeführt (zwei Patienten mit endogener Depression, eine Patientin mit einer neurotisch überlagerten, chronifizierten endogenen Depression). Die Patienten wurden während der Dauer der Untersuchung stationär in der Psychiatrischen Universitätsklinik Heidelberg (in erster Linie medikamentös) behandelt (s. Richter 1982). Um mögliche Periodizitäten in Antrieb und Stimmung festzustellen, füllten die Patienten den ESTA III und nachfolgend zu nennende Skalen der Eigenschaftswörterliste (EWL-K / Janke und Debus 1978) dreimal täglich (9.00 h, 12.00 h, 16.30 h) über einen Zeitraum von etwa einem Monat aus<sup>1</sup>. Folgende Skalen der Eigenschaftswörterliste kamen dabei zur Anwendung: Depressivität, Desaktiviertheit, Müdigkeit, Gehobene Stimmung und Aktiviertheit<sup>2</sup>. Zu Beginn und am Ende der Untersuchung bearbeiteten die Patienten den BDI (Beck et al. 1961).

### Zur konkurrierenden Validität von ESTA III

#### Auswahl des konkurrierenden Instruments

Unter einem konkurrierenden Test versteht man ein Instrument mit einem Validitätsanspruch, der dem Gültigkeitsanspruch des zu überprüfenden Tests vergleichbar ist. Als konkurrierendes Instrument wurden folgende Skalen der Eigenschaftswörterliste (EWL-K / Janke und Debus 1978) herangezogen: Aktiviertheit, Desaktiviertheit, Müdigkeit, Gehobene Stimmung und Depressivität.

Bei der EWL-K handelt es sich wie bei ESTA III um einen Selbstbeurteilungsbogen, der verschiedene Dimensionen der momentanen Befindlichkeit abbildet. Von den insgesamt 14 Skalen der EWL-K wurden diejenigen 5 Skalen gewählt, die Aktiviertheits- und Stimmungsparameter erfassen. Die Validität der EWL-K ist in zahlreichen Untersuchungen überprüft worden, so auch bei Studien zum Nachweis von Tagesschwankungen (Janke und Debus 1978).

#### Zur Auswertungsmethode

Die Frage der konkurrierenden Validität von ESTA III wird anhand von Kreuzkorrelationen (mit lag 0, also ohne Zeitverzögerung) zwischen ESTA III und den oben genannten Skalen der EWL-K überprüft. Wie bereits oben dargelegt, ist bei seriellen Messungen mit einer Überlagerung der Kreuzkorrelationen durch Autokorrelationen zu rechnen, die durch geeignete Filterungsmaßnahmen (z.B. durch Arima-Modelle)

<sup>1</sup> Die Auswahl der Meßzeitpunkte ist bei Richter (1982) begründet.

<sup>2</sup> Die Auswahl der Skalen orientiert sich an Überlegungen zur hypostasierten konkurrierenden Validität des ESTA III.

ausgeschlossen werden kann. Filterungsmaßnahmen beseitigen sowohl vorgefundene Abhängigkeiten in den Daten (und damit auch Periodizitäten) als auch ermittelte Trends. Da es jedoch für die Beurteilung eines Testinstruments, das vorgibt, Prozesse abzubilden, wichtig ist zu wissen, inwieweit das Instrument sensibel für die Erfassung von Periodizitäten und Trends ist, kann im vorliegenden Fall die Frage der Übereinstimmung zwischen ESTA III und den Skalen der EWL-K nicht allein anhand der Kreuzkorrelationen zwischen gefilterten Meßwertreihen entschieden werden. Um die Übereinstimmung in Trend und Periodizität zu berücksichtigen, ist es daher sinnvoll, bei der Berechnung der Zusammenhänge zwischen den Instrumenten mehrstufig vorzugehen. In einem ersten Schritt werden die Korrelationen der Rohwerte berechnet. Anschließend werden die Abhängigkeitsmodelle der Meßwertreihen identifiziert (hier Arima-Modelle) und die Meßwertreihen gefiltert. In einem dritten Schritt werden die Kreuzkorrelationen zwischen den gefilterten Meßwertreihen berechnet. Da lediglich in den Rohwerten die Varianzquellen Trend und Periodizität wirksam sind, werden primär die Kreuzkorrelationen der Rohwerte interpretiert, jedoch auf dem Hintergrund der identifizierten Arima-Modelle und der Korrelationen zwischen den gefilterten Meßwertreihen.

## Ergebnisse

In den Tabellen 1 bis 3 werden die Korrelationen zwischen den Rohwerten, die Abhängigkeitsmodelle und die Zusammenhänge nach Filterung der Daten (n.F.) aufgeführt.

Wie Tabelle 1 zu entnehmen ist, sind sowohl in der Einzelfallstudie F als auch H die Korrelationen zwischen den Skalen der EWL-K und ESTA III überraschend niedrig. Die Korrelationen zwischen den Skalen der EWL-K und ESTA III unterscheiden sich kaum von Null und weisen z.T. Vorzeichen auf, die der Meßintention widersprechen (positive Korrelation zwischen der Skala Desaktiviertheit und Antrieb in der Einzelfallanalyse H). Die Skalen Aktiviertheit und Gehobene Stimmung wurden aufgrund der zu geringen Änderungssensibilität (Varianz von 0) in dieser Einzelfallanalyse wie

**Tabelle 1.** Zusammenhänge zwischen Antrieb und Stimmung (ESTA III) und Skalen der EWL-K. Einzelfallanalyse Frau F ( $N = 75$ )

	Antrieb	Stimmung	DEP	DESM	STD	Filtermodelle
Antrieb	1.0	0.53	-0.17	-0.18	1.64	Arima (0,0,0)
n.F.		0.43	-0.17	-0.18		
Stimmung		1.0	0.03	-0.18	1.57	Arima (0,1,1)
n.F.			0.06	-0.06		
DEP			1.0	0.05	0.50	Arima (0,0,0)
n.F.				0.05		
DESM				1.0	1.36	Arima (0,0,0)

N: Anzahl der Beobachtungen

Antrieb und Stimmung: Skalen von ESTA III (nach Definition von Supprian 1975a)

DEP: Deprimiertheit  
(Summenwert der gleichnamigen Skala der EWL-K)

DESM: Desaktiviertheit + Müdigkeit  
(Summenwert/Skalen der EWL-K)

n.F.: In dieser Zeile werden die Zusammenhänge nach Filterung durch entsprechende Arima-Modelle aufgeführt.

STD: Standardabweichung

**Tabelle 2.** Einzelfallanalyse Herr H ( $N = 93$ )

	Antrieb	Stimmung	DEP	DESM	STD	Filtermodelle
Antrieb	1.0	0.24	0.17	0.33	4.31	Arima (0,1,1)
n.F.			-0.16	-0.17		
Stimmung		1.0	-0.16	0.20	2.63	Arima (1,0,0)
n.F.			-0.14	-0.09		M.K.
DEP			1.0	0.31	1.03	Arima (0,1,1)
n.F.				0.02		
DESM				1.0	1.65	Arima (0,1,1)

Zur Tabellenbeschriftung s. Tabelle 1

M.K.: mit Konstante

**Tabelle 3.** Einzelfallanalyse Frau S ( $N = 72$ )

	Antrieb	Stimmung	AKT	GST	STD	Filtermodelle
Antrieb	1.0	0.24	0.08	0.25	1.31	Arima (0,0,0)
n.F.		0.06	0.07	-0.01		
Stimmung		1.0	-0.08	0.73	3.24	Arima (0,1,1)
n.F.			0.06	-0.17		
AKT			1.0	-0.02	0.63	Arima (0,0,0)
n.F.				-0.02		
GST				1.0	2.74	Arima (0,1,0)
						(0,0,1) <sup>8</sup>

Zur Tabellenbeschriftung s. Tabelle 1

AKT: Aktiviertheit

(Summenwert der gleichnamigen Skala der EWL-K)

GST: Gehobene Stimmung (Summenwert / EWL-K)

auch in der Einzelfallstudie von Herrn H nicht weiter berücksichtigt.

Ein Vergleich der Ergebnisse der Einzelfallanalysen gibt Hinweise auf mögliche Ursachen für die niedrigen Korrelationskoeffizienten. Vergleicht man die Zusammenhänge der Einzelfallstudien, so ergibt sich zwanglos folgende Reihenfolge entsprechend der Höhe der Korrelationskoeffizienten. Die geringste Übereinstimmung zwischen den Skalen der beiden Instrumente herrscht bei Frau F. Geringfügig höhere Korrelationen zwischen den Skalen findet man in der Einzelfallstudie von Herrn H. Die zweifelloso höchsten Koeffizienten finden sich in der Einzelfallstudie von Frau S. Diese Patientin erlebte einen erstaunlichen Heilungsprozeß, so daß sich bereits nach etwa einer Woche Bodeneffekte in den Skalen Desaktiviertheit und Deprimiertheit einstellten. Bei dieser Patientin wurden die Daten der ersten Woche daher getrennt ausgewertet (s. Richter 1982). Hier werden nur die Ergebnisse der übrigen vier Untersuchungswochen mitgeteilt.

Zur Interpretation dieser Ergebnisse ist ein Vergleich der Einzelfälle aufschlußreich. Die Patienten unterscheiden sich u.a. in der Diagnose. Bei Frau F wird die Diagnose „neurotisch überlagerte, chronifizierte endogene Depression“ gestellt. Herr H und Frau S leiden an „endogener Depression.“ Abweichungen in der Diagnose erklären nicht die unterschiedlichen Ergebnisse bei Frau S und Herrn H, auch nicht die oben skizzierte Reihenfolge in der Höhe der Korrelationskoeffizienten. Diese Reihenfolge hingegen spiegelt sich in den Verlaufsbildern der Erkrankungen wider, die durch Verlaufskurven der applizierten Skalen, die Angaben im

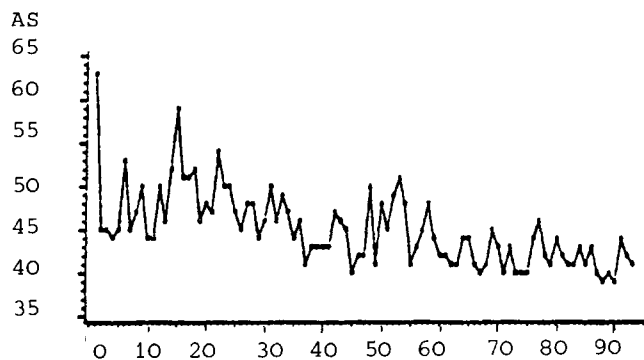


Abb. 1. Antrieb (ESTA III) Einzelfallanalyse H. Zeitachse (1.-93. Beobachtung); AS: Antrieb (Summenwert)

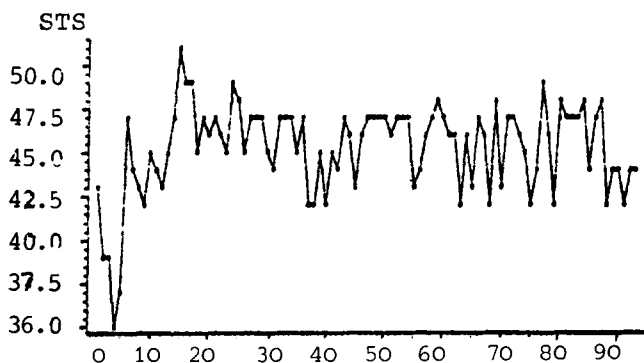


Abb. 2. Stimmung (ESTA III) Einzelfallanalyse H. Zeitachse (1.-93. Beobachtung); STS: Stimmung (Summenwert)

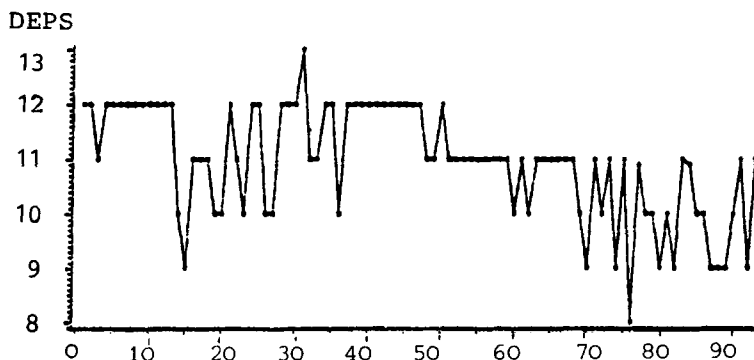


Abb. 3. Deprimiertheit (EWL-K) Einzelfallanalyse H. Zeitachse (1.-93. Beobachtung); DEPS: Deprimiertheit (Summenwert)

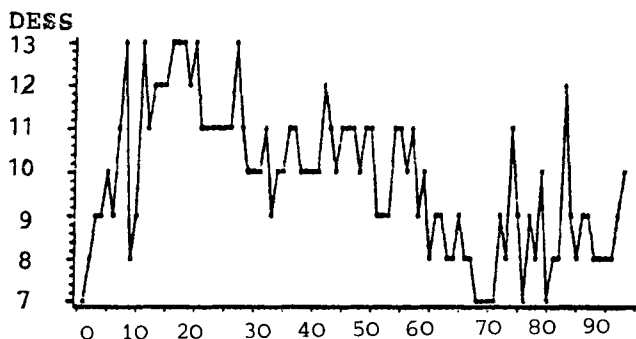


Abb. 4. Desaktiviertheit (EWL-K) Einzelfallanalyse H. Zeitachse (1.-93. Beobachtung); DESS: Desaktiviertheit (Summenwert)

BDI, die Dauer des stationären Aufenthaltes und die Entlassungsdiagnose gestützt werden. Während Frau F offensichtlich an einer chronischen Depression leidet, erfreut sich Frau S einer sehr raschen Genesung. Herr H hingegen zeigt das Bild eines sich nur langsam erholenden Patienten. Dabei darf jedoch nicht übersehen werden, daß sich bei Frau S die hohen Zusammenhänge lediglich in den Stimmungsskalen (Stimmung und Gehobene Stimmung) zeigen. In den Antriebsskalen (Antrieb und Aktiviertheit) liegen möglicherweise Deckeneffekte vor, da die Angaben in diesen Skalen mit vergleichsweise niedriger Varianz ( $v = 1.7$  bzw.  $v = 0.4$ ) um einen Maximalwert schwanken.

Obleich die Variabilität der Skalen in den Tabellen 1 bis 3 (in der Spalte STD) dokumentiert ist, wird sie zusätzlich beispielhaft anhand der Verlaufskurven der Einzelfallanalyse H in den Abbildungen 1 bis 4 dargestellt (sämtliche Verlaufskurven sind bei Richter (1982) aufgeführt).

Wie den Abbildungen 1 bis 4 zu entnehmen ist, weisen die Skalen eine Variabilität auf, die einen korrelationsstatistischen Vergleich der Instrumente rechtfertigen. Skalen, die keine oder nur eine sehr geringe Variation aufweisen (in den Einzelfallanalysen F und H die Skalen Aktiviertheit und Gehobene Stimmung), werden bei der konkurrierenden Validierung nicht berücksichtigt.

Was die Sensibilität der Instrumente in bezug auf die Abbildung von Periodizitäten anbelangt, so zeigt ein Vergleich der Tests eine weitgehende Übereinstimmung. Abgesehen von der Skala Gehobene Stimmung bei Frau S werden in den Antriebs- und Stimmungsskalen keine saisonalen (periodischen) Arima-Modelle identifiziert. Das Ausmaß der Parallelität der Verlaufskurven (beurteilt anhand der Höhe der Korrelationskoeffizienten) hingegen ist offensichtlich abhängig von der Stärke der auftretenden Veränderung. Je deutlicher sich gleichgerichtete Trends bemerkbar machen, um so eher sind hohe Zusammenhänge zu erwarten.

#### *Zur Validität bei der Vorhersage psychotischer Verläufe*

Supprian (1975d, 1977, 1978) verwendet ESTA III zur Überprüfung seines chronopathologischen Modells (Supprian 1975c, 1977). Ausgehend von dem Postulat einer Eigenrhythmicität von Antrieb und Stimmung bei endogenen Depressionen behauptet Supprian (1975d, 1977), daß sich aufgrund der Koinzidenzperiodik beider Systeme (dem Produkt der Perioden von Antrieb und Stimmung) Phasenbeginn und -ende bei Zylothymien berechnen lassen. Während Supprian (1977) Antrieb und Stimmung als „zwei cerebrale Funktionen“ und damit als biologische Größen definiert, operationalisiert er in seinen empirischen Studien zur Modellüberprüfung Antrieb und Stimmung mithilfe von ESTA III und damit mithilfe von psychischen Indikatoren. Angesichts der Kovariationsproblematik von psychologischen und physiologischen Indikatoren (s. Fahrenberg et al. 1979), ist eine Operationalisierung biologischer Konstrukte mithilfe psychologischer Indikatoren fragwürdig. Unsere Validierungsbemühungen von ESTA III auf der Grundlage des chronopathologischen Modells (Supprian 1975c, 1977) stützen sich daher auf die von Supprian (1975a, 1975b, 1975d, 1977, 1978) vorgenommene operationale Definition der Konstrukte. Antrieb und Stimmung werden dabei gemäß einer Skalenfestlegung von Supprian (1975a) definiert.

Wie bereits andernorts (Richter 1984) dargelegt, ist der Nachweis der Eigenrhythmicität eine notwendige Voraus-

setzung für die Berechnung von Phasenbeginn und -ende auf der Grundlage des chronopathologischen Modells von Supprian (1975c, 1977). Wie aus den unter der Überschrift Filtermodelle in den Tabellen 1 bis 3 aufgeführten Arima-Modellen ersichtlich, wird in keiner der Einzelfallanalysen ein saisonales Arima-Modell und damit eine Periodizität in den Variablen Antrieb und Stimmung identifiziert. Die zentrale Annahme einer Eigenrhythmicität konnte somit in den eigenen Einzelfallanalysen nicht bestätigt werden.

Eine daraufhin durchgeführte Reanalyse der Daten der Einzelfallstudie einer bipolar depressiven Patientin von Supprian (1977) konnte den von Supprian behaupteten Nachweis einer 6-Tage-Periodik von Stimmung und eines 7-Tage-Rhythmus von Antrieb nicht nachvollziehen. Wie andernorts (Richter 1982) dargestellt ist, sind die postulierten Periodizitäten auf Rechenfehler und der Verletzung von Modellvoraussetzungen (bestehende Trends wurden nicht beseitigt) zurückzuführen. In der Reanalyse der Daten (Supprian 1977) wurde für die Variable Antrieb eine Periodizität von vier Tagen ermittelt. Für die Variable Stimmung konnte lediglich ein reguläres, nicht saisonales Arima-Modell identifiziert und somit kein Rhythmus nachgewiesen werden. Gemäß der Definition von Supprian (1975c) ist in dem Falle die Koinzidenzperiodik, d.h. das Produkt der Rhythmen von Antrieb und Stimmung Null und daher eine Vorhersage von Phasenbeginn und -ende nicht möglich.

Da sich das von Supprian postulierte Modell auf „phasische Psychosen“ (Supprian 1977, S. 598), mithin auch auf den Verlauf monopolarer Depressionen bezieht, sind die eigenen Studien zur Beurteilung der Konstruktvalidität des ESTA III relevant. Wie die Reanalyse der Daten des Testautors (Supprian 1977) zeigt, läßt sich das chronopathologische Modell auch nicht anhand der Verlaufsdaten (erhoben mit Hilfe von ESTA III) einer bipolar depressiven Patientin bestätigen.

## Diskussion

Wie Lienert (1961) bemerkt, ist es nicht sinnvoll, von der Validität eines Testes zu sprechen, sondern nur von verschiedenen Zugängen zur Validität, die durchaus zu unterschiedlichen Ergebnissen führen, sich jedoch bei reflektierter Anwendung sinnvoll ergänzen können. Faktorenanalytische Validitätsüberprüfungen, wie sie vom Testautor durchgeführt wurden, erscheinen im vorliegenden Fall ungeeignet, da die Faktorenanalyse sich zur statistischen Abbildung von Prozessen nicht eignet (s.o.). Ein kritischer Vergleich der Faktorenlösung und der vom Testautor vorgenommenen Faktoreninterpretationen (Supprian 1977, 1978) weist darüber hinaus auf die fehlende Stabilität der postulierten faktoriellen Strukturen hin (s.o.).

Da das Instrument primär zur Erfassung von Verläufen konzipiert wurde, ist es ratsam, anstelle faktorenanalytischer Methoden folgende Vorgehensweise zur Validitätsprüfung einzuschlagen: Die eigenen Validierungsbemühungen orientieren sich an Fragen zum Testverständnis der Probanden, einem Vergleich von ESTA III mit Skalen der EWL-K und dem Versuch einer Konstruktvalidierung im Rahmen der Überprüfung notwendiger Annahmen eines chronopathologischen Modells des Testautors.

Wie die Exploration der befragten Patienten zeigte, sind einige Items schwer verständlich. Die vom Testautor angenommene Aufteilung der Skala in einen „manischen“ und einen „depressiven“ Pol (Supprian 1975a) wird nicht von allen

Patienten nachvollzogen. Bei allen befragten Patienten kommt es zu einer Verschiebung des „neutralen“ in Richtung des postulierten „manischen“ Pols. Einzelantworten der Befragten sind offensichtlich nur im Rahmen eines intraindividuellen Vergleichs interpretierbar.

Zur Überprüfung der konkurrierenden Validität von ESTA III wurden Skalen der Eigenschaftswörterliste (EWL-K / Janke und Debus 1978) herangezogen, eines Instrumentes, das in zahlreichen Untersuchungen zur Erfassung subjektiver Befindlichkeit bisher Verwendung fand und im Vergleich zur Eppendorfer Stimmungs-Antriebs-Skala sicherlich testtheoretisch gut dokumentiert ist. Ein Vergleich der untersuchten Einzelfälle legt die Annahme nahe, daß die Höhe der Korrelationen zwischen den Skalen der EWL-K und ESTA III dem Betrag nach offensichtlich von der Stärke der beobachteten Veränderungen abhängt. Fraglich ist, inwieweit ESTA III geeignet ist, auch minimale Veränderungen, wie sie im Rahmen eines primär chronischen Verlaufes auftreten können, abzubilden. Eine Aufteilung der Skala in 17 Stufen ist u.E. kein geeignetes Mittel, die Sensibilität eines Instrumentes für Verlaufsänderungen zu erhöhen. Bei einer derartig feinen Einteilung der Skala bei vergleichsweise geringer semantischer Differenzierung (z.B. äußerst stark/sehr stark) ist eher eine kognitive Überforderung der Probanden zu vermuten, wie z.B. die Untersuchung von Miller (1956) nahelegt. Wenngleich eine kognitive Überforderung bei unipolaren Adjektivskalen wie der EWL-K nicht zu befürchten ist, so zeigen auch diese Skalen unbestreitbare Schwächen in der Abbildung von Verläufen, wie sie sich in den Bodeneffekten in einer der Einzelfallanalysen dokumentieren.

Unter dem Begriff Konstruktvalidierung werden in der vorliegenden Arbeit Validierungsbemühungen verstanden, die auf der Basis des chronopathologischen Modells und einer Überprüfung zentraler Annahmen dieses Ansatzes durchgeführt wurden. Als zentrale Annahme des Modells wurde in einem anderen Zusammenhang (Richter 1984) das Postulat der Eigenrhythmicität von Antrieb und Stimmung abgeleitet. Auf die Problematik einer Diskrepanz zwischen biologischer Definition der Konstrukte Antrieb und Stimmung und ihrer psychologischen Operationalisierung wurde bereits oben im Zusammenhang mit dem Kovariationsproblem (Fahrenberg et al. 1979) verwiesen. Die Konstrukte Antrieb und Stimmung wurden hier wie auch bei Supprian (1975a–d, 1977, 1978) mithilfe von ESTA III operationalisiert. Die zentrale Annahme des Modells konnte mithilfe des zu validierenden Tests weder anhand eigener Daten noch anhand von veröffentlichten Daten des Testautors (Supprian 1977) bestätigt werden. Offensichtlich ist die Aussage des Testautors, er habe sein Modell bestätigen können (Supprian 1977), auf Rechenfehler und ein Mißverständnis zeitreihenanalytischer Methodologie zurückzuführen (Richter 1985). Da eine kritische Durchsicht weiterer Publikationen des Testautors ähnliche Fehler auch dort vermuten lassen, kann konstatiert werden, daß bislang kein methodenkritisch abgesicherter Beleg für die Validität von ESTA III bei der Vorhersage endogen depressiver Phasen erbracht wurde.

Wie Lienert (1961) ausführt, stehen Test und Konstrukt in Wechselwirkung miteinander. Fehlschlüsse bei einer Konstruktvalidierung stellen nicht nur den Test, sondern auch die Theorie in Frage. Auf die Diskrepanz zwischen theoretischer Konstruktdefinition und Operationalisierung wurde bereits im Zusammenhang mit dem Kovariationsproblem verwiesen. Widersprüche innerhalb des chronopathologischen Modells

sind bereits andernorts dargelegt (Richter 1982). Eine Diskussion dieser Widersprüche würde den Rahmen der vorliegenden Arbeit sprengen. Generell ist zu fragen, ob nicht der Anspruch, „ein Modell für den Gesamtablauf manisch-depressiver Psychosen“ (Supprian 1975c) zu formulieren, angesichts der vorliegenden Operationalisierung (ESTA III) überzogen ist.

## Literatur

- Beck AT (1976) Cognitive therapy and the emotional disorders. International University Press, New York
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiatry 4:53–63
- Bortz J (1977) Lehrbuch der Statistik für Sozialwissenschaftler. Springer, Berlin Heidelberg New York
- Box GEP, Jenkins GM (1976) Time series analysis: Forecasting and control. Holden Day, San Francisco
- Fahrenberg J, Walschburger P, Förster F, Myrtek M, Müller W (1979) Psychophysiologische Aktivierungsforschung. Ein Beitrag zu den Grundlagen der multivariaten Emotions- und Streß-Theorie. Minerva, München
- Janke W, Debus G (1978) Die Eigenschaftswörterliste (EWL-K). Ein Verfahren zur Erfassung der Befindlichkeit. Hogrefe, Göttingen
- Jenkins GM, Watts DG (1968) Spectral analysis and its applications. Holden Day, San Francisco
- Lienert GA (1961) Testaufbau und Testanalyse. Beltz, Weinheim
- McCleary RM, Hay RA (1981) Applied time series analysis for the social sciences. Sage Publications, London
- Miller GA (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol Rev 63:81–91
- Richter P (1982) Einzelfallanalysen depressiver Patienten – Vergleichende Betrachtung von Stimmungs- und Antriebsverläufen unter kritischer Verwendung von Arima-Modellansätzen. Unveröff Diplomarbeit, Heidelberg
- Richter P (1984) Zur Chronopathologie endogener Depressionen. Nervenarzt 55:154–161
- Richter P (1985) Berechnung des Phasenkalenders endogener Depressionen nach Supprian – ein statistisches Artefakt? Fortschr Neurol Psychiatr (im Druck)
- Supprian U (1975a) Eppendorfer Stimmungs-Antriebs-Skala ESTA. Ein Instrument zur Abbildung des Verlaufs manisch-depressiver Psychosen. Pharmakopsychiatrie 1:8–25
- Supprian U (1975b) Zur chronopathologischen Struktur der depressiven Tagesschwankung. Arch Psychiatr Nervenkr 220:9–22
- Supprian U (1975c) Ein theoretisches Modell für den Gesamtablauf manisch-depressiver Psychosen. Fortschr Neurol Psychiatr 43:358–379
- Supprian U (1975d) Über eine strenge Periodizität im Ablauf einer manisch-depressiven Psychose. Nervenarzt 46:440–446
- Supprian U (1976) Eppendorfer Stimmungs-Antriebs-Skala. Beltz, Weinheim
- Supprian U (1977) Verlaufsstudien an einer phasischen Psychose. Fortschr Neurol Psychiatr 45:596–614
- Supprian U (1978) Stimmung und Antrieb. Interne Desynchronisation einer zylothymen Tagesschwankung. Thieme Verlag, Stuttgart
- Zerssen D v (1976) Die Befindlichkeits-Skala. Parallelförmige Bf-S und Bf-S'. Beltz, Weinheim

Eingegangen am 28. August 1984/1. Februar 1985