EVALUATION DER DEUTSCHEN FASSUNG DER REVISION DES DIAGNOSTISCHEN INTERVIEWS FÜR BORDERLINE-PATIENTEN

DISSERTATION

zur Erlangung der Würde eines Doktors der Philosophie der Universität Hamburg

vorgelegt von

Dennis Fred Brodbeck

aus

Freiburg im Breisgau

Hamburg 2009

REFERENT: Prof. Dr. Jochen Eckert
Universität Hamburg

KOREFERENT: Prof. Dr. Reinhold Schwab
Universität Hamburg

Tag der mündlichen Prüfung: 23. Januar 2009

DANKSAGUNG

Prof. Dr. Jochen Eckert
für die fachliche und persönliche Unterstützung sowie das langjährige, vertrauensvolle Arbeitsverhältnis.

Dipl.-Psych. Eva-Maria Biermann-Ratjen,
Dr. Michael Schödlbauer,
Dr. Michael Wuchner,
Dipl.-Psych. Ruth Ladendorf
und den weiteren zeitweiligen Mitarbeitern der Arbeitsgruppe DIB-R-Evaluation für die Unterstützung u.a. bei der diagnostischen Untersuchung der einhundert Patienten, den vielen Stunden Video-Rating und der Diskussion vieler Fragen rund um das Projekt.

Dr. Peter Figge
vom Zentrum für Studienberatung der Universität Hamburg für die Unterstützung bei der Planung und Strukturierung, als es Hindernisse in der Fertigstellung der Arbeit zu überwinden und einen neuen Anfang zu machen galt.

Dr. Dr. Klaudia Odreitz für die große Hilfe und freundschaftliche Unterstützung in der Endphase und bei der Korrektur der vorliegenden Arbeit.

Mag. Lenka Alžběta Dušková, für den Zugang zu einer Kraft, der sich mir durch ihre Verbundenheit und ihr Interesse eröffnet und den Abschluss der Arbeit getragen hat.

Meiner Mutter Lonny Brodbeck für die Unterstützung des Projekts über die gesamte Laufzeit in so ziemlich jeder Hinsicht.

In der Rückschau ist für mich erkennbar:
Ein Projekt – sei es auch klein – ist niemals ohne die Unterstützung anderer Menschen zu schaffen. Um diese Hilfe zu erhalten ist es notwendig, aufrichtig und auch rechtzeitig um sie zu bitten. Stolz und seine Freunde helfen hier nicht weiter.

INHALTSVERZEICHNIS

TABELLENVERZEICHNIS

ABBILDUNGSVERZEICHNIS

ABKÜRZUNGSVERZEICHNIS

APA	American Psychiatric Association
BPS	Borderline-Persönlichkeitsstörung
DIB	Diagnostisches Interview für Borderline-Patienten
DIB-R	Revision des Diagnostischen Interview für Borderline-Patienten
DSM-III	Diagnostisches und Statistisches Manual Psychischer Störungen, 3. Ausgabe (APA, 1980)
DSM-III-R	Diagnostisches und Statistisches Manual Psychischer Störungen, Revision der 3. Ausgabe (APA, 1987, 1989)
DSM-IV	Diagnostisches und Statistisches Manual Psychischer Störungen, 4. Ausgabe (APA, 1994, 1996)
DSM-IV-TR	Diagnostisches und Statistisches Manual Psychischer Störungen, Textrevision der 4. Ausgabe (APA, 2000, 2003)
ICC	Intraclass-Korrelationskoeffizient (auch bezeichnet als Intraklassen-Korrelationskoeffizient)
ICD-9	Diagnosenschlüssel und Glossar psychiatrischer Krankheiten: Dtsch. Ausg. der internationalen Klassifikation der Krankheiten der WHO. 9. Revision, Kapitel V (Weltgesundheitsorganisation, 1980)
ICD-10	Internationale Klassifikation der Krankheiten 10. Revision (Weltgesundheitsorganisation, 2006)
n.b.	nicht beantwortbar
NNB	nicht näher bezeichnet
PS	Persönlichkeitsstörung
S.	im Zusammenhang mit dem DIB-R: "Statement" statt "Seite"
SKID	Strukturiertes Klinisches Interview für das DSM-III-R bzw. -IV
SKID-I	Strukturiertes Klinisches Interview für Achse I des DSM-III-R bzw. -IV
SKID-II	Strukturiertes Klinisches Interview für Achse II des DSM-III-R bzw. -IV
SPSS	Statistik- und Datenanalyse-Programm der SPSS Inc. bzw. SPSS GmbH
UKE	Universitätskrankenhaus Eppendorf
WHO	World Health Organization / Weltgesundheitsorganisation

ZUSAMMENFASSUNG

Bisher war für die deutsche Fassung der Revision des Diagnostischen Interviews für Borderline-Patienten (DIB-R) keine umfassende Evaluationsstudie durchgeführt worden. Mit der vorliegenden Studie wurden nun Validität und Reliabilität des DIB-R überprüft.

Validität

Die Validität des DIB-R wurde anhand von 100 zum größten Teil stationären Patienten der Klinik und Poliklinik für Psychiatrie und Psychotherapie des Universitätsklinikums Hamburg-Eppendorf untersucht. Mit sämtlichen Patienten wurde das DIB-R und zum Erhalt eines Außenkriteriums bzw. einer Vergleichsdiagnose auch eine komplette Diagnostik mittels des SKID-I- und des SKID-II-Interviews nach DSM-III-R bzw. -IV durchgeführt. Anhand der erhaltenen Daten konnten die konvergente und die diskriminante Validität des DIB-R überprüft werden.

Als zentrales Ergebnis ist festzuhalten, dass das DIB-R einen Validitätskoeffizienten von r_{tc}=.60 für die Übereinstimmung mit der Borderline-Diagnose nach SKID-II als Außenkriterium erhalten hat (konvergente Validität). Da dem SKID-II nicht das gleiche Konzept der Borderline-Persönlichkeitsstörung zugrunde liegt, ist dieser Koeffizient als gut zu bewerten. Weiter wurde die diskriminante Validität des DIB-R zunächst mittels der Korrelationen des Vorliegens der Borderline-Diagnose mit dem der übrigen Persönlichkeitsstörungen nach SKID-II überprüft. Mit keiner dieser Achse-II-Störungen wurde eine substanzielle Korrelation ≥.30 festgestellt, was ebenfalls positiv zu bewerten ist. Anhand von nach den DSM-III-R bzw. -IV-Diagnosen auf Achse I und II gebildeten Störungsgruppen konnte außerdem gezeigt werden, dass das DIB-R auch von diesen gut zu differenzieren vermag. Das DIB-R konnte auf allen relevanten Kennwerten zwischen diesen DSM-Diagnosegruppen differenzieren. "Verwechslungen" der Borderline-Persönlichkeitsstörung nach DIB-R mit anderen Achse-I- oder Achse-II-Störungen nach DSM-III-R bzw. -IV traten praktisch nicht auf.

Reliabilität

Die Reliabilitäts-Überprüfung wurde anhand von 19 videographierten DIB-R-Interviews, die von einer Gruppe mehrerer (sowohl in dessen Anwendung erfahrener als auch mit der Symptomatik von Borderline-Patienten vertrauter) Rater in Gruppensitzungen gesehen und unabhängig geratet wurden, vorgenommen. Zusätzlich wurden (bei gleichem Vorgehen) von vier in der Anwendung des DIB-R geschulten studentischen Ratern 12 videographierte DIB-R-Interviews gesehen. Aus

den so erhaltenen Interviews beider Ratergruppen wurden Interrater-Reliabilitäten mittels des Intraclass-Korrelationskoeffizienten (ICC) und des Finn-Koeffizienten berechnet.
Die zur Überprüfung der Reliabilität relevante Gruppe der Experten-Rater erreichte für die DIB-R-Diagnose mit einem ICC von .74 eine gute, im Gesamt-Score des DIB-R mit einem ICC von .91 sogar eine sehr gute Reliabilität. In den Bereichswerten des DIB-R (Skalierte Section-Scores) wurden mit ICC-Koeffizienten zwischen .57 und .89 überwiegend gute Reliabilitäten gefunden.
Die Studentengruppe erreichte insgesamt vergleichbare Kennwerte. Es besteht aber die Vermutung, dass ein konzeptgemäßes Einschätzen der DIB-R-Statements nicht immer sichergestellt werden konnte. Dieses Ergebnis zeigt auf, dass nach einem relativ kurzen Rater-Training durchaus reliable Einschätzungen zu erzielen sind, es für eine wirklich valide Einschätzung der Borderline-Symptomatik auch mittels des DIB-R aber einer soliden klinischen Erfahrung bedarf.

Insgesamt erscheint das DIB-R in der deutschen Fassung als ein reliables, valides und auch praktikables Instrument zur Stellung der Ein- bzw. Ausschlussdiagnose Borderline-Persönlichkeitsstörung.

1. DARLEGUNG DES FORSCHUNGSSTANDES

1.1. KURZER ABRISS DER GESCHICHTE DES BORDERLINE-KONZEPTS

1.1.1. Historische Entwicklungslinien des Borderline-Konzepts

Der Begriff der Borderline-Persönlichkeitsstörung geht historisch gesehen auf das Ende des 19. Jahrhunderts zurück, als der amerikanische Psychiater C.H. Hughes in eher allgemeiner Form den Begriff „Borderland of Insanity" einführte: *„The borderland of insanity is occupied by many persons who pass their whole life near that line, sometimes on one side, sometimes on the other."* (Hughes, 1884, zit. nach Grinker, 1977).

Als nosologischer Begriff wurde „Borderline" erstmalig 1938 von dem Psychoanalytiker Stern eingeführt und verwandt (Stern, 1938, zit. nach Wuchner, 1997, S. 1).

Nach Herpertz und Saß (2000) lassen sich im Wesentlichen vier Entwicklungslinien bis zum heutigen Konzept der Borderline-Persönlichkeitsstörung (mit BPS abgekürzt) ausmachen, die im Folgenden umrissen werden:

1.1.1.1. Die Borderline-Persönlichkeitsstörung als sub-schizophrene Störung

Die bereits bei Kraepelin (1904, zit. nach Herpertz & Saß, 2000, S. 115) zu findende Auffassung eines möglichen Zwischen- oder Übergangsbereichs zwischen abnormen Persönlichkeiten und Schizophrenien erlangt besondere Bedeutung im Konzept der *pseudoneurotischen Schizophrenie* von Hoch und Polatin (1949): Angenommen wird eine untypische Form der Schizophrenie, die durch spezifische Merkmale charakterisiert ist.

Hierzu gehören die primären Symptome Autismus, Ambivalenz, Störungen des Denkens und der Affekte sowie sekundäre Symptome, die multiple und variationsreiche klinische Symptome sind. „Pan neurosis" für multiple neurotische, „pan anxiety" für multiple Angstsymptome und auch „pan sexuality" für polymorph-perverse Sexualität sind zugehörige Begriffe. Weiter wurde die diagnostische Kategorie gekennzeichnet durch „mikro-psychotische Episoden" mit vorübergehenden Depersonalisations- und Derealisationserscheinungen, sowie durch Beziehungsideen und hypochondrische Befürchtungen.

Jene sekundären Symptome gingen später in psychoanalytische Konzeptionen zum *Borderline-Syndrom* ein, z.B. bei Grinker, Werble und Drye (1968) bei Kernberg (1967) und bei Rohde-Dachser (1979).

Heute lassen sich diese Strömungen vor allem als Vorläufer des DSM-III- bzw. DSM-IV-Konzepts zur Schizotypischen Persönlichkeitsstörung identifizieren. Jedoch fanden einige dieser

Symptome auch Eingang in die Diagnostik der eigentlichen Borderline-Persönlichkeitsstörung im Diagnostischen Interview für Borderline-Patienten (Gunderson, Kolb & Austin, 1981) und im DSM-IV (s.u.).

1.1.1.2. Die BPS als sub-affektive Störung

Eine zweite Entwicklungslinie geht historisch zurück auf die Beschreibung von Persönlichkeiten mit instabiler und rasch wechselnder Stimmungslage sowie Erregbarkeit. Klassische Beschreibungen hierzu finden sich z.B. bei Falret (1854, zit. nach Herpertz & Saß, 2000, S. 116) als „Folie circulaire". Hier werden in Fallbeschreibungen neben Patienten mit periodischen Stimmungsänderungen auch solche mit abrupten und unvorhersehbaren Stimmungswechseln zwischen Depression, Wut, Euphorie und Langeweile dargestellt. Zu den psychopathischen Zuständen rechnet auch Kraepelin ab der 5. Auflage seines Lehrbuches (1896, zit. nach Herpertz & Saß, 2000, S. 117) andauernde „konstitutionelle Verstimmungen", die dieser gleichsam als „Verdünnungsformen" der manisch-depressiven Erkrankung angesehen hat, oder als einen Grundzustand, aus dem heraus sich affektive Störungen entwickelten. In dieser Tradition sieht sich z.B. auch Akiskal (2000), der die Borderline-Persönlichkeitsstörung in einen biologischen Zusammenhang mit bipolaren affektiven Erkrankungen stellt.

Die beschriebenen, andauernd betrübten oder auch anhaltend erregten Verfassungen erinnern stark an die affektive Instabilität der Borderline-Persönlichkeitsstörung.

Schneiders Beschreibung der „stimmungslabilen" oder „explosiblen Psychopathen" besitzt eine große konzeptionelle Ähnlichkeit zur Borderline-Persönlichkeitsstörung (Schneider, 1923). Diese Personen werden beschrieben durch triebhafte Sucht nach Veränderung und Neuem sowie durch Stimmungsauslenkungen insbesondere reizbarer und depressiver Art. Aus unvermutet auftretenden mürrischen Stimmungen könnten dann auch Triebentladungen, z.B. in Form von Trinkexzessen, kriminellem und gewalttätigem Verhalten, folgen. Aus dieser Sichtweise werden Verluste der Impulskontrolle also auf Stimmungslagen zurückgeführt.

In einer Symptomatologie des typischen Borderline-Syndroms von Grinker, Werble und Drye (1968) werden entsprechend v.a. die affektiven Merkmale Ärger als Hauptaffekt, Störung emotional-affektiver Beziehungen, instabile Selbstidentität und Schwierigkeiten im zwischenmenschlichen Kontakt und in diesem Zusammenhang Einsamkeit und Depression hervorgehoben. Hierdurch wird deutlich eine affektive Symptomatik der Borderline-Patienten in den Vordergrund gestellt und eine Abgrenzung zum Ansatz der schizophrenienahen Erkrankung getroffen.

Aus heutiger Sicht führt diese Entwicklungslinie vor allem zur DSM-III und DSM-IV-Konzeption der Borderline-Persönlichkeitsstörung als auch zur im ICD-10 enthaltenen

„emotional-instabilen Persönlichkeitsstörung“. Auch in der DIB-Konzeption findet sich diese Linie in den verschiedenen Bereichen wieder.

1.1.1.3. Die BPS als Impulskontrollstörung

Von besonderer klinischer Bedeutung und Auffälligkeit ist die Neigung von Borderline-Patienten insbesondere zu selbst- aber auch zu fremddestruktiven Impulshandlungen. Entsprechend stehen im DSM-IV-Kriterien-Katalog für die Borderline-Persönlichkeitsstörung Merkmale einer Impulskontrollstörung im Vordergrund: allein drei der neun Kriterien einer Borderline-Persönlichkeitsstörung beschreiben selbst- und fremdschädigende Verhaltensweisen. Auch im Diagnostischen Interview für Borderline-Patienten ist, sowohl in der ursprünglichen Version als auch in der Revision, eine eigene Testsektion den Impulskontrollverlusten gewidmet.
In dieser Perspektive wird nicht den Affekten die Funktion als Auslöser für impulsives Verhalten zugeschrieben. Vielmehr wird Impulsivität über die reine Verhaltensebene hinaus als eine persönlichkeitseigene Tendenz, auf Reize plötzlich und heftig zu reagieren aufgefasst (Buss & Plonin, 1975). So betrachtet liegt es dann nahe, die affektive Instabilität bei Borderline-Patienten umgekehrt als Folge der eingeschränkten Impulskontrolle zu betrachten (siehe z.B. Herpertz et al., 1997, Herpertz & Saß, 1997).
(Die oben angesprochenen Typen von Psychopathen nach Schneider lassen sich entsprechend auch im Sinne einer Impulskontrollstörung verstehen, hier stünde dann die „explosible Persönlichkeitsstörung“ im Vordergrund.)

1.1.1.4. Die Borderline-Persönlichkeitsstörung als Posttraumatische Belastungsstörung

Ein erweitere Perspektive zur Erklärung und zum Verständnis der Borderline-Persönlichkeitsstörung ist die Interpretation als Posttraumatische Belastungsstörung.
Diese Überlegungen fußen zum Ersten auf der hohen Komorbidität von Borderline-Persönlichkeitsstörung und Posttraumatischer Belastungsstörung. So weisen bis zu 30% der Borderline-Patienten (nach DSM-III) auch die Symptome einer Posttraumatischen Belastungsstörung auf (Zimmerman & Coryell, 1989).
Ein weiterer Hinweis ist die konzeptuelle und phänomenologische Ähnlichkeit beider Störungen. So weisen die diagnostischen Kriterien *Instabilität des Selbstbildes* und *dissoziative Erlebnisse* aus dem DSM-IV auf relevante psychische Zustände nach gravierenden traumatischen Erlebnissen hin. Gunderson und Sabo (1993) geben jedoch an, dass diese Zustände bei Borderline-Patienten von kürzerer Dauer und geringerer Intensität sind.
Ebensolche Hinweise sind Berichten über die Bedeutung von Kindheitstraumata in der Geschichte von Borderline-Patienten (z.B. van der Kolk, 1999) zu entnehmen.

Vor einer vorschnellen und einseitigen kausalen Interpretation im Sinne einer Auslösung der Entwicklung einer Borderline-Persönlichkeitsstörung durch kindliche Traumata sei hingegen gewarnt. Denn erstens scheint die Borderline-Persönlichkeitsstörung zum Entwickeln einer Posttraumatischen Belastungsstörung zu disponieren, da die Bewältigung von potentiell traumatischen Erlebnissen erschwert ist (Gunderson & Sabo, 1993). Zum anderen ist das Vorliegen von Traumata für die Borderline-Persönlichkeitsstörung *weder spezifisch* – es finden sich auch Häufungen von Traumata bei anderen psychischen Erkrankungen (Zanarini et al., 1997) – *noch sensitiv* – es finden sich auch Borderline-Patienten ohne erkennbare traumatische Kindheitserlebnisse.

1.1.2. Aktuelle kriterienorientierte Definitionen

1.1.2.1. Die Borderline-Persönlichkeitsstörung im DSM-System

1.1.2.1.1. Diagnose-Kriterien

Die Borderline-Persönlichkeitsstörung ist als eingegrenztes und vor allem anhand operationaler Kriterien beschriebenes Konzept erstmals im DSM-III (1980) niedergelegt. Als Vorläufer der DSM-III-Diagnose können der im DSM-I vorkommende aggressive Typ der passiv-aggressiven Persönlichkeitsstörung sowie die Explosible/Epileptoide Persönlichkeitsstörung im DSM-II gelten.

Die Operationalisierung der Diagnose „Borderline-Persönlichkeitsstörung“ ist seit der Definition im DSM-III im Grundsatz unverändert geblieben, obwohl seit dem Erscheinen des DSM-IV einige Änderungen vorgenommen wurden (s.u.).

Herpertz und Saß (2000) gruppieren im "Handbuch der Borderlinestörungen" die 9 einzelnen Kriterien in **vier Bereiche**:

1) **affektive Instabilität**
2) **Impulshandlungen**
3) **Identitätsstörung**
4) **dissoziative bzw. (pseudo-)psychotische Symptome**

Im gleichen Buch komprimieren allerdings Clarkin und Dammann (2000) die Kriterien in lediglich drei Gruppen: Sie fassen die *dissoziativen und (pseudo-)psychotischen Phänomene* mit den *Identitätsstörungen* zum Bereich *„kognitive Probleme“* zusammen.

Der vierte Bereich verdient hier besondere Beachtung – denn erst in der aktuellen Version DSM-IV wurde das entsprechende Kriterium 9 – nach einer langen Validitätskontroverse bzgl. der Frage einer Verbesserung oder Verschlechterung der Validität durch Einbeziehung dieser

Symptome – in die Diagnose eingefügt (Gunderson & Zanarini, 1987; Widiger, Miele & Tilly, 1992).

Aus der Sicht der Forschergruppe um Gunderson und die DIB-Entwicklung wurde das Vorkommen kurzer, nicht-psychosetypischer paranoider Vorstellungen und Halluzinationen sowie dissoziativer Phänomene als pathognomisch für die Borderline-Persönlichkeitsstörung angesehen (Zanarini, Gunderson & Frankenburg, 1990) und nahm in DIB (und mehr noch im DIB-R) schon immer einen hohen Stellenwert ein. Andere Autoren hingegen bevorzugten die oben beschriebene historisch entstandene Trennung von Borderline-Persönlichkeitsstörung und Schizotypischer Persönlichkeitsstörung und wollten diese nicht wieder „aufweichen" (Serban, Conte & Plutchik, 1987).

Clarkin und Dammann (2000) führen als klare Grenzen des DSM-Systems allerdings auf, dass es keine Grundlage für eine Therapieplanung bereithalte und dass letztlich die rein kategoriale Diagnose keine Aussage über den Schweregrad der Störung (auch des subklinisch gestörten Patienten) bereithalte.

1.1.2.1.2. Diskussion der Änderungen der DSM-IV-Kriterien im Vgl. zum DSM-III-R

Die aktuellen diagnostischen Kriterien nach DSM-IV sind aus Abb. 1 zu ersehen.

Kriterien von Borderline-Persönlichkeitsstörungen (BPS) nach DSM-IV

1. Verzweifeltes Bemühen, tatsächliches oder vermutetes Verlassenwerden zu vermeiden.
Beachte: Hier werden keine suizidalen oder selbstverletzenden Handlungen berücksichtigt, die in Kriterium 5 enthalten sind.

2. Ein Muster instabiler, aber intensiver zwischenmenschlicher Beziehungen, das durch einen Wechsel zwischen den Extremen der Idealisierung und Entwertung gekennzeichnet ist.

3. Identitätsstörung: Ausgeprägte und andauernde Instabilität des Selbstbildes oder der Selbstwahrnehmung.

4. Impulsivität in mindestens zwei potentiell selbstschädigenden Bereichen [...].
Beachte: Hier werden keine suizidalen oder selbstverletzenden Handlungen berücksichtigt, die in Kriterium 5 enthalten sind.

5. Wiederholte suizidale Handlungen, Selbstmordandeutungen oder -drohungen oder Selbstverletzungsverhalten.

6. Affektive Instabilität infolge einer ausgeprägten Reaktivität der Stimmung (z.B. hochgradige episodische Dysphorie, Reizbarkeit oder Angst, [...]).

7. Chronische Gefühle von Leere.

8. Unangemessene, heftige Wut oder Schwierigkeiten, die Wut zu kontrollieren (z.B. häufige Wutausbrüche, andauernde Wut, wiederh. körperliche Auseinandersetzungen).

9. Vorübergehende, durch Belastungen ausgelöste paranoide Vorstellungen oder schwere dissoziative Symptome.

(Mindestens fünf der o.g. Kriterien müssen erfüllt sein. Aus DSM-IV, S. 739; leicht gekürzt, Auslassungen durch [...] gekennzeichnet)

Abb. 1: *Kriterien der BPS nach DSM-IV*

Ein Versuch, die Veränderungen von DSM-III-R zu DSM-IV übersichtlich zu gestalten wurde von Gunderson (2005) unternommen (s. Abb. 2) – leichte Ungenauigkeiten sind hier allerdings nicht zu vermeiden. Die tiefgreifendste Neuerung ist die Hinzufügung eines neuen *neunten Kriteriums „Vorübergehende, durch Belastungen ausgelöste paranoide Vorstellungen oder schwere dissoziative Symptome"*. Aus der Beibehaltung des bisherigen Schwellenwertes (fünf Kriterien müssen für eine Diagnosestellung uneingeschränkt erfüllt sein) folgt automatisch eine geringfügig veränderte Gewichtung in der Bedeutung der ursprünglichen acht Kriterien sowie eine Erleichterung der Diagnosestellung.

Außerdem wurden die Kriterien in der Formulierung und der Reihenfolge ein wenig verändert. So wurde z.B. im DSM-III-R noch bezüglich des Kriteriums *Identitätsstörung* eine „ausgeprägte und andauernde Identitätsstörung *in mind. zwei Lebensbereichen*" gefordert, während im DSM-IV lediglich eine „ausgeprägte und andauernde Instabilität des Selbstbildes *oder der Selbstwahrnehmung*" gefordert wird. In einem anderen Fall wird beim „*verzweifelten Bemühen des Patienten, Alleinsein zu verhindern*", dieses durch „*Verlassenwerden*" ersetzt.

Bei einem anderen Kriterium gilt nur noch ein „*chronisches Gefühl der Leere*" als borderlinetypisch, nicht aber mehr die „*chronische Langeweile*".

Es ist kaum zu ermessen, inwiefern sich durch die Vielzahl geringfügiger Formulierungsänderungen de facto auch Veränderungen in der Diagnosestellung ergeben haben, und welche dieser Veränderungen wirkliche Präzisierungen sind oder aber nur „Kosmetik" darstellen. Insgesamt scheint es im DSM-IV gelungen, eine hinreichend klare Definition des Störungsbildes festzulegen.

Mit der Veröffentlichung der Textrevision des DSM-IV-TR in deutscher Sprache im Jahr 2003 ergaben sich übrigens keine wesentlichen Veränderungen bzgl. der Diagnose einer Borderline-Persönlichkeitsstörung.

Veränderte Kriterien der BPS von DSM-III-R zu DSM-IV nach Gunderson (2005, S. 36, Tab. 1-1)*

1. Verzweifeltes Bemühen, tatsächliches oder vermutetes *Verlassenwerden* [Alleinsein] zu vermeiden.
(Beachte: Hier werden keine suizidalen oder selbstverletzenden Handlungen berücksichtigt, die in Kriterium 5 enthalten sind.)

2. Ein Muster instabiler, aber intensiver zwischenmenschlicher Beziehungen, das durch einen Wechsel zwischen den Extremen der [Über-] Idealisierung und Entwertung gekennzeichnet ist.

3. Identitätsstörung [Unsicherheit bzgl. mind. zwei der folgenden: Selbstbild, sexuelle Orientierung, Ziele oder Berufswahl, Typ von Freunden, allg. Werten]; ausgeprägte und andauernde Instabilität des *Selbstbildes* und/oder der *Selbstwahrnehmung (a).*

4. Impulsivität in mindestens zwei potentiell selbstschädigenden Bereichen (Geldausgeben, Sexualität, Substanzmissbrauch, rücksichtsloses Fahren, "Fressanfälle").
Beachte: Hier werden keine suizidalen oder selbstverletzenden Handlungen berücksichtigt, die in Kriterium 5 enthalten sind.

5. Wiederholtes suizidales Verhalten, Suizidandeutungen oder -drohungen oder Selbstverletzungsverhalten.

6. Affektive Instabilität [deutliche Schwankungen von ausgeglichener Grundstimmung zu Depression, Reizbarkeit oder Angst] *infolge einer ausgeprägten Reaktivität der Stimmung* (*z.B. hochgradige episodische Dysphorie,* Reizbarkeit oder Angst, wobei diese Verstimmungen gewöhnlich einige Stunden und nur selten mehr als einige Tage andauern).

7. Chronische Gefühle von Leere [oder Langeweile].

8. Unangemessene, heftige Wut oder Schwierigkeiten, die Wut zu kontrollieren
(z.B. häufige Wutausbrüche, andauernde Wut, wiederh. körperliche Auseinandersetzungen)

9. *Vorübergehende, durch Belastungen ausgelöste paranoide Vorstellungen oder dissoziative Symptome (b).*

Kursivschrift zeigt an, welche Textteile *nicht* in DSM-III-R, aber in DSM-IV vorkommen. [Text in Klammern] zeigt an, welche Teile in DSM-III-R, aber *nicht* in DSM-IV vorkommen.

Anmerkungen (Gunderson):
(a) z.B. er oder sie hat das Gefühl, dass er oder sie nicht existiert oder Schlimmes droht.
(b) oder Gefühle von Depersonalisation, Derealisation oder hypnagogenen Illusionen

** Die Reihenfolge der Kriterien wurde aus Gründen der Vergleichbarkeit – abweichend von Gunderson – dem Tab. 1 zugrunde liegenden DSM-IV angepasst. Die Änderungen in Krit. 1 waren von Gunderson offenbar vergessen worden und wurden vom Autor ergänzt. Darüber hinaus fallen gelegentlich minimale Abweichungen in der Übersetzung auf.*

Abb. 2: *Veränderungen der Kriterien der BPS von DSM-III-R zu DSM-IV nach Gunderson (2005)*

1.1.2.2. Die Borderline-Persönlichkeitsstörung im ICD-System

Im ICD-System findet die Borderline-Persönlichkeitsstörung erst in der aktuellen Version ICD-10 Eingang in Form des *Borderline-Typus* der *"emotional-instabilen Persönlichkeitsstörung" (F60.31)*. Ein ähnlicher Vorläufer ist im ICD-9 die Beschreibung des *"erregbaren Psychopathen"*.

Abweichungen zwischen ICD-10 und DSM-IV finden sich trotz konzeptueller Ähnlichkeiten im Bereich Impulsivität, die im ICD-10 im Mittelpunkt der Betrachtung steht und neben der Beschreibung konkreter impulsiver bzw. aggressiver Verhaltensweisen auch als Planlosigkeit und als Unfähigkeit, Belohungsaufschub zu tolerieren oder aversive Verhaltenskonsequenzen zu antizipieren, aufgefasst wird. Weiter fehlen die (pseudo-)psychotischen und dissoziativen Symptome gänzlich. Die Verlassenheitsängste, denen sowohl im DSM-IV als auch in den DIB-Versionen relativ hohes Gewicht verliehen wurde, finden sich nur in den erweiterten Forschungskriterien des ICD-10.

1.1.2.3. Die Borderline-Persönlichkeitsstörung als abgegrenztes Störungsbild

Aus den obigen Ausführungen wird ersichtlich, dass die Borderline-Persönlichkeitsstörung heute keine verwirrende Restkategorie psychiatrischer Störungen, sondern ein hinreichend abgrenzbares Störungsbild darstellt. Den Entwicklungen der Vergangenheit, in der die Borderline-Persönlichkeitsstörung oft sogar als „Mülleimer-Diagnose“ für schwierige, schlecht einzuordnende Patienten galt, wurde durch die relativ klare, wenn auch vielleicht noch zu offene Operationalisierung in den Systemen DSM-IV und ICD-10 sowie das Vorliegen relativ zuverlässiger Diagnoseverfahren entgegengewirkt.

1.2. DIAGNOSTIK DER BORDERLINE-PERSÖNLICHKEITSSTÖRUNG

1.2.1. Typen psychometrischer Instrumente

Hier sind zunächst der Übersicht halber einige allgemeine definitorische Unterscheidungen und Einordnungen zu treffen.

Vorschläge für Hauptkategorien zur Einteilung von psychometrischen Instrumenten hält z.B. Millon (1995) bereit.

Er unterscheidet:

a) Selbstbeurteilungsinstrumente
b) strukturierte klinische Interviews
c) persönlichkeitsorientierte Checklisten
d) projektive Tests

Bronisch (1999) führt hierzu aus, dass Fragebögen nur als Screening-Instrumente zur Voruntersuchung gut geeignet seien, da sie viele falsch-positive Diagnosen produzieren würden, sie also sensitiv, nicht aber spezifisch seien. Checklisten stellten verbunden mit einem freien Interview z.B. die Grundlage für die Einschätzung von Diagnosekriterien oder anderen Merkmalen dar. Weiter trifft er die Unterscheidung zwischen strukturierten und standardisierten Interviews, die andernorts auch der Unterscheidung zwischen "halb-standardisierten" und "voll-standardisierten" Interviews entspricht. Bei strukturierten Interviews bleibt es dem Interviewer vorbehalten, neben verbindlichen, im Interview vorgegebenen Fragen, selbstformulierte Fragen zu stellen, während dies bei standardisierten Verfahren nicht gestattet ist. Bronisch führt außerdem weiter aus, dass gerade die Entwicklung von strukturierten Diagnostikinstrumenten die Basis für die Erhöhung der diagnostischen Reliabilität von Persönlichkeitsstörungen war, keineswegs allein die Entwicklung und Verwendung operationalisierter Kriterien (Mellsop et al., 1982)!

Das Interesse der vorliegenden Untersuchung ist das Anliegen der Diagnosestellung: Im Brennpunkt des Interesses stehen somit kategoriale Diagnoseansätze, nicht etwa dimensionale Ausprägungen.

Zum einen gibt es Instrumente, die zur „simultanen“ Diagnose verschiedener Persönlichkeitsstörungen verwendet werden können (z.B. SKID-II). Zum anderen gibt es Verfahren, die nur zur Erfassung *einer* Persönlichkeitsstörung geeignet sind (z.B. DIB-R für die Borderline-Persönlichkeitsstörung). Außerdem existieren Instrumente, die nicht zur Diagnose der Borderline-Persönlichkeitsstörung im engeren Sinne konzipiert sind. Eine herausragende Rolle nimmt hier das „Strukturelle Interview“ nach Kernberg (s. Kernberg, 1996, Kernberg et al., 1981) zur Diagnose der Borderline-Persönlichkeits*struktur* ein. Es handelt sich um ein Instrument, welches Patienten nicht auf der deskriptiven Ebene ihres Verhaltens oder Erlebens unterscheidet, sondern auf der Ebene des Struktur-Niveaus, auf dem sie „funktionieren“.

Im Folgenden wird ein kurzer Überblick gegeben über gegenwärtig relevante Interviewverfahren, die zur Diagnosestellung der Borderline-Persönlichkeits*störung* im engeren Sinne geeignet sind.

1.2.2. Erste Instrumente zur strukturierten Erfassung der BPS

Als die erste systematische Untersuchung auf dem Gebiet der Borderline-Persönlichkeitsstörung kann die Studie von Grinker, Werble und Drye (1968) gelten. Die Basis für diese Studie bildeten deskriptive psychiatrische Ansätze (s.o.), sowie Untersuchungen, die testpsychologische Verfahren einschlossen (Rapaport, Gill & Schafer, 1946). Grinker, Werble und Drye (1968)

untersuchten 53 stationäre Patienten im Hinblick auf 93 Verhaltenskriterien und fanden hierbei vier Grundcharakteristika des „Borderline-Syndroms“:

- chronische Wut
- gestörte zwischenmenschliche Beziehungen
- Identitätsstörung
- Depression auf der Basis von Einsamkeitsgefühlen

Diese Untersuchung wurde trotz der noch ausstehenden Skalenbildung und Festlegung klarer Ein- und Ausschlusskriterien zu einem Meilenstein der Entwicklung des Borderline-Konzepts.
Trotz dieser Untersuchung bestand die Tendenz zur Einordnung des „Borderline-Zustands“ als Variante der Schizophrenie oder der affektiven Störungen fort, bis Gunderson und Singer (1975) eine Zusammenfassung der gesamten verfügbaren Literatur zum Thema vornahmen. Im gleichen Jahr erschien auch das Buch von Kernberg (1975). Mit diesen Arbeiten konnte die Borderline-Persönlichkeitsstörung als eigenständige nosologische Entität und beschreibbares Störungsbild wieder in den Vordergrund gerückt werden.
In der o.g. Arbeit stellten Gunderson und Singer zunächst eine Liste von sechs Kriterien zusammen, um die jetzt auch von ihnen so genannte „Borderline-Persönlichkeitsstörung“ zu charakterisieren und zu objektivieren:

- intensive Wut
- Impulsivität
- eine gestörte Identität hinter oberflächlicher Anpassung
- kurze psychotische Erfahrungen
- bizarre Antworten in projektiven Tests
- intensive und schwankende zwischenmenschliche Beziehungen, die von Idealisierungen und Entwertungen gekennzeichnet sind

Auf dieser Basis wurde von den Autoren das "Diagnostische Interview für Borderlinepatienten" entwickelt, welches 1976 erstmalig vor der Jahresversammlung der American Psychiatric Association vorgestellt und 1981 veröffentlicht wurde (Gunderson, Kolb & Austin, 1981). Es handelt sich hierbei also um das erste strukturierte diagnostische Interview zur Diagnose der Borderline-Persönlichkeitsstörung überhaupt. (Das Diagnostische Interview für Borderline-Patienten (DIB) und v.a. dessen grundlegende Revision DIB-R werden als Hauptgegenstand dieser Arbeit weiter unten vorgestellt.)

Das DIB stellt damit mehr dar als nur ein Diagnoseverfahren: Es ist eines der ersten systematischen Konzepte der Borderline-Persönlichkeitsstörung überhaupt, weswegen z.B.

Leichsenring (2003, S. 212 ff.) der "Borderline-Persönlichkeitsstörung nach Gunderson" sogar ein eigenes (dem zu DSM und ICD gleichrangiges) Kapitel widmet.

Die Weiterentwicklung spezifischer Kriterien wurde durch die Arbeitsgruppe um Gunderson stark befördert (s. Grinker & Werble, 1977; Grinker, 1979; Spitzer, Endicott & Gibbon, 1979). Die erste Aufnahme der so von Gunderson et al. geprägten Diagnose "Borderline-Persönlichkeitsstörung" in das DSM-III erfolgte dann im Jahr 1980 unter Einbeziehung vieler wesentlicher diagnostischer Aspekte aus dem DIB (s. Leichsenring, 2003, S. 16).

1.2.3. Interviews zur Diagnose von Persönlichkeitsstörungen nach dem DSM- bzw. ICD-System

Im deutschsprachigen Raum sind heute einige strukturierte diagnostische Interviews zur Diagnose von Persönlichkeitsstörungen (insbesondere Borderline-Persönlichkeitsstörung) verbreitet. Das Handbuch "Klinische Interviews und Ratingskalen" (Strauß & Schumacher, 2005) nennt drei Interviews, denen die psychiatrischen Definitionen von Störungsbildern nach dem DSM- bzw. ICD-Diagnose-System zugrunde liegen. Diese werden im Folgenden kurz vorgestellt.

1.2.3.1. International Personality Disorder Examination (IPDE)

Das IPDE ist zunächst durch Loranger in Zusammenarbeit mit der WHO in englischer Sprache entwickelt worden (s. Loranger, Janca & Sartorius, 1997). Die deutsche Übersetzung erfolgte durch eine Arbeitsgruppe um Mombour (Loranger & WHO, 1996).

Das Verfahren ist zur Diagnose von Persönlichkeitsstörungen nach ICD-10 und DSM-IV bei Erwachsenen einsetzbar. Die Diagnostik hat zum Einen kategoriale Ergebnisse. Es handelt sich um die Diagnose spezifischer Störungen beim Erreichen der jeweils geforderten Anzahl diagnostischer Kriterien. Zum Anderen erhält man mehrdimensionale Scores – hier werden die einzelnen Persönlichkeitsstörungen als Persönlichkeitsdimensionen aufgefasst und entsprechende Scores vergeben (allerdings nur, falls der entsprechende störungsspezifische Schwellenwert nicht erreicht wurde).

Das IPDE besteht aus 63 vorgegebenen Fragen und aus drei Verhaltensbeschreibungen, wobei frei zu stellende Zusatz- und Verständnisfragen erlaubt sind. Es handelt sich somit um ein halbstrukturiertes Interview.

Die Interviewdauer variiert in Abhängigkeit von Schwere und Anzahl der Störungen zwischen 1,5 und vier Stunden und kann auf mehrere Termine verteilt werden.

Als Anwendungsvoraussetzung nennen die Autoren psychiatrische bzw. psychologische Praxis und diagnostisches Differenzierungsvermögen insbesondere hinsichtlich Persönlichkeitsstörungen und lang andauernden anderen Störungen (Neurosen, Persönlichkeitsstörungen bei Psychosen) sowie Erfahrung in der Anwendung strukturierter Interviews. Außerdem fordern Sie ein Anwendertraining in Form von mind. 10 Interviews unter diagnostischer Supervision und empfehlen dies in Form eines Seminartrainings.

Das vollständige Interview besteht aus einem Screening-Fragebogen für den Patienten, einem Interviewer-Manual mit Interviewleitfaden sowie einem Protokollheft. Jeder Störung sind die entsprechenden Fragen und Kriterien zugeordnet, welche jeweils mit "2" = *voll ausgeprägt*, "1" = *unterschwellig ausgeprägt* oder "0" = *nicht vorhanden* sowie "?" = *fraglich* einzuschätzen sind. Diese Einschätzungen werden zur Auswertung dann in das Protokollheft übertragen.

Die Gütekriterien des IPDE bzgl. der Interrater-Reliabilität werden mit über .70 (meist über .80) für die meisten Kriterien angeben und wurden anhand von gemeinsam durchgeführten (Live-)Interviews zweier Psychiater bzw. Psychologen erhoben.

Bezüglich der Validität betonen die Autoren die Schwierigkeit, ein brauchbares Außenkriterium zu finden und führen hier lediglich eine (zwar nicht quantifizierte, aber durch die beteiligten Forscher festgestellte) Augenschein- bzw. Expertenvalidität an.

1.2.3.2. Diagnostisches Interview bei psychischen Störungen (DIPS)

Das DIPS wurde auf der Basis eines englischsprachigen Interviews für Angststörungen (Revidierte Fassung des Anxiety Disorders Interview Schedule, ADIS-R) als Erweiterung 1988 in Marburg von Margraf (s.u.) konstruiert. In kontinuierlichen Überarbeitungen wurde es an die jeweils aktualisierten DSM-Kriterien angepasst (Margraf, Schneider & Ehlers, 1994).

Das DIPS erzielt als Ergebnis eine kategoriale Diagnostik der für den psychotherapeutischen Bereich wichtigsten Störungen, es legt aber auch besonderen Wert auf die Sammlung für die Therapieplanung relevanter Informationen.

Die aktuelle DSM-IV-TR-Version (Schneider & Margraf, 2006) erfasst die folgenden Störungen: Angststörungen, gemischte Angst-Depressionsstörungen, Affektive Störungen, Somatoforme Störungen, Essstörungen, Schlafstörungen, Substanzmissbrauch und -abhängigkeit sowie die Borderline-Persönlichkeitsstörung als einzige aus dem Bereich der Persönlichkeitsstörungen.

Nach einem strukturierten Leitfaden werden alle für die DSM-Diagnose notwendigen Fragen erhoben. Die Fragen, insbesondere die durch Screening-Fragen gesteuerten Sprungregeln, sind genau vorgegeben. Ein vertiefendes Nachfragen und ggf. Konfrontieren des Patienten mit Widersprüchen ist erlaubt, es handelt sich demnach um ein halbstrukturiertes Interview. Am Ende jedes Störungsbereichs sind die genauen DSM-IV-Kriterien wiedergegeben. Neben dem

Interviewleitfaden liegt ein Protokollbogen zur Dokumentation und Kodierung der Patientenantworten vor. Die übliche Interviewdauer liegt bei ca. 60-90 Minuten. Die Auswertung erfolgt nach dem Interview anhand einer Kriterienchecklliste.

Vor einer Anwendung empfehlen die Autoren zur Sicherung hochwertiger Ergebnisse dringend eine ausführliche Beurteilerschulung anhand der im Testhandbuch gegebenen Trainings-, Durchführungs- und Auswertungshinweise.

Im Handbuch "Klinische Interviews und Ratingskalen" (Strauß & Schumacher, 2005) geben die Autoren Reliabilitäts-Kennwerte der ausführlicheren Forschungsversion F-DIPS für DSM-IV an. Sie berichten für die Oberklassen Angststörungen und Affektive Störungen für aktuelle und lifetime Diagnosen gute bis sehr gute Interrater-Reliabilitäten mit Kappa-Werten (κ) zwischen .64 und 1.0. Die Validität wurde anhand einer Reihe von Fragebögen aus international verbreiteten Standardverfahren überprüft: In jedem Fall ergaben sich die vorhergesagten Unterschiede zwischen den DIPS-Diagnosegruppen *Zieldiagnose*, *andere Diagnose* und *keine Diagnose*, daher bezeichnen die Autoren die Validität als "zufriedenstellend".

1.2.3.3. SKID-II

Das SKID-II-Interview wird an dieser Stelle überblicksweise vorgestellt. In der Anwendung muss es im Zusammenhang mit dem SKID-I gesehen werden, da eine integrierte Diagnosestellung beider Interviews und Achsen das Ziel ist (s. hierzu Kap. 3.4.2.2).

Die SKID-Interviews wurden nach den Achsen I und II des DSM-III zunächst in einer englischsprachigen Vorversion konstruiert, die 1984 veröffentlicht wurde (s. Spitzer, Endicott & Robins, 1984). Die Ursprungsversionen wurde 1990 (Spitzer et al., 1990a, 1990b) an das revidierte DSM-III-R und 1996 an das DSM-IV (First et al., 1996) angepasst.

Das SKID-II stellt dabei ein eigenständiges Interview dar, welches speziell zur Diagnose von Persönlichkeitsstörungen geschaffen wurde. Dem Interview vorgeschaltet ist ein sensitiver aber relativ unspezifischer Screening-Fragebogen, v.a. um bei der Diagnose Zeit einzusparen. Im eigentlichen Interview werden m.E. nur noch die Persönlichkeitsstörungen detailliert erhoben, bei denen nach den Antworten im Screening-Bogen eine Diagnose möglich ist. Zweifelhafte oder fehlende Antworten sowie Ja-Antworten in möglichen Störungsbereichen sollen überprüft werden.

Das SKID-II- Interview erfasst die zehn auf der Achse-II genannten Persönlichkeitsstörungen, die in der folgenden Tabelle, nebst der unten besprochenen Reliabilitäten, wiedergeben sind. Außerdem können noch die nur im Anhang des DSM-IV genannten Persönlichkeitsstörungen Negativistische und Depressive Persönlichkeitsstörung erhoben werden.

Tab. 1: *SKID-II-Persönlichkeitsstörungen mit jeweiligen Interrater-Reliabilitäten (Kappa-Koeffizienten)*

Studie / Sprache / DSM-Version	Maffei et al. (1997) engl. Vers.; DSM-IV	Fydrich et al. (1996) dtsch. Vers.; DSM-III-R
	κ	**κ***
Selbstunsichere Persönlichkeitsstörung	.971	.60
Dependente Persönlichkeitsstörung	.863	.81 [13]
Zwanghafte Persönlichkeitsstörung	.834	.82 [7]
Paranoide Persönlichkeitsstörung	.928	.55 [5]
Schizotypische Persönlichkeitsstörung	.912	-
Schizoide Persönlichkeitsstörung	.905	-
Histrionische Persönlichkeitsstörung	.916	.55
Narzisstische Persönlichkeitsstörung	.981	-
Borderline-Persönlichkeitsstörung	.909	.79 [3]
Antisoziale Persönlichkeitsstörung	.945	-

***** Eckige Klammern hinter dem κ bedeuten, dass hier eine Berechnung auf der Basis von Grundraten unter 10% erfolgte. Die Fallzahl der Diagnosen, die von wenigstens einem der Beurteiler gegeben wurde, ist dann in der eckigen Klammer angegeben. Bei einem Strich konnten kein Koeffizient berechnet werden.

Bezüglich der Reliabilität des SKID-II liegen für die Originalausgabe und die deutsche Übersetzung insgesamt befriedigende Ergebnisse vor. So berichten z.B. Maffei et al. (1997) für den englischsprachigen SKID-II für DSM-IV ausgezeichnete Interrater-Reliabilitäten von κ=.83 und höher. Für die deutsche SKID-II-Version für DSM-III-R berichten Fydrich et al. (1996) Interrater-Reliabilitäten in einem Bereich von κ=.55 bis hin zu κ=.82. Die Borderline-Persönlichkeitsstörung erreicht hier einen Wert von κ=.79, der allerdings auf einer geringen Fallzahl beruht (siehe Tab. 1). Der Median der Werte lag mit κ= .70 in einer guten Höhe.

1.2.4. DIB und DIB-R: Interviews speziell zur Diagnose der BPS

Als speziell zur Diagnose der Borderline-Persönlichkeitsstörung konzipierte Interview-Verfahren sind derzeit nur DIB und DIB-R bekannt und verbreitet. Im Folgenden wird, weitgehend übernommen aus einer zusammengefassten und überarbeiteten Veröffentlichung der Hamburger Arbeitsgruppe zur DIB-R-Evaluation ("Zur Revision des »Diagnostischen Interviews für Borderlinepatienten« (DIB)"; Schödlbauer et al., 1997), die aktuelle Version des DIB-R detailliert und (teilweise im Vergleich mit der Ursprungsversion DIB) beschrieben (die verwendete DIB-R-Version befindet sich im Anhang dieser Studie).

1.2.4.1. Die Revision des Diagnostischen Interviews für Borderlinepatienten

Die erste Fassung des "Diagnostic Interview for Borderline Patients" wurde 1976 auf dem 129. Kongress der American Psychiatric Association von Gunderson und Kolb (deutsche Fassung s. Gunderson, 1985) vorgestellt und später zweimal revidiert: 1982 (Zanarini et al., 1989) und geringfügig noch einmal 1992. Die zweite Revision 1992 (im Folgenden abgekürzt: DIB-R) liegt der deutschen Übersetzung von Rohde-Dachser et al. zugrunde (Rohde-Dachser, 1995, S. 225-237).

Der von Zanarini, die in den 80er-Jahren eng mit Gunderson zusammenarbeitete, genannte Grund zur Revision des DIB-R ist die Steigerung des Unterscheidungsvermögens des DIB zwischen Borderline-Patienten und Patienten mit anderen Persönlichkeitsstörungen (Zanarini, Frankenburg & Vujanovic, 2002).

Die schwierige Differenzierung der Borderline-Persönlichkeitsstörung von Achse-I-Störungen (insbesondere von Depressiven und Psychotischen Störungen) konnte laut Zanarini, Frankenburg und Vujanovic (2002) zwar bereits mit dem ursprünglichen DIB ausreichend demonstriert werden, war dabei aber unbedingt beizubehalten oder sogar auszubauen, da es viele Kontroversen in dieser Hinsicht gab und gibt. Zum Zusammenhang mit bzw. zur Differentialdiagnose von Affektiven Störungen siehe z.B. Akiskal (2000, S. 259 ff.).

1.2.4.1.1. Der Aufbau des Interviews

Im Anschluss an allgemeine Angaben zum Alter, Familienstand, Schulabschluss, Beruf etc. erhebt das halbstrukturierte Interview in 97 Fragen borderlinetypische Erlebens- und Verhaltensweisen, die in vier Bereiche gegliedert sind:

- **Affekte**
- **Kognition**
- **Impulshandlungen**
- **Zwischenmenschliche Beziehungen**

Die Fragen beziehen sich (bis auf wenige Ausnahmen) auf die vergangenen zwei Jahre. Wenn nicht anders angegeben, werden die Antworten des Patienten auf die Fragen kodiert mit:

"2" = *Ja*

"l" = *Wahrscheinlich*

"0" = *Nein*

n.b. = *nicht beantwortbar*

Die Informationen aus der Beantwortung der Einzelfragen gehen in insgesamt 22 Statements (im Folgenden abgekürzt: S.) ein, die in der gleichen Weise kodiert werden. Die Statements jedes

Bereichs werden anschließend zum Section-Score (z.T. auch als Summenwert bezeichnet) aufsummiert. Dieser Punktwert wird nach vorgegebenen Regeln in einen Skalierten Section-Score (z.T. auch als Bereichswert bezeichnet) transformiert. Dabei können in den Bereichen Affektivität und Kognition 0, l oder 2 Punkte, in den Bereichen Impulsivität und Zwischenmenschliche Beziehungen 0, 2 oder 3 Punkte erreicht werden. Die letzten beiden Fragenkomplexe werden damit stärker gewichtet. Die vier Skalierten Section-Scores werden anschließend zum Gesamt-Score (z.T. auch als Gesamtwert bezeichnet) addiert. Liegt dieser zwischen 8 und 10 Punkten, so gilt dies als hinreichender Hinweis auf das Vorliegen einer Borderline-Persönlichkeitsstörung und zur Stellung einer solchen Diagnose. Ein Punktwert von ≤7 spricht dagegen. Ergeben die Section-Scores der beiden ersten Bereiche – Affekte und Kognition – zusammengerechnet nicht mehr als l Punkt, so kann das Interview theoretisch bereits abgebrochen werden, da dann das Erreichen eines Gesamt-Scores von 8 Punkten und somit eine Diagnosestellung ausgeschlossen ist.

Affekte

Die Fragen aus dem Bereich Affektivität erfassen Depressivität (S. 1: Chronische depressive Verstimmung oder Episoden einer Major Depression), andauernde Gefühle von Hilflosigkeit, Hoffnungslosigkeit, Wertlosigkeit oder Schuld (S. 2). Das Erleben chronischer Gefühle von Ärger, Wut beziehungsweise entsprechende Verhaltensweisen (S. 3), chronische Ängstlichkeit oder körperliche Äquivalente von Angst wie Kopfschmerzen oder Herzklopfen (S. 4) sowie dysphorische Affekte (S. 5).

Anhand der Fragen zu dysphorischen Stimmungen lässt sich zeigen, wie die Einzelitems in die Einschätzung des Statements eingehen:

(20.) Haben Sie sich während der letzten zwei Jahre häufig sehr einsam gefühlt? (2, l, 0)

(21.) Haben Sie sich gelangweilt? (2, l, 0)

(22.) Oder innerlich leer gefühlt? (2, l, 0)

(23.) **S. 5. Der Patient erlebte chronische Gefühle von Einsamkeit, Langeweile oder Leere. (2, l, 0)**

Die »oder«-Verbindung im Statement 5 weist darauf hin, dass eines der drei Gefühle – sofern es chronisch vorhanden ist – genügt, um das Kriterium zu erfüllen: In diesem Fall wird es mit »2« kodiert. Sind die Fragen 20-22 nur wahrscheinlich zutreffend, so wird auch S. 5 max. mit »l« eingeschätzt.

Die Exploration des Bereichs der Affektivität wird mit Fragen nach einem anderen Merkmal, nämlich nach hypomanischen oder manischen Episoden abgeschlossen. Da mehrere hypomanische Episoden von den Autoren als Indiz gegen eine Borderlinestörung gewertet werden, wird bei ihrem Vorliegen der Skalierte Affekt-Section-Score auf »0« gesetzt.

Kognition

Alle Fragen nach Eigentümlichkeiten des Denkens und der Wahrnehmung betreffen Erlebensweisen, die nicht auf den Einfluss von Drogen zurückgehen. Als borderlinetypisch gelten Aberglaube, Magisches Denken, Sechster Sinn, Telepathie, bestimmte Arten von Sinnestäuschungen sowie Depersonalisations- und Derealisationserleben (S. 6).

Derealisationserleben ist wie folgt operational definiert:

(35.) Hatten Sie oft das Gefühl, als seien die Dinge um Sie herum unwirklich, so als ob sie fremd wären oder ihre Größe oder Form veränderten? Als ob Sie in einem Traum wären? Als ob etwas wie eine Glasscheibe zwischen Ihnen und der Welt wäre? (Derealisation) (2, 1, 0)

Übertriebenes Misstrauen, flüchtige Beziehungsideen und vorübergehende nicht-wahnhafte Vorstellungen von Verfolgtwerden fließen in das 7. Statement ein.

Für die anschließenden Fragen zu Pseudopsychotischen Symptomen (S. 8) gilt eine andere Kodierung der Antworten: Echt psychotische Wahnvorstellungen und Halluzinationen werden mit »2«, pseudopsychotische mit »l« kodiert. Pseudopsychotische Symptome unterscheiden sich von psychotischen dadurch, dass sie zeitlich flüchtig, umschrieben, für Psychosen atypisch und nicht systematisiert sind. Während pseudopsychotische Symptome borderlinetypisch sind, spricht eine früher aufgetretene psychotische Episode, aber auch eine ausgeprägte manische Episode im bisherigen Leben, gegen eine Störung dieses Typs. In diesen Fällen beträgt der Skalierte Kognitions-Section-Score »0« Punkte.

In diesen differentialdiagnostischen Anweisungen spiegelt sich zum einen das ursprüngliche Konzept der Borderline-Persönlichkeitsstörung wieder: Man ging davon aus, dass die Borderlineerkrankung als Grenzfall zwischen klassischen Neurosen und psychotischen Erkrankungen anzusehen sei. Die besondere Berücksichtigung maniformer Phasen ist im Rahmen der Diskussion der Frage zu verstehen, ob es eine Gruppe von Patienten gibt, die deskriptiv betrachtet zwar borderlinetypisch ist, aber eigentlich den Affektiven Störungen zuzurechnen wäre. Bei diesen Patienten soll etwa eine Medikation mit Lithium nicht nur die affektiven Symptome, sondern auch die Impulsivität reduzieren (Davison & Neal, 1996, S. 301-302).

Impulshandlungen

Im Abschnitt zur »Impulsivität« wird in der Regel die Häufigkeit entsprechender Verhaltensmuster kodiert (2 = 5mal oder mehr; 1 = 3 bis 4mal: 0 = 2mal oder weniger). Symptome, die nur »wahrscheinlich« vorhanden sind, werden damit nicht gewertet. Inhaltlich werden erfasst: Ernsthafter Drogenmissbrauch (S. 9) sowie sexuelle Deviation in Form promiskuitiver oder perverser Tendenzen (S. 10). In den Statements 11 und 12 spricht man

bereits von einem Handlungsmuster, wenn das entsprechende Verhalten 2mal oder häufiger aufgetreten ist:

(67.) Haben Sie sich absichtlich selbst verletzt, ohne dass Sie sich damit umbringen wollten (z. B. sich geschnitten, verbrannt oder geschlagen, Ihre Hand durch die Scheibe gestoßen, gegen die Wand geschlagen, Ihren Kopf angeschlagen)?
(Selbstbeschädigung) (2 = 2mal oder öfter, l = l mal. 0 = niemals) (2, 1, 0)

(68.) **S. 11. Der Patient zeigte ein Muster von körperlicher Selbstschädigung. (2, 1, 0)**

Ein sensibles Vorgehen erfordert die Klärung der Frage, ob der Patient zu manipulativen Suizid-drohungen, -versuchen oder -gesten neigt (S. 12). Die zugehörige Frage (Nr. 69) ist in ihrer Direktheit nicht unproblematisch. Sie wirkt leicht als Unterstellung und birgt die Gefahr, dass der Eindruck entsteht, der Interviewer nähme die Suizidalität nicht ernst.

Eine Liste weiterer impulsiver Handlungsmuster schließt diesen Bereich ab. Hier genügt bereits ein ausgeprägt vorhandener Handlungstyp, um im Statement 13 eine »2« zu raten. Als Beispiele sind zu nennen: Fressanfälle, Kaufrausch, lautstarke oder körperliche Auseinandersetzungen, antisoziale Handlungen.

Zwischenmenschliche Beziehungen:

Die Fragen zu den interpersonalen Beziehungen des Patienten nehmen den breitesten Raum ein. Als borderlinetypisch gilt der Versuch, Alleinsein unbedingt zu vermeiden (z.B. durch stundenlanges Telefonieren) oder auf Alleinsein extrem dysphorisch zu reagieren (S. 14). Verlassenheitsängste, bis hin zur Befürchtung, im Falle des Verlassenwerdens völlig zerstört zu werden, beziehungsweise das Gefühl, bei zu großer Nähe vom anderen verschlungen zu werden, gelten ebenfalls als borderlinespezifisch (S. 15).

Bei den Items, die nach der Abwehr von **Abhängigkeitswünschen** beziehungsweise nach Ambivalenzen bezüglich des Versorgens und Versorgtwerdens (S. 16) fragen, ist die Einschätzung des Statements schwierig:

(94.) Waren Sie irgendwo beschäftigt, wo eine Ihrer Hauptaufgaben darin bestand, sich um andere Menschen oder Tiere zu kümmern? (2, l, 0)

(95.) Haben Sie Freunden, Verwandten oder Kollegen ständig Hilfe angeboten? (2, 1, 0)

(96.) Hat es Sie in den vergangenen zwei Jahren besonders gestört, wenn andere Menschen Ihnen helfen wollten oder sich um Sie zu kümmern versuchten? (2, 1, 0)

Zur Einschätzung des diagnostischen Merkmals müssen unter Umständen die Antworten auf mehrere der Fragen berücksichtigt werden. Eine Abwehr im Sinne von S. 16 liegt etwa nur bei einer »2« in Frage 95 *und* 96 vor.

Die Beziehungen von Borderline-Patienten sind typischerweise charakterisiert durch die Tendenz zu intensiven, aber instabilen engen Beziehungen (S. 17), chronischen Problemen mit

Abhängigkeit oder Masochismus (S. 18), Abwertung, Manipulation anderer Personen oder Sadismus (S. 19) sowie durch forderndes und ansprüchliches Verhalten (S. 20).
Das diagnostische Interview schließt mit Fragen nach Komplikationen in therapeutischen Beziehungen, nämlich deutlich regressiven Tendenzen während einer stationären oder ambulanten Therapie (S. 21) sowie typischen Gegenübertragungsproblemen während der Behandlung (z.B. Spaltung von Behandlungsteams, Freundschaften oder Affären mit einem Behandler (vgl. S. 22). Wie im Bereich der Impulsivität kann der Skalierte Beziehungs-Section-Score »0«, »2« oder »3« Punkte betragen.

1.2.4.1.2. Die Änderungen in den Revisionen des DIB

Eine Vielzahl von Fragen wurde ausgetauscht, umformuliert und teilweise in andere Bereiche eingegliedert. Die wichtigsten Änderungen werden hier erwähnt:

Formale Änderungen

Handhabung: Während die Anordnung der Statements in der ersten DIB-Version uneinheitlich war, werden die Statements in den Revisionen direkt im Anschluss an die einschlägigen Fragen geratet, womit lästiges Blättern entfällt.
Der Interviewer wurde im DIB durch viele Leerzeilen für Bemerkungen und für die nähere Beschreibung der Symptome (z.B. Häufigkeit und Inhalte pseudopsychotischer Phänomene) zu einer weiterführenden Exploration angehalten, was in den Revisionen nicht mehr der Fall ist.
Trennwert: Während man im DIB ab einem Gesamt-Score von sieben Punkten vom Vorliegen einer Borderlinestörung ausging, wurde der Trennwert in den Revisionen auf acht Punkte angehoben. Damit ist die Diagnose »strenger« geworden.
Referenzzeitraum: Die Fragen bezogen sich im DIB auf unterschiedliche Zeiträume (die vergangenen drei Monate bis hin zu den letzten drei Jahren); dagegen gilt in den Revisionen ein einheitlicher Referenzzeitraum von zwei Jahren.

Inhaltliche Änderungen

Anzahl der Bereiche: Ein im ursprünglichen DIB enthaltener fünfter Bereich (Soziale Anpassung) wurde in den Revisionen fallengelassen. Dieser Bereich hatte sich auch in der Validierung der deutschen DIB-Fassung als nicht ausreichend trennscharf erwiesen (Eckert et al., 1991). Die Anzahl der Statements ist im revidierten DIB um sieben reduziert worden. Durch diese Änderungen wurde die Durchführungszeit etwas verkürzt.

Affekte: Die Einschätzung von Affekten, die man *während* des Interviews beim Patienten beobachten konnte, entfällt in den Revisionen. Auf ausführliche Fragen zur Major Depression im

DIB wurde in den Revisionen (leider) verzichtet (Fragen nach Gewichtsveränderung, Schlafstörungen, Suizidgedanken etc.). Die Frage nach somatischen Angstäquivalenten kommt in den Revisionen hinzu.

Kognition: In den Revisionen sind einige Fragen zu Magischem Denken, Telepathie etc. hinzugekommen (vgl. S. 6). Psychotische Erlebnisse unter Drogeneinwirkung gehen in den Revisionen aber nicht mehr in die Wertung des Bereichs ein!

Impulshandlungen: Die Kriterien »Homosexualität« und »Inzest« fehlen in der zweiten Revision.

Zwischenmenschliche Beziehungen: Der Beurteiler wird in den Revisionen nicht mehr dazu aufgefordert, seine eigenen Gegenübertragungsgefühle, die der Patient in der Interviewsituation auslöst, diagnostisch zu verwerten.

Differentialdiagnostische Merkmale: Da man annahm, dass hypomanische Episoden gegen eine Borderlinestörung sprechen, war bei ihrem Vorliegen im DIB ein Punktabzug bei den Rohwerten im Bereich Affektivität vorgesehen. In den beiden Revisionen wurden die Fragen zur Erfassung einer fraglichen Hypomanie deutlich erweitert und dem DSM-III-R angeglichen. Die Revisionen begnügen sich aber nicht mit einem Punktabzug, sondern im Falle wiederholter hypomanischer Phasen wird der Bereich Affekte auf »0« Punkte gesetzt. Das gleiche gilt für den gesamten Skalierten Section-Score Kognition, wenn jemals eine Manie oder eine »echte«, verzweigte Psychose längerer Dauer aufgetreten ist.
Zur besseren Abgrenzung von psychotischen Erkrankungen wurden im DIB flache Affektivität beziehungsweise soziale Isolation ebenfalls mit Punktabschlägen gewertet (in den Bereichen Affektivität bzw. Beziehungen). In den Revisionen wird der gesamte interpersonale Section-Score auf »0« gesetzt, wenn der Patient sozial isoliert ist und im Interview ein seltsames Sozialverhalten zeigt.

Andere Merkmale: Die erste Revision führte noch 9 weitere Merkmale auf, die *nicht* verbindlich in die zahlenmäßige Wertung eingingen, aber entweder differentialdiagnostisch relevant erschienen (zum Beispiel *Psychotische Sprache*) oder als zumindest als borderlinetypisch galten und weiter beobachtet werden sollten (zum Beispiel *Affektive Instabilität*, *Idealisierung*, *Ernsthafte Identitätsstörung*). Diese Merkmale wurden in die zweite Revision aber *nicht* übernommen.

1.2.4.2. Evaluations-Studien zum DIB-R

Zum DIB-R liegen trotz seines Alters wenige international bekannte Studien zur Evaluation vor. Die beiden wesentlichen Studien wurden von Zanarini und Kollegen bzw. Mitarbeitern veröffentlicht.

1.2.4.2.1. Die Validitätsstudie von Zanarini et al. (1989)

Die erste Studie mit dem Titel *"The revised Diagnostic Interview for Borderlines: Discriminating BPD from other Axis II Disorders"* wurde mit der Vorstellung der Revision veröffentlicht (Zanarini et al., 1989).

Sie behandelt v.a. die Güte der differentialdiagnostischen Abgrenzung der Borderline-Persönlichkeitsstörung von anderen Achse-II-Störungen, die anhand einer Stichprobe von 237 Patienten (105 ambulant, 132 stationär) geprüft wurde. Jeder dieser Patienten hatte von seinem Behandler die Diagnose eine Persönlichkeitsstörung nach DSM-III erhalten, die dann als Referenzkriterium verwendet wurde, gegen das die DIB-R-Diagnose geprüft werden konnte. Insgesamt hatten nach diesem Vorgehen 95 Patienten die Diagnose *Borderline-Persönlichkeitsstörung* und 142 eine andere Achse-II-Diagnose erhalten. Außer dem erheblich größeren Frauenanteil (82,1% vs. 41,6%) in der Gruppe der Borderline-Persönlichkeitsstörung waren beide Gruppen in keiner relevanten Variable signifikant unterschiedlich (Alter, sozioökonomischer Status, Rasse).

Als erstes für die vorliegende Arbeit relevantes Ergebnis ist anzugeben, dass Patienten mit Borderline-Persönlichkeitsstörung bei einem Mittelwert von 8,52 wie erwartet deutlich höhere Werte im Gesamt-Score erhielten, als Patienten mit anderer Achse-II-Diagnose (Mittelwert 5,89, der Test auf Mittelwertsunterschiede wird mit $p < .0001$ als hochsignifikant angegeben).

Außerdem werden weitere Kennwerte zur Einschätzung der Testgüte angegeben. Für verschiedene Cutoff-Werte durchgespielt werden anhand der Stichprobe die Konstellationen von Spezifität und Sensitivität sowie dem Positiven (PPW) und dem Negativen Prädiktiven Wert (NPW, beide prävalenz-adjustiert) angegeben. Beim vorgeschlagenen Cutoff von 8 werden die im Zusammenhang insgesamt besten Kennwerte gefunden.

Tab. 2: *Cutoff-Werte des DIB-R, Evaluation der Originalversion von Zanarini et al. (1989)*

N=237	"6"	"7"	"8"	"9"	"10"	Wahrscheinlichkeit, dass bei einem Cutoff von ...
Sensitivität	.92	.88	**.82**	.70	.31	DIB-R-BPS, wenn klinisch Borderline-Persönlichkeitsstörung
Spezifität	.42	.55	**.80**	.86	.96	*nicht* DIB-R-BPS, wenn *nicht* klinisch BPS
PPW	.51	.57	**.74**	.77	.83	klinisch BPS, wenn DIB-R-Borderline-Persönlichkeitsstörung
NPW	.88	.88	**.87**	.81	.67	klinisch *nicht* BPS, wenn *nicht* DIB-R-BPS

Über dem Cutoff von 8 würden zunehmend Sensitivität und Negativer Prädiktiver Wert geopfert, während unter dem Cutoff Spezifität und Positiver Prädiktiver Wert auf der Strecke blieben. Insgesamt wird von einer richtigen Zuordnung der Patienten zu der gegebenen Referenz-Diagnose in 80% der Fälle gesprochen.

1.2.4.2.2. Die Reliabilitätsstudie von Zanarini, Frankenburg & Vujanovic (2002)

Im Jahr 2002 wurde schließlich die bereits im Artikel von 1989 angekündigte Reliabilitätsstudie mit dem Titel "Inter-Rater and Test-Retest Reliability of the Revised Diagnostic Interview for Borderlines" (Zanarini, Frankenburg & Vujanovic, 2002) veröffentlicht. In dieser Studie werden sowohl die klassische Interrater-Reliabilität als auch die Test-Retest-Reliabilität und die eher ungewöhnlichen Größen "Follow-Up-Interrater-Reliabilität" und "Follow-Up-Longitudinal-Reliabilität" angegeben. Die Ergebnisse zur Interrater-Reliabilität werden im Folgenden berichtet.

Für die Berechnung der Interrater-Reliabilität wurden die Werte der in der DIB-R-Anwendung durch die Erstautorin geschulten Rater verwendet, die jeweils gemeinsam 45 stationäre Patienten interviewt hatten. Leider werden in der Studie keine näheren Angaben über die Patienten der Stichprobe (z.B. Anteil der Patienten mit und ohne Borderline-Persönlichkeitsstörung, Co-Diagnosen etc.) gemacht.

Auf der Ebene der Symptome (Statements) sowie für die Borderline-Diagnose wurden *Kappa*-Werte verwendet, für die "dimensionalen" Kennwerte der Intraclass-Korrelationskoeffizient. Dabei werden sowohl Kappa- als auch Intraclass-Korrelationskoeffizienten ab einem Wert von über .75 als exzellent und solche zwischen .40 und .75 als "fairly good" bezeichnet (unter Bezugnahme auf die Richtwerte von Fleiss, 1981).

In den beiden folgenden Tabellen werden die Koeffizienten wiedergegeben. Die Ergebnisse fallen auffallend gut aus. Bei 18 der 22 Statements sowie bei der Diagnose werden exzellente Werte erzielt. Die vier übrigen Statements (grau unterlegt) erreichen immerhin noch gute Kappa-Koeffizienten.

Tab. 3: *Kappa-Koeffizienten für Statements und Diagnose der Borderline-Persönlichkeitsstörung aus der Studie von Zanarini, Frankenburg & Vujanovic (2002)*

	S.	Der Patient ...	κ
Affekte	**1**	litt an einer chronischen depressiven Verstimmung oder hatte eine oder mehrere Perioden von Major Depression	1.0
	2	hatte anhaltende Gefühle von Hilf-, Hoffnungs-, Wertlosigkeit oder Schuld	1.0
	3	hatte chronische Gefühle von Ärger, Wut oder verhielt sich häufig ärgerlich wütend	1.0
	4	hat sich chronisch sehr ängstlich gefühlt oder litt häufig unter körperlichen Angstsymptomen	1.0
	5	erlebte chronische Gefühle von Einsamkeit, Langeweile oder Leere	1.0
Kognition	**6**	neigte zu seltsamen Denken oder ungewöhnlichen Wahrnehmungserlebnissen	.74
	7	hatte häufig flüchtige, nicht-wahnhafte paranoide Erlebnisse	.85
	8	hatte wiederholte pseudo-psychotische Wahnvorstellungen oder Halluzinationen	1.0
Impulsbereich	**9**	betrieb einen ernsthaften Drogenmissbrauch	.91
	10	hatte ein Muster sexuell abweichenden Verhaltens	.90
	11	zeigte ein Muster von körperlicher Selbstbeschädigung	.90
	12	zeigte ein Muster von manipulativen Selbstmorddrohungen, -gesten oder –versuchen	.85
	13	zeigte ein anderes Muster impulsiven Verhaltens	.83
Zwischenmenschliche Beziehungen	**14**	hat typischerweise versucht, das Alleinsein zu vermeiden oder fühlte sich extrem dysphorisch, wenn er allein war	.93
	15	hat (wiederholt) Verlassenheits-, Verschlingungs- oder Vernichtungsängste erlebt	.88
	16	hat seine Abhängigkeitswünsche stark abgewehrt oder befand sich in einem ernsten Konflikt zwischen Versorgen und Versorgtwerden	.80
	17	neigte zu intensiven instabilen engen Beziehungen	.94
	18	hatte in engen Beziehungen immer wieder Probleme mit Abhängigkeit oder Masochismus	1.0
	19	hatte in engen Beziehungen wiederkehrende Probleme mit Abwertung, Manipulation oder Sadismus	.73
	20	hatte in engen Beziehungen immer wieder Probleme mit seiner Forderungs- oder Anspruchshaltung	.84
	21	zeigte während der Therapie oder der psychiatrischen Hospitalisierung eine deutliche Regression	.55
	22	hat auf der psychiatr. Station oder in einer Psychotherapie auffallende Gegenübertragungsreaktionen ausgelöst oder ist mit einem profess. Helfer eine ganz besondere Beziehung eingegangen	.73
	DIB-R –Diagnose (Cutoff ≥8)		**.94**

Bezüglich der Intraclass-Korrelationskoeffizienten werden nur exzellente Werte berichtet, die am höchsten im Bereich der Affekte und am niedrigsten bei den Zwischenmenschlichen Beziehungen liegen.

Tab. 4: *Intraclass-Korrelationskoeffizienten aus der Reliabilitäts-Studie von Zanarini, Frankenburg & Vujanovic (2002)*

N=45	**DIB-R-Bereiche**	**ICC**
Summen-Scores	**Affekte** (Range=0–10)	.99
	Kognition (Range=0–6)	.80
	Impulsivität (Range=0–10)	.92
	Zwischenm. Beziehungen (Range=0–18)	.80
Gesamt-Summen-Score	(Range=0–44)	.84

Trotz ihrer herausragenden Stellung wegen des Fehlens anderer Reliabilitätsstudien zum DIB-R weist die vorliegende Veröffentlichung aber vermutlich einen Mangel auf, der ihre Verwendbarkeit zum Vergleich mit den Ergebnissen der Evaluation der deutschen Studie erheblich einschränkt.

Die Autoren geben an, dass sie während der gesamten Studie immer wieder Gruppenmeetings hatten, bei denen die Interviews detailliert durchgegangen und auftauchende Differenzen diskutiert wurden. Bei diesen Meetings wurden auch "consensus rules that determined future ratings" festgelegt. "This praxis may, in part, explain our good reliability findings" erklären die Autoren hierzu. Dies legt eigentlich nahe, dass es sich nicht um ein vollständig *unabhängiges* Rating gehandelt hat. Auf Nachfrage des Autors teilte Zanarini per E-Mail (2007) jedoch mit:

"The ratings were always independent but we did discuss each case so that we developed a coherent way of thinking about these diagnostic issues as time went on."

Die Reliabilität der Ratings dürfte durch diese enge Abstimmung bei der geringen Zahl von zwei Ratern für den Zweck einer Generalisierung *in einem nicht näher bestimmbaren Maß überschätzt* worden sein. Es muss auf eine sehr enge (über eine detaillierte Schulung im DIB-R weit hinausgehende) Absprache und Diskussion zurückzuführen sein, dass für über 45 Fälle in ca. einem Drittel aller Kennwerte *perfekte Übereinstimmung* erzielt werden konnte. Entsprechendes gilt für den äußerst hohen Wert von κ=.94 für eine Übereinstimmung bzgl. des Vorliegens bzw. Nichtvorliegens der Borderline-Persönlichkeitsstörung.

Eine weitere Schwierigkeit in der Verwendung der o.g. Studie ist, dass die Autoren zusätzlich zu den o.g. Werten nur die eher unrelevanten *Summen-Scores,* nicht aber die *Skalierten Section-Scores der Bereiche* überprüft haben sowie entsprechend den unbedeutenden und in der

Testanweisung nicht definierten Gesamt-Summen-Score (aufsummierte Statement- bzw. Bereichs-Summen-Scores; mgl. Range 0–44), nicht aber den DIB-R-Gesamt-Score (mgl. Range 0–10), von dem die Diagnosestellung abhängt. Der Gesamt-Summen-Score ist a priori viel reliabler einschätzbar als der Gesamt-Score, da keinerlei möglicherweise Differenzen zwischen den Ratern schaffende Skalierungsregeln der vier Bereiche ins Spiel kommen. Die hierzu eher uneindeutige Formulierung in der Veröffentlichung bzgl. der o.g. Werte wurde ebenfalls durch die erwähnte persönliche Mitteilung geklärt.

1.2.4.2.3. Die spanische Übersetzung

Für das DIB-R liegt nunmehr ebenfalls eine spanische Adaption vor, für die 2005 eine Evaluationsstudie veröffentlicht wurde (Szerman et al., 2005), die Fragen der Validität und der Reliabilität berücksichtigt. (Bei dieser Studie war John G. Gunderson einer der Co-Autoren.)

Insgesamt waren 111 Patienten untersucht worden, davon 84 mit einer Borderline-Persönlichkeitsstörung nach DSM-IV. 27 Patienten hatten als Kontrollgruppe andere psychiatrische Störungen.

Zur **Reliabilitätsuntersuchung** waren 31 Patienten ausgewählt worden und von zwei Ratern jeweils unabhängig (mit Wiederholung des jeweils von einem Rater einzeln durchgeführten Interviews) mittels des DIB-R eingeschätzt worden. Die Interviewer kannten dabei die Ausgangsdiagnose des Patienten nach DSM-IV nicht.

Die Interraterübereinstimmung bzgl. des DIB-R war, neben einigen anderen Kennwerten, für den Gesamt-Score, die Skalierten Section-Scores und die gefundene Diagnose (dichotome diagnostische Kategorie) anhand des Kappa-Koeffizienten überprüft worden. Alle Werte waren signifikant und bis auf den Kognitionsbereich *gut bis sehr gut*.

Tab. 5: *Kappa-Koeffizienten für DIB-R; Skalierte Section-Scores , Gesamt-Score und -Diagnose (spanische Übersetzung, Szerman et al., 2005)*

N=31	**DIB-R-Bereiche**	**κ**
Skalierte Section-Scores	**Affekte** *(Range 0-2)*	.953
	Kognition *(Range 0-2)*	.630
	Impulsivität *(Range 0-3)*	.961
	Zwischenmenschl. Beziehungen *(Range 0-3)*	.958
Gesamt-Score	*(Range 0-10)*	.904
DIB-R-Diagnose* Cutoff ≥7	**BPS vs. Non-BPS**	**.783**

* zu beachten ist der abweichende Cutoff, der der spanischen Version zu Grunde gelegt wurde

Der Validitätsteil der Studie folgt eng dem Vorgehen der oben genannten Studie von Zanarini et al. (1989) und überprüft anhand der Stichprobe Spezifität und Sensitivität sowie den Positiven (PPW) und den Negativen Prädiktiven Wert (NPW). Geprüft wird ebenfalls die Wirkung verschiedener Cutoff-Werte auf die genannten Kennwerte.

Tab. 6: *Überprüfung der Cutoff-Werte des DIB-R (spanische Übersetzung Szerman et al., 2005)*

N=31	**"6"**	**"7"**	**"8"**	**"9"**	**"10"**	**Wahrscheinlichkeit, dass bei einem Cutoff von ...**
Sensitivität	.964	**.964**	.869	.571	.179	DIB-R-BPS, wenn klinisch Borderline-Persönlichkeitsstörung
Spezifität	.741	**.889**	1	1	1	*nicht* DIB-R-BPS, wenn *nicht* klinisch Borderline-Persönlichkeitsstörung
Positiver Prädiktiver Wert	.920	**.964**	1	1	1	klinisch BPS, wenn DIB-R-Borderline-Persönlichkeitsstörung
Negativer Prädiktiver Wert	.870	**.889**	.711	.290	.281	klinisch *nicht* BPS, wenn *nicht* DIB-R-Borderline-Persönlichkeitsstörung
κ *	.742	**.853**	.764	.393	.096	Diagnoseübereinstimmung zwischen DIB-R / DSM-IV

** alle signifikant; Auszug aus Tab. 4, Szerman et al. (2005)*

Überraschenderweise finden die Autoren aber nicht, wie Zanarini et al. (1989) für die Originalversion, beim gültigen Cutoff-Wert von "8" den besten Zusammenhang mit dem als Außenkriterium verwendeten DSM-IV. Stattdessen stellen sie fest, dass das Optimum, auch erkennbar am höchsten Kappa-Wert für die Übereinstimmung beider Diagnosen, bei einem Cutoff von "7" erzielt wird. Dieser wurde dann offenbar der gesamten spanischen Übersetzung abweichend zur Originalversion zu Grunde gelegt.

1.2.4.2.4. Die französische Übersetzung

Eine französische Übersetzung des DIB-R wurde im Jahr 1992 durch Chaine et al. (1995) veröffentlicht. In der hier besprochenen Veröffentlichung wurden die Ergebnisse einer Evaluationsstudie wiedergegeben, die u.a. das DIB-R betreffen.

Zu diesem Zweck waren 36 von französischen Klinikern als Borderline-Persönlichkeiten eingeschätzte Patienten untersucht worden. Mit diesen Patienten wurde zu Vergleichszwecken neben dem DIB-R auch das International Personality Disorder Examination (IPDE) und das Minnesota Multiphasic Personality Inventory (MMPI) durchgeführt. Vergleichsdiagnosen wurden nach den ICD-10- und DSM-III-R-Systemen (beide anhand IPDE) sowie dem DIB-R erstellt.

Für die untersuchte Stichprobe geben die Autoren an, dass das DIB-R weniger inklusiv diagnostiziert als das DSM-III-R, jedoch mehr Patienten die Borderline-Diagnose zuschreibt als das ICD-10.

Eine Untersuchung der Diagnose-Übereinstimmungen in der Patienten-Stichprobe ergab, dass 69,5% der Patienten in zumindest zwei der drei Diagnosen eine Übereinstimmung erzielten und immerhin 41,5% in allen dreien. Diese 15 Patienten, die nach allen Systemen die Borderline-Diagnose erhalten hatten, wurden zu weiteren Vergleichen als "Borderline-Kerngruppe" definiert.

Ein Gesamt-Score von ≥8 schien den Autoren ein gutes Zeichen der Zugehörigkeit zu dieser Gruppe zu sein. Ein höherer Gesamt-Score von 9 oder 10 verbesserte die Qualität der Vorhersage aber *nicht* mehr (siehe Tab. 7).

Tab. 7: *Zugehörigkeit zur Kerngruppe der Borderline-Patienten anhand des DIB-R-Gesamt-Scores aus der Studie von Chaine et al. (1995; Auszug aus Tableau IX)*

	Kerngruppe (N=15)	**Rest der Stichprobe** (N=21)	**Gesamt-Stichprobe** (N=36)
≥ 8	*15 (100%)*	*10 (47,5%)*	25 (69,5%)
≥ 9	8 (55,5%)	5 (24%)	14 (39%)

Eine Untersuchung zur Reliabilität der französischen Version wird von den Autoren weder durchgeführt noch erwähnt.

1.2.4.3. Verbreitung des DIB-R

Zanarini bezeichnete 1989 das DIB als das derzeit bekannteste und am meisten eingesetzte, speziell zur Diagnose der Borderline-Persönlichkeitsstörung verwendete Instrument.

Dies dürfte auch für das DIB-R gelten, welches inzwischen in vielen Studien, insbesondere aus dem Umfeld der Autoren, vielfältige Verwendung gefunden hat. Beispielhaft seien aufgezählt einige Langzeitstudien (Zanarini et al., 2003 & 2006), eine Studie zur Wirksamkeit der Behandlung mit Omega-3-Fettsäuren bei Borderline-Patientinnen (Zanarini & Frankenburg, 2003) oder zur Komorbidität mit Achse-II-Störungen (Zanarini et al., 2004). Die o.g. französische Fassung wird offenbar auch in Kanada verwendet – hierzu liegt z.B. eine Studie von Gagnon et al. (2006) zu Borderline-Persönlichkeitstraits nach traumatischen Hirnverletzungen vor.

2. EVALUATION VON TESTS UND INTERVIEWS

Objektivität, Validität und Reliabilität stellen die Hauptgütekriterien psychologischer Tests dar und sind somit auch als die Grundlagen einer ordnungsgemäßen Testdurchführung zu verstehen. Im Folgenden werden diese Themenbereiche überblicksweise dargelegt. Im Anschluss folgt ein Überblick über die Grundlagen der Bestimmung der Interrater-Reliabilität, da sie für die vorliegende Studie von zentraler Bedeutung ist.

2.1. OBJEKTIVITÄT

Unter der Objektivität eines wissenschaftlichen Tests ist die Unabhängigkeit der Testergebnisse von den Rahmenbedingungen zu verstehen. Diese gliedern sich in Durchführungsobjektivität, Auswertungsobjektivität und die Interpretationsobjektivität. Je objektiver ein Test ist, desto weniger anfällig gegenüber äußeren Bedingungen ist er.

2.2. VALIDITÄT

Validität bedeutet Gültigkeit einer Untersuchung oder Messung. Im Folgenden werden die einzelnen Teilaspekte der Validität im Zusammenhang mit typischen Fehlerquellen und Gefährdungen vorgestellt. Daraufhin werden einige Berechnungsmöglichkeiten und ein idealerweise zu erreichender Richtwert genannt.

2.2.1. Externe Validität

Diese Validitätsform wird auch als Allgemeingültigkeit, Verallgemeinerungsfähigkeit oder ökologische Validität bezeichnet. Sie bedeutet die Übereinstimmung von tatsächlichem und intendiertem Untersuchungsgegenstand. Grundidee ist hier die Frage nach der *Generalisierbarkeit*. Regelmäßig führt man zuerst Studien an kleinen und leicht zu erreichenden Gesamtheiten durch, etwa an Studenten oder Patienten. Das korrekte Vorgehen ist, nach einer solchen explorativen Studie eine repräsentative durchzuführen, was in jedem Falle aufwändig und bisweilen auch sehr schwierig ist.

Stichprobenbias bezeichnet in diesem Zusammenhang die Abweichung einer konkreten Stichprobe von dem Ideal einer streng zufälligen Auswahl aus der richtigen Grundgesamtheit.

2.2.2. Interne Validität

Ein Messinstrument oder Experiment ist in dem Maß intern valide, in dem es gelungen ist, potentielle Störvariablen zu kontrollieren.

Für Experimente sind wichtige Störungen der internen Validität v.a. Effekte, die aus der Gruppenzuteilung herrühren (z.B. mangelnde Randomisierung), Selektionseffekte bei der

Auswahl der Versuchspersonen, Reifungseffekte und Einflüsse zwischenzeitlichen Geschehens, Messeffekte (z.B. bei Testwiederholung erworbene Erfahrung oder Sensibilisierung) sowie das Ausscheiden von Versuchspersonen und Testabbrüche.

2.2.3. Inhaltsvalidität

Die Inhaltsvalidität bedeutet als gewissermaßen einfachstes Validitätskonzept, dass die Gültigkeit der Messung mehr oder weniger unmittelbar einsichtig aus den einzelnen Teilen des Messinstruments selbst hervorgeht. Sie beruht auf der Kenntnis von "Experten" über den betreffenden Gegenstand (wobei offen bleibt, wie genau die Definition eines Experten zu lauten hat). Inhaltsvalidität ist eng verknüpft mit dem Begriff der "face validity", d.h. der "Augenschein-Validität".

Die Idee der Inhaltsvalidität ist dennoch sehr wichtig – hierzu Ludwig-Mayerhofer (2008) in seinem ILMES-Internet-Lexikon: *"Es geht letztlich darum, dass eine Messung das relevante Phänomen möglichst in allen Aspekten erfasst, und dies kann nur durch Forschen, Nachdenken und Kommunikation zwischen Wissenschaftlern herausgefunden werden und nicht durch bestimmte 'Techniken'."*

2.2.4. Konstruktvalidität

Ein Test, der vorgibt, z.B. Intelligenz zu messen, könnte in Wirklichkeit auch nur eine Vertrautheit mit dem soziokulturellen Hintergrund messen, dem dieser Test entstammt. Ebenso wäre es möglich, dass ein solcher Test eher die Lesefähigkeit oder die Kreativität eines Probanden misst als dessen Intelligenz. Diese korrelieren zwar beide mit der Intelligenz, sind aber unabhängige Konstrukte.

Die Konstruktvalidität kann dementsprechend in zwei Validitätsformen aufgeteilt werden:

- Die **konvergente Validität** prüft die Übereinstimmung eines Testergebnisses mit dem eines anderen, bereits angewandten und validierten konstrukt*nahen* Tests. Unter dieser Voraussetzung müssten die Ergebnisse des zu validierenden Tests mit den Ergebnissen bekannter Tests hoch korrelieren. Auf das o.g. Beispiel bezogen heißt das, dass ein neuer Intelligenztest mit älteren Tests, je nach konzeptueller Ähnlichkeit, in einem substanziellen Zusammenhang stehen *muss*. *Auf die vorliegende Untersuchung bezogen bedeutet es, dass das DIB-R-Interview mit anderen Interviews zur Borderline-Persönlichkeitsstörung mittlere bis hohe Korrelationen aufweisen sollte.*
- Wie bei der konvergenten Validität steht bei der **diskriminanten Validität** (auch divergente Validität) die Überprüfung des korrelativen Zusammenhanges der Ergebnisse des zu

validierenden Tests mit den Ergebnissen anderer Testverfahren im Zentrum. An dieser Stelle soll allerdings sichergestellt werden, dass nicht konstrukt*ferne* Merkmale lediglich in einem neuen Konstrukt verpackt erneut gemessen werden. Auf das o.g. Beispiel bezogen bedeutet das, dass ein reiner Intelligenztest möglichst nicht oder nur mäßig mit den Werten der Probanden z.B. in einem Kreativitätstest korrelieren sollte. *Im Falle der vorliegenden Untersuchung gilt z.B. die Forderung, dass das DIB-R nur gering mit Interview-Testwerten anderer Persönlichkeitsstörungen korrelieren sollte.*

Sowohl konvergente als auch diskriminante Validität müssen gegeben sein, um einen vollständigen Nachweis der Konstruktvalidität zu gewährleisten.

2.2.5. Kriterienbezogene Validität

Die Ergebnisse eines Messinstruments werden mit den Werten eines zuvor festgelegten (beobacht- oder messbaren) Außenkriteriums verglichen. Die Kriteriumsvalidität kann in kongruente und prädiktive Validitäten unterteilt werden:

- *Kongruente bzw. Übereinstimmungs-Validität*

Sofern der Test zeitgleich zu einem anderen Test (als Außenkriterium) durchgeführt wird, kann eine Übereinstimmungsvalidität in Form einer Korrelation berechnet werden. (Diese Übereinstimmungs-Validität entspricht im vorliegenden Anwendungsfall weitestgehend der konvergenten Validität und wird m.E. synonym verwendet.)

- *Prädiktive/prognostische Validität*

Für den Fall, dass die Daten des zu validierenden Tests zu einem früheren Zeitpunkt erhoben werden als die Daten eines anderen Tests (als Außenkriterium), so könnten die Messdaten das Ergebnis des folgenden Tests vorhersagen. Es kann also eine Prognose anhand der Testdaten erfolgen und überprüft werden (z.B. Verlauf einer Störung oder Therapie in Abhängigkeit von der Diagnose).

2.2.6. Anforderungen an die Validität eines Tests

Für die in der vorliegenden Studie verwendete und untersuchte *konvergente Validität* (entsprechend der *Übereinstimmungs-Validität)* können (wie oben beschrieben) Validitätskoeffizienten in Form von Korrelationen berechnet werden.

Aber ab wann ist die Höhe dieser Koeffizienten ausreichend? Lienert und Raatz (1998, S. 269) nehmen hierzu Stellung:

"Hier lassen sich keinerlei starre Normen einführen, sondern nur Richtlinien aufzeigen, nach denen ein empirischer Validitätskoeffizient zu bewerten ist. Es wird dabei deutlich werden, dass ein relativer Maßstab mehr für sich hat als ein absoluter."

Dabei hänge die Höhe stark vom Verwendungszweck ab: Für statistische Vorhersagen, insbesondere im Individualfall, müssten nach Lienert und Raatz Validitätskoeffizienten von $r_{tc} \geq .70$ verlangt werden. Dies gelte insbesondere dann, wenn *nur dieser eine Test* zur individuellen Vorhersage verwendet werde – in der Praxis würden so hohe Werte aber nur im Einzelfall erreicht. Mit einem Koeffizienten um .60 sei man schon sehr zufrieden.

Sollte eine *Kombination von Einzeltests* vorliegen, oder, wie das in der diagnostischen Praxis üblich ist, noch weitere Informationen zur Beurteilung von Persönlichkeitsmerkmalen, Eignung etc. herangezogen werden, reduziere sich die Anforderung an den Validitätskoeffizienten auf Werte von $r_{tc} \geq .50$.

Für die vorliegende Studie ist zu betonen, dass aufgrund der Konzeptabhängigkeit der Diagnose der Borderline-Persönlichkeitsstörung *kein wirklich valides Außenkriterium* vorliegen kann. Es bleibt nur die Korrelation mit einem anderen diagnostischen Merkmal, hier einem Interview, dem ein verwandtes, aber keineswegs identisches Konzept der Borderline-Persönlichkeitsstörung zu Grunde liegt, und welches zudem naturgemäß selbst den Fehlerquellen einer unperfekten Validität und auch Reliabilität unterworfen ist.

Lienert und Raatz (1998, S. 271) schreiben hierzu im Kap. 11.9.2 "Allgemeine Richtlinien über die Auswahl von Validitätskoeffizienten" (Auszug aus der Aufzählung):

- *"Ein Test der an einem zulänglichen und reliablen Kriterium validiert wurde muss eine höhere Reliabilität haben als ein Test, bei dem dies nicht möglich war."*
- *"Ein Validitätskoeffizient ist auch danach zu beurteilen, ob das durch den Test geprüfte Persönlichkeitsmerkmal ohne Verwendung von Testverfahren leicht oder schwer zu erfassen ist."*
- *"Tests mit Validitätskoeffizienten unter 0,3 sind – auch wenn sie statistisch gesichert sind – für eine allgemeine Verwendung nahezu nutzlos. [...]"*

Dieses zusammenfassend wird $r_{tc} \geq .50$ als eine für die vorliegende Studie sinnvolle *Untergrenze für die konvergente Validität bzw. Übereinstimmungsvalidität* definiert. Eine Korrelation $r_{tc} < .30$ wird entsprechend als noch zu tolerierende *Obergrenze für die divergente Validität angesehen*.

2.3. RELIABILITÄT

Unter Reliabilität versteht man die Zuverlässigkeit wissenschaftlicher Untersuchungen. Nimmt man, wie in der klassischen Testtheorie vorausgesetzt, an, dass es tatsächlich einen (wenn auch

nicht direkt beobachtbaren) "wahren Wert" gibt, so beschreibt die Reliabilität den Grad der Übereinstimmung zwischen diesem wahren Wert und dem gemessenen Wert. Im Idealfall sind Messwert und wahrer Wert identisch. Das eingesetzte Messverfahren misst das Kriterium exakt und die Reliabilität hat dann den Wert „1“.

Bei der praktischen Umsetzung dieser Vorgaben stößt man meist auf das Problem, dass der "wahre Wert" nicht zu quantifizieren ist. Erstens ist er nicht notwendigerweise zeitstabil – vor allem aber ist er nicht *unmittelbar* quantifizierbar, sondern muss seinerseits erst durch eine Messung erschlossen werden. Die klassische Testtheorie löst dieses kaum auflösbare Erkenntnisproblem mit einer einfachen Annahme: Man erhalte dann den wahren Wert einer Person, wenn man mit ihr einen Test unendlich oft wiederhole und die Ergebnisse mittele.

Die Reliabilität einer Untersuchung wird entsprechend zumeist dadurch überprüft, dass getestet wird, ob man bei der Durchführung unter "gleichen" Bedingungen zu denselben Ergebnissen kommt. In Abweichung von der ursprünglichen Definition (Übereinstimmung zwischen dem wahren und dem gemessenen Wert) versteht man im wissenschaftlich-empirischem Kontext unter Reliabilität zumeist jene Übereinstimmung. Man beachte, dass dabei das jeweils verwendete Messinstrument durchaus immer systematisch die gleichen *falschen* Werte, also einen Messfehler liefern kann. Der so definierten Reliabilität liegt *nicht* der Zusammenhang zwischen dem wahren und dem gemessenen Wert zugrunde.

Hierzu Lienert und Raatz (1998, S. 9):

"Unter der Reliabilität oder Zuverlässigkeit eines Tests versteht man den Grad der Genauigkeit, mit dem er ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal mißt, gleichgültig, ob er dieses Merkmal auch zu messen beansprucht (welche Frage ein Problem der Validität ist). [...] Der Grad der Reliabilität wird durch einen Reliabilitätskoeffizienten bestimmt, der angibt, in welchem Maße unter gleichen Bedingungen gewonnene Messwerte über ein und denselben Probanden übereinstimmen, in welchem Maße also das Testergebnis reproduzierbar ist."

Zur Überprüfung der Reliabilität unterscheidet man v.a. zwei verschiedene Methoden:

- **Test-Retest-Verfahren:** Hier wird geprüft, ob eine Wiederholung der Messung bei anzunehmender Konstanz der zu messenden Eigenschaft die gleichen Messwerte liefert.
- **Paralleltestung** (Paralleltest-Verfahren)**:** Hier wird geprüft, ob ein "vergleichbares" Messverfahren identische Ergebnisse liefert. Anstelle gleichwertiger Testverfahren können auch *Parallelformen* des Tests verwendet werden.

Für viele Tests ist eine Wiederholung entsprechend dem Test-Retest-Verfahren nur theoretisch möglich, da die damit einhergehenden Übungs- oder Gewöhnungseffekte das Ergebnis beeinflussen (z.B. Übungseffekte in der Problemlösung von Aufgaben zum räumlichen Denken

in Intelligenztests). Um in diesen Fällen trotzdem Angaben über die Zuverlässigkeit des Tests machen zu können, werden die Items eines Tests aufgeteilt und miteinander korreliert (Split-Half-Reliabilität).

Im Falle der Reliabilitätsprüfung eines diagnostischen Testverfahrens in Interviewform ist keines der beiden o.g. Verfahren zur Reliabilitätstestung vielversprechend. Erstens können insbesondere die Symptome einer Persönlichkeitsstörung (insbesondere der Borderline-Persönlichkeitsstörung) nicht als zeitstabil gelten, so dass ein Test-Retest-Verfahren als Mittel der Wahl ausscheidet.

Eine Paralleltestung durch Testhalbierung ist ebenfalls nicht möglich. Diese häufige Problematik führte zur Entwicklung einer Art Spezialform der Paralleltestung: der **Interrater-Reliabilität**. Bei dieser Form der Reliabilitäts-Überprüfung wird simultan von mehreren prinzipiell vergleichbaren Personen eine Einschätzung vorgenommen.

Dieses Vorgehen ist neben seiner Praktikabilität sehr nah am praktischen Anwendungsfall orientiert. Es wird im folgenden Kap. 2.4 detailliert beschrieben.

2.4. GRUNDLAGEN DER BESTIMMUNG DER INTERRATER-RELIABILITÄT

In diesem Kapitel werden nun (Bezug nehmend auf Kap. 2.3) die relevanten Grundlagen der Bestimmung der Messgenauigkeit von Ratingskalen bei mehreren Ratern in ihren verschiedenen Erscheinungsformen bestimmt. Wesentliche Grundlagen des Kapitels sind die Übersichtsarbeiten von Tinsley und Weiss (1975) sowie von Asendorpf und Wallbott (1979).

2.4.1. Interrater-Übereinstimmung vs. Interrater-Reliabilität

Um zur Klarheit beizutragen, werden Methoden zur Beurteilung der Genauigkeit von Ratingskalen nach dem Interrater-Verfahren zunächst anhand einer von Tinsley und Weiss (1975) explizierten *Differenzierung zwischen Reliabilität und Übereinstimmung* (Agreement) von Ratern vorgenommen.

Wenn auch manche Autoren (wie Bortz & Döring, 1995) nicht nach einer solchen Unterscheidung vorgehen, wird diese Differenzierung aufgrund ihrer Veranschaulichung der Problematik als wertvoll erachtet.

Interrater-Übereinstimmung bedeutet das Ausmaß, in dem Rater exakt *dieselbe* Einschätzung über ein zu beurteilendes Merkmal treffen. Im Falle einer numerischen Skala wäre dies die Wahl desselben Zahlenwertes.

Interrater-Reliabilität bedeutet hingegen eine *Verhältnismäßigkeit*. Sie beschreibt den Grad, in dem die Urteile verschiedener Rater *zueinander proportional* sind, und zwar wenn sie in Abweichung von Mittelwerten bzw. Varianzen gesehen werden.

Es sei vorab darauf hingewiesen, dass die häufig (und auch in der vorliegenden Untersuchung verwendeten) nicht-adjustierten Intraclass-Korrelationskoeffizienten eine Art Zwischenstellung zwischen Reliabilität und Agreement einnehmen: es geht um verhältnismäßige Übereinstimmung. Ein Reliabilitätskoeffizient im eigentlichen Sinne von Tinsley und Weiss (1975) wäre der *adjustierte Reliabilitätskoeffizient*, wie er z.B. bei Bortz und Döring (1995, Tafel 27) beschrieben wird.

Wie in Tab. 8 aus dem von Tinsley und Weiss (1975, S. 359) vorgelegten Beispieldatensatz ersichtlich, sind Konstellationen möglich, die in Agreement und Reliabilität erheblich differieren: So geht eine hohe Reliabilität typischerweise durchaus mit einem hohen Grad an Übereinstimmung einher (Fall 1). Ebenso ist aber eine hohe Reliabilität gepaart mit einer niedrigen Übereinstimmung der Rater oder sogar umgekehrt möglich (Fälle 2 und 3).

Tinsley und Weiss (1975) kommen hier zu dem Schluss, dass die bloße Angabe der Interrater-*Reliabilität* nur dann ausreiche, wenn lediglich eine Übereinstimmung der Raterurteile in der relativen Ordnung die Anforderung sei. Sobald aber eine absolute Werteübereinstimmung gefordert werde, müsse auch die Interrater-*Übereinstimmung* berichtet werden.

Für die vorliegende Untersuchung erscheint neben der Erhebung von Reliabilitätsmaßen auch die von Übereinstimmungsmaßen wesentlich – und zwar aus einem weiteren von Tinsley und Weiss (1975) berichteten Grund, welcher anhand von Fall 3 aus dem Beispieldatensatz erläutert wird.

Dort ist eine *niedrige* Interrater-*Reliabilität* gepaart mit einer *hohen* Interrater-*Übereinstimmung*. In diesem Fall ist die Reliabilität – trotz der recht hohen Übereinstimmung – niedrig, und zwar v.a. wegen des *stark eingeschränkten Range der Raterurteile* über die Zeilen bzw. Beobachtungen (Range nur 3-5 statt möglicher 1-9) sowie wegen der unsystematischen Abweichungen der Raterurteile über die Fälle.

Dieser schmale Range könnte z.B. aus einer unangemessenen Verwendung der Ratingskalen z.B. im Sinne einer Antworttendenz ungeachtet der zu beurteilenden Objekte oder deren mangelnder Validität herrühren. Entscheidend ist für die vorliegende Untersuchung jedoch der Umstand, dass dieser Range auch aus einer großen Ähnlichkeit der *wahren Werte* der zu beurteilenden Objekte, also einer wirklichen Homogenität der Probanden auf dem zu messenden Merkmal, herrühren könnte. Für die Bestimmung der Interrater-Reliabilität wäre eine solche Stichprobe dann eigentlich nicht geeignet (obwohl dies ein in der Sozialwissenschaft häufiger Umstand ist).

Tab. 8: *Hypothetical Ratings of Accurate Empathy Illustrating Different Levels of Interrater Agreement and Interrater Reliability for Interval-Scaled Data (Tinsley & Weiss, 1975, S. 359; ergänzend werden Übereinstimmungskoeffizienten* angegeben)*

Counselor	**Case 1:** *High* interrater agreement and *high* interrater reliability			**Case 2:** *Low* interrater agreement and *high* interrater reliability			**Case 3:** *High* interrater agreement and *low* interrater reliability		
	Rater			**Rater**			**Rater**		
	1	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
A	1	1	1	1	3	5	5	4	4
B	2	2	2	1	3	5	5	4	3
C	3	3	3	2	4	6	5	4	5
D	3	3	3	2	4	6	4	4	5
E	4	4	4	3	5	7	5	4	3
F	5	5	5	3	5	7	5	5	4
G	6	6	6	4	6	8	4	4	5
H	7	7	7	4	6	8	5	5	4
I	8	8	8	5	7	9	4	5	3
J	9	9	9	5	7	9	5	5	5
Mean	4,8	4,8	4,8	3,0	5,0	7,0	4,7	4,4	4,1
SD	2,7	2,7	2,7	1,5	1,5	1,5	0,5	0,5	0,9
IC_a	**1,0**			**1,0**			**-0,08**		
IC_u	**1,0**			**0,18**			**-0,10**		

* Die Reliabilität ist für das obige Beispiel nur unter Verwendung des (adjustierten) Reliabilitätskoeffizienten IC_a für ein "fixed set of raters" (Tinsley & Weiss, 1975, S. 364) als hoch einzustufen. Bei diesem Koeffizienten wird der Mittelwertsunterschied *zwischen* den Ratern *nicht* berücksichtigt. Unter Verwendung des in der vorliegenden Studie verwendeten nicht-adj. IC_u (s. Kap. 2.4.6.1) würde aufgrund der vorliegenden Unterschiede im Urteilsniveau der Rater im Case 2 nur eine eher geringe Reliabilität erreicht. Die Logik der von Tinsley und Weiss vorgenommenen Unterscheidung wird dadurch aber nicht beeinträchtigt.

Immer ist bei einer hohen Übereinstimmung und einer niedrigen, konkret festgestellten Reliabilität nämlich die Möglichkeit gegeben, dass eine in Wahrheit hohe Reliabilität der Ratingskala lediglich nicht erkannt werden konnte, da der verwendete Datensatz mit den einzuschätzenden Subjekten für die Reliabilitätsschätzung ungeeignet war. Keinesfalls kann daher *zwangsläufig* aus einer niedrigen Reliabilität auf eine rein zufällige Übereinstimmung der Rater geschlossen werden!

Somit ergibt sich ein gestuftes Vorgehen: Wenn sowohl die Interrater-Reliabilität als auch die Interrater-Übereinstimmung hoch sind, *muss* von einer hohen Güte der Ratingskala ausgegangen werden. Ist lediglich die Interrater-Übereinstimmung hoch, die Reliabilität aber niedrig, sollte durch weitere Untersuchungen an einem geeigneteren Datensatz geklärt werden, ob eine hohe Reliabilität nur aus untersuchungstechnischen Gründen verborgen geblieben war.

Im Falle einer hohen Interrater-Reliabilität gepaart mit geringem Agreement muss die Frage gestellt werden, ob evtl. trotz formal ausreichender Reliabilität die *Qualität* der Urteile, z.B. für klinische Zwecke bei der Diagnosestellung, ausreichend ist. Wenn sowohl Interrater-Reliabilität als auch Interrater-Übereinstimmung niedrig sind, fällt die Entscheidung leicht: Die Ratingskala ist als unbrauchbar zu betrachten.[1]

2.4.2. Zum Skalenniveau von Ratingskalen

Bekanntermaßen liegt das Skalenniveau von Ratingskalen üblicherweise zwar *über einem Ordinalskalenniveau*, *erreicht aber nie vollständig Intervallskalenniveau*: Von einer exakten Gleichheit der Abstände zwischen den Skalenpunkten kann nicht ausgegangen werden, während aber in der Regel dennoch eine Vergleichbarkeit der Abstände gegeben ist, die deutlich über eine reine „größer bzw. kleiner als“-Bedingung hinausgeht.

Zu diesem Thema liefern Bortz und Döring (1995, S. 168 f.) interessante Informationen. Sie legen insbesondere den verbreiteten Trugschluss bloß, dass parametrische Tests mit Daten unter Intervallskalenniveau zu falschen Signifikanz-Ergebnissen führen würden und daher nicht angewandt werden dürften. Vielmehr zeigen Bortz und Döring, dass solche Tests mit Daten beliebigen Skalenniveaus zu korrekten Signifikanzprüfungen führen, wenn die Grundbedingungen wie z.B. Normalverteilung erfüllt sind. Dies betrifft jedoch nur den Signifikanztest selbst! Um auch zu *inhaltlich* interpretierbaren Ergebnissen eines solchen

[1] Eine Unterscheidung zwischen Interrater-Übereinstimmung und Interrater-Reliabilität ist bei Nominaldaten übrigens nicht möglich, da es dann nur noch „richtig“ und „falsch“, absolute oder fehlende Übereinstimmung gibt. Da Urteilsunterschiede sich somit in ihrem Schweregrad nicht mehr unterscheiden können, entfällt auch das Konzept einer Proportionalität von Raterurteilen.

Mittelwertvergleiches gelangen zu können, muss selbstverständlich erst einmal in sinnvoller Weise ein Mittelwert gebildet werden können, und die Skala natürlich auch sorgfältig konstruiert und angewandt worden sein.

Da dies bei den meisten Ratingskalen möglich ist und in der Praxis meist auch ohne Bedenken gemacht wird (ohne z.B. wegen vermeintlichen Ordinalskalenniveaus statt Mittelwerten in den Skalen eines Persönlichkeitsinventars nur mittlere Rangplätze zu berechnen), kann auch eine Testung mit Verfahren parametrischer Art vorgenommen werden. Bortz und Döring raten an o.a. Stelle sogar ausdrücklich dazu, sich nicht unnötig „eines wichtigen, für die Urteiler relativ einfach zu handhabenden Erhebungsinstrumentes“ zu berauben.

Die obigen Ausführungen haben selbstverständlich auch für die Bestimmung der Reliabilität Gültigkeit (Bortz & Döring, 1995, S. 252). So gibt es zur Berechnung der Interrater-Reliabilität bzw. Interrater-Übereinstimmung Verfahren, die verschiedene Skalenniveaus voraussetzen. Im Folgenden werden nach der Erörterung einiger Grundfragen die relevanten Verfahren mit Ordinal- und Intervallskalenniveau vorgestellt.

2.4.3. Grundfragen zur Bestimmung der Interrater-Reliabilität

An dieser Stelle werden für die vorliegende Untersuchung relevante Fragen zur Auswahl der geeigneten Reliabilitätsbestimmung vorgestellt. Prämisse ist hierbei, dass aufgrund der vorangegangenen Ausführung prinzipiell von der Anwendbarkeit von Verfahren für Intervalldaten auf das Diagnostische Interview für Borderlinepatienten ausgegangen werden kann, obwohl dessen dreistufige Ratingskalen sicher kein Intervallskalenniveau erreichen.

Schwerpunkt der Betrachtung wird hier und im Folgenden die Anwendung des *Intraclass-Korrelationskoeffizienten* (auch „Korrelationskoeffizient in Klassen“) in seinen vielfältigen Anwendungsformen sein. Im Folgenden werden außerdem Belege geliefert, die dessen Verwendung rechtfertigen.

Obschon die Mehrzahl der dem Autor bekannten Studien zur Reliabilitätsbestimmung des Diagnostischen Interviews für Borderlinepatienten für Nominaldaten den *Kappa*- und für Ordinaldaten (wie Statements des DIB-R) den *weighted Kappa*-Kennwert benutzt haben (z.B. Zanarini, Frankenburg und Vujanovic, 2002), liegen z.B. mit einer Untersuchung von Gunderson, Kolb und Austin (1981) auch Studien vor, in denen der Intraclass-Korrelationskoeffizient für Intervalldaten erfolgreich Anwendung fand. Für eine allgemeine Übersicht über Reliabilitäts- und Übereinstimmungskoeffizienten verschiedener Skalenniveaus

wird auf die beiden bereits eingangs erwähnten Arbeiten von Tinsley und Weiss (1975) und Asendorpf und Wallbott (1979) verwiesen.[2]

Bei der Messung der Reliabilität auf Intervallskalenniveau sind zunächst einige, unten aufgelistete, Grundfragen zu klären, um den geeigneten Koeffizienten wählen zu können. Diese Fragen werden zunächst erläutert und dann im Hinblick auf die vorliegende Untersuchung beantwortet.

1. *Sollen Unterschiede zwischen den Ratermittelwerten berücksichtigt werden, d.h. spielen systematische, über die Ratings gleichbleibende Abweichungen der Rater voneinander eine Rolle oder nicht?*

Solche Unterschiede gehen im Wesentlichen auf einen unterschiedlichen Bias (durch Antworttendenzen der Rater) zurück, also wenn z.B. Rater A immer höhere Werte angibt als Rater B. Sind bei einem Rating nur die Ordnung der Werte oder deren Intervalle von Interesse, also z.B. im Sinne einer Rangreihenbildung der Subjekte oder einer Ipsatierung, können solche Unterschiede der Rater untereinander getrost vernachlässigt werden und sollten nicht in die Reliabilitäts-Berechnung einbezogen werden, da sonst die Reliabilität unterschätzt wird. Diese Form der Reliabilitätsberechnung wird von Asendorpf und Wallbott (1979) als *adjustierte* Reliabilität bezeichnet und auch von Bortz und Döring (1995, S. 252) vorgeschlagen.

Im Standardfall der Reliabilitätsmessung, der auch auf die vorliegende Untersuchung zutrifft, kann die o.g. Berechungsvariante aber nicht angewandt werden, da Interesse an den *absoluten* Werten besteht. So werden bei Ratingskalen meist aus Items Skalenmittelwerte berechnet, die dann z.B. zu Normwerten oder anderen Stichproben in Beziehung gesetzt werden. In anderen Fällen, wie bei der vorliegenden Untersuchung, werden diagnostische Entscheidungen anhand von festen „Cutting Points“ getroffen, so dass oben beschriebene Antworttendenzen die Diagnosestellung massiv verändern würden. Aus diesem Grunde ist hier unbedingt eine *unjustierte* Reliabilität zu berechnen.[3]

---

[2] Für jede Anforderung sind in der Regel mehrere Berechnungsvarianten des ICC verfügbar, so dass sich ein komplexes und recht unübersichtliches Gefüge möglicher Varianten ergibt (siehe z.B. Armstrong, 1981, dessen Artikel einige der folgenden Fragen entlehnt sind). Leider ist außerdem festzustellen, dass die in der Literatur vorkommende Vielzahl von Berechnungsvarianten stets noch je nach Autor mit verschiedenen Termini, divergierenden Benennungen der Koeffizienten und Bestimmungsgrößen sowie abweichenden zu bedenkenden Problemen und Einschränkungen versehen ist.

[3] Auch aus einem weiteren Grund ist im Regelfall bei der Berechnung einer *adjustierten* Reliabilität Vorsicht geboten: Wird die Entwicklung eines Ratingsystems angestrebt, welches auch von anderen Ratern, als den in der Ratergruppe enthaltenen genutzt werden soll, ist der Erhalt eines generalisierbaren Reliabilitätskoeffizienten das Ziel. In diesem Fall muss die Ratergruppe als eine repräsentative Zufallsauswahl aus allen *möglichen* Ratern betrachtet werden. (Obwohl diese Bedingung genau genommen in der Regel nicht vollständig erfüllt werden kann und zumeist völlig außer Acht gelassen wird, liegt sie einer Generalisierung prinzipiell doch zugrunde.) Asendorpf und Wallbott (1979, S. 245 f.) weisen daraufhin, dass die Generalisierbarkeit bei der üblichen Berechnungsvariante

2. *Wurden die Ratings der Subjekte von denselben identischen Ratern vorgenommen oder variierte das „Set of Judges" ganz oder teilweise?*

Diese vorgenommene Unterscheidung betrifft ebenfalls die Frage der Generalisierung des Ergebnisses. Je nachdem ob die Ratings von denselben oder unterschiedlichen Ratern vorgenommen wurde sind unterschiedliche Berechnungsvarianten im Hinblick auf die Generalisierbarkeit vorzunehmen. Im Falle der vorliegenden Untersuchung liegt klar ein stets (teilweise) unterschiedliches Set von Ratern vor. Hierdurch wird die Wahl von Verfahren zur Interrater-Reliabilität bzw. -übereinstimmung eingeschränkt, da nicht alle Verfahren diese Bedingung tolerieren.

3. *Interessiert die Reliabilität eines einzelnen durchschnittlichen Raters einer Ratergruppe oder die Reliabilität eines über eine Ratergruppe gemittelten Ratings?*

Hier ist zuerst der Zweck der erhobenen Reliabilität zu bedenken. Handelt es sich um eine Studie, die ein Messinstrument einsetzt um inhaltliche Fragestellungen zu klären oder um Hypothesen anhand einer Ratingskala zu testen? In einem solchen Fall ist es, wenn vom Aufwand her vertretbar, immer sinnvoll, aus Gründen der Genauigkeit mit mehreren Ratern zu arbeiten und deren Ratings zwecks Hypothesentestung zu einem Mittelwert zu aggregieren. Dann kann ein Reliabilitätskoeffizient ermittelt werden, der die Reliabilität des gemittelten, durchschnittlichen Raterurteils angibt. Entsprechend der Spearman-Brown-Formel zur Testverlängerung ist dieser Koeffizient abhängig von der Rateranzahl *immer* höher als der eines einzelnen Raters. Bei solchen Untersuchungen sollte der Untersucher nicht einfach nur eine Reliabilität angeben, die er aus dem Testhandbuch entnommen hat, sondern die in der jeweiligen Untersuchung konkret ermittelten Werte angeben, die durchaus höher sein können! Im Falle der Bestimmung der Reliabilität eines Messinstrumentes im Rahmen einer Evaluationsstudie hingegen, wie im vorliegenden Fall, ist immer die zu erwartende Reliabilität eines durchschnittlichen Einzelraters (unter der obigen Bedingung der Generalisierbarkeit) anzugeben.

---

des adjustierten Reliabilitätskoeffizienten nicht gegeben ist: Er ist prinzipiell nur für genau die Ratergruppe gültig, an der er auch erhoben wurde. Dieses „fixed set of judges" stellt dann sozusagen die Population der Rater selbst dar (Armstrong, 1981). Um dennoch verallgemeinern zu können empfehlen Asendorpf und Wallbott (1975, Formel 7) eine Berechnungsmethode von Bartko (1966), in der die Rater als Zufallsfaktor in die Formel einfließen.

4. *Nahmen zwei oder mehrere Rater am Rating teil?*

Ein ebenfalls für die Auswahl eines geeigneten Verfahrens zu berücksichtigender Faktor ist die Rateranzahl. So können einige Verfahren nicht bzw. nur unter erheblich komplizierten Rechenoperationen mit einer Rateranzahl größer als zwei durchgeführt werden (z.B. *weighted Kappa*).

5. *Sind die Ratings vollständig oder gibt es fehlende Werte einzelner Rater?*

In vielen sozialwissenschaftlichen Untersuchungen – und auch im Rahmen der vorliegenden Studie – ist es nicht möglich, solche Ausfälle zum Beispiel aufgrund fehlender Teilnehmer einer Ratergruppe zu vermeiden. In diesen Fällen wurden manche Subjekte oder einzelne Fragen von manchen Ratern nicht eingeschätzt. Dies bedeutet in der Regel unterschiedliche Rateranzahlen pro Subjekt. Die Wahl geeigneter Koeffizienten wird durch diese Tatsache erheblich eingeschränkt. Eine einfache Ergänzung fehlender Werte (z.B. durch Mittelung), wie gelegentlich berichtet wird, ist nicht statthaft, da hierdurch die Reliabilitätsschätzung in der Regel unkontrollierbar verfälscht wird.

2.4.4. Zur Bestimmung des Agreements

Entsprechend der von Tinsley und Weiss (1975) vorgeschlagenen Differenzierung zwischen Reliabilität und Übereinstimmung von Ratern (Agreement) wird hier ein kurzer Überblick über relevante Verfahren gegeben.

Nach der Arbeit von Tinsley und Weiss gibt es nur wenige speziell für die Prüfung der Übereinstimmung entwickelte Verfahren. Ursprünglich wurden hierfür meist eigentlich zu anderen Zwecken entwickelte Verfahren eingesetzt, insbesondere der Anteil prozentualer Übereinstimmung, paarweise Interrater-*Korrelation* sowie verschiedene Chi-Quadrat-Indizes.

Cohen (1960) und Robinson (1957) kritisierten die Verwendung prozentualer Übereinstimmungswerte v.a. aus zwei Gründen: Erstens könne hierbei grundsätzlich *nur* eine absolute Übereinstimmung („all-or-none fashion“) Berücksichtigung finden. Zweitens gebe es keine Kontrolle zufällig zu erwartender Übereinstimmung.

Auch die paarweise Interrater-Korrelation wird von Tinsley und Weiss (1975) für diesen Zweck abgelehnt. Sie wird als Reliabilitäts-, nicht aber als Übereinstimmungsmaß gewertet, da sie rein proportionale Übereinstimmung zeige. (Außerdem gebe es auch hier keinerlei zufallskritische Überprüfung, so dass die einfache paarweise Korrelation als mangelhaft und dem Intraclass-Korrelationskoeffizienten generell unterlegen dargestellt wird.)

Der Einsatz von Chi-Quadrat-Verfahren wird ebenfalls grundsätzlich kritisiert, da hier in der Regel eine Übereinstimmung nicht direkt quantifiziert werden kann. Lediglich kann eine signifikante Abweichung einer Gleichverteilung bzw. Übereinstimmung ausbleiben, die dann aber keinen direkten Schluss auf das Vorliegen und schon gar nicht auf den Grad einer Übereinstimmung zulässt.

2.4.5. Zusammenfassende Beurteilung von Agreement und Reliabilität nach Tinsley und Weiss

Insgesamt scheint die klare Unterteilung von Übereinstimmung und Reliabilität (zumindest in der Darstellung der Autoren) inzwischen wieder aufgeweicht worden zu sein durch Weiterentwicklungen der Verfahren selbst. So sind z.B. für die Interrater-Reliabilität Berechnungsvarianten verfügbar, die nur tatsächliches Agreement berücksichtigen, während gerade auch Verfahren, die von den Autoren unter der Agreement-Kategorie aufgeführt wurden, sich inzwischen durch Gewichtungen von mehr oder weniger großen Abweichungen auszeichnen (s. z.B. weighted Kappa, Koeffizient nach Lawlis & Lu, 1972).

Nicht zuletzt scheint es fragwürdig, den *weighted* Kappa als Agreement-Maß einzustufen, da dieser bei großen Fallzahlen in den ICC-Koeffizienten übergeht und somit zumindest als äquivalent zu betrachten ist (s. u.).

Bemerkenswert und inhaltlich wesentlich bleibt aber die – offenbar selten berücksichtigte und oben dargelegte – zusammenhängende Interpretation von eher Reliabilität oder Übereinstimmung messenden Verfahren; zum einen im Hinblick auf die häufig vorkommende Homogenität von Datensätzen aber auch auf die Frage, ob die möglicherweise formal reliablen Ergebnisse auch wirklich übereinstimmend genug sind, um klinisch-praktisch brauchbar zu sein. (Ähnlich dem Umstand, dass gefundene signifikante Unterschiede nur eine so geringe Effektstärke aufweisen, dass sie keinerlei "klinische Signifikanz" oder Bedeutsamkeit besitzen.)

2.4.6. Geeignete Reliabilitäts-Koeffizienten für die vorliegende Studie

2.4.6.1. Intraclass-Korrelationskoeffizient (ICC)

Im Endeffekt fällt nach der Erwägung aller o.g. Fragen die Beantwortung eindeutig aus.

Der Intraclass-Korrelationskoeffizient nach der Formel

$IC_u = (MS_b - MS_w) / (MS_b + MS_w * (N - 1))$

(in der von Asendorpf und Wallbott [1979, Formel 2] verwendeten Schreibweise, wobei MS_b die Varianz zwischen den Items, MS_w die Varianz zwischen den Raterurteilen und N die Anzahl der Rater wiedergibt) ist die einzige Berechnungsvariante, die von allen bisher in diesem Kap.

genannten Autoren zur Verwendung bei quantitativen Daten *und* zur Generalisierung bei mehreren unterschiedlichen Ratern zum Erhalt der Reliabilität eines einzelnen durchschnittlichen Raters empfohlen wird.

So schreiben Tinsley und Weiss (1975, S. 373):

„The intraclass correlation (R, Equation 4) [entspricht obiger Formel des IC_u; Anm. d. Autors] is recommended as the best measure of interrater reliability available for ordinal and interval level measurement."

Auch Bartko und Carpenter (1976) empfehlen diese Berechnungsform des ICC für quantitative Daten bei zwei oder mehr Ratern und empfehlen ihn *ebenfalls bei dichotomen Daten im Mehr-Rater-Fall,* was bei der vorliegenden Untersuchung für die durch das Diagnostische Interview für Borderlinepatienten gestellte Ein- bzw. Ausschluss-Diagnose zutrifft.

Insbesondere aber ist der Intraclass-Korrelationskoeffizient der einzige heute in vielen Studien verwendete Reliabilitätskoeffizient, der nach Anwendung einer einfachen Korrekturformel, fehlende Ratings pro Subjekt toleriert. Hierzu ist die Anzahl der Rater N einfach nach Bartko und Carpenter (1976, S. 316) zu korrigieren (vgl. Horst-Koeffizient, Kap. 2.4.6.4).
Zur Korrektur wird N_0 nach der folgenden Formel E5 (bei Bartko und Carpenter als R_0 bezeichnet) berechnet:

$$N_0 = \left[M - \sum_{i=1}^{N} n_i^2 / M \right] / (N-1)$$

Somit handelt es sich bei diesem Intraclass-Korrelationskoeffizienten um den idealen Reliabilitäts-Koeffizienten für die Reliabilitätsbestimmung im Falle der vorliegenden Untersuchung für alle Statements, Summen- und Skalierten Section-Scores sowie für die Stellung der Ein- bzw. Ausschlussdiagnose Borderline-Persönlichkeitsstörung selbst.
Da die Korrekturformel von Bartko und Carpenter nicht im SPSS (bis Version 10.1) verfügbar war, musste vom Autor ein Excel-Programm angefertigt werden, mit welchem die Berechnung vorgenommen werden konnte.

2.4.6.2. Zum Verhältnis „weighted Kappa" und ICC

Zur Verwendung des ICC seien zusätzlich zur eingangs erwähnten Stellungnahme von Bortz und Döring (1995) zur Verwendung des ICC bei Ordinaldaten noch zwei weitere Quellen angegeben.
Fleiss und Cohen (1973) gehen der Frage nach, in wieweit weighted Kappa-Werte von Ordinaldaten, die ein Verhältnis der Übereinstimmung darstellen, mit ICC-Koeffizienten verglichen werden können, da diese ja ein Varianzenverhältnis widerspiegeln. Die Autoren stellen zu diesem Bemühen fest:

„An affirmative answer would provide a useful bridge over the gap between those two different levels of measurement.“ (Fleiss & Cohen, 1973, S. 615)

Sie kommen zur Antwort, dass der ICC einen Spezialfall des weighted Kappa für den Intervallskalenfall darstellt.

So legt auch Rae (1988) in der Form eines allgemeinen Beweises dar, dass der weighted Kappa-Koeffizient im Mehr-Rater-Fall für quantitative Daten bei zunehmendem N kontinuierlich in den ICC übergeht, ihm also äquivalent ist. (Voraussetzung für die o.g. Äquivalenzen von Rae sowie Fleiss und Cohen ist die Verwendung der üblichen *quadratischen* Gewichtung der Abweichungen beim weighted Kappa.)

Es sei noch einmal angemerkt, dass der ICC im Mehrraterfall viel einfacher als der weighted Kappa zu berechnen ist, da keine Mittelung aller einzeln zu berechnenden paarweisen Inter-Rater-Übereinstimmungen nötig ist.

2.4.6.3. Finn-Koeffizient

Da, wie bereits dargelegt, ein Intraclass-Korrelationskoeffizient insbesondere bei Varianzeinschränkung des Merkmals aufgrund zu hoher Homogenität zur Interpretation nicht ausreichend erscheint, sollte zu Vergleichszwecken zusätzlich wie von Tinsley und Weiss (1975) empfohlen ein *Finn-Koeffizient* berechnet werden. (Die Berechnungsvorschrift wurde ebenfalls in das Excel-Programm integriert.)

Finn (1970, zit. nach Tinsley & Weiss, 1975, Formel 1 und 2) hat für Ordinaldaten folgenden Index zur Bestimmung der Interrater-Reliabilität vorgeschlagen:

(1) $$r = 1.0 - \frac{Observed\ Variance}{Chance\ Variance}$$

Die „zufällig zu erwartende Varianz“ (s_c^2) ist die bei einem von den Ratern rein zufällig ausgeführten Rating zu erwartende Varianz. Sie kann wie folgt berechnet werden:

(2) $$s_c^2 = \frac{\left(k^2 - 1\right)}{12}$$

(wobei k die Anzahl der möglichen Skalenwerte bzw. -kategorien darstellt)

Die beobachtete Varianz ist die Varianz innerhalb bzw. zwischen den Ratern.

Der Grad, in dem die beobachtete Varianz der Ratings kleiner als die Zufallsvarianz ist, erlaubt den Rückschluss auf den Anteil der nicht-zufälligen Varianz der Ratings: So gibt das Verhältnis der beobachteten und der Zufallsvarianz von 1 subtrahiert den Anteil der nicht-zufälligen Varianz der Ratings an.

Der Finn's *r* hat einen Range zwischen 1 und 0, wobei 0 *völlige* Unreliabilität bedeutet, d.h. die Ratings variieren genauso stark wie rein zufällige Ratings.

Vorher war festgestellt worden, dass Homogenität im einzuschätzenden Merkmal (also eine reduzierte Varianz der Ratings) üblicherweise eine verringernde Wirkung auf die Reliabilitätsschätzung hat. Diesen Nachteil weist der Finn-Koeffizient nicht auf: da wie oben beschrieben die Schätzung der Zufallsvarianz völlig unabhängig vom tatsächlichen Rating und den Daten erfolgt und einzig auf der Basis der Kategorienanzahl des Ratingsystems ermittelt wird, ist dessen Höhe hiervon nicht berührt.

Diese Unabhängigkeit hat jedoch ihren Preis: Die Schätzung der Zufallsvarianz ist im Normalfall etwas überhöht, da sie eine gleichwahrscheinliche Verwendung aller Kategorien der Ratingskala durch die Rater und z.B. ein völliges Fehlen von Antworttendenzen voraussetzt. Dieses ist in der Regel nicht gegeben und so führt die Überschätzung der Zufallsvarianz nach der Formel zur Berechnung des Finn-Koeffizienten zu einer möglichen Erhöhung der Reliabilitätsschätzung. Tinsley und Weiss (1975) weisen aber darauf hin, dass Finn's *r* in den von Finn (1970, 1972) durchgeführten Vergleichsstudien – bis auf Ausnahmen – nicht zu Überschätzungen sondern zu mit dem Intraclass-Korrelationskoeffizient vergleichbaren Ergebnissen geführt habe.

So raten Tinsley und Weiss (1975) zur Verwendung des Finn-Koeffizienten zusätzlich zum ICC – insbesondere um Fälle eingeschränkter Varianz erkennen und sowohl Auswirkung als auch Ursachen beurteilen zu können.

Auch bezüglich fehlender Ratings pro Subjekt, also Raterausfällen, ist der Finn-Koeffizient ein für diese Studie geeignetes Maß: Es kann dennoch eine Berechnung durchgeführt werden, da lediglich die Varianz zwischen den Ratern beim Finn-Koeffizienten aus den konkreten Rating-Daten zu berechnen ist. Die Berechnung hängt dabei *nur* von der Anzahl der untersuchten Objekte, der *Anzahl* der Ratings *pro* Subjekt und den jeweils vergebenen Werten ab. Er bedarf aber nicht einer gleichen Anzahl der Ratings pro Subjekt.

Diese Eigenschaften machen ihn ideal zur ergänzenden Verwendung zusätzlich zum ICC.

Beim Beispielsdatensatz in Tab. 8, Fall 3 wäre die Reliabilität unter Verwendung des Finn-Koeffizienten mit r=.93 übrigens ebenfalls hoch ausgefallen (in Fall 1: r=1.0, Fall 2: r=.40).

2.4.6.4. Horst-Koeffizient

Der Horst-Koeffizient war in der Studie von Eckert et al. (1987) in einem ähnlichen Anwendungsfall benutzt worden. Er wird wie der ICC aus Varianzenverhältnissen berechnet. In der vorliegenden Studie wird er zu Vergleichszwecken im Anwendungsfall mitberechnet. Die Formel und weitere Informationen sind einer Veröffentlichung von Langer und Schulz v. Thun (1974, Formel 42 auf S. 91) zu entnehmen.

2.4.6.5. Prozentuale Übereinstimmung

Zur Quantifizierung der *absoluten* Übereinstimmung mehrerer Rater in der Einschätzung des dichotomen Merkmals "Vorliegen einer Borderline-Diagnose" nach DIB-R (in Kap. 6.1.3.3 und 6.2.3.3) schien die Berechnung einer beobachteten Übereinstimmung in Prozent (ohne Relativierung an der zufällig zu erwartenden Übereinstimmung) sinnvoll. Hierbei war wiederum ein Verfahren zu finden, welches auch bei einer *wechselnden* Anzahl von Ratern anwendbar ist. Tinsley und Weiss (1975) beschreiben eine solche Berechnungsmöglichkeit auf S. 372 ihrer Veröffentlichung (Formel 16).

2.4.7. Klassifizierung der Höhe von Reliabilitätskoeffizienten

Bei gründlicher Literaturdurchsicht lassen sich wenig eindeutige Festlegungen, dafür aber eine Vielzahl unterschiedlicher Verwendungen ausmachen. Eine durchgehende und einhellige Meinung darüber, ab wann *genau* eine Reliabilität als ausreichend zu betrachten sei, existiert offenbar nicht.

Dies liegt u.a. daran, dass es sehr unterschiedliche Berechnungsmethoden, Erhebungsdesigns und auch Verwendungszwecke gibt. So fallen die Anforderungen an die Beurteilung von Gruppenmittelwerten anders aus als bei der Beurteilung von Testwerten im Einzelfall.

Bortz und Döring (1995, Tafel 27) bezeichnen das Interrater-Reliabilitätsmaß Cohen's Kappa bereits ab κ=.70 als gut – nach einer persönlichen Mitteilung von Bortz (2000, per E-Mail) gelte dies auch für den Intraclass-Korrelationskoeffizienten.

Nach Lienert und Raatz (1998, S. 269) wird zur Beurteilung individueller Differenzen vorsichtiger ein Reliabilitätskoeffizient von $r_{tt} \geq .70$ als "eben noch ausreichend" bezeichnet. Fleiss (1981) bezeichnet hingegen schon Reliabilitäten von Cohen's $\kappa \geq .40$ als ausreichend für die psychiatrische Diagnosestellung und solche von $\kappa \geq .70$ bereits als ausgezeichnet. Auch Wegener (1976, S. 123) bezeichnet Reliabilitäten bereits ab .60 als hoch. Allgemein werden Reliabilitäten <.40 aber als nicht genügend bezeichnet.

*Für die Schilderung und Einordnung der Ergebnisse in Kap. 6 "Ergebnisse zur Reliabilität des DIB-R" der vorliegenden Studie wird hiermit folgender **einheitlicher Maßstab** festgelegt:*

- **Reliabilitäten <.50** werden als **ungenügend** oder **unbefriedigend** bezeichnet
- **Reliabilitäten ≥.50** sind **ausreichend** bzw. **befriedigend**
- **Reliabilitäten ≥.70** werden als **gut** klassifiziert
- **Reliabilitäten ≥.90** gelten als **ausgezeichnet** bzw. **sehr gut**

3. METHODISCHES VORGEHEN

3.1. DAS PROJEKT "DIB-R-EVALUATION"

Die Durchführung der Untersuchung begann Ende 1995 durch die Hamburger Arbeitsgruppe zur DIB-R-Evaluation, bestehend aus Prof. Dr. J. Eckert, Dipl.-Psych. D. Brodbeck, Dipl.-Psych. M. Schödlbauer und Dipl.-Psych. M. Wuchner (Psychologisches Inst. III, Universität Hamburg) sowie Dipl.-Psych. E.-M. Biermann-Ratjen und Dipl.-Psych. R. Ladendorf (Klinik und Poliklinik für Psychiatrie und Psychotherapie, Universitätsklinikum Hamburg-Eppendorf).
Die Projektleitung lag bei Prof. Dr. J. Eckert und Dipl.-Psych. D. Brodbeck, dem auch die Planung und Organisation der Untersuchungen und Gesprächstermine oblag. Die Datenerhebung selbst (in Form der SKID-II und DIB-R-Interviews sowie der Videoaufnahmen und Gruppenratings) wurde in unterschiedlichen Anteilen von sämtlichen Mitgliedern der Hamburger Arbeitsgruppe bewältigt. Weitere an der Erhebung beider Interviews beteiligte Personen waren Dipl.-Psych. N. Lanzoni und Dipl.-Psych. B. Collmann.
Die Projektdurchführung erfolgte in enger Abstimmung mit Frau Prof. Dr. Chr. Rohde-Dachser des Institutes für Psychoanalyse der JWG Universität Frankfurt, die auch die deutsche Übersetzung des DIB-R in der 5. Auflage ihres Buches "Das Borderline-Syndrom" 1995 veröffentlicht hat (Rohde-Dachser, 1995).

3.2. GEPLANTER AUFBAU DES PROJEKTS

3.2.1. Validitätsprüfung

Zum Zweck der Vailiditätsüberprüfung des DIB-R sollen 100 psychiatrische Patienten der Klinik und Poliklinik für Psychiatrie und Psychotherapie des Universitätsklinikums Hamburg-Eppendorf mittels des DIB-R untersucht und zum Vergleich eine komplette Diagnostik mittels der SKID-Interviews für die Achsen I und II des DSM-III-R bzw. (nach dessen erwartetem Erscheinen) des DSM-IV durchgeführt werden. Diagnosen der Behandler – wie z.B. in der Validitätsstudie zum DIB von Eckert et al. (1991) als Kriterium verwandt – werden nicht berücksichtigt. Über das Ergebnis der Untersuchung soll in der Regel ein schriftlicher diagnostischer Bericht verfasst werden, der den behandelnden Ärzten bzw. Psychologen zur Verfügung gestellt wird.

Bezüglich der Umstellung auf die neue Version DSM-IV werden keine substanziellen Änderungen des DSM-Systems insbesondere im Bereich der Persönlichkeitsstörungen erwartet.

Eine Umstellung auf die neue Version soll auf jeden Fall gleich nach Erscheinen der Interviewleitfäden vorgenommen werden, da die Patientenrekrutierung und das Interesse der behandelnden Ärzte und Psychologen von der Aktualität der Diagnose abhängt.

Die Validierungs-Stichprobe muss bereits bei der Rekrutierung wegen ihrer nötigen Eignung zur differentialdiagnostischen Einschätzung des DIB-R und der hierzu nötigen Bildung von Diagnosegruppen anhand von Verdachtsdiagnosen seitens der Behandler vorselegiert werden. Der angestrebte Diagnoseschlüssel soll nach der SKID-Diagnose auf den Achsen I und II ca. ein Drittel Patienten mit der DSM-Diagnose *Borderline-Persönlichkeitsstörung* betragen, die weiteren Patienten sollen etwa zu gleichen Teilen *Andere* (d.h. non-borderline) *DSM-Persönlichkeitsstörungen*, *Psychotische Störungen* und *(möglichst gleichermaßen sowohl bipolare als auch depressive) Affektive Störungen* umfassen.

Eine Validierung der DIB-R-Diagnose wird v.a. angestrebt durch eine

a) Überprüfung der Diagnoseübereinstimmung mit der SKID-II-Diagnose Borderline-Persönlichkeitsstörung

b) Überprüfung der Güte der durch das DIB-R erreichten differentialdiagnostischen Abgrenzung der Patienten mit der DIB-R-Diagnose "Borderline-Persönlichkeitsstörung" von den Patienten der nach unten beschriebenem Vorgehen (Kap. 3.5.3) nach DSM-Diagnosen gebildeten übrigen (also non-borderline) DSM-Diagnosegruppen

3.2.2. Reliabilitätsprüfung

Die Güte des DIB-R soll in Form der Interrater-Reliabilität (bzw. -Übereinstimmung) anhand von videographierten DIB-R-Interviews von insgesamt 30 Patienten in einer Gruppe von Experten überprüft werden. Hierzu werden die Aufnahmen in der Ratergruppe gemeinsam gesehen. Von jedem Rater wird dabei anhand des Videos *unabhängig* ein DIB-R-Interview-Bogen ausgefüllt.

Eine entsprechende Überprüfung der Ergebnisse soll außerdem vergleichsweise an einer in der Anwendung des DIB-R geschulten Gruppe von studentischen Ratern vorgenommen werden.

Die Übereinstimmung der Rater untereinander wird dann (entsprechend der detaillierten Ausführungen im Kap. 2.4.6) v.a. anhand des Intraclass-Korrelationskoeffizienten und einiger zu einem besseren Verständnis zusätzlich erhobener Koeffizienten quantifiziert werden.

Der Anteil der mittels des DIB-R als Patient mit Borderline-Persönlichkeitsstörung diagnostizierten Personen an den beurteilten Videos sollte etwa ein Drittel erreichen, um eine

ausreichende Variabilität der Merkmale (Diagnose Borderline-Persönlichkeitsstörung sowie diagnostische Kriterien) zur Reliabilitätsprüfung sicherzustellen.

3.2.3. Auswertung der Daten

Die gesamte Eingabe und Auswertung der Daten soll durch Herrn Dipl.-Psych. D. Brodbeck erfolgen, der hierin z.T. durch Diplomanden unterstützt wird. Die Auswertung wird in erster Linie mit dem Analyseprogramm SPSS in der jeweils aktuellsten erhältlichen Version erfolgen.

3.3. FRAGESTELLUNG UND HYPOTHESEN

In diesem Kapitel werden die grundlegenden Fragestellungen der Arbeit gruppiert, d.h. nach Validität, Reliabilität und allgemeinen Fragen zusammengefasst, vorgestellt. Im Anschluss werden die formulierbaren spezifischen Hypothesen dargelegt.

3.3.1. Validität

3.3.1.1. Fragestellungen zur Validität

Im Vordergrund der Validitätsüberprüfung des DIB-R steht die Überprüfung der Konstruktvalidität (konvergente und diskriminante Validität) des dem DIB-R zugrundeliegenden Borderline-Konzepts.

Die relevanten Fragestellungen hierzu lauten:

- Stimmen die Testergebnisse des DIB-R mit denen des SKID-II als einem alternativen, bereits überprüften und verbreiteten Instrument zur Diagnose der Borderline-Persönlichkeitsstörung überein? Weist das DIB-R also eine ausreichende *konvergente* Validität (bzw. Übereinstimmungsvalididät) auf, so dass von einer prinzipiellen Übereinstimmung beider Messinstrumente ausgegangen werden kann?[4]
- Unterscheiden sich die Testergebnisse des DIB-R deutlich von denen konstruktfernerer diagnostischer Instrumente wie z.B. den Testergebnissen der weiteren spezifischen Persönlichkeitsstörungen nach SKID-II? Weist das DIB-R also eine ausreichende *diskriminante* Validität auf, so dass von einer hinreichenden Unterscheidungskraft im Hinblick auf andere Konstrukte gesprochen werden kann?

[4] Zur Klarheit sei noch einmal erwähnt, dass eine perfekte Übereinstimmung grundsätzlich *nicht* erwartet wird, da die den Tests zugrundeliegenden Konstrukte der Borderline-Persönlichkeitsstörung nicht identisch sind und kein objektives Außenkriterium "Borderline-Persönlichkeitsstörung" bestimmt werden kann (siehe Kap. 2.2.6). Ebenso divergieren die Ein- und Ausschlüsse einzelner Aspekte der abzugrenzenden spezifischen anderen Persönlichkeitsstörungen im DIB-R und DSM-Konstrukt der Borderline-Persönlichkeitsstörung erheblich.

3.3.1.2. Spezifische Hypothesen zur Validität

Hypothese 1.1: Die Übereinstimmung der Borderline-Diagnose nach DIB-R und SKID-II ist berechnet nach Cohen's Kappa (κ) signifikant überzufällig häufig.

Hypothese 1.2: Das Vorliegen der Borderline-Diagnosen nach DIB-R und SKID-II korreliert, berechnet mittels des φ -Koeffizienten, signifikant und erreicht einen Wert von $\varphi \geq .50$.

Hypothese 2.1: Die Verteilung der Patienten mit und ohne DIB-R-Borderline-Persönlichkeitsstörung ist signifikant unterschiedlich über die definierten DSM-Diagnosegruppen. Die häufigste Kombination ist dabei die der *BPS nach DIB-R* und *BPS nach SKID-II*, die zweithäufigste *BPS nach DIB-R* und *Andere Persönlichkeitsstörungen*.

Hypothese 2.2.a): Die DIB-R-Gesamt-Scores der Patienten unterscheiden sich über die verschiedenen Diagnosegruppen nach SKID-II signifikant voneinander.

Hypothese 2.2.b): Die Patienten der SKID-II Diagnosegruppe Borderline-Persönlichkeitsstörung unterscheiden sich im DIB-R-Gesamt-Score jeweils signifikant von den Patienten jeder anderen Diagnosegruppe und haben den höchsten Wert.

Hypothese 2.3.a): Für jeden der vier Skalierten Section-Scores (SSS) der Bereiche des DIB-R gilt: Über die verschiedenen Diagnosegruppen nach SKID-II sind die Mittelwerte des SSS der Patienten signifikant unterschiedlich.

Hypothese 2.3.b): Für jeden der vier Skalierten Section-Scores (SSS) der Bereiche des DIB-R gilt: Die Patienten der SKID-II-Diagnosegruppe Borderline-Persönlichkeitsstörung unterscheiden sich im Mittelwert des SSS signifikant von den Patienten aller anderen Diagnosegruppen und haben jeweils den höchsten Wert.

3.3.2. Reliabilität

3.3.2.1. Fragestellungen zur Reliabilität

Der Überprüfung der Zuverlässigkeit des DIB-R liegt die folgende Fragestellung zugrunde: Kommen verschiedene Diagnostiker (Rater) bei denselben Patienten auch zu gleichen bzw. hinreichend ähnlichen Ergebnissen in der Diagnosestellung? Mit anderen Worten formuliert lautet die Frage: Ist das diagnostische Ergebnis des DIB-R stark von den jeweiligen Patienten, aber nur unwesentlich vom durchführenden Diagnostiker abhängig?

In diesem Zusammenhang taucht außerdem die Frage auf, inwiefern sich eine quantitativ und qualitativ unterschiedliche klinische Erfahrung und Ausbildung in der Anwendung des DIB-R auf die Genauigkeit der DIB-R-Ergebnisse auswirkt, die durch eine vergleichende Gegenüberstellung der Ergebnisse der Rater der Studentengruppe mit denen der Experten untersucht wird.

Im Rahmen der Studie werden außerdem noch eine Vielzahl einzelner – vorab nicht spezifizierter – Fragestellungen zur allgemeinen Anwendbarkeit und Praktikabilität des DIB-R sowie zu Gemeinsamkeiten und Unterschieden der DIB-R- und SKID-II-Diagnosen untersucht werden.

3.3.2.2. Spezifische Hypothesen zur Reliabilität

Hypothese 3.1: Die Interraterübereinstimmung ist in der Expertengruppe, gemessen mit dem Intraclass-Korrelationskoeffizienten, für die DIB-R-Diagnose signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Hypothese 3.2: Die Interraterübereinstimmung ist in der Expertengruppe, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den DIB-R-Gesamt-Score signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Hypothese 3.3.a): Die Interraterübereinstimmung in der Expertengruppe ist, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den Skalierten Section-Score des DIB-R-Bereichs "Affekte" signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Hypothese 3.3.b): Die Interraterübereinstimmung in der Expertengruppe ist, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den Skalierten Section-Score des DIB-R-Bereichs "Kognition" signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Hypothese 3.3.c): Die Interraterübereinstimmung in der Expertengruppe ist, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den Skalierten Section-Score des DIB-R-Bereichs "Impulsivität" signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Hypothese 3.3.d): Die Interraterübereinstimmung in der Expertengruppe ist, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den Skalierten Section-Score des DIB-R-Bereichs "Zwischenmenschliche Beziehungen" signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Für die Gruppe der studentischen Rater werden keine spezifischen Hypothesen formuliert.

3.4. VERWENDETE DIAGNOSEINSTRUMENTE

3.4.1. DIB-R

Das zu validierende halbstrukturierte Interview **DIB-R** wurde bereits im Kap. 1.2.4.1 zur Geschichte und dem aktuellen Stand der Borderline-Persönlichkeitsstörung eingehend beschrieben.

Für die vorliegende Studie wurde eine geringfügig durch Kommentare erweiterte Projektversion verwendet, die im Anhang einzusehen ist. Bei diesen Kommentaren handelt es sich v.a. um Präzisierungen bzgl. der Einschätzung sowie um einige Beispiele. Diese Kommentare wurden ohne Eingriff in das eigentliche Interview auf den jeweils mit "a" versehenen Rückseiten eingefügt. (Die Kommentare z.B. zur Seite 13 des Interviews finden sich auf Seite 13a wieder.) Die Kommentare greifen in keiner Weise in die rechnerische Auswertung des DIB-R ein.

Zusätzlich eingefügt worden waren auch zwei Skalen am Ende des Interviews:

- Sicherheit der Einschätzung
- Vollständigkeit der Information bzw. Güte der Exploration

Beide Skalen erwiesen sich als wenig praktikabel und schlecht einschätzbar. Deren Einschätzung wurde im Laufe der Studie fallengelassen und nicht systematisch ausgewertet.

3.4.2. SKID: Strukturiertes Klinisches Interview für das DSM-III-R bzw. -IV

Das SKID ist ein **halbstrukturiertes Interviewverfahren** zur Diagnostik psychischer Störungen (Achse I) und Persönlichkeitsstörungen (Achse II) nach DSM-III-R bzw. aktuell DSM-IV. Die folgenden Erläuterungen unterscheiden sich für beide DSM-Versionen nicht relevant. Das Verfahren kann sowohl im ambulanten als auch im stationären Bereich eingesetzt werden. Zusätzlich zu den genau festgelegten Fragen des Leitfadens steht es dem Beurteiler frei, eigene Fragen zur Vertiefung zu stellen.

Eine Durchführung bei akut psychotischen sowie dementen Patienten ist meist nicht manualgerecht möglich. Das Verfahren wurde für Erwachsene ab 18 Jahren entwickelt. Das SKID-Verfahren liegt neben dem amerikanischen Original für DSM-III-R (Spitzer et al., 1990a, 1990b) und DSM-IV (First et al., 1996, 1997) in Übersetzungen in zahlreiche Sprachen vor. Die in der vorliegenden Studie verwendete deutsche Fassung des DSM-III-R wurde von Wittchen et al. (1991, 1993), die des DSM-IV von Wittchen, Zaudig und Fydrich (1997) veröffentlicht.

Neben den beiden o.g. Hauptachsen gibt es noch *drei* weitere zu beurteilende Skalen, um eine umfassende Einschätzung des Patienten zu ermöglichen. Insbesondere wird eine Skala zur

globalen Erfassung des Funktionsniveaus (Achse V) vorgegeben, anhand derer die derzeit vorhandene Beeinträchtigung beurteilt werden soll. Zusätzlich sollen evtl. relevante **körperliche Erkrankungen** (Achse III) und psychosoziale Beeinträchtigungen (Achse IV) benannt werden. Außerdem sollen die **Achse I-Hauptdiagnose** und der geschätzte **Sicherheitsgrad** der diagnostischen Entscheidung angegeben werden.
Als Voraussetzung zur Anwendung werden von den Autoren im Testhandbuch klinisch-psychiatrische Erfahrung, die Kenntnis des Klassifikationsmanuals, die Ausbildung als Psychologe oder Psychiater und die Teilnahme an einer zweitägigen SKID-Schulung angegeben.

Das SKID liegt in zwei vollständig getrennten Fassungen bzw. Bereichen vor:

- **SKID-I:** für psychische Störungen (Achse I-Störungen des DSM-III-R bzw. -IV) (angegebene Durchführungszeit 100 min.)
- **SKID-II:** für Persönlichkeitsstörungen (Achse II-Störungen des DSM-III-R bzw. -IV) (angegebene Durchführungszeit 30 min.)

3.4.2.1. SKID-I

Das SKID-I beginnt mit einem wenig strukturierten Teil der **Exploration**, in welchem allgemeine Informationen über den Patienten erhoben werden. Mittels eines kurzen Explorationsleitfadens wird ein Überblick über derzeitige und frühere Beschwerden bzw. Symptome des Patienten gewonnen; dies kann für eine bessere Bewertung und Kodierung der in den folgenden Sektionen erhobenen Informationen hilfreich sein.
Anschließend wird das halbstrukturierte Interview zur Achse I durchgeführt. Dabei durchläuft der Interviewer die **Sektionen A bis J**, die jeweils Fragen zu verschiedenen Störungen enthalten. Zusätzlich zu den genau festgelegten und den ergänzend frei zu stellenden Fragen können auch Angaben von Angehörigen, Verhaltensbeobachtungen u.ä. zur Einschätzung der genau formulierten diagnostischen Kriterien herangezogen werden.

Anhand der Antworten auf die Fragen überprüft der Interviewer, ob ein **diagnostisches Kriterium** erfüllt ist; je nach Antwort findet er einen Verweis zu den nächsten Fragen.

Folgende Sektionen bzw. Störungsbereiche werden im SKID-I (nach DSM-IV) erfasst:

- *Sektion A – Affektive Syndrome:* Kodierung affektiver Symptome und Syndrome (Major Depression, Manie, Dysthymie), noch keine Diagnosestellung (außer Dysthymie)
- *Sektion B – Psychotische Symptome:* Kodierung psychotischer Symptome, Registrierung des zeitlichen Verlaufs, noch keine Diagnosestellung
- *Sektion C – Differentialdiagnose psychotischer Störungen:* Diagnosestellung psychotischer Störungen (z.B. Schizophrenie, schizoaffektive Störung, Wahn)
- *Sektion D – Differentialdiagnose affektiver Störungen:* Diagnosestellung rein affektiver Störungen und affektiver Störungen mit psychotischen Merkmalen
- *Sektion E – Missbrauch und Abhängigkeit von psychotropen Substanzen:* Beurteilung der DSM-IV Kriterien für Missbrauch und Abhängigkeit von Alkohol, Drogen und Medikamenten, Fragen zum Störungsverlauf
- *Sektion F – Angststörungen:* Diagnosestellung von Angststörungen (z.B. Panikstörung, Agoraphobie, Soziale Phobie, Generalisierte Angststörung, Zwangsstörung), zeitlicher Verlauf
- *Sektion G – Somatoforme Störungen:* Beurteilung der DSM-IV Kriterien für Somatoforme Störungen (Somatisierungsstörung, Hypochondrie, Schmerzstörung, Körperwahrnehmungsstörung)
- *Sektion H – Essstörungen:* Beurteilung der DSM-IV Kriterien für Essstörungen (z.B. Anorexia Nervosa, Bulimia Nervosa)
- *Sektion I – Anpassungsstörungen:* Beurteilung der DSM-IV Kriterien für Anpassungsstörungen (z.B. depressive Verstimmung, ängstliche Gehemmtheit, Verhaltensstörung)
- *Sektion J – Optionale Störungen:* z.B. Stressstörungen, manische Episode (seit 1997 im SKID)

Die Gesamtbewertung des Interviewers wird kodiert; es gibt **vier Möglichkeiten der Kodierung**:

- ? = unsicher / zu wenig Informationen
- 1 = nein / nicht vorhanden
- 2 = vorhanden, aber nicht kriteriumsmäßig
- 3 = sicher vorhanden und kriteriumsmäßig

Um jedes Störungsbild abzuklären, sollten *alle* Sektionen durchgegangen werden. Exakt definierte Sprungregeln sorgen dafür, dass nur diagnostisch relevante Fragen gestellt und z.B. a priori auszuschließende Störungen oder Störungsbereiche nicht exploriert werden.

Auf dem umfangreichen Kodierblatt erfolgt die abschließende Bestimmung der Diagnosen, zu denen häufig auch Verlaufstyp, Schweregrad und Dauer angegeben werden können. Es können nach dem Komorbiditätsprinzip grundsätzlich mehrere Diagnosen gestellt werden, der Interviewer ist aber gehalten, schließlich zu *einer Achse-I-Hauptdiagnose* zu kommen.

3.4.2.2. SKID-II

Das SKID-II-Interview wurde bereits im Kap. 1.2.3.3 vorgestellt. Es ist – als grundsätzlich eigenständiges Verfahren zur Erfassung von Persönlichkeitsstörungen – nur dann sinnvoll einsetzbar, wenn differenzierte Informationen zu Störungen des Patienten auf der Achse I vorliegen. Daher sollte, wenn irgend möglich, das SKID-I-Interview immer *vorab* durchgeführt werden.

Den SKID-Interviews liegt entsprechend dem DSM-III-R- und DSM-IV-System das Prinzip eines deskriptiven Ansatzes mit mehrdimensionaler Sichtweise zu Grunde. Eine abschließende diagnostische Einschätzung muss daher sowohl die Achse-I- als auch die Achse-II-Diagnosen in einen sinnvollen Zusammenhang stellen und dabei auch die o.g. weiteren Achsen III – V berücksichtigen. Von einer isolierten Diagnosestellung mittels des SKID-II ist abzuraten.

3.5. DURCHFÜHRUNG DER UNTERSUCHUNG ZUR VALIDITÄT

Die Durchführung der Untersuchung erfolgte an der Klinik und Poliklinik für Psychiatrie und Psychotherapie des Universitätsklinikums Hamburg-Eppendorf.

3.5.1. Erhebung der Stichprobe

Im Vorfeld der Studie waren Ärzte und Psychologen der Klinik hausintern per Rundschreiben sowie in persönlichen Gesprächen unserer in der Klinik tätigen Projektmitarbeiter über das Forschungs-Projekt informiert worden. Um die verschiedenen Ärzte und Psychologen der Stationen zur Mitarbeit anzuregen, boten wir unsere standardmäßige Durchführung der kompletten SKID-I- und -II-Diagnostik als eine im Klinikalltag willkommene, unterstützende "Dienstleistung" an. Zu diesem Zwecke fertigte der jeweilige Untersucher für die meisten Patienten ein ausführliches diagnostisches Gutachten an. Diese Unterstützung wurde in erheblichem Umfang in Anspruch genommen und war ein starker Anreiz, dem Projekt Patienten zukommen zu lassen.

Außerdem führten unsere im Hause tätigen Projektmitarbeiter in möglichst vielen Fällen, in denen eine ausführliche Diagnostik angezeigt schien, selbst die Probanden dem Projekt zu. Sie führten dann meist auch eines der beiden Interviews durch.

Darüber hinaus wurden z.T. die in der Klinik durchgeführten Fallseminare zur Diagnostik der Borderline-Persönlichkeitsstörung im Rahmen des Studentenunterrichts des Psychologischen Institutes III als Rekrutierungsquelle genutzt. Alle in den Fallseminaren untersuchten Patienten werden dort von den Seminarleitern standardmäßig mittels des DIB-R interviewt. Nach Möglichkeit (und wenn für die Behandlung sinnvoll), wurde versucht, mit dem Patienten auch eine SKID-Diagnostik durchzuführen.

Für die Validitätsstudie verwertet werden konnten aber nur Probanden, bei denen eine vollständige SKID-I- und -II-Diagnostik vorlag. Z.B. in Folge von Entlassungen, Terminproblemen oder Verweigerung einer weiteren Untersuchung konnten etwa 20% der begonnenen Untersuchungen nicht abgeschlossen werden.

Die Untersuchungen zur Überprüfung der Validität begannen im September 2005 und endeten im Januar 2001 mit dem Erreichen des 100. Patienten.

3.5.2. Das Setting der Untersuchung und Diagnosestellung im UKE

Die Untersuchung der Patienten nach SKID und DIB-R erfolgte aus Objektivitätsgründen immer unabhängig durch je zwei Projektmitarbeiter ohne Kenntnis der jeweiligen Diagnose im anderen

Verfahren. Die Reihenfolge der Interviews war hierbei aus Gründen der Praktikabilität nicht festgelegt.

Im Rahmen der SKID-Untersuchung wurde zuerst mit dem SKID-I-Interview begonnen. Dieses erforderte in der Regel eine, manchmal auch zwei bis max. drei Sitzungen in der Länge von in der Regel 1 – 2 Stunden, je nach Anzahl der zu explorierenden Störungsbereiche, der Geschwindigkeit der diagnostischen Abklärung und der Belastbarkeit des Patienten.

Vor Beginn der SKID-II-Untersuchung wurde dem Patienten der SKID-II-Fragebogen, nach Möglichkeit einige Tage vor dem Gespräch, mitgegeben, den er dann eigenständig auszufüllen hatte. Vor Beginn des SKID-II-Interviews wurde dieser Screening-Bogen ausgewertet. Entsprechend der Anweisungen in den Test-Manualen erfolgte die Erhebung einzelner Persönlichkeitsstörungen nur, wenn im Fragebogen bereits relevante Hinweise gegeben worden waren. Ohne vorliegenden Screening-Fragebogen waren grundsätzlich alle Persönlichkeitsstörungen zu überprüfen.

Auch wenn nach dem Fragebogen eine Untersuchung der Borderline-Persönlichkeitsstörung (nach SKID-II-Manual) nicht angezeigt war, wurden dennoch zur Sicherheit und zum Zwecke vollständiger Vergleichsdaten die diagnostischen Kriterien zur Borderline-Persönlichkeitsstörung erhoben. Lediglich in zwei Fällen musste hierauf verzichtet werden: Im ersten Fall schien wegen offensichtlicher, heftiger psychotischer Symptome bei einem Patienten mit per SKID-I diagnostizierter Schizophrenie die Durchführung des SKID-II nicht angezeigt. Im zweiten Fall bestand der Patient darauf, das Interview so kurz wie möglich zu halten. In beiden Fällen wurde daher auf die vom SKID-II nicht geforderte und vom Diagnostiker auch nicht für nötig befundene diagnostische Abklärung der Borderline-Persönlichkeitsstörung verzichtet.

Alle Beurteiler waren in der Verwendung der diagnostischen Instrumente geschulte Diplom-Psychologen mit klinisch-psychiatrischer Erfahrung hinsichtlich der Diagnose der Borderline-Persönlichkeitsstörung. In der Anwendung des DIB-R waren alle Untersucher durch Prof. Dr. J. Eckert und Dipl.-Psych. E.-M. Biermann-Ratjen geschult worden.

Die Anwender des SKID-I und -II hatten keine formale Beurteilerschulung bei den Autoren der deutschen Übersetzung durchlaufen. Sie waren jedoch von erfahrenen SKID-Anwendern in das Verfahren eingewiesen worden, hatten eine Reihe von Trainings-Interviews unter Supervision durchgeführt und glichen regelmäßig ihre diagnostischen Ergebnisse mit den Behandlern ab, so dass der Sicherheitsgrad und die Qualität der erhaltenen Diagnosen für den vorliegenden Zweck als völlig ausreichend erschienen.

3.5.3. Bildung der Diagnosegruppen

Die Stichprobe wurde nach den vorliegenden DSM-III-R und DSM-IV-Diagnosen auf Achse I und Achse II in Diagnosegruppen eingeteilt.

Dabei wurden die Gruppen zunächst in Anlehnung an folgende in den SKID-Interviews erfassten diagnostischen DSM-Kategorien gebildet:

- Affektive Störungen
  (mit Unterteilung in Depressive und Bipolare Störungen)
- Psychotische Störungen
- Substanzmissbrauch und -abhängigkeit
- Angststörungen
- Somatoforme Störungen
- Essstörungen
- Anpassungsstörungen u.a. DSM-Störungen
- Persönlichkeitsstörungen
  (mit Unterscheidung der Borderline-Persönlichkeitsstörungen von der Gruppe Andere Persönlichkeitsstörungen)

Die Zuordnung zu den Diagnosegruppen erfolgte dem Untersuchungsziel angemessen z.T. *hierarchisch* unter teilweiser Ignorierung des Komorbiditätskonzepts des DSM-Systems:

- Lag *derzeit* eine akute "typische" psychotische Störung vor (also "Schizophrenie", "Schizoaffektive Störung", "Schizophreniforme Störung" oder "Wahnhafte Störung") wurde der Patienten *grundsätzlich* der Diagnosegruppe *Psychotische Störungen* zugeordnet.
- Beim Vorliegen lediglich einer *Lifetime*-Diagnose "Psychotische Störung" wurde, falls vorliegend, einer anders lautenden *Derzeit*-Diagnose der Vorzug gegeben.
- Da das Vorkommen atypischer, meist kurzer psychotischer Phänomene und substanzinduzierter Psychosen bei verschiedenen Persönlichkeitsstörungen (darunter die Borderline-Persönlichkeitsstörung) typisch zu sein scheint, wurden die in der Stichprobe vorkommenden Diagnosen "Psychotische Störung NNB", "Kurze psychotische Störung" und "Substanzinduzierte psychotische Störung" einer evtl. vorliegenden Diagnose einer Persönlichkeitsstörung *untergeordnet*.

- War eine Borderline-Persönlichkeitsstörung nach DSM-III-R oder DSM-IV diagnostiziert worden, erfolgte die Zuordnung *vorrangig* vor allen anderen Diagnosen zur Diagnosegruppe *Borderline-Persönlichkeitsstörung* – es sei denn, es lag eine *typische derzeitige* (also akute) "Psychotische Störung" wie oben beschrieben vor.
- Entsprechend diesem Vorgehen wurden beim Vorliegen der Diagnose einer "Anderen Persönlichkeitsstörung" nach DSM-III-R oder DSM-IV die Patienten immer vorrangig der Kategorie *Andere Persönlichkeitsstörung* zugeordnet.
- Lagen neben der Borderline-Persönlichkeitsstörung noch weitere Persönlichkeitsstörungen vor, wurden die Patienten dennoch der Diagnosegruppe *Borderline-Persönlichkeitsstörung* zugeordnet, da ja v.a. die Übereinstimmung der Borderline-Diagnosen nach SKID-II und DIB-R zu überprüfen war.
- Die Kategorie *Affektive Störungen* wurde grundsätzlich nachrangig zu den Persönlichkeitsstörungs-Kategorien (*Andere Persönlichkeitsstörungen* und *Borderline-Persönlichkeitsstörung*) besetzt, da bei vielen Persönlichkeitsstörungen affektive Symptomatik zumindest begleitend vorkommt – und dies nach DSM-III-R und -IV keinen Widerspruch darstellt.
- "Depressive Störungen" bzw. "Depressive Episoden" gelten insbesondere bei der Borderline-Persönlichkeitsstörung als typisch. "Bipolare Störungen" sprechen aber nach dem DIB-R-Konzept *gegen* das Vorliegen einer Borderline-Persönlichkeitsstörung, während dies nach der Konzeption des DSM-III-R oder DSM-IV gleichgültig ist. Aus diesem Grunde erschien eine Aufteilung der Gruppe der *Affektiven Störungen* in die Unterkategorien *Bipolare Störungen* und *Depressive Störungen* für die Evaluation des DIB-R als notwendig und sinnvoll.
- Die DSM-Kategorien "Substanzmissbrauch und –abhängigkeit", "Somatoforme Störungen", "Angststörungen", "Essstörungen" sowie "Anpassungsstörungen" und weitere DSM-Diagnosen wurden absolut nachrangig behandelt und nur dann besetzt, wenn nach den o.g. Regeln noch keine anderweitige Zuordnung getroffen werden konnte (dies war nur bei insgesamt fünf Patienten der Fall).
- Beim Vorliegen *mehrerer* Diagnosen gab die im SKID-I vom Interviewer angegebene Hauptdiagnose den Ausschlag. Bei drei Fällen war dies die Diagnose einer "Angststörung", bei einem Fall die alleinige Diagnose "Substanzmissbrauch und -abhängigkeit", bei einem weiteren Fall die Diagnose "Anpassungsstörung". Wegen der

geringen Häufigkeiten wurden die vorgenannten Diagnosen dann zusammengefasst zur Diagnosegruppe *Andere DSM-Störungen*.

Insgesamt wurden also folgende Diagnosegruppen für die vorliegende Studie besetzt:

- *Borderline-Persönlichkeitsstörungen*
- *Andere Persönlichkeitsstörungen*
- *Psychotische Störungen*
- *Depressive Störungen*
- *Bipolare Störungen*
- *Andere DSM-Störungen*

3.5.4. Wechsel von DSM-III-R zu DSM-IV

Nach dem Erscheinen der neuen SKID-I- und SKID-II-Interviews wurde von uns im Februar 1999 der Wechsel von der DSM-III-R- auf die DSM-IV-Diagnosestellung vollzogen.

In der Struktur der Diagnosestellungen auf der Achse I liegen für das Evaluationsprojekt keinerlei relevante Änderungen vor. Die regulären Persönlichkeitsstörungen auf Achse II sind ebenfalls namentlich und von der Grundkonzeption her identisch geblieben (Änderungen ergaben sich nur in den Forschungsversionen).

In Bezug auf die vorliegende Studie relevant sind gewisse Änderungen des Wortlauts: Es handelt sich um geringfügige Umformulierungen bei den diagnostischen Kriterien einzelner Persönlichkeitsstörungen. In Bezug auf die Borderline-Persönlichkeitsstörung ergab sich v.a. eine auffallende Veränderung, nämlich das Hinzufügen eines neuen Kriteriums (Kriteriums 9 "*vorübergehende, stressabhängige paranoide Vorstellungen oder eindeutige dissoziative Symptome"*) bei Beibehaltung des Schwellenwerts "fünf" für die Diagnosestellung.

In den Monaten vor unserer Umstellung auf das neue DSM-IV führten wir zwecks einer Augenscheinüberprüfung bei einigen Patienten beide SKID-II-Versionen parallel durch, um Informationen darüber zu erhalten, welche Unterschiedlichkeiten bzgl. der Diagnosestellung zu erwarten sein würden. Hinsichtlich der leicht unterschiedlich formulierten Kriterien stellten wir fest, dass nur mit sehr geringen Veränderungen zu rechnen sein würde.

Bzgl. des zusätzlichen Kriteriums 9 bei der Borderline-Persönlichkeitsstörung nahmen wir eine geringe Zunahme der Borderline-Prävalenz an. (Aus den von uns überblicksweise gesichteten Fällen ergaben sich durch das Kriterium 9 *keine* unterschiedlichen Diagnosen.)

3.6. DURCHFÜHRUNG DER UNTERSUCHUNG ZUR RELIABILITÄT

3.6.1. Erhebung der Stichprobe

Die Rekrutierung der Stichprobe erfolgte in erster Linie über die bereits beschriebenen Fallseminare zur Diagnostik der Borderline-Persönlichkeitsstörung für Psychologiestudenten an der Klinik, da hier üblicherweise mit den Patienten die DIB-R-Interviews nicht nur durchgeführt, sondern zum Zwecke der weiteren Schulung und Forschung (Einverständniserklärung vorausgesetzt) auf Video aufgezeichnet werden. Außerdem wurden auch von im Rahmen des diagnostischen Service zugeführten Patienten bei Einwilligung und nach Möglichkeit die DIB-R-Interviews videographiert.

Falls der Ratinggruppe gerade kein neues DIB-R-Video vorlag, wurden bereits vorhandene, ältere Aufzeichnungen eines Mitglieds der Ratergruppe aus dem Archiv der Klinik verwendet.

3.6.2. Die Gruppenratings zur Schätzung der Interrater-Reliabilität

Ein Rating zur Reliabilitätsbestimmung muss gewissen Anforderungen genügen, um eine hinreichend verallgemeinerbare Reliabilitätsberechnung zu ermöglichen. Die wichtigste ist die der *Unabhängigkeit* in der Einschätzung der einzelnen Rater. Eine andere ist die, dass die Rater das Rating unter *vergleichbaren Bedingungen* durchführen.

Diesen beiden Bedingungen trugen wir durch das Abhalten eines Gruppenratings nach bestimmten Regeln Rechnung. Die Ratergruppe traf sich regelmäßig, d.h. in der Regel ein Mal pro Woche für ca. 1,5 - 2 Stunden, um ein DIB-R-Interview einzuschätzen. Das Video wurde über ein für alle gleichermaßen sichtbares TV-Gerät wiedergegeben.

Es wurde dann üblicherweise die Aufzeichnung der Exploration *mehrerer DIB-R-Items* zur Einschätzung *eines DIB-R-Statements* am Stück abgespielt (z.B. Items 4 – 7 zur Einschätzung des Statements 2). Gelegentlich wurde auch ein (meist kurzer) *Unterbereich* eines DIB-R-Bereichs mit mehreren Items und mehr als einem einzuschätzenden Statement am Stück abgespielt (z.B. Unterbereich "Depression", Items 1 – 7, S. 1 – S. 2 des Bereichs "Affektivität"), um ein Rating im Zusammenhang zu ermöglichen und um Zeit zu sparen. Jeder Bereich wurde realistischerweise nur ein einziges Mal abgespielt. Nur im Falle von Problemen mit der Bild- oder Tonqualität konnte eine Passage wiederholt werden.

Entscheidend für ein unabhängiges Rating war, dass die Rater sich unter keinen Umständen untereinander austauschen durften, bis alle die entsprechenden Statements und auch Items ausgefüllt hatten. Unklarheiten und Fragen konnten in einer kurzen Besprechung *nach* dem Ausfüllen besprochen werden. Änderungen durften in keinem Falle mehr vorgenommen werden.

In der Regel konnte ein Band nicht während einer einzigen Zusammenkunft zu Ende geratet werden. Es waren meist zwei, manchmal drei Sitzungen nötig. Dieses führte leider dazu, dass nicht immer alle Rater das Band komplett sehen konnten. Zum Teil konnte die Lücke durch ein individuelles Nacharbeiten geschlossen werden. In mehreren Fällen blieben aber Ratings unvollständig, so dass für die betreffenden Rater nur die Statements und die komplett vorliegenden Skalierten Section-Scores Eingang in die Stichprobe fanden. Daher ist die Stichprobengröße für die einzelnen DIB-R-Kennwerte unterschiedlich und für die Statements grundsätzlich höher als für die Diagnosen.

3.6.3. Die Raterstichproben

Für die Studie wurden insgesamt zwei Stichproben von Ratern gebildet: Eine Stichprobe der Experten und eine Stichprobe erfahrener Studenten, die im Folgenden beschrieben werden.

3.6.3.1. Experten-Stichprobe

Die Gruppe der Experten bestand sämtlich aus den Mitgliedern der Arbeitsgruppe. Alle waren zumindest Dipl.-Psychologen mit zumindest grundlegender klinischer Erfahrung. Alle hatten Vorkenntnisse im Umgang mit Menschen mit Borderline-Persönlichkeitsstörung aus eigener Erfahrung und hatten das DIB bereits in der unrevidierten Fassung selbst mehrfach in diesem Zusammenhang angewandt.

Um die Ratergruppe in der Anwendung des DIB-R auszubilden und auf einen gemeinsamen Stand zu bringen, hatten die Rater über den Zeitraum von einem halben Jahr bis zum Dezember 1995 in loser Folge insgesamt sechs Videos von Patienten mit und ohne Borderline-Persönlichkeitsstörung gemeinsam in Form eines *Konsensratings* geratet und gründlich Statement für Statement diskutiert (Reliabilitäten der Trainingsbänder wurden nicht berechnet, da bei einem Konsensrating wegen der Abhängigkeit der Ratings nicht sinnvoll).

Vom Januar 1996 bis zum Juni 1997 konnte dann die Ratergruppe insgesamt die Bänder von 19 Patienten gemeinsam *unabhängig* raten. Das ursprünglich angestrebte Ziel von 30 unabhängigen Ratings wurde u.a. aufgrund der hohen Aufwändigkeit nicht erreicht.

3.6.3.2. Studenten-Stichprobe

Die zu Vergleichszwecken gebildete Studenten-Stichprobe bestand aus vier Studentinnen der Psychologie (Diplomstudiengang), die bereits alle klinischen Basisseminare und v.a. das Fallseminar zur Diagnostik der Borderline-Persönlichkeitsstörung (unter Verwendung des DIB-R) besucht hatten. Der Arbeits- und Zeitaufwand der Rater wurde über studentische Hilfskraftgelder vergütet.

Die Rater wurden von den Projektmitgliedern Dipl.-Psych. D. Brodbeck und Dipl.-Psych. M. Schödlbauer im Gruppenrating (nach denselben Regeln wie die Expertengruppe) unterwiesen. Die Einhaltung der Regeln wurde in allen Gruppenratings durch Anwesenheit eines Projektmitgliedes überwacht. Ausnahme waren hier einige wenige Nachholtermine einzelner Studentinnen, die eines der Gruppenratings verpasst hatten.

Die Ratings fanden im Zeitraum von April bis September 1996 statt. Nach dem "Konsensrating" zweier Bänder (s.o.) zum Auffrischen des Wissens und zum Gewöhnen an die Regeln schätzten die Studentinnen 12 Patienten anhand der Videoaufnahmen ein.

4. ALLGEMEINE ERGEBNISSE

4.1. SOZIODEMOGRAPHISCHE DATEN

4.1.1. Geschlecht

Von den 100 untersuchten Patienten waren 59 Frauen und 41 Männer. Die vorgefundene stärkere Vertretung des weiblichen Geschlechts in der Stichprobe entspricht ungefähr der üblichen klinischen Verteilung von ca. zwei zu eins.

4.1.2. Alter

Die Patienten waren im Durchschnitt 32,0 Jahre alt (s=8,95). Die Bandbreite des Lebensalters erstreckte sich von 18-61 Jahren.

Die männlichen Patienten waren hierbei mit 30,5 Jahren (s=8,4) geringfügig und insignifikant jünger als die weiblichen Patienten, die im Durchschnitt 33,0 Jahre (s=9,3) alt waren.

4.1.3. Familienstand

Über den Familienstand liegen von 93 der 100 Patienten Angaben vor.
Demnach sind 68 Patienten (73,1%) ledig. Verheiratet oder geschieden sind 25 (26,9%) der untersuchten Personen. Eine weitere Differenzierung wird im DIB-R, aus dem diese Angaben stammen, nicht vorgenommen.

4.1.4. Nationalität

Angaben über die Nationalität liegen von 96 Patienten vor. Die deutsche Staatsangehörigkeit besitzen 89 Patienten (92,7%). Eine andere Staatsangehörigkeit haben nur 7 Patienten (7,3%).

4.1.5. Schulabschluss

Informationen über das Bildungsniveau liegen uns von 90 Patienten in Form von Angaben über den erreichten Schulabschluss vor:

Tab. 9: *Bildungsniveau der Patienten*

	N	***%***
kein Schulabschluss	4	4,4
Hauptschule oder abgeschl. Lehre	16	17,8
Mittlere Reife	27	30,0
Fachhochschulreife	6	6,7
Abitur	37	41,1
Gesamt	90	100,0

Insgesamt verfügen fast alle Patienten über einen Schulabschluss, mit 47,8% hatte fast die Hälfte der Patienten eine Studienbefähigung in Form von Fachhochschulreife oder Abitur.

4.1.6. Psychiatrische Hospitalisierung

Von 98 Patienten waren 86 noch aktuell in stationärer psychiatrischer Behandlung. Lediglich 12 der untersuchten Personen waren ambulante psychiatrische Patienten. Es handelte sich in diesen Fällen meist um unmittelbar zuvor hospitalisierte Personen, die gerade entlassen worden waren, ein Teil dieser Patienten befand sich noch in tagesklinischer Behandlung.

4.2. DIAGNOSESTELLUNG NACH DSM-III-R UND DSM-IV

4.2.1. Achse I-Diagnosen

4.2.1.1. Überblick über DSM-Diagnosekategorien

Viele Patienten haben mehr als eine Achse I-Diagnose erhalten, ein Umstand der v.a. dem Komorbiditätskonzept des DSM-Systems geschuldet ist, sowie dem Umstand, dass sowohl *Derzeit*- als auch *Lifetime*-Diagnosen vergeben werden.

Dadurch ergibt sich eine insgesamt sehr uneinheitliche und unübersichtliche Verteilung. Es ist nicht möglich, ein leicht überschaubares Bild zu vermitteln.[5]

Dennoch wird folgend der Versuch unternommen, die Diagnosen überblicksweise darzustellen:

Die Patienten der Stichprobe hatten insgesamt bis zu vier *Derzeit*- und ebenfalls bis zu vier *Lifetime*-Diagnosen erhalten.

Diese Diagnosen werden nicht einzeln aufgeführt, sondern nur nach den DSM-Oberkategorien wiedergeben, da zu viele Unterkategorien (sowie Zusatzkodierungen) und damit zu viele Kombinationen vorkommen.

Da inhaltlich im Zusammenhang mit der Fragestellung und dem Verständnis der Diagnosegruppen wichtig, wird lediglich für die Kategorien *Psychotische Störungen* und *Affektive Störungen* nachfolgend eine detailliertere Aufgliederung vorgenommen.

Wiedergegeben wird jeweils die Anzahl der Diagnosen, die die Gesamtheit der Patienten in der jeweiligen Störungsgruppe erhalten haben.

---

[5] Ein geringfügig komplizierender Sachverhalt ist zusätzlich die Verwendung der zwei unterschiedlichen Versionen des DSM-Manuals: Zwischen der Version DSM-III-R und DSM-IV liegen auch auf der Ebene der gängigen Diagnosen leichte Unterschiede vor. So ist z.B. die Störung *Bipolar II* im DSM-III-R anders als im DSM-IV noch nicht als eigene Kategorie vorhanden, stattdessen wird sie unter der Diagnose *Bipolar NNB* optional erfasst.

Anzumerken ist, dass die Diagnose *Dysthyme Störung* ihrer anhaltenden Definition wegen nur Zeitraum übergreifend und nicht *derzeit* und/oder *lifetime* gegeben werden kann. Im SKID-I nach DSM-III-R wird sie als *lifetime,* im DSM-IV jedoch als *derzeit* kodiert. Inhaltlich hat dies keine Bedeutung. Nachfolgend wird sie einheitlich nur in der *Lifetime*-Tabelle aufgeführt werden.

Tab. 10: *Vergebene Lifetime-Achse-I-Diagnosen nach DSM-Kategorien; % von Gesamtdiagnosen*

DSM-III-R bzw. -IV Störungsgruppen	***N***	***%***
Psychotische Störungen	18	12,9
Affektive Störungen	68	48,6
Substanzmissbrauch und -abhängigkeit	19	13,6
Angststörungen	23	16,4
Somatoforme Störungen	1	0,7
Essstörungen	8	5,7
Anpassungsstörungen u.a. DSM-Diagnosen	3	2,1
Gesamt-Diagnosen *lifetime*	140	100
keine *Lifetime*-Diagnose gestellt	12	

Insgesamt wurden 140 *Lifetime*-Diagnosen vergeben, pro Patient also durchschnittlich 1,4 (s=0,88) Diagnosen. 12 Patienten erhielten keine *Lifetime*-Diagnose.

Lifetime stellen Diagnosen aus dem Bereich Affektive Störungen die häufigste Diagnosegruppe dar, gefolgt von Angststörungen, Substanzmissbrauch bzw. -abhängigkeit sowie Psychotischen Störungen.

Insgesamt wurden 142 *Derzeit*-Diagnosen vergeben, pro Patient also durchschnittlich 1,4 (s=0,94) derzeitige Diagnosen. 10 Patienten erhielten *derzeit* keine Diagnose.

Auch *derzeit* wurden am häufigsten Diagnosen im affektiven Bereich vergeben, zweithäufigste Kategorie war dann aber die Gruppe der Psychotischen Störungen, gefolgt von Angststörungen und Substanzmissbrauch bzw. -abhängigkeit.

Tab. 11: *Vergebene Derzeit-Achse-I-Diagnosen nach diagnostischen DSM-Kategorien; % von Gesamtdiagnosen*

Diagnostische Kategorien	***N***	***%***
Psychotische Störungen	29	20,4
Affektive Störungen	62	43,7
Substanzmissbrauch und –abhängigkeit	17	12,0
Angststörungen	20	14,1
Somatoforme Störungen	4	2,8
Essstörungen	5	3,5
Anpassungsstörungen u.a. DSM-Diagnosen	5	3,5
Gesamt-Diagnosen *derzeit*	142	100
keine *Derzeit*-Diagnose gestellt	10	

4.2.1.2. Aufschlüsselung der diagnostischen Kategorie *Affektive Störungen*

Insgesamt zeigt sich im Bereich affektive Störungen eine größere Häufigkeit der depressiven Störungen im Vergleich zu den Bipolaren Störungen. Dies geht zumindest teilweise auf eine häufige Komorbidität und daher doppelte Auflistung der Major Depression und der Dysthymen Störung zurück.

Tab. 12: *Affektive Störungen – Vergebene Lifetime-Diagnosen; % von Gesamtdiagnosen*

Affektive Störungen	***N***	***%***
Bipolare Störung	4	5,9
Bipolare Störung II bzw. NNB	14	20,6
Major Depression	27	39,7
Dysthyme Störung	17	25,0
Depressive Störung NNB	6	8,8
Gesamt-Diagnosen *lifetime*	68	100

Tab. 13: *Affektive Störungen – Vergebene derzeit- Diagnosen; % von Gesamtdiagnosen*

Affektive Störungen	***N***	***%***
Bipolare Störung	6	9,7
Bipolare Störung II bzw. NNB	7	11,3
Major Depression	23	37,1
Dysthyme Störung	19	30,6
Depressive Störung NNB	7	11,3
Gesamt-Diagnosen *derzeit*	62	100

4.2.1.3. Aufschlüsselung der diagnostischen Kategorie *Psychotische Störungen*

Diagnosen einer Psychotischen Störung sind häufiger *derzeit* als *lifetime* in der Stichprobe zu finden. Dieser Umstand geht z.T. sicher darauf zurück, dass akute Psychosen ein häufiger Grund sind, eine Klinik aufzusuchen.

Tab. 14: *Psychotische Störungen – Vergebene lifetime Diagnosen; % von Gesamtdiagnosen*

Psychotische Störungen	***N***	***%***
Schizophrenie	6	33,3
Schizophreniforme Störung	0	0
Schizoaffektive Störung	1	5,6
Wahnhafte Störung	4	22,2
Kurze Psychotische Störung	0	0
Substanzinduzierte Psychotische Störungen	1	5,6
Psychotische Störung NNB	6	33,3
Gesamt-Diagnosen *lifetime*	18	100

Tab. 15: *Psychotische Störungen – Vergebene derzeit Diagnosen; % von Gesamtdiagnosen*

Psychotische Störungen	***N***	***%***
Schizophrenie	9	31,0
Schizophreniforme Störung	2	6,9
Schizoaffektive Störung	2	6,9
Wahnhafte Störung	4	13,8
Kurze Psychotische Störung	1	3,5
Substanzinduzierte Psychotische Störungen	1	3,5
Psychotische Störung NNB	10	34,5
Gesamt-Diagnosen *derzeit*	29	100

4.2.2. Achse II-Diagnosen

In diesem Kapitel wird ein allgemeiner Überblick über das Vorliegen und die jeweilige Häufigkeit von Persönlichkeitsstörungen nach DSM-III-R und DSM-IV gegeben.

Aufgeführt werden nur die in DSM-III-R und DSM-IV in der diagnostischen Kategorie Persönlichkeitsstörung angegebenen Störungen. Die in den jeweiligen DSM-Versionen (voneinander abweichend) nur im Anhang zur weiteren Forschung aufgeführten Kriterien für Persönlichkeitsstörungen wurden im Rahmen der Studie nicht ausgewertet.[6]

Die im DSM-IV aufgeführte *Vermeidend-Selbstunsichere Persönlichkeitsstörung* wird im Folgenden zur Vereinfachung und Verkürzung gemäß der Schreibweise im SKID-II *Selbstunsichere Persönlichkeitsstörung* genannt.

Ebenso wird die bei beiden SKID-II-Versionen nach Clustern gruppierte Reihenfolge der aufgeführten Persönlichkeitsstörungen (bei beiden Versionen von der Reihenfolge im DSM-Manual abweichend) der Einfachheit halber übernommen.

---

[6] Es handelt sich hier z.B. um die *Depressive Persönlichkeitsstörung* und die *Negativistische Persönlichkeitsstörung* aus DSM-IV (Anhang B, S. 791 f.) oder die *Selbstschädigende Persönlichkeitsstörung* aus DSM-III-R (Anhang A, S. 503). Im Falle des Erfüllens der Kriterien einer solchen Forschungs-Persönlichkeitsstörung wurde entsprechend der allg. diagnostischen Richtlinien eine Persönlichkeitsstörung NNB diagnostiziert, insofern nicht bereits eine reguläre Persönlichkeitsstörung diagnostiziert worden war.

Tab. 16: *Anzahl der DSM-Persönlichkeitsstörungen pro Patient (N=100)*

Anzahl	Häufigkeit
0	43
1	26
2	21
3	8
4	2

Insgesamt habe von den 100 untersuchten Patienten 57 eine oder mehrere Persönlichkeitsstörungs-Diagnosen erhalten, während 43 Patienten keine Achse-II-Diagnose erhielten. (Das sind im Mittel 1,75 Persönlichkeitsstörungs-Diagnosen pro Patient. Dass dabei insgesamt exakt 100 Diagnosen einer Persönlichkeitsstörung vergeben wurden, war nicht geplant und ist dem Zufall zuzuschreiben.)

Tab. 17: *DSM-Persönlichkeitsstörungen der 57 Patienten; % von sämtlichen Achse-II-Diagnosen*

Persönlichkeitsstörungen N=57		***Kriterien erfüllt***		***unterschwellig***	
		N	***%***	***N***	***%***
Cluster C	Selbstunsichere Persönlichkeitsstörung	21	21	4	7,1
	Dependente Persönlichkeitsstörung	15	15	5	8,9
	Zwanghafte Persönlichkeitsstörung	5	5	13	23,2
Cluster A	Paranoide Persönlichkeitsstörung	9	9	6	10,7
	Schizotypische Persönlichkeitsstörung	5	5	3	5,4
	Schizoide Persönlichkeitsstörung	4	4	5	8,9
Cluster B	Histrionische Persönlichkeitsstörung	5	5	4	7,1
	Narzisstische Persönlichkeitsstörung	3	3	3	5,4
	Borderline Persönlichkeitsstörung	28	28	12	21,5
	Antisoziale Persönlichkeitsstörung	3	3	1	1,8
	Persönlichkeitsstörung NNB	2	2	0	0
	Gesamt-Diagnosen	100	100	56	100

4.3. TATSÄCHLICHE BESETZUNG DER DSM-DIAGNOSEGRUPPEN

Die tatsächliche Besetzung der Diagnosegruppen erfolgte wegen eingeschränkter Möglichkeiten bei der Auswahl der Patienten (wie in Kap. 3.5 beschrieben) etwas vom ursprünglich angestrebten Schlüssel abweichend wie folgt:

Die Gruppe der Patienten mit *Borderline-Persönlichkeitsstörung* erreicht mit 28 das angestrebte Drittel nicht ganz. Die Gruppe der *Affektiven Störungen* ist mit 20 Patienten (13 *depressiven* und 7 *bipolaren*) statt den geplanten 22 Patienten geringfügig kleiner, die der *Anderen Persönlichkeitsstörungen* mit 26 statt 22 etwas größer als vorgesehen ausgefallen.

Tab. 18: *Besetzung der DSM-Diagnose-Gruppen zur Validitätsprüfung des DIB-R*

Diagnostische Kategorien	***N***
Borderline-Persönlichkeitsstörung	28
Andere Persönlichkeitsstörungen	26
Bipolare Störungen	7
Depressive Störungen	13
Psychotische Störungen	21
andere DSM-Diagnosen	5
Gesamt	100

Außerplanmäßig liegen bei fünf Patienten Diagnosen vor, die nicht aus einer der gewünschten Diagnose-Gruppen stammen und zur "Restkategorie" *Andere DSM-Diagnosen* zusammengefasst werden mussten.

Für die geplante differentialdiagnostische Untersuchung sind die erhaltenen Diagnosegruppen dennoch uneingeschränkt verwendbar.

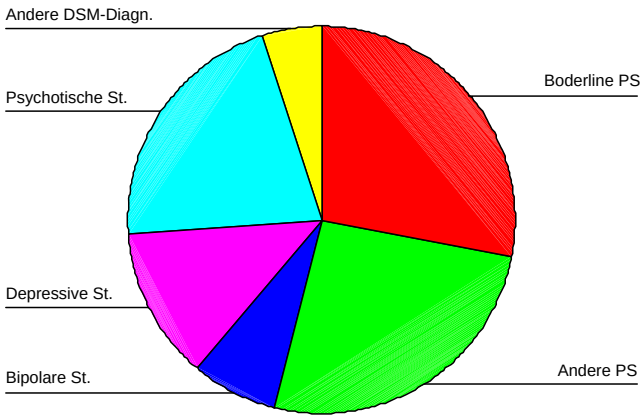

Abb. 3: *Besetzung der DSM-Diagnosegruppen*

5. ERGEBNISSE ZUR VALIDITÄT DES DIB-R

5.1. ABSOLUTE DIAGNOSEN-ÜBEREINSTIMMUNGEN DER BPS NACH DIB-R UND SKID-II

In diesem Kapitel werden die Ergebnisse zur konvergenten bzw. Übereinstimmungs-Validität wiedergegeben. Nach der SKID-II-Diagnostik erhielten, wie im Kapitel 4.2 dargestellt, von 100 Patienten 28 die Diagnose einer Borderline-Persönlichkeitsstörung (z.T. zusammen mit weiteren Persönlichkeitsstörungs-Codiagnosen).

Die diagnostische Untersuchung mit dem DIB-R brachte größenordnungsmäßig ein gleiches Ergebnis: mit 26 von 100 Patienten ist der Anteil der Patienten mit Borderline-Diagnose insgesamt fast gleich groß.

Die Borderline-Diagnosen stimmen individuell – auf die jeweiligen Patienten bezogen – jedoch nur zu einem geringeren Teil überein, als es zunächst scheint. Dies ist in der folgenden Tabelle erkennbar.

Tab. 19: *Kreuztabelle, Anzahl Patienten mit Borderline-Diagnosen nach SKID-II vs. DIB-R (Cutoff=8)*

Cutoff=8		**Borderline nach DIB-R**		Total
		nein	**vorhand.**	
Borderline nach SKID-II	**nein**	65	7	72
	vorhand.	9	19	28
Total		74	26	100

Von den insgesamt 35 Patienten mit einer *Borderline-Diagnose nach wenigstens einem der beiden Instrumente* haben nur 19 (54,3%) auch die übereinstimmende Diagnose nach dem jeweils anderen Interview erhalten! Eine entsprechende Übereinstimmung in der *Nicht*-Diagnose einer Borderline-Persönlichkeitsstörung in *beiden* Interviews liegt in 65 von insgesamt 81 Fällen *ohne Borderline-Diagnose nach wenigstens einem der beiden Instrumente* vor (80,3%).

Insgesamt kann (bei Addition des linken oberen und des rechten unteren Feldes in Entsprechung zum Gesamten Prädiktiven Wert, s.u.) von einer diagnostischen Gesamt-Übereinstimmung in 84 der 100 Fälle gesprochen werden. Diese Übereinstimmung ist m.E. als relativ hoch anzusehen. Sie entspricht einem korrelativen Zusammenhang beider Diagnosen in Höhe von $\varphi = .60$ ($p=.000$, siehe Tab. 24 im Kap. 5.5.1). Der Phi-Koeffizient ist nach Bortz (1993, S. 210) in der Deutung einem Pearson-Korrelations-Koeffizienten äquivalent.

Der zufallskritische Übereinstimmungskoeffizient Kappa nach Cohen (s. Bortz & Döring, 1995, Tafel 27) wurde ebenfalls berechnet. Dieser liegt bei κ=.59 (ebenfalls hochsignifikant).

Im Kap. 3.3.1.2 waren folgende spezifische Hypothesen zur Übereinstimmungsvalidität formuliert worden:

Hypothese 1.1: Die Übereinstimmung der Borderline-Diagnose nach DIB-R und SKID-II ist berechnet nach Cohen's Kappa signifikant überzufällig häufig.

Hypothese 1.2: Das Vorliegen der Borderline-Diagnosen nach DIB-R und SKID-II korreliert, berechnet mittels des φ -Koeffizienten, signifikant und erreicht einen Wert von $\varphi \geq .50$.

Beide Hypothesen können nach den vorliegenden Ergebnissen beibehalten werden.

5.2. ÜBEREINSTIMMUNG UNTER VARIATION DER DIAGNOSEN-SCHWELLENWERTE IM DIB-R

In der im Kap. 1.2.4.2.1 genannten Validitätsstudie von Zanarini et al. (1989) war ebenfalls eine Überprüfung der Übereinstimmungsgüte zwischen DIB-R und klinischen Achse-II-Diagnosen (Außenkriterium) bei einer Variation des Diagnose-Cutoff-Wertes von üblicherweise "8" auf alle Werte zwischen "6" und "10" durchgeführt worden. Die Autoren waren zu dem Schluss gekommen, dass der Wert von "8" die insgesamt besten Ergebnisse hinsichtlich Spezifität, Sensitivität, sowie Positivem und Negativem Prädiktiven Wert erbrachte. In der Studie von Szerman et al. (2005) war für die spanische Version hingegen eher ein Cutoff-Wert von "7" als optimal für die entsprechende Übereinstimmung zwischen DIB-R- und DSM-III-R-Diagnose vorgeschlagen worden.

In der vorliegenden Studie wird dies ebenfalls anhand einer Variation des Cutoffs überprüft, allerdings lediglich mit den als möglich erscheinenden Werten "7" und "9", die in den folgenden beiden Tabellen wiedergegeben sind. Die Ergebnisse für Sensitivität und die folgende Werte werden im folgenden Kapitel wiedergegeben.

Tab. 20: *Kreuztabelle, Anzahl Patienten mit Borderline-Diagnosen nach SKID-II vs. DIB-R (Cutoff=7)*

Cutoff=7		**Borderline nach DIB-R**		Total
		nein	**vorhand.**	
Borderline nach SKID-II	**nein**	61	11	72
	vorhand.	6	22	28
Total		67	33	100

Beim Absenken des Cutoffs auf "7" steigt zunächst die Anzahl der Borderline-Patienten nach DIB-R auf den recht groß erscheinenden Wert von 33% an.
Die Zahl der Patienten mit einer Borderline-Diagnose, die nach wenigstens einem der beiden Instrumente eine solche erhalten haben, steigt dabei von 35 bei einem Cutoff von "8" auf die Zahl von 39 an. Auch die Zahl der nach beiden Interviews übereinstimmend diagnostizierten Patienten steigt von vorher 19 (54,3%) auf 22 deutlich an.

Eine Übereinstimmung in der *Nicht*-Diagnose einer Borderline-Persönlichkeitsstörung liegt dagegen nur noch in 61 statt vorher 65 Fällen vor. Insgesamt senkt sich die diagnostische Gesamt-Übereinstimmung auf 83 von hundert Fällen (statt vorher 84).

Tab. 21: *Kreuztabelle, Anzahl Patienten mit Borderline-Diagnosen nach SKID-II vs. DIB-R (Cutoff=9)*

Cutoff=9		**Borderline nach DIB-R**		Total
		nein	**vorhand.**	
Borderline nach SKID-II	**nein**	71	1	72
	vorhand.	12	16	28
Total		83	17	100

Beim Steigern des Cutoffs auf "9" sinkt der Anteil der Borderline-Patienten nach DIB-R erwartungsgemäß auf den deutlich niedrigeren Wert von 17%.
Die Zahl der Patienten, die nach wenigstens einem der beiden Instrumente eine Borderline-Diagnose erhalten haben, sinkt dabei von 35 bei einem Cutoff von "8" stark auf die Zahl von 29 ab. Auch die Zahl der nach beiden Interviews übereinstimmend diagnostizierten Patienten sinkt von vorher 19 auf 16 ab.

Eine Übereinstimmung in der *Nicht*-Diagnose einer Borderline-Persönlichkeitsstörung steigert sich dafür auf nunmehr 71 statt vorher 65 Fälle. Insgesamt steigert sich (entsprechend dem Gesamten Prädiktiven Wert, s.u.) die diagnostische Gesamt-Übereinstimmung von 84% auf 87%.

5.3. WECHSELSEITIGE BEZIEHUNGEN DER BORDERLINE-DIAGNOSEN NACH DIB-R UND SKID-II

Die deskriptiven statistischen Kriterien Sensitivität, Spezifität sowie Positiver und Negativer Prädiktiver Wert werden im Folgenden zur differenzierteren Darstellung der o.g. Beziehungen zwischen den diagnostischen Instrumenten DIB-R und SKID-II eingesetzt. Zur

Vergleichsmöglichkeit der o.g. Werte mit der Validitäts-Studie von Zanarini et al. (1989) werden diese Werte ebenfalls für einen variierten Cutoff-Wert angegeben.
Ausgegangen wird bei diesen Kennwerten immer von *einer* feststehenden "wahren" Diagnose: Im vorliegenden Fall ist dies die SKID-II-Diagnose als äußeres Validitätskriterium (während in der Studie von Zanarini et al. eine Expertendiagnose verwendet worden war). Die Richtung der Beziehung ist dabei zum Verständnis der Kennwerte und der unten stehenden Definition entscheidend.

Die Definitionen der Kennwerte – mit SKID-II als "wahrem" Außenkriterium – lauten wie folgt (siehe auch Zanarini et al., 1989):

- Die *Sensitivität* bezeichnet den Anteil der SKID-II-Borderliner, die vom DIB-R korrekt als Borderline-Patienten identifiziert werden. Mit anderen Worten gibt die *Sensitivität* die Wahrscheinlichkeit an, einen "echten" Borderline-Patienten mittels des DIB-R auch zu entdecken.

- Die *Spezifität* bezeichnet im Gegenteil den Anteil der *Nicht*-SKID-II-Borderliner unter den Patienten, die die Diagnose auch im DIB-R *nicht* bekommen. Mit anderen Worten gibt die *Spezifität* die Wahrscheinlichkeit an, einen "echten" *Nicht*-Borderline-Patienten mittels des DIB-R auch als solchen zu erkennen.

- Der *Positive Prädiktive Wert (PPW)* wird aus einem anderen Blickwinkel als das obige Wertepaar berechnet. Er gibt den Anteil der "wahren" Borderline-Patienten unter allen vom DIB-R als Borderline-Patienten klassifizierten Menschen an. Er bezeichnet sozusagen die Wahrscheinlichkeit des Trägers einer Borderline-Diagnose, nach DIB-R auch "wirklich" ein Borderline-Patient zu sein.

- Der *Negative Prädiktive Wert (NPW)* gibt entsprechend den Anteil der "wahren" *Nicht*-Borderliner unter allen Patienten an, die im DIB-R *keine* Borderline-Diagnose bekommen haben. Er bezeichnet die Wahrscheinlichkeit, mit der ein *Nicht*-Borderline-Patient nach DIB-R auch "in Wirklichkeit" *kein* Borderline-Patient ist.

- Zusätzlich kann ein entsprechender *Gesamter Prädiktiver Wert (GPW)* gebildet werden, der den Anteil aller richtig (positiv und negativ) diagnostizierten Patienten angibt, ausgehend davon ob das jeweilige Symptom vorliegt oder nicht vorliegt. Er bezeichnet also die Wahrscheinlichkeit, mit der aufgrund der Präsenz oder Absenz der Borderline-Diagnosen nach DIB-R eine "wahre" Aussage gemacht wird.

Festzuhalten bleibt an dieser Stelle, dass die SKID-II-Diagnose selbstverständlich, wie schon erläutert, kein wahres Außenkriterium darstellt. Die folgende Tab. 22 gibt die gefundenen Kennwerte wieder.

Die *Sensitivität*, also der Anteil der SKID-II-diagnostizierten Borderliner, die auch die DIB-R-Diagnose erhalten haben, sinkt mit der Erhöhung des Cutoffs von 78,6% bei "7" auf 57,1% bei "9" kontinuierlich ab. Die *Spezifität* steigt im Gegenteil von 84,7% auf 98,6% an.

Ähnlich liegt der *Positive Prädiktive Wert*, also der Anteil der DIB-R-diagnostizierten Borderliner, die auch die SKID-II-Diagnose erhalten haben, für den Cutoff von "7" bei 66,7% und steigt für den Wert "9" 94,1% an. Der Negative Prädiktive Wert sinkt wiederum vom Cutoff "7" mit 91,0% auf 85,5% beim Wert "9".

Tab. 22: *Wechselseitige Beziehungen der SKID-II und DIB-R–BPS-Diagnosen bei Cutoffs "7" – "9"*

Cutoff:	**"7" %**	**"8" %**	**"9" %**	**Wahrscheinlichkeit, dass ...**
Sensitivität	78,6	67,9	57,1	DIB-R-BPS, wenn SKID-II BPS
Spezifität	84,7	90,3	98,6	*nicht* DIB-R-Borderline-Persönlichkeitsstörung, wenn *nicht* SKID-II BPS
PPW	66,7	73,1	94,1	SKID-II-BPS, wenn DIB-R-BPS
NPW	91,0	87,8	85,5	*nicht* SKID-II-BPS, wenn *nicht* DIB-R-BPS
GPW	83	84	87	Diagnosen (pos. oder neg.) für SKID-II und DIB-R übereinstimmend
Prävalenz nach SKID-II		28		
Prävalenz nach DIB-R	33	26	17	
κ*	.60	.59	.63	Übereinstimmung SKID-II und DIB-R
φ*	.60	.60	.67	

* alle signifikant

Wie bereits oben dargestellt variiert der *Gesamte Prädiktive Wert*, der den Anteil der beiderseitigen Übereinstimmung der Interviews im Vorhandensein bzw. *Nicht*-Vorhandensein der Diagnose der Borderline-Persönlichkeitsstörung angibt, zwischen 83% und 87%.

Die gefunden Kennwerte liegen sämtlich in einer Höhe, die wie erwartet eine mittlere, aber nicht sehr hohe Übereinstimmung der DIB-R-Ergebnisse mit denen des Außenkriteriums SKID-II wiedergeben.

Insgesamt erhält man für den Cutoff-Wert von "8" nicht nur eine – auch im Verhältnis zu den Ergebnissen nach SKID-II – plausible Prävalenz einer Borderline-Persönlichkeitsstörung, sondern auch eine gute Ausgewogenheit der Kennwerte.[7]

5.4. ÜBEREINSTIMMUNG UNTER VARIATION DER DIAGNOSEN-SCHWELLENWERTE IM SKID-II

Nach dem SKID-II sind neben vollgültigen Persönlichkeitsstörungsdiagnosen auch *unterschwellige* Diagnosen möglich (mit einem Diagnose-Index von 2 statt 3). Werden die *unterschwelligen* SKID-II-Borderline-Diagnosen mittels des dreistufigen Diagnose-Index in die Auswertung einbezogen, ist festzustellen, dass drei von sieben der allein mittels des DIB-R identifizierten Borderline-Patienten immerhin *unterschwellig* auch nach dem SKID-II eine Borderline-Persönlichkeitsstörung aufweisen.

Tab. 23: *Anzahl BPS nach SKID-II mit unterschwelliger Diagnose vs. Borderline nach DIB-R*

		Borderline nach DIB-R		Total
		nein	**vorhand.**	
Borderline nach SKID-II	**nein**	56	4	60
	unterschw.	9	3	12
	vorhand.	9	19	28
Total		74	26	100

Eine vollständige Berücksichtigung der unterschwelligen SKID-II-Borderline-Persönlichkeitsstörung würde die Übereinstimmung beider Interviews im Endeffekt jedoch sogar verringern, da insgesamt 9 der vom DIB-R als non-borderline identifizierte Patienten ebenfalls eine *unterschwellige* Diagnose nach SKID-II aufweisen. Insgesamt würden bei einer solchen Zählweise im Vorliegen und Nicht-Vorliegen der Borderline-Persönlichkeitsstörung also nur

[7] Dass sowohl der κ- als auch der φ-Koeffizient beim Cutoff von "8" – wenn auch nur mit einem geringfügigen Unterschied – nicht die höchsten der je drei Werte sind, dürfte zu vernachlässigen sein. Dieser Umstand könnte mit der stark veränderten Prävalenz der Borderline-Persönlichkeitsstörung nach DIB-R zusammenhängen, die sich bei Veränderung des Cutoffs in Richtung einer Über- bzw. Unterinkludierung verschiebt und so die Verteilung verändert. In der o.g. Studie von Szerman et al. war für den Cutoff von "7" mit .85 der höchste Wert gefunden worden, für "8" φ=.76 und für "9" nur φ=.39. Eine abschließende Beurteilung ist auf der Basis der vorliegenden Ergebnisse nicht möglich.

noch 78% statt vorher 84% Übereinstimmungen vorliegen. Eine verbesserte Übereinstimmung beider Instrumente ist so offenbar nicht zu erzielen.

5.5. KONVERGENTE UND DISKRIMINANTE VALIDITÄT: ZUSAMMENHÄNGE DER BPS NACH DIB-R MIT DEN SKID-II-DIAGNOSEN

5.5.1. Univariate Untersuchung

Oben wurde bereits der univariate Zusammenhang zwischen der DIB-R- und der SKID-II-Diagnose Borderline-Persönlichkeitsstörung (konvergente Validität) wiedergegeben. In diesem Kapitel werden nun auch die Zusammenhänge mit den verschiedenen anderen SKID-II-Diagnosen für Persönlichkeitsstörungen (diskriminante Validität) überprüft.

Für folgende Kennwertkombinationen wurden die genannten Korrelationen berechnet:

- Vorliegen/Nicht-Vorliegen einer Borderline-Persönlichkeitsstörung nach DIB-R mit dem Vorliegen/Nicht-Vorliegen der Persönlichkeitsstörungen nach SKID-II (Phi-Koeffizient)
- Gesamt-Score (Range 0-10) des DIB-R mit dem SKID-II-*Diagnose-Index* (Range 1-3) der Persönlichkeitsstörungen (Produkt-Moment-Korrelation)
- Gesamt-Score (Range 0-10) des DIB-R mit der *Anzahl* erfüllter SKID-II-Kriterien (Range 0- max. 9) der jeweiligen Persönlichkeitsstörung (Produkt-Moment-Korrelation)

Eine Übersicht der Ergebnisse ist in der folgenden Tab. 24 wiedergeben.
Für alle o.g. Korrelationen der verschiedenen DIB-R-Werte ergeben sich bzgl. der Borderline-Persönlichkeitsstörung nach SKID-II signifikante Zusammenhänge. Es handelt sich stets um mittlere bis hohe Korrelationen.
Für die Korrelationen der verschiedenen DIB-R-Werte mit denen anderer SKID-II-Persönlichkeitsstörungen (also außer der Borderline-Persönlichkeitsstörung) ergeben sich jeweils nur eine bis zwei signifikante Korrelationen. Diese sind allesamt nur als gering einzustufen und wesentlich (um mehr als .30) kleiner als die entsprechende Korrelation mit der SKID-II-Borderline-Persönlichkeitsstörung.
Diese geringen Korrelationen sind, wie in der Tabelle ersichtlich, bei der *Schizotypischen*, *Dependenten*, der *Selbstunsicheren* und der *Paranoiden Persönlichkeitsstörung* zu finden.
Zu beachten sind die unterschiedlichen Fallzahlen der Koeffizienten. Die Angabe der tatsächlich erfüllten Kriterienanzahl der verschiedenen Persönlichkeitsstörungen nach SKID-II ist nur für die Störungsbereiche möglich, die komplett erhoben worden sind. Dies ist dem Screening-

Fragebogen folgend (der dem eigentlichen SKID-II-Interview vorgeschaltet ist) nur bei einem Teil der Störungen passiert. Außerdem kann nach der SKID-II-Anweisung die Exploration einzelner Persönlichkeitsstörungen abgebrochen werden, wenn klar wird, dass eine zur Diagnose ausreichende Anzahl erfüllter Kriterien nicht mehr erreicht werden kann.
War entsprechend eine Persönlichkeitsstörung nicht oder nur teilweise erfragt worden, wurde der Diagnoseindex regelgemäß auf '1' für 'nicht vorhanden' gesetzt. War 'fraglich' angekreuzt worden, wurde dieser Fall nicht in Korrelationen, die sich auf den Diagnose-Index beziehen, einbezogen. Für die endgültige Diagnosestellung einer Persönlichkeitsstörung, also das dichotome Merkmal 'vorhanden/nicht-vorhanden', gelten natürlich auch fragliche Fälle als 'nicht-vorhanden'.
Außerdem war die Persönlichkeitsstörung NNB wegen der Ausschließlichkeit dieser Diagnose in die folgenden Analysen nicht einbezogen worden: Eine Komorbidität ist prinzipiell nicht möglich, da die Diagnose Persönlichkeitsstörung NNB nur vergeben werden *darf*, wenn trotz erheblicher Auffälligkeiten im Persönlichkeitsbereich *keine* spezifische Persönlichkeitsstörungs-Diagnose zu stellen war.
Zur Übersicht und zum besserem Verständnis sind auch die der Berechnung der φ-Koeffizienten zugrunde liegenden Häufigkeiten des Vorliegens bzw. Nicht-Vorliegens der Borderline-Persönlichkeitsstörung nach DIB-R und der SKID-II-Persönlichkeitsstörungen (in Form von Kreuztabellen) in Tab. 25 wiedergegeben.

Zusammenfassend kann festgehalten werden, dass die verschiedenen Kennwerte der Borderline-Persönlichkeitsstörung nach SKID-II als äußeres Validitätskriterium, ganz im Sinne des Konzepts der konvergenten Validität, substanziell mit denen der DIB-R-Diagnose korreliert sind, während die der anderen SKID-II-Persönlichkeitsstörungen im Sinne einer hohen diskriminanten Validität nur geringe Korrelationen aufweisen.

Tab. 24: *Korrelationen DIB-R- mit SKID-II-Kennwerten, φ-Koeffizienten und Produkt-Moment-Korrelationen bei zweiseit. Signifikanztestung**

SKID-II-Persönlichkeitsstörungen			*Vorl. DIB-R- mit SKID-II-Diagnose (φ-Koeffizient)*	*DIB-R-Gesamt-Score mit SKID-II-Diagnoseindex*	*DIB-R-Gesamt-Score mit SKID-II-Kriterienanzahl*
Cluster C	**Selbstunsichere**	**r**	,03	,17	,07
		p	,76	,09	,60
		N	100	98	59
	Dependente	**r**	,13	***,33****	***,34****
		p	,18	***,001***	***,004***
		N	100	100	70
	Zwanghafte	**r**	-,03	,06	,08
		p	,75	,58	,50
		N	100	99	71
Cluster A	**Paranoide**	**r**	,05	,19	***,34****
		p	,60	,07	***,003***
		N	100	97	71
	Schizotypische	**r**	***,28****	***,21****	,20
		p	***,005***	***,04***	,16
		N	100	97	53
	Schizoide	**r**	-,01	-,03	,01
		p	,96	,79	,94
		N	100	97	38
Cluster B	**Histrionische**	**r**	-,03	,07	-,01
		p	,75	,50	,98
		N	100	99	40
	Narzisstische	**r**	,03	,04	,07
		p	,77	,73	,67
		N	100	100	41
	Borderline	**r**	***,60****	***,74****	***,77****
		p	***,000***	***,000***	***,000***
		N	100	100	96
	Antisoziale	**r**	,16	,17	
		p	,10	,09	
		N	100	100	

** signifikante Korrelationen sind durch Fettdruck hervorgehoben*

Tab. 25: *Kreuztabelle Borderline-Diagnose DIB-R * SKID-II-Diagnosen*

		BPS nach DIB-R		
SKID-II-Störungen		**nein**	**vorliegend**	**Total**
Selbstunsichere Persönlichkeitsstörung	**nein**	**59**	**20**	**79**
	Zeilen-%	79,7%	76,9%	79,0%
	vorlieg.	**15**	**6**	**21**
	Zeilen-%	20,3%	23,1%	21,0%
	Total	**74**	**26**	**100**
Dependente Persönlichkeitsstörung	**nein**	**65**	**20**	**85**
	Zeilen-%	87,8%	76,9%	85,0%
	vorlieg.	**9**	**6**	**15**
	Zeilen-%	12,2%	23,1%	15,0%
	Total	**74**	**26**	**100**
Zwanghafte Persönlichkeitsstörung	**nein**	**70**	**25**	**95**
	Zeilen-%	94,6%	96,2%	95,0%
	vorlieg.	**4**	**1**	**5**
	Zeilen-%	5,4%	3,8%	5,0%
	Total	**74**	**26**	**100**
Paranoide Persönlichkeitsstörung	**nein**	**68**	**23**	**91**
	Zeilen-%	91,9%	88,5%	91,0%
	vorlieg.	**6**	**3**	**9**
	Zeilen-%	8,1%	11,5%	9,0%
	Total	**74**	**26**	**100**
Schizotypische Persönlichkeitsstörung	**nein**	**73**	**22**	**95**
	Zeilen-%	98,6%	84,6%	95,0%
	vorlieg.	**1**	**4**	**5**
	Zeilen-%	1,4%	15,4%	5,0%
	Total	**74**	**26**	**100**
Schizoide Persönlichkeitsstörung	**nein**	**71**	**25**	**96**
	Zeilen-%	95,9%	96,2%	96,0%
	vorlieg.	**3**	**1**	**4**
	Zeilen-%	4,1%	3,8%	4,0%
	Total	**74**	**26**	**100**
Histrionische Persönlichkeitsstörung	**nein**	**70**	**25**	**95**
	Zeilen-%	94,6%	96,2%	95,0%
	vorlieg.	**4**	**1**	**5**
	Zeilen-%	5,4%	3,8%	5,0%
	Total	**74**	**26**	**100**
Narzisstische Persönlichkeitsstörung	**nein**	**72**	**25**	**97**
	Zeilen-%	97,3%	96,2%	97,0%
	vorlieg.	**2**	**1**	**3**
	Zeilen-%	2,7%	3,8%	3,0%
	Total	**74**	**26**	**100**
Borderline Persönlichkeitsstörung	**nein**	**65**	**7**	**72**
	Zeilen-%	87,8%	26,9%	72,0%
	vorlieg.	**9**	**19**	**28**
	Zeilen-%	12,2%	73,1%	28,0%
	Total	**74**	**26**	**100**
Antisoziale Persönlichkeitsstörung	**nein**	**73**	**24**	**97**
	Zeilen-%	98,6%	92,3%	97,0%
	vorlieg.	**1**	**2**	**3**
	Zeilen-%	1,4%	7,7%	3,0%
	Total	**74**	**26**	**100**

5.5.2. Diskriminanzanalyse: Multivariate Testung der Zusammenhänge von DIB-R- und SKID-II-Diagnosen

Die Unterschiedlichkeit zweier Gruppen oder Stichproben im Hinblick auf verschiedene Merkmale (bzw. die Analyse der Zusammenhänge der Gruppenzugehörigkeit mit den Merkmalen) kann wie bisher *univariat* für jedes Merkmal separat erfolgen.

Nach Bortz (1993, S. 559 ff.) kann aber letztendlich die Bedeutsamkeit der verschiedenen gefundenen Merkmalszusammenhänge auf dieser Ebene noch nicht abschließend beurteilt werden.

Eine simultane statt einer unabhängigen Betrachtung von Merkmalen ist erforderlich, damit z.B. Suppressions-Effekte aufgedeckt werden können: Interkorrelationen der Variablen könnten sonst zu völlig falschen Schlüssen führen. Hierfür leistet z.B. die Diskriminanzanalyse einen wertvollen Beitrag.

Mit der Diskriminanzanalyse kann man herausfinden, welchen Beitrag die unabhängigen Variablen für die Unterscheidung der zwei (oder auch mehr) Gruppen leisten, ob eine Unterscheidung möglich ist und welche Qualität diese hat.

Entsprechend der Terminologie der Varianzanalyse werden die Merkmale als unabhängige Variablen und die Gruppenzugehörigkeit als abhängige Variable bezeichnet.

Hierzu werden (entsprechend der β-Gewichte bei der multiplen Regressionsrechnung) Gewichtskoeffizienten ermittelt, die angeben, in welchem Ausmaß die unabhängigen Variablen am Zustandekommen des Gesamtunterschiedes beteiligt sind.

Die Gewichtskoeffizienten (standardisierte kanonische Diskriminanzfunktions-Koeffizienten) geben an, wie die einzelnen Merkmale zu gewichten sind, damit eine maximale Separierung der Stichproben erreicht werden kann.

Außerdem können Strukturkoeffizienten berechnet werden. Diese sind die über die Gruppen gemittelten Korrelationen zwischen den unabhängigen Variablen und den für die jeweilige Person vorhergesagten Diskriminanzwerten. Auch diese Koeffizienten erlauben eine Einschätzung des Beitrags der Merkmale zur Gruppenunterscheidung.

(Die Diskriminanzanalyse kann auch der Zuordnung "neuer" Individuen noch unbekannter Gruppenzugehörigkeit allein aufgrund der unabhängigen Variablen dienen. Eine zuordnende Diskriminanzfunktion kann errechnet werden. Dieser Aspekt der Diskriminanzanalyse spielt bei diesem Untersuchungsansatz aber *keine* Rolle.)

In die Analyse fanden 93 von 100 Fällen Eingang, für die alle verwendeten und im Folgenden dargelegten Diagnose-Indizes vorlagen.

Vorgenommen wird also der Versuch einer Vorhersage der Borderline-Diagnosen nach DIB-R aus den vorhandenen SKID-II-Diagnosen (bzw. den jeweiligen Indizes). Die aus den Diagnose-Indizes aller SKID-II-Persönlichkeitsstörungen pro Patient errechneten Diskriminanzwerte korrelieren mit r_{Pbis}=.69 relativ hoch mit der Gruppenzugehörigkeit. Auch unterscheiden sich die Diskriminanzwerte nach Wilks' Lambda hochsignifikant über die beiden Gruppen (p=.000, χ^2=55,9, df=10), sodass von einer Unterscheidbarkeit der Patienten-Gruppen mit bzw. ohne Borderline-Persönlichkeitsstörung nach DIB-R anhand der SKID-II-Diagnosen gesprochen werden kann.

Die Diskriminanzkoeffizienten, die die Bedeutung der unabhängigen Variablen bei der Gruppenzuordnung angeben, sind dabei entsprechend der Ergebnisse im vorigen Kapitel eindeutig verteilt.

Tab. 26: *Ergebnisse der Diskriminanzanalyse: Diskriminanz- und Strukturkoeffizienten der einzelnen SKID-II-Persönlichkeitsstörungen (bezogen auf das Vorliegen einer BPD nach DIB-R)*

	Diskriminanzkoeffizienten	***Strukturkoeffizienten***
Selbstunsichere PS	,12	,07
Dependente PS	**,42**	,15
Zwanghafte PS	-,21	-,01
Paranoide PS	***-,32***	,08
Schizotypische PS	**,47**	,27
Schizoide PS	,02	,06
Histrionische PS	-,18	,04
Narzisstische PS	**,38**	,16
Borderline PS	**1,12**	**,79**
Antisoziale PS	,14	,15

Den mit Abstand größten Beitrag leistet den *Diskriminanzkoeffizienten* zufolge der *SKID-II-Diagnose-Index der Borderline-Persönlichkeitsstörung*. Alle weiteren Diagnose-Indizes erreichen nur eine geringe Höhe und sind z.T. sogar leicht negativ. Nennenswerte positive Beiträge leisten nur noch die Indizes der Schizotypischen, der Dependenten und in geringerem Maße der Narzisstischen Persönlichkeitsstörung. Das Vorliegen einer Paranoiden

Persönlichkeitsstörung scheint bei einem Index von -.32 sogar in gewissem Maße gegen das Vorliegen einer Borderline-Persönlichkeitsstörung zu sprechen.

Die Strukturkoeffizienten zeigen nur für die SKID-II-Borderline-Persönlichkeitsstörung eine hohe Korrelation von r=.79 auf, alle anderen Korrelationen sind als sehr niedrig zu bezeichnen. Negative Korrelation treten praktisch nicht auf.

Zusammenfassend ist festzustellen, dass *nur die Diagnose-Indizes der Borderline-Persönlichkeitsstörung einen erheblichen Beitrag zur Gruppenzuordnung leisten*. Diese Ergebnisse entsprechen in vollem Umfang den Ergebnissen aus dem vorigen Kapitel.

Die guten Ergebnisse zur diskriminanten und konvergenten Validität haben damit auch bei einer multivariaten Prüfung Bestand.

5.6. KONVERGENTE UND DISKRIMINANTE VALIDITÄT: ZUORDNUNG DER PATIENTEN MIT UND OHNE BPS NACH DIB-R ZU DEN GEBILDETEN DSM-DIAGNOSEGRUPPEN

Die Analyse der Konstruktvalidität anhand der konvergenten und diskriminanten Validität wird nun unter zusätzlicher Einbeziehung des SKID-I anhand der gebildeten DSM-Diagnosegruppen fortgesetzt.

5.6.1. Analyse der Häufigkeitsverteilungen

Es war angenommen worden, dass sich die Patienten mit Borderline-Persönlichkeitsstörung nach DIB-R zum größten Teil in der Diagnosegruppe SKID-II-Borderline-Persönlichkeitsstörung wiederfinden würden, ein geringer Teil in der Gruppe Andere Persönlichkeitsstörungen und möglichst *kein* Patient in einer anderen Diagnosegruppe.

Hierzu war die **Hypothese 2.1.** gebildet worden:

Die Verteilung der Patienten mit und ohne Borderline-Persönlichkeitsstörung ist signifikant unterschiedlich über die definierten DSM-Diagnosegruppen. Die häufigste Kombination ist dabei die der *BPS nach DIB-R* und *BPS nach SKID-II*, die zweithäufigste *BPS nach DIB-R* und *Andere Persönlichkeitsstörungen*.

Die vorgefundenen Ergebnisse sind in Abb. 4 sowie in Tab. 27 wiedergegeben.

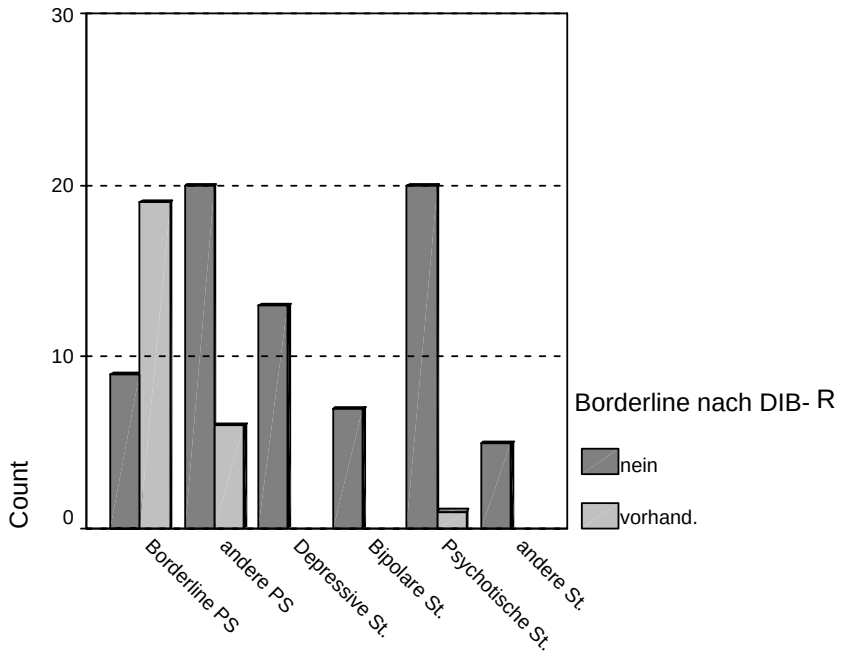

Abb. 4: *Verteilung der Patienten mit DIB-R-Diagnosen über die DSM-Diagnosegruppen (N=100)*

Entsprechend dem Ergebnis zur Kriteriumsvalidität (Kap. 5.1) findet sich der größte Teil der Borderline-Diagnosen in der DSM-Diagnosegruppe SKID-II-Borderline-Persönlichkeitsstörung wieder: Mit 19 Patienten sind 73,1% der DIB-R-Borderliner in dieser Diagnosegruppe.

Von den übrigen sieben Patienten mit DIB-R-Borderline-Persönlichkeitsstörung sind sechs (23,1%) in der Gruppe der Anderen Persönlichkeitsstörungen aufzufinden. Bei einem einzigen Patienten mit Borderline-Diagnose nach DIB-R (also rechnerisch 3,8% der Patienten) tritt eine große Diskrepanz der Diagnosen nach DIB-R und DSM auf: diese Person hat statt der Diagnose einer Persönlichkeitsstörung die einer Schizophrenie erhalten und war somit in der DSM-Diagnosegruppe Psychotische Störungen einzuordnen.

Die dargestellte Verteilung ist mit einem asymptotischen χ^2-Test gerechnet hochsignifikant von einer zufälligen zu erwartenden Verteilung verschieden (χ^2=39,3/df=5).

Die sechs Patienten mit Borderline-Diagnosen nach DIB-R, die nicht die SKID-II-Diagnose einer Borderline-Persönlichkeitsstörung, sondern (bis zu drei) andere Persönlichkeitsstörungs-Diagnosen erhalten haben, sind in Tab. 28 mit ihren Codiagnosen der Anschaulichkeit halber aufgeführt.

Tab. 27: *Kreuztabelle Borderline-Diagnose DIB-R * DSM-Diagnosegruppen (2x6-Felder)*

DSM-Diagnosegruppen		BPS nach DIB-R		Total
		nein	vorhand.	
Borderline PS	**N**	**9**	**19**	**28**
	Zeilen-%	32,1%	67,9%	100,0%
	Spalten-%	12,2%	73,1%	28,0%
Andere PS	**N**	**20**	**6**	**26**
	Zeilen-%	76,9%	23,1%	100,0%
	Spalten-%	27,0%	23,1%	26,0%
Depressive St.	**N**	**13**	**0**	**13**
	Zeilen-%	100,0%		100,0%
	Spalten-%	17,6%		13,0%
Bipolare St.	**N**	**7**	**0**	**7**
	Zeilen-%	100,0%		100,0%
	Spalten-%	9,5%		7,0%
Psychotische St.	**N**	**20**	**1**	**21**
	Zeilen-%	95,2%	4,8%	100,0%
	Spalten-%	27,0%	3,8%	21,0%
Andere St.	**N**	**5**	**0**	**5**
	Zeilen-%	100,0%		100,0%
	Spalten-%	6,8%		5,0%
Total	**N**	**74**	**26**	**100**
	Zeilen-%	74,0%	26,0%	100,0%
	Spalten-%	100,0%	100,0%	100,0%

Tab. 28: *Persönlichkeitsstörungs-Codiagnosen der Patienten mit einer Diagnoseabweichung zwischen SKID-II und DIB-R*

	Persönlichkeitsstörungen nach SKID-II		
Patient 1	Schizotypische	Selbstunsichere	
Patient 2	Antisoziale		
Patient 3	Selbstunsichere		
Patient 4	NNB		
Patient 5	Schizotypische	Schizoide	Zwanghafte
Patient 6	Narzisstische		

Wegen der geringen Fallzahl ist eine Untersuchung der Frage, ob die Codiagnosen den Unterschied zwischen der DIB-R- und der SKID-II-Diagnostik bedingt haben könnten, nicht möglich.

5.6.2. Interindividuelle Unterschiede in den DIB-R-Kennwerten der Patienten der Diagnosegruppen

Es war angenommen worden, dass sich die Patienten der verschiedenen Diagnosegruppen nach SKID-II nicht nur in der DIB-R-Diagnose sondern auch in den verschiedenen Kennwerten des DIB-R voneinander unterscheiden würden.

Hierzu waren folgende Hypothesen formuliert worden:

2.2.a) Die DIB-R-Gesamt-Scores der Patienten unterscheiden sich über die verschiedenen Diagnosegruppen nach SKID-II signifikant voneinander.

2.2.b) Die Patienten der SKID-II Diagnosegruppe Borderline-Persönlichkeitsstörung unterscheiden sich im DIB-R-Gesamt-Score signifikant von den Patienten jeder anderen Diagnosegruppe und haben den höchsten Wert.

2.3.a) Für jeden der vier Skalierten Section-Scores (SSS) der Bereiche des DIB-R gilt: Über die verschiedenen Diagnosegruppen nach SKID-II sind die Mittelwerte des SSS der Patienten signifikant unterschiedlich.

2.3.b) Für jeden der vier Skalierten Section-Scores (SSS) der Bereiche des DIB-R gilt: Die Patienten der SKID-II-Diagnosegruppe Borderline-Persönlichkeitsstörung unterscheiden sich im Mittelwert des SSS signifikant von den Patienten aller anderen Diagnosegruppen und haben jeweils den höchsten Wert.

5.6.2.1. Ergebnisse zu den DIB-R-Gesamt-Scores

Die in Abb. 5 und Tab. 29 wiedergegebenen Werte zeigen, dass die mittleren DIB-R-Gesamt-Scores in der Tat deutlich voneinander abweichen. Die Gruppe der Patienten mit SKID-II-Borderline-Persönlichkeitsstörung erreicht einen Gesamt-Score von 8,1, während die nächst niedriger liegende Gruppe, die der Anderen Persönlichkeitsstörungen, mit einem Wert von 4,58 schon deutlich niedriger liegt. Die geringsten Werte weisen die Patienten mit einer Bipolaren Störung auf. Diese Unterschiede im Niveau sind mit einer einfaktoriellen Varianzanalyse über alle Gruppen hinweg getestet hochsignifikant. Ein Scheffé-Test zur Prüfung auf signifikante Einzelunterschiede der Gruppen im Gesamt-Score erbringt ebenso eindeutige Ergebnisse: Nur die Gruppe Borderline-Persönlichkeitsstörung nach SKID-II unterscheidet sich im Einzelvergleich hinsichtlich des Gesamt-Scores signifikant von jeder anderen Gruppe. Weitere signifikante Einzelunterschiede hinsichtlich des Gesamt-Scores liegen nicht vor. Die Hypothesen 2.2.a) und 2.2.b) können somit beibehalten werden.

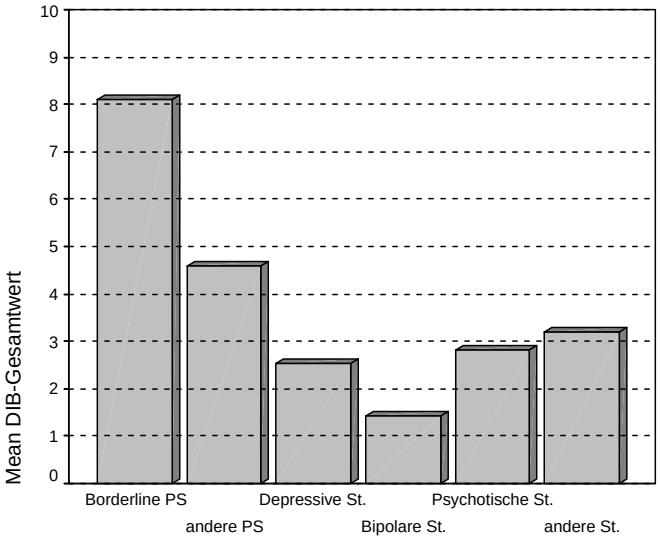

Abb. 5: *Mittelwerte des DIB-R-Gesamt-Scores nach Diagnosegruppen*

Tab. 29: *Mittelwerte und Streuungen des DIB-R-Gesamt-Scores nach Diagnosegruppen; Ergebnisse der One-Way-Anova und der Scheffé-Tests auf signifikante Einzelunterschiede zwischen den Diagnosegruppen*

	M	**s**
Borderline PS (N=28)	8,11	1,97
Andere PS (N=26)	4,58	2,61
Depressive St. (N=13)	2,54	2,07
Bipolare St. (N=7)	1,43	2,51
Psychotische St. (N=21)	2,81	2,48
Andere St. (N=5)	3,20	2,59
F (bei df 5/94)	20,01	
p	.000	
Scheffé-Test (auf signifik. Diff.)	BPS von allen anderen	
$\acute{\eta}^2$	,52	
f	1,03	

5.6.2.2. Ergebnisse zu den Skalierten Section-Scores des DIB-R

In den einzelnen DIB-R-Bereichen zeigt sich ein ebenfalls klares, wenn auch weniger einheitliches Bild. Die Ergebnisse sind in den folgenden Abb. 6–9 sowie in Tab. 30 wiedergegeben.

Vorab ist festzustellen, dass in jedem DIB-R-Bereich der maximal mögliche Wertebereich der Skalierten Section-Scores von 0-2 in den Bereichen Affekte und Kognition und von 0-3 in den

Bereichen Impulsivität und Zwischenmenschliche Beziehungen von den Beurteilern mit den entsprechenden Minima und Maxima in jeder Diagnosegruppe voll ausgeschöpft worden ist.

Im **Affektbereich** ergeben sich mit p=.015 signifikante Unterschiede im Skalierten Section-Score über die Diagnosegruppen. (Von allen DIB-R- Bereichen ist diese die niedrigste Signifikanz.)
Im Affektbereich sind sich, im Vergleich mit den übrigen DIB-R-Bereichen, die Werte der Skalierten Section-Scores aller Diagnose-Gruppen am ähnlichsten.

Die höchsten Werte haben wie erwartet mit 1,50 die Gruppe der Borderline-Persönlichkeitsstörung und mit 1,27 die Patienten mit Anderen Persönlichkeitsstörungen. Den niedrigsten Wert hat mit 0,71 die Gruppe der Patienten mit Bipolarer Störung erhalten. Signifikante Einzelunterschiede zwischen den Diagnosegruppen ergeben sich nach dem Scheffé-Test im Affektbereich in keinem Fall. Lediglich ist ein statistischer Trend zwischen den Gruppen Borderline-Persönlichkeitsstörung und Psychotische Störung erkennbar. Der mit einer Differenz von 0,79 numerisch größere Unterschied zwischen der Borderline-Gruppe und der Gruppe Bipolarer Störungen ist insignifikant (da bei nahezu gleicher Streuung die Gruppe kleiner ist).

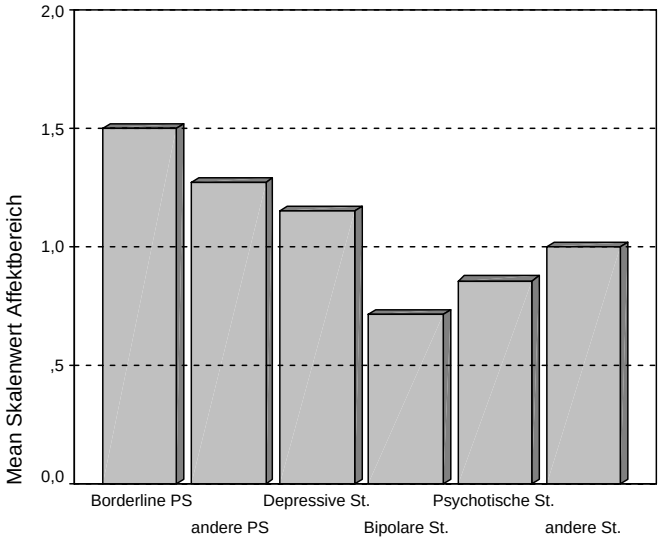

Abb. 6: *Mittelwerte der Diagnose-Gruppen in den Skalierten Section-Scores des Affektbereichs*

Tab. 30: *Mittelwerte und Streuungen der Skalierten Section-Scores aller DIB-R-Bereiche; Ergebnisse der One-Way-Anovas und der Scheffé-Tests auf signifikante Einzelunterschiede zwischen den Diagnosegruppen*

DIB-R-Bereiche	**Affekte**		**Kognition**		**Impulshandl.**		**Beziehungen**	
Diagnosegruppen	**M**	**s**	**M**	**s**	**M**	**s**	**M**	**s**
Borderline PS (N=28)	1,50	,69	1,61	,63	2,39	1,07	2,61	,83
andere PS (N=26)	1,27	,67	,69	,68	1,12	1,28	1,50	1,27
Depressive St. (N=13)	1,15	,55	,54	,66	,31	,75	,54	1,05
Bipolare St. (N=7)	,71	,76	,29	,76	,00	,00	,43	1,13
Psychotische St. (N=21)	,86	,73	,33	,66	,67	1,11	,95	1,16
andere St. (N=5)	1,00	,71	,60	,89	,80	1,10	,80	1,10
F (bei df 5/94)	2,98		11,44		11,74		10,58	
p	,015		,000		,000		,000	
Scheffé-Test (auf signifik. Differenzen)	*keine* Differenzen (statist. Trend *BPS* vs. *Psychot. St.*)		*BPS* vs. alle Gruppen außer *Andere St.*		*BPS* vs. alle Gruppen außer *Andere St.*		*BPS* vs. alle Gruppen	
η^2	,14		,38		,38		,36	
f	,40		,78		,79		,75	

Im **Kognitionsbereich** fallen die Unterschiede zwischen den Diagnosegruppen deutlich größer aus. Die Borderline-Gruppe weist mit im Mittel 1,61 von 2 möglichen Punkten im Skalierten Section-Score den höchsten Wert auf, mit großem Abstand gefolgt von der Gruppe der Anderen Persönlichkeitsstörungen mit einem Mittelwert von ,69. Die niedrigsten Werte haben wiederum die Gruppen Bipolare Störung (,29) und Psychotische Störung (,33).

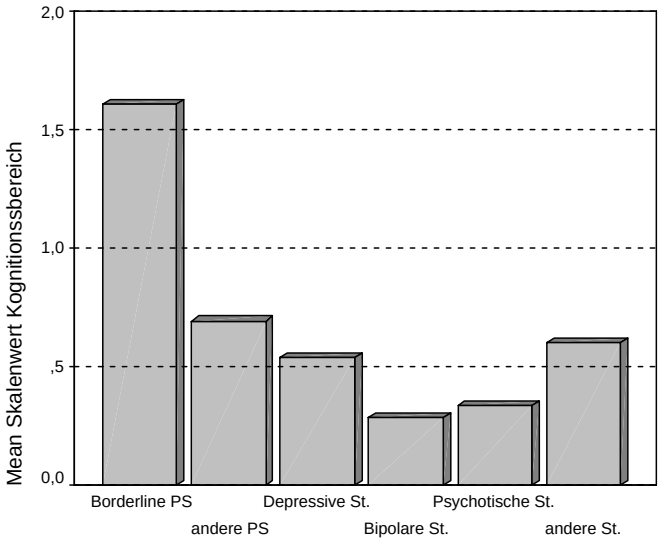

Abb. 7: *Mittelwerte der Diagnose-Gruppen in den Skalierten Section-Scores des Kognitionsbereichs*

Diese Gruppenunterschiede sind, getestet mit einer einfaktoriellen Varianzanalyse über die Gruppen hinweg, hochsignifikant. In den Einzelvergleichen unterscheidet sich die Gruppe Borderline-Persönlichkeitsstörung signifikant von allen anderen Gruppen, ausgenommen der zusammengefassten Gruppe der Anderen DSM-Störungen. (Diese hat zwar einen den anderen Gruppen vergleichbar niedrigen Mittelwert, ist aber sehr klein und weist eine höhere Streuung auf. Daher ist hier ein signifikanter Unterschied nicht zu erwarten.) Weitere signifikante Einzelunterschiede zwischen den Diagnosegruppen liegen nicht vor.

Im Bereich **Impulshandlungen** sind die Unterschiede zwischen den Diagnosegruppen ebenfalls groß. Die Gruppe der SKID-II-Borderline-Persönlichkeitsstörung weist mit im Mittel 2,39 von 3 möglichen Punkten im Skalierten Section-Score den höchsten Wert auf. Mit großem Abstand folgen die Anderen Persönlichkeitsstörungen mit einem Mittelwert von 1,12. Den niedrigsten Wert weisen wieder die bipolaren Störungen auf: kein Patient hat hier Punkte bekommen, so dass der Skalierte Section-Scores dieser Gruppe 0 beträgt. Den zweitniedrigsten Mittelwert von 0,31 haben die Patienten mit Depressiven Störungen.

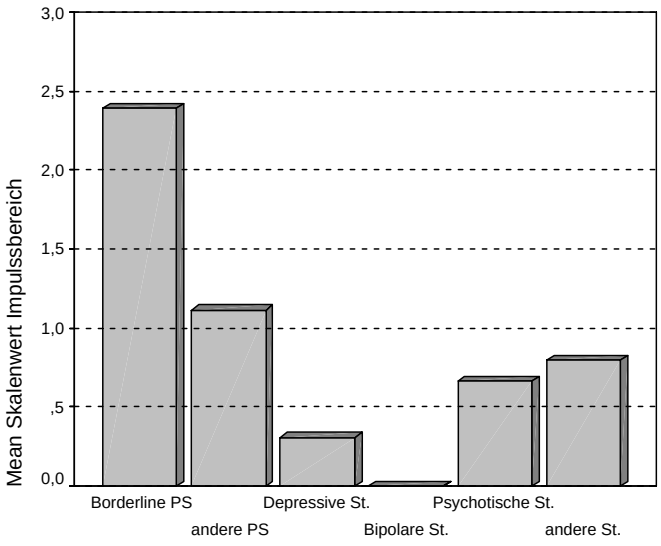

Abb. 8: *Mittelwerte der Diagnose-Gruppen in den Skalierten Section-Scores des Bereichs Impulshandlungen*

Die Gruppenunterschiede sind getestet mit der einfaktoriellen Varianzanalyse über alle Gruppen hinweg hochsignifikant. In den Einzelvergleichen unterscheidet sich die Gruppe Borderline-Persönlichkeitsstörung, entsprechend dem Kognitions-Bereich, signifikant von allen anderen Gruppen, ausgenommen von der der Anderen DSM-Störungen. Weitere signifikante Unterschiede zwischen den Diagnosegruppen im Scheffé-Test wurden nicht gefunden.

Auch im Bereich **Zwischenmenschliche Beziehungen** sind große Differenzen in den Skalierten Section-Scores vorzufinden. Die Borderline-Persönlichkeitsstörung weist mit im Mittel 2,61 von 3 möglichen Punkten im Skalierten Section-Score erwartungsgemäß wieder den höchsten Wert auf, mit deutlichem Abstand gefolgt von den Anderen Persönlichkeitsstörungen bei einem Mittelwert von 1,50. Die niedrigsten Werte der Stichprobe haben die Bipolaren Störungen mit 0,43 und die Depressiven Störungen mit 0,54 inne.

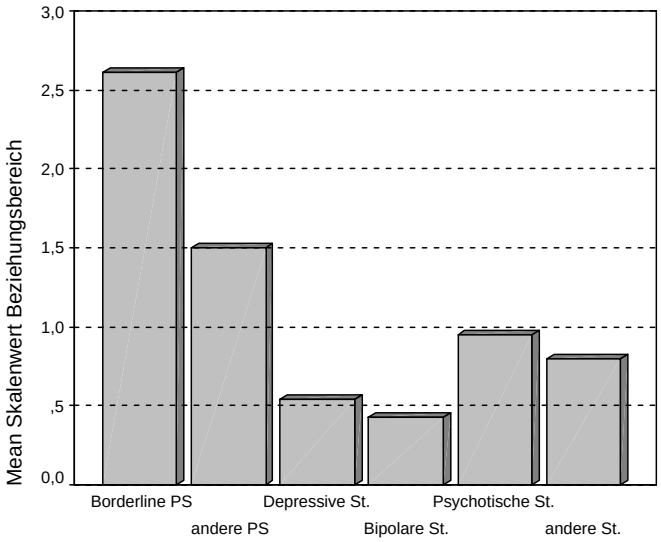

Abb. 9: *Mittelwerte der Diagnose-Gruppen in den Skalierten Section-Scores des Beziehungsbereichs*

Die Gruppenunterschiede sind getestet mit der einfaktoriellen Varianzanalyse über die Gruppen hinweg wieder hochsignifikant. In den Einzelvergleichen unterscheidet sich die Gruppe Borderline-Persönlichkeitsstörung diesmal signifikant von ausnahmslos allen anderen Diagnosegruppen. Weitere signifikante Einzelunterschiede zwischen den Diagnosegruppen liegen nicht vor.

Für alle Bereiche kann zusammenfassend festgestellt werden, dass in der Stichprobe – wie schon beim DIB-R-Gesamt-Score – die Gruppe der Patienten mit Borderline-Persönlichkeitsstörung in allen Skalierten Section-Scores die höchsten, die der Patienten mit anderen Persönlichkeitsstörungen die zweithöchsten und die der Patienten mit bipolaren Störungen die niedrigsten Werte aufweist. Die Rangreihe variiert über die DIB-R-Bereiche hinweg nur für die Diagnosegruppen Depressive Störungen, Psychotische Störungen und Andere DSM-Störungen.

Unter Einbeziehung der Ergebnisse der Scheffé-Tests kann gesagt werden, dass die Scores der Gruppe Borderline-Persönlichkeitsstörung

- im DIB-R-Gesamt-Score und im Bereich Zwischenmenschliche Beziehungen signifikant höher sind als die aller anderen Diagnosegruppen

- in den Bereichen Kognition und Impulsivität signifikant höher liegen als die aller anderen Diagnosegruppen mit Ausnahme der Anderen DSM-Störungen
- sich im Affektbereich nur als statistischer Trend von der Gruppe der Psychotischen Störungen unterscheiden

Zur besseren Einschätzbarkeit der praktischen Bedeutsamkeit der vorgefundenen Unterschiede wurden Effektgrößen berechnet. Die vom Statistikprogramm SPSS bei der Berechnung der Varianzanalyse anforderbare Schätzung der Effektgröße Eta-Quadrat wurde nach folgender

Formel 9.11 $f = \sqrt{\frac{\eta^2}{1-\eta^2}}$

nach Bortz und Döring (1995, S. 571) für die einzelnen Analysen in die vorgeschlagene Effektgröße f umgerechnet.

Der Bereich Affekte hat entsprechend der ausbleibenden Unterschiede im Scheffé-Test die mit Abstand niedrigste Effektstärke von f=.40, alle anderen liegen in einer Höhe zwischen f=.75 und .79. Im Gesamt-Score wird sogar eine Effektstärke von 1,03 erreicht.

Nach Bortz und Döring (1995, S. 568) gelten Effektgrößen f der Varianzanalyse in der Größenordnung von etwa .10 als klein, von .25 an als mittel und ab .40 als hoch. Dementsprechend wurden für alle Bereiche hohe bis sehr hohe Effektstärken erreicht.

5.6.3. Beurteilung der Hypothesen

Insgesamt sind die Vorhersagen der Hypothese 2.3.a) für jeden der vier DIB-R Bereiche eingetroffen, die Hypothese kann daher beibehalten werden.

Die Ergebnisse entsprechen der Hypothese 2.3.b) aber nicht vollständig:

Für den Bereich Zwischenmenschliche Beziehungen kann die Hypothese beibehalten werden.

Auch für die Bereiche Kognition und Impulshandlungen ist die Hypothese 2.3.b) beizubehalten, wenn auch mit einem gewissen Erklärungsbedarf: Der ausbleibende signifikante Unterschied der Gruppe Borderline-Persönlichkeitsstörung gegenüber der Gruppe der Anderen DSM-Störungen ist bzgl. der Hypothese ohne Bedeutung und nicht interpretierbar, da die Gruppe der Anderen DSM-Störungen eine im Studiendesign und bei der Formulierung der Hypothesen nicht vorgesehene und zudem sehr kleine bzgl. der Diagnosen heterogene Restgruppe ist.

Für den Affektbereich ist die Hypothese 2.3.b) aber eindeutig zurückzuweisen, denn in diesem Bereich unterscheiden sich die Gruppen zu wenig und offenbar unsystematisch voneinander.

6. ERGEBNISSE ZUR RELIABILITÄT DES DIB-R

In diesem Teil der Untersuchung wird v.a. die Interrater-Reliabilität des DIB-R anhand von Videoaufzeichnungen der mit psychiatrischen Patienten durchgeführten Interviews überprüft. Diesen Interviews folgend wurden von Ratergruppen unabhängig DIB-R-Interviewbögen ausgefüllt. Es handelte sich dabei um eine Expertengruppe und um eine Gruppe vorher in der Anwendung des DIB-R trainierter Psychologie-Studenten (siehe Kap. 3.6.3.2).

Diese beiden Ratergruppen haben nur zu einem kleineren Teil dieselben Bänder gesehen, die Studentengruppe außerdem insgesamt weniger. Daher sind die gefundenen Gütemaße beider Gruppen *nicht unmittelbar* vergleichbar.

Im Folgenden wird für jede der Ratergruppen zunächst ein "globaler" Mittelwert jedes Statements oder DIB-R-Scores über alle Patienten und Rater vorgestellt. Dieser stellt also ein an der Anzahl der Rater gewichtetes mittleres Rating dar: Zuerst erfolgte fallweise eine Mittelwertsbildung *aller* jeweiligen Rater-Einschätzungen. Diese mittleren Raterurteile wurden dann wiederum über alle Fälle gemittelt. (Durch die in der Expertengruppe schwankende Rateranzahl war ein solches Vorgehen nötig. Bei konstanten Rateranzahlen wie in der Studentengruppe können die Raterurteile über alle Patienten auch in einem Schritt insgesamt gemittelt werden.)

Weiter wird die durchschnittliche fallweise Häufigkeit der einzelnen verwendeten Skalenstufen pro Statement oder Skaliertem Section-Score angegeben.

Für die Reliabilitätseinschätzung sind die o.a. Werte deswegen bedeutsam, da sehr einförmig verwendete Statements *a priori nur wenig Variation über die Probanden aufweisen können*. Statements mit wenig Variation über die Probanden sind, wie im Kap. 2.4 dargelegt, für eine Reliabilitätstestung mittels des *Intraclass-Korrelationskoeffizienten (ICC)* aber weniger geeignet, da die Reliabilität dann voraussichtlich unterschätzt wird.

Um eine differenziertere Einschätzung des Sachverhalts zu ermöglichen, wurde zu Vergleichszwecken außer dem ICC auch der nicht von solch fehlender Variation beeinflusste *Finn-Koeffizient* berechnet. Zur Vergleichsmöglichkeit mit der Studie von Eckert et al. (1987) zur Reliabilität der ersten DIB-Version wurde außerdem auch ein *Horst-Koeffizient* berechnet.

Zur Klassifizierung der Höhe der gefundenen Reliabilitäts-Koeffizienten wurde in Kap. 2.4.7 ein Einordnungs- und Benennungsschema geschildert.

6.1. EXPERTEN-STICHPROBE

6.1.1. Beschreibung der Stichprobe der untersuchten Patienten

Von den Experten wurden insgesamt 19 psychiatrische Patienten auf Video gesehen und geratet, komplett mit Diagnose liegen aber nur 18 DIB-R-Interviews vor. Zehn der untersuchten Patienten waren männlichen, neun weiblichen Geschlechts. Das Durchschnittsalter lag bei 28,6 Jahren (s=6,8; Range 18-44 J.). Von den 19 Patienten waren 12 im Rahmen der Studie zur DIB-R-Validierung auch mit den SKID-I- und -II-Interviews untersucht worden, von sieben Patienten lagen nur Videoaufnahmen des DIB-R vor. Die DSM-Diagnose-Kategorien der 12 SKID-untersuchten Patienten waren sechs Mal Borderline-Persönlichkeitsstörung, zwei Mal Andere Persönlichkeitsstörung, zwei Mal Depressive Störung, einmal Bipolare Störung und einmal Psychotische Störung.

6.1.2. Ergebnisse der Experten zu den DIB-R-Statements

6.1.2.1. Rateranzahl, Mittelwerte und Verteilung

Die durchschnittliche Rateranzahl ist in der folgenden Tab. 31 angegeben. Sie variiert über die Statements nur geringfügig zwischen 4,5 und 4,7.

Die globalen Statement-Mittelwerte schwanken zwischen einem Minimum von M=0,45 und einem Maximum von M=1,71. Die geringsten Werte haben das Statement S. 8 aus dem Kognitionsbereich (M=0,45) und das Statement S. 10 (M=0,55) aus dem Bereich Impulsivität.

Die Statements (S.) mit den jeweils höchsten Werten liegen sämtlich im Affektbereich: Es handelt sich um S.1 (M=1,6), S. 2 (M=1,71) und S. 5 (M=1,62). (Die textlichen Statementbeschreibungen sind zum besseren Verständnis in den folgenden Tabellen angegeben.)

Über die Patienten und Rater besonders einseitig verteilte Statements sind ebenfalls S. 1 und S. 2: Bei S. 1 wurden 66% aller Ratings mit "2" bewertet und nur 6% mit "0". Beim S. 2 wurde sogar in 74% die "2" vergeben und in nur 4% eine "0" geratet.

Zwei ebenfalls auffallende, wenn auch gleichförmiger verteilte Statements sind das S. 16 und das S. 8. Beim S. 16 haben immerhin 59% der Rater eine "2" und nur 11% eine "0" vergeben. Beim S. 8, dem Statement mit dem geringsten Mittelwert, vergaben umgekehrt 68% der Rater eine "0" und nur 13% eine "2" (Übrige, oben nicht erwähnte Ratings waren bei einem möglichen Wertebereich von 0 bis 2 jeweils "1": Diese mittlere Kategorie war in jedem Fall die zweithäufigste Besetzung.)

Tab. 31: *Experten-Rating; Mittelwerte der DIB-R-Statements, Nennungshäufigkeiten der Skalenstufen in Prozent*

	S.	Der Patient ...	Rater-anzahl	M	*Nennungsanteile in %*		
					"0"	*"1"*	*"2"*
Affekte	**1**	litt an einer chronischen depressiven Verstimmung oder hatte eine oder mehrere Perioden von Major Depression	4,6	1,60	6	28	66
	2	hatte anhaltende Gefühle von Hilf-, Hoffnungs-, Wertlosigkeit oder Schuld	4,6	1,71	4	22	74
	3	hatte chronische Gefühle von Ärger, Wut oder verhielt sich häufig ärgerlich wütend	4,6	1,34	18	30	52
	4	hat sich chronisch sehr ängstlich gefühlt oder litt häufig unter körperlichen Angstsymptomen	4,6	1,50	16	18	66
	5	erlebte chronische Gefühle von Einsamkeit, Langeweile oder Leere	4,6	1,62	13	13	74
Kognition	**6**	neigte zu seltsamen Denken oder ungewöhnlichen Wahrnehmungserlebnissen	4,5	0,96	39	26	35
	7	hatte häufig flüchtige, nicht-wahnhafte paranoide Erlebnisse	4,5	1,40	18	24	58
	8	hatte wiederholte pseudo-psychotische Wahnvorstellungen oder Halluzinationen	4,6	0,45	68	19	13
Impulsbereich	**9**	betrieb einen ernsthaften Drogenmissbrauch	4,6	1,14	31	25	44
	10	hatte ein Muster sexuell abweichenden Verhaltens	4,7	0,55	68	10	22
	11	zeigte ein Muster von körperlicher Selbstbeschädigung	4,6	1,04	43	9	47
	12	zeigte ein Muster von manipulativen Selbstmorddrohungen, -gesten oder –versuchen	4,6	1,05	38	19	43
	13	zeigte ein anderes Muster impulsiven Verhaltens	4,6	1,11	35	19	46
Zwischenmenschliche Beziehungen	**14**	hat typischerweise versucht, das Alleinsein zu vermeiden oder fühlte sich extrem dysphorisch, wenn er allein war	4,5	1,34	23	21	57
	15	hat (wiederholt) Verlassenheits-, Verschlingungs- oder Vernichtungsängste erlebt	4,5	1,41	23	14	64
	16	hat seine Abhängigkeitswünsche stark abgewehrt oder befand sich in einem ernsten Konflikt zwischen Versorgen und Versorgtwerden	4,5	1,47	11	30	59
	17	neigte zu intensiven instabilen engen Beziehungen	4,5	0,88	46	21	34
	18	hatte in engen Beziehungen immer wieder Probleme mit Abhängigkeit oder Masochismus	4,5	1,22	32	15	53
	19	hatte in engen Beziehungen wiederkehrende Probleme mit Abwertung, Manipulation oder Sadismus	4,5	0,93	41	24	34
	20	hatte in engen Beziehungen immer wieder Probleme mit seiner Forderungs- oder Anspruchshaltung	4,6	1,06	40	14	46
	21	zeigte während der Therapie oder der psychiatrischen Hospitalisierung eine deutliche Regression	4,5	0,80	56	9	36
	22	hat auf der psychiatr. Station oder in einer Psychotherapie auffallende Gegenübertragungsreaktionen ausgelöst oder ist mit einem profess. Helfer eine ganz besondere Beziehung eingegangen	4,5	0,86	42	30	28

6.1.2.2. Reliabilitäts-Kennwerte nach ICC und Finn-Koeffizient

Nur drei der 22 Statements liegen im Bereich ungenügender Reliabilität: Diese sind die Statements S. 1 und S. 2 aus dem Affektbereich und das S. 16 aus dem Bereich Zwischenmenschliche Beziehungen, bei denen zuvor bereits eine extrem einseitige und daher varianzarme Verteilung der Ratings festgestellt worden war.

Eine Betrachtung der Reliabilitäten nach dem Finn-Koeffizienten erbringt hier wie erwartet andere Ergebnisse. Auffallend ist zunächst, dass die Finn-Koeffizienten für alle Statements, die bisher wenigstens befriedigend abgeschnitten hatten, größenordnungsmäßig gleich geblieben sind.

Für die Mehrzahl der nach dem ICC *unbefriedigend* abgeschnittenen Statements traten aber Verbesserungen auf, für wenige auch Verschlechterungen.

Insgesamt liegen nach dem Finn-Koeffizienten berechnet nun die Reliabilitäten von 16 Statements im Bereich guter Reliabilität. Nur fünf Statements weisen noch wenig befriedigende Kennwerte auf. Beim Statement S. 19 ist die Reliabilität nach dem Finn-Koeffizienten mit einem Wert von .49 sogar in den Bereich ungenügender Reliabilität <.50 abgefallen.

Die Statements S. 1 und S. 2 aber, die nach dem ICC noch ungenügend abgeschnitten hatten, liegen jetzt mit Finn-Werten von r >.70 im guten Bereich. Auch das Statement S. 8 (das den geringsten Mittelwert aller Statements aufwies) hat sich vom befriedigenden ICC=.60 zum guten Finn-Koeffizienten von .71 verbessert.

Diese genannten Verbesserungen im Vergleich mit dem ICC-Wert können als Zeichen dafür gewertet werden, dass die Varianz der Ratings zwischen den Fällen zu niedrig war, um trotz eigentlich akzeptabler Beurteilerübereinstimmungen einen befriedigenden ICC ermöglichen zu können.

Beim Statement S. 16 kann dieser Schluss jedoch nicht gezogen werden: Die Reliabilität hat sich von ICC=.44 lediglich auf r=.56 verbessert. Dies deutet darauf hin, dass *neben* oder *zusätzlich* zu einer tatsächlich geringen Beurteilerübereinstimmung eine geringe Variation des Statements über die Fälle bestanden hat.

Tab. 32: *Experten-Rating; Reliabilitäts-Kennwerte für die Statements des DIB-R*

	S.	Der Patient ...	ICC*	Finn's *r*	Horst
Affekte	**1**	litt an einer chronischen depressiven Verstimmung oder hatte eine oder mehrere Perioden von Major Depression	**0,49**	0,72	**0,48**
	2	hatte anhaltende Gefühle von Hilf-, Hoffnungs-, Wertlosigkeit oder Schuld	**0,39**	0,74	**0,37**
	3	hatte chronische Gefühle von Ärger, Wut oder verhielt sich häufig ärgerlich wütend	0,57	0,60	0,55
	4	hat sich chronisch sehr ängstlich gefühlt oder litt häufig unter körperlichen Angstsymptomen	0,54	0,59	0,52
	5	erlebte chronische Gefühle von Einsamkeit, Langeweile oder Leere	0,80	0,85	0,79
Kognition	**6**	neigte zu seltsamen Denken oder ungewöhnlichen Wahrnehmungserlebnissen	0,65	0,60	0,63
	7	hatte häufig flüchtige, nicht-wahnhafte paranoide Erlebnisse	0,72	0,72	0,71
	8	hatte wiederholte pseudo-psychotische Wahnvorstellungen oder Halluzinationen	0,60	0,71	0,58
Impulsbereich	**9**	betrieb einen ernsthaften Drogenmissbrauch	0,81	0,78	0,81
	10	hatte ein Muster sexuell abweichenden Verhaltens	0,86	0,85	0,86
	11	zeigte ein Muster von körperlicher Selbstbeschädigung	0,96	0,95	0,96
	12	zeigte ein Muster von manipulativen Selbstmorddrohungen, -gesten oder –versuchen	0,84	0,80	0,83
	13	zeigte ein anderes Muster impulsiven Verhaltens	0,89	0,86	0,89
Zwischenmenschliche Beziehungen	**14**	hat typischerweise versucht, das Alleinsein zu vermeiden oder fühlte sich extrem dysphorisch, wenn er allein war	0,79	0,78	0,78
	15	hat (wiederholt) Verlassenheits-, Verschlingungs- oder Vernichtungsängste erlebt	0,83	0,81	0,82
	16	hat seine Abhängigkeitswünsche stark abgewehrt oder befand sich in einem ernsten Konflikt zwischen Versorgen und Versorgtwerden	**0,44**	0,56	**0,42**
	17	neigte zu intensiven instabilen engen Beziehungen	0,84	0,80	0,83
	18	hatte in engen Beziehungen immer wieder Probleme mit Abhängigkeit oder Masochismus	0,86	0,82	0,85
	19	hatte in engen Beziehungen wiederkehrende Probleme mit Abwertung, Manipulation oder Sadismus	0,54	**0,49**	0,53
	20	hatte in engen Beziehungen immer wieder Probleme mit seiner Forderungs- oder Anspruchshaltung	0,84	0,79	0,83
	21	zeigte während der Therapie oder der psychiatrischen Hospitalisierung eine deutliche Regression	0,93	0,91	0,93
	22	hat auf der psychiatr. Station oder in einer Psychotherapie auffallende Gegenübertragungsreaktionen ausgelöst oder ist mit einem profess. Helfer eine ganz besondere Beziehung eingegangen	0,67	0,67	0,66

* Sämtliche ICC sind signifikant von Null verschieden. Kennwerte, die den angestrebten Wert ≥.70 für gute Reliabilitäten verfehlten, sind grau unterlegt dargestellt, ungenügende Reliabilitäten ≤.50 zusätzlich fett gedruckt.

Zusammenfassend können die Ergebnisse der Analyse des Intraclass-Korrelationskoeffizienten und des Finn-Koeffizienten wie folgt wiedergegeben werden:

- Die Mehrzahl der DIB-R-Statements, nämlich 13, ist nach dem Kriterium ICC und $r \geq .70$ als mindestens gut bis sehr gut zu bezeichnen und befriedigt voll.
- Drei Statements, nämlich S. 1, S. 2 und S. 8 schnitten nur im Finn-Koeffizienten mit einem Kennwert ≥.70 gut ab, waren im ICC aber nur ausreichend (S. 8) bzw. ungenügend (S. 1, S. 2). Dies war v.a. auf eine mangelnde Variabilität der Raterurteile über die Patienten zurückgeführt worden – dies kann für diese Statements auf der Basis der vorliegenden Untersuchung aber nur vermutet und nicht sichergestellt werden.
- Bei den Statements S. 3, S. 4, S. 6 und S. 22 bestehen seitens der Expertenrater geringfügige Schwierigkeiten in einer übereinstimmenden Beurteilung der DIB-R-Statements. Die Ergebnisse liegen sowohl für den ICC- als auch für den Finn-Koeffizienten aber noch im ausreichenden Reliabilitätbereich ≥.50 aber <.70.
- Mit geringer Genauigkeit wurden in der Gruppe der Expertenrater die Statements S. 16 und S. 19 eingeschätzt. Bei S. 16 war mit einem ICC <.50 die Reliabilität unbefriedigend, eine Berechnung des Finn-Koeffizienten verbesserte mit .56 das Ergebnis nur minimal. Bei S. 19 war der ICC mit .54 nur noch knapp ausreichend, der Finn-Koeffizient fiel hingegen mit .49 ungenügend aus, was tendenziell auf eine Überschätzung im ICC hindeutet.

6.1.3. Ergebnisse der Experten zu den DIB-R-Scores und Diagnosen

6.1.3.1. Rateranzahl, Mittelwerte und Verteilung

Die Rateranzahl bei den Bereichsscores des DIB-R (Summen-Scores und Skalierte Section-Scores) schwankt zwischen 4,5 und 4,6. Die Raterzahl beim Gesamt-Score des DIB-R und der Ein- bzw. Ausschluss-Diagnose Borderline-Persönlichkeitsstörung liegt bei durchschnittlich 4,2 wie aus Tab. 33 ersichtlich.

Die Rateranzahl nimmt mit der Komplexität der Kennwerte kontinuierlich geringfügig ab: auf Grund der Raterfluktuation wurden nicht immer *alle* Bereiche des DIB-R von den beteiligten Ratern vollständig eingeschätzt, so dass dann keine höheren, darauf aufbauenden Kennwerte wie der Gesamt-Score berechnet werden konnten.

Die Mittelwerte der Summen-Scores der DIB-R-Bereiche über alle Ratings liegen in etwa in der Mitte der möglichen Bandbreite, mit Ausnahme des Affektbereichs: Mit einem Rating-

Durchschnitt von 7,49 von 10 möglichen Punkten zeigt sich auch hier die durchgehend relativ einförmige, starke Ausprägung des Affektbereichs.

Im Mittel liegen die Skalierten Section-Scores zwischen 0,9 und 2,1. (Achtung: Es ist hier ein *über die Bereiche differierender Wertebereich* zu beachten: Der Range für den Affekt- und den Kognitionsbereich beträgt 0 bis 2, der Wertebereich für die Bereiche Impulshandlungen und Zwischenmenschliche Beziehungen reicht aber bis 3, wobei der Wert 1 in diesen Bereichen nicht vorgesehen ist.)

Neben den wiederum hohen Werten im Affektbereich ist die vermeintlich starke Ausprägung von 2,1 im Bereich Zwischenmenschliche Beziehungen v.a. eine Folge dieser Akzentuierung.

Tab. 33: *Experten-Rating; Mittelwerte der Summen-Scores, Skalierten Section-Scores, des Gesamt-Scores und der DIB-R-Diagnose, Nennungshäufigkeiten der Skalenstufen in Prozent*

		Rater-anzahl	M*	*Nennungsanteile der Skalierten Section-Scores in %*			
				"0"	*"1"*	*"2"*	*"3"*
Summen-scores	**Affekte** *(Range 0-10)*	4,6	7,5				
	Kognition *(Range 0-6)*	4,5	2,7				
	Impulsivität *(Range 0-10)*	4,6	4,8				
	Zwischenm. Beziehungen *(Range 0-18)*	4,5	9,8				
Skalierte Section-Scores	**Affekte** *(Range 0-2)*	4,6	1,3	16	44	41	-
	Kognition *(Range 0-2)*	4,5	0,9	35	32	33	-
	Impulsivität *(Range 0-3)*	4,6	1,8	33	-	23	44
	Zwischenm. Beziehungen *(Range 0-3)*	4,5	2,1	26	-	10	64
Gesamt-Score	*(Range 0-10)*	4,2	6,2				
Diagnose	***Borderline-Persönlichkeitsstörung*** *(dichotom)*	4,2	39%				

** Globale Mittelwerte über alle Rater und Patienten. M für Diagnose wird als prozentualer Anteil der Borderline-Diagnosen angegeben.*

Die Verteilung der Werte der Skalierten Section-Scores über die mögliche Bandbreite erscheint in den Bereichen Kognition und Impulsivität recht ausgewogen. Mit der Vergabe einer "0" im Affektbereich in 16% und des Ratings einer "3" im Bereich Zwischenmenschliche Beziehungen

in 64% der Ratings sind diese Bereiche ungleichmäßiger, wenn auch nicht extrem einseitig verteilt.

Der DIB-R-Gesamt-Score liegt im Mittel über alle Videos und Rater bei 6,2 und damit deutlich unter dem Grenzwert von "8" für die Stellung der Borderline-Diagnose.

Dieser Cutoff-Wert wurde in 39% der vorliegenden Ratings erreicht und somit die Diagnose Borderline-Persönlichkeitsstörung nach DIB-R gestellt, entsprechend war der Gesamt-Score in 61% der Ratings ≤7.

6.1.3.2. Reliabilitäts-Kennwerte nach ICC und Finn-Koeffizient

Die Reliabilitäten der Bereiche fallen, wie in Tab. 34 dargestellt, in jeweils deutlich unterschiedlicher Höhe aus:

6.1.3.2.1. Affektbereich

Im Affektbereich, in dem von fünf Statements vier wenig reliabel waren, liegt der Summen-Score als reine Aufsummierung der Bereichs-Statements ohne Anwendung weiterer Transformationsregeln mit einem ICC von 0,65 knapp unter dem Grenzwert von ,70 für gute Reliabilitäten. Der Intraclass-Korrelationskoeffizient des Skalierten Section-Scores liegt mit .57 noch etwas niedriger und ist somit nur noch befriedigend.

Die jeweiligen Finn-Koeffizienten liegen, wie bei den Statements des Affektbereichs auch, höher: Der Finn-Koeffizient des Summen-Scores liegt mit .76 im angestrebten Bereich, der Wert des Skalierten Section-Scores erreicht mit .67 immerhin beinahe die angestrebte Marke von .70 für einen uneingeschränkt guten Reliabilitäts-Kennwert.

6.1.3.2.2. Kognitionsbereich

Die Reliabilität des Kognitionsbereichs, in dem zwei von drei Statements wenig reliabel waren, erreicht sowohl für den Summen-Score als auch für den Skalierten Section-Score den als gut zu bewertenden Reliabilitätskoeffizienten von ICC=.77.

Der Finn-Koeffizient liegt mit r=.83 für den Summen-Score und .76 für den Skalierten Section-Score in einer ähnlichen Höhe.

6.1.3.2.3. Bereich Impulshandlungen

Der ICC des Summen-Scores des DIB-R-Bereichs Impulshandlungen, der durchgehend gute Statements aufwies, erreicht den sehr guten Wert von .89. Der ICC des Skalierten Section-Scores des Bereichs liegt bei ebenfalls reliablen ICC=.80.

Die zugehörigen Finn-Koeffizienten liegen mit .93 für den Summen-Score und .80 für den Skalierten Section-Score in vergleichbarer Höhe.

6.1.3.2.4. Bereich Zwischenmenschliche Beziehungen

Der Beziehungsbereich, in dem sechs von insgesamt neun Statements eine gute Reliabilität aufwiesen, ist insgesamt der reliabelste DIB-R-Bereich.

Der ICC des Summen-Scores erreicht genauso wie der entsprechende Finn-Koeffizient ausgezeichnete .95. Der Skalierte Section-Score weist mit einem ICC-Koeffizienten von .89 und einem Finn-Koeffizienten von .87 gute Reliabilitäts-Kennwerte auf.

Tab. 34: *Experten-Rating; Reliabilitätsmaße der Summen-Scores, der Skalierten Section-Scores, des Gesamt-Scores und der DIB-R-Diagnose*

		ICC*	Finn's *r*	Horst
Summen-Scores	Affekte	0,65	0,76	0,64
	Kognition	0,77	0,83	0,75
	Impulsivität	0,89	0,93	0,89
	Zwischenmenschl. Beziehungen	0,95	0,95	0,95
Skalierte Section-Scores	Affekte (Range 0-2)	0,57	0,67	0,55
	Kognition (Range 0-2)	0,77	0,76	0,76
	Impulsivität (Range 0-3)	0,80	0,80	0,79
	Zwischenmenschl. Beziehungen (Range 0-3)	0,89	0,87	0,88
Gesamt-Score	(Range 0-10)	0,91	0,92	0,90
DIB-R Diagnose	(dichotom)	0,74	0,74	0,72

* Sämtliche ICC sind signifikant von Null verschieden. Kennwerte, die den angestrebten Wert ≥.70 für gute Reliabilitäten verfehlten, sind grau unterlegt dargestellt, ungenügende Reliabilitäten ≤.50 zusätzlich fett gedruckt.

6.1.3.2.5. Zusammenfassung zu den Bereichen

Die Reliabilitäten aller Bereiche mit Ausnahme der Affektbereichs sind als uneingeschränkt gut zu beurteilen. Zu beachten ist dabei die weitgehende Übereinstimmung der Finn- und der Intraclass-Korrelationskoeffizienten.

Nur der Bereich Affekte fällt aus dem einheitlichen Bild heraus: Er ist offenbar der am wenigsten reliable Bereich des DIB-R. Die beiden ICC-Koeffizienten liegen aber im Bereich noch befriedigender Reliabilitäten zwischen ≥.50 und .70. Die Finn-Koeffizienten liegen für die Affekte über den ICC-Werten und erreichen für den Skalierten Section-Score mit .67 sogar einen annähernd guten Reliabilitäts-Kennwert.
Diese Differenz zwischen dem ICC und dem Finn-Koeffizienten in diesem Bereich liegt, wie im Methodenkapitel erläutert, vermutlich daran, dass der Finn-Koeffizient unabhängig von der *tatsächlichen* Varianz über die beurteilten Fälle bzw. Videobänder berechnet wird.
Zusammenfassend darf der Schluss gezogen werden, dass eine geringe Variabilität der Raterurteile aufgrund der relativen Gleichförmigkeit der zu beurteilenden affektiven Merkmale über die Fälle in den Statements des Affektbereichs neben einer eher geringen Interraterübereinstimmung eine wichtige Ursache für das relativ mäßige Abschneiden im ICC ist.

Für die o.g. Reliabilitäten des DIB-R in der Expertengruppe waren im Kap. 3.3.2.2 folgende spezifischen Hypothesen formuliert worden:

Hypothese 3.3.a): Die Interraterübereinstimmung in der Expertengruppe ist, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den Skalierten Section-Score des DIB-R-Bereichs "Affekte" signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Hypothese 3.3.b): Die Interraterübereinstimmung in der Expertengruppe ist, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den Skalierten Section-Score des DIB-R-Bereichs "Kognition" signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Hypothese 3.3.c): Die Interraterübereinstimmung in der Expertengruppe ist, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den Skalierten Section-Score des DIB-R-Bereichs "Impulsivität" signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Hypothese 3.3.d): Die Interraterübereinstimmung in der Expertengruppe ist, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den Skalierten Section-Score des DIB-R-Bereichs "Zwischenmenschliche Beziehungen" signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Auf der Basis der vorliegenden Ergebnisse können die Hypothesen 3.3.b), c) und d) beibehalten werden. Die Hypothese 3.3.a) ist jedoch zurückzuweisen.

6.1.3.3. Reliabilitäts-Kennwerte und Verteilung der DIB-R-Gesamtwerte und der Diagnosen

Die Reliabilität des Gesamt-Scores des DIB-R, der aus den vier Section-Scores aufsummiert wird, ist mit einem ICC von .91 sehr reliabel. Die Reliabilität der aus dem Gesamt-Score anhand des Schwellenwerts "8" (s.o.) gestellten Diagnosen erreicht mit einem ICC=.74 immer noch einen guten Wert (siehe Tab. 34).
Für die Reliabilität des DIB-R in der Expertengruppe waren im Kap. 3.3.2.2 folgende spezifischen Hypothesen formuliert worden:

Hypothese 3.1: Die Interraterübereinstimmung ist in der Expertengruppe, gemessen mit dem Intraclass-Korrelationskoeffizienten, für die DIB-R-Diagnose signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Hypothese 3.2: Die Interraterübereinstimmung ist in der Expertengruppe, gemessen mit dem Intraclass-Korrelationskoeffizienten, für den DIB-R-Gesamtwert signifikant und erreicht mit einer Höhe von wenigstens .70 ein befriedigendes Niveau.

Beide Hypothesen können auf der Basis der vorliegenden Ergebnisse beibehalten werden.

Auch absolut betrachtet ist über alle 18 Videobänder, in denen Diagnosen vorliegen, eine weitgehende Übereinstimmung festzustellen: in nur vier der 18 Fälle traten überhaupt unterschiedliche diagnostische Einschätzungen der Rater auf. Nach einer im Kapitel 2.4.6.5 zu den Grundlagen der Reliabilitätbestimmung geschilderten Rechenmethode konnte eine mittlere prozentuale Übereinstimmung der Diagnosen in Höhe von 88% bestimmt werden.
Zusammenfassend kann die Stellung der Ein- bzw. Ausschlussdiagnose Borderline-Persönlichkeitsstörung anhand des DIB-R als reliabel gelten.

Die von den Ratern tatsächlich vergebenen Gesamt-Scores und Diagnosen sind zur besseren Vorstellbarkeit und Übersichtlichkeit in der folgenden Tabelle einzeln über die 18 Fälle wiedergegeben. Beinahe immer ist eine Streuung der Gesamt-Scores vorhanden: In nur zwei von 18 Fällen sind die Scores über alle Rater identisch. In den übrigen Fällen treten Differenzen zwischen einem und drei Punkten auf. Grob zusammenfassend kann man die größten

Unterschiede in den Fällen beobachten, in denen der Gesamt-Score insgesamt niedrig ist – in diesen Fällen sind die Unterschiede für eine Diagnosestellung irrelevant.

Tab. 35: *Experten-Rating; Raterurteile im Gesamt-Score und Diagnosen über alle Fälle*

Fälle*	**Raterurteile***										
	Gesamt-Scores					**M**	**Diagnosen**				
1	**7**	**5**	**6**	**7**	**8**	**6,6**	**0**	**0**	**0**	**0**	**1**
2	9	10	10	9	-	9,5	1	1	1	1	-
3	6	5	5	5	5	5,2	0	0	0	0	0
4	0	0	0	-	-	,0	0	0	0	-	-
5	6	6	6	-	-	6,0	0	0	0	-	-
6	5	3	4	-	-	4,0	0	0	0	-	-
7	1	1	3	1	1	1,4	0	0	0	0	0
8	10	10	9	9	8	9,2	1	1	1	1	1
9	6	5	7	5	3	5,2	0	0	0	0	0
10	9	9	10	-	-	9,3	1	1	1	-	-
11	10	10	9	9	10	9,6	1	1	1	1	1
12	**6**	**6**	**8**	**6**	**-**	**6,5**	**0**	**0**	**1**	**0**	**-**
13	9	8	9	9	-	8,8	1	1	1	1	-
14	9	9	-	-	-	9,0	1	1	-	-	-
15	**7**	**9**	**7**	**8**	**7**	**7,6**	**0**	**1**	**0**	**1**	**0**
16	**9**	**9**	**7**	**7**	**-**	**8,0**	**1**	**1**	**0**	**0**	**-**
17	5	5	6	4	5	5,0	0	0	0	0	0
18	3	3	3	0	2	2,2	0	0	0	0	0

** Die Rater sind über die Fälle nur z.T. dieselben Personen.*
Fälle mit Diagnoseabweichungen sind in Fettdruck dargestellt.

Kleinere, aber wegen ihrer Nähe zum Cutoff-Wert für die Diagnosestellung entscheidende Unterschiede sind bei den Fällen mit hohem Gesamt-Score zu finden: Diagnoseabweichungen zwischen den Ratern traten in der Stichprobe immer dann auf, wenn der mittlere Gesamt-Score der Rater $\geq 6{,}5$ und ≤ 8 war.

Bei allen anderen gemittelten Gesamt-Scores traten keinerlei Diagnose-Abweichungen auf, so dass, wie nicht anders zu erwarten, ein "Bereich der Unsicherheit" um den Cutoff-Wert herum vorliegt. Die auffallende Absenkung der Reliabilität vom Gesamt-Score zum dichotomen

Merkmal „Diagnose“ ist dabei offensichtlich auf den Übergang von einem dimensionalen zu einem kategorialen Merkmal an einem bestimmten Schwellenwert zurückzuführen.

6.1.3.4. Zu den Ergebnissen der Horst-Koeffizienten

Diese Koeffizienten entsprechen in der Experten-Stichprobe bei den DIB-R-Statements, den Summen-Scores und den Skalierten Section-Scores in allen Fällen größenordnungsmäßig den ICC-Werten, in vielen Fällen sind sie bis auf die zweite Stelle nach dem Komma sogar identisch. Insofern kann dieser offenbar nur geringfügig abweichende Koeffizient als dem ICC äquivalent gewertet werden. Unsere Ergebnisse sind von der Berechnung her mit denen der Studie von Eckert et al. (1987) zur ersten Version des DIB uneingeschränkt vergleichbar.

6.2. STUDENTEN-STICHPROBE

6.2.1. Beschreibung der Stichprobe der untersuchten Patienten

Die Stichprobe der studentischen Rater besteht (wie im Kap. 3.6.3.2) beschrieben, aus Psychologiestudenten im Hauptstudium, die das DIB-R im Rahmen eines Fallseminares kennen gelernt und danach eine kompakte Raterschulung mit zwei der Expertenrater durchlaufen haben.
Von den Studenten wurden 12 psychiatrische Patienten untersucht, je zur Hälfte Männer und Frauen. Das Durchschnittsalter lag bei 28,2 Jahren in (s=6,2; Range 19-36 J.)
Von den 12 Patienten waren 9 im Rahmen der Studie auch mittels SKID-I und -II untersucht worden, von drei Patienten lagen nur Videoaufnahmen des DIB-R *ohne* Außendiagnose vor.
Die DSM-Diagnose-Kategorien der SKID-untersuchten Patienten waren in fünf Fällen *Borderline-Persönlichkeitsstörung*, in drei Fällen *Andere Persönlichkeitsstörung* und in einem Fall *Depressive Störung*.
Die studentischen Rater haben nur in vier Fällen die gleichen Videobänder wie die Experten untersucht, so dass ein direkter Vergleich der Ergebnisse nicht möglich ist.

6.2.2. Ergebnisse der Studenten zu den DIB-R-Statements

6.2.2.1. Rateranzahl, Mittelwerte und Verteilung

Die in Tab. 36 angegebenen globalen Statement-Mittelwerte schwanken auch bei der Studenten-Stichprobe erheblich. Den mit Abstand geringsten Wert hat dabei das Statement S. 10 aus dem Bereich Impulsivität mit einem Mittel von M=0,17. Die höchsten Werte liegen alle im Affektbereich: Den höchsten Wert weist mit 1,94 das Statement S. 2 auf, darauf folgen S. 1 (1,88), S. 4 (1,85) und S. 5 (1,81).

Außerdem fällt das Statement 15 aus dem Beziehungsbereich mit einem Mittelwert von 1,75 auf. Diese aufgrund ihrer extremen Werte auffallenden Statements sind auch über die Patienten und Rater sehr *unausgewogen verteilt.* Herausragend sind S. 1 und S. 2: so wurden 87% bzw. 94% aller Ratings mit "2" bewertet und in keinem Fall eine "0" für nicht vorhanden vergeben. Umgekehrt wurde bei S. 10 in 88% der Fälle ein Nichtvorliegen des Symptoms eingeschätzt und nur in 4% der Fälle eine "2" vergeben.

6.2.2.2. Reliabilitäts-Kennwerte nach ICC und Finn-Koeffizient

Die Interrater-Reliabilitäten der DIB-R-Statements der Studenten (siehe Tab. 37) liegen sehr verteilt zwischen .17 und .95 (sämtlich signifikant von Null verschieden).

Von den 22 Statements liegen acht im guten bis sehr guten Bereich ICC≥.70. Sieben Statements weisen noch einen befriedigenden ICC≥.50 auf.

Sieben der 22 Statements liegen im Bereich ungenügender Reliabilität <.50 – dies betrifft bei der Studenten-Stichprobe u.a. den *gesamten Affektbereich* sowie die Statements S. 10 zu sexuell abweichendem Verhalten und S. 19 zu Beziehungsproblemen mit Abwertung, Manipulation oder Sadismus. Es handelt sich um genau die Statements, bei denen sich vorher eine extrem unausgewogene Verteilung der Nennungen gezeigt hatte.

Eine Betrachtung der Reliabilitäten nach dem Finn-Koeffizienten erbringt in diesem Fall erwartungsgemäß meist bessere Ergebnisse. Auffallend ist (wie in der Experten-Stichprobe), dass die Finn-Koeffizienten für alle Statements, die bisher ausreichend abschnitten, größenordnungsmäßig gleich geblieben sind. Für mehrere der unbefriedigend abgeschnittenen Statements traten wieder substanzielle Verbesserungen auf, während gelegentlich auch geringfügige Verschlechterungen zu beobachten waren.

Insgesamt liegen hier die Reliabilitäten von 13 Statements im Bereich guter oder sehr guter Werte, sieben Statements im Bereich befriedigender Reliabilitäten. Zwei Statements liegen mit einem Finn-Koeffizienten von .26 (S. 19) und .47 (S. 3) noch immer im Bereich ungenügender Reliabilität, diese Statements wurden also *eindeutig unreliabel eingeschätzt*.

Tab. 36: *Studenten-Rating; Mittelwerte der DIB-R-Statements, Nennungshäufigkeiten der Skalenstufen in Prozent*

	S.	Der Patient ...	M*	Nennungsanteile in %		
				„0“	„1“	„2“
Affekte	**1**	litt an einer chronischen depressiven Verstimmung oder hatte eine oder mehrere Perioden von Major Depression	1,88	0	13	87
	2	hatte anhaltende Gefühle von Hilf-, Hoffnungs-, Wertlosigkeit oder Schuld	1,94	0	6	94
	3	hatte chronische Gefühle von Ärger, Wut oder verhielt sich häufig ärgerlich wütend	1,52	10	27	63
	4	hat sich chronisch sehr ängstlich gefühlt oder litt häufig unter körperlichen Angstsymptomen	1,85	2	10	88
	5	erlebte chronische Gefühle von Einsamkeit, Langeweile oder Leere	1,81	4	10	85
Kognition	**6**	neigte zu seltsamen Denken oder ungewöhnlichen Wahrnehmungserlebnissen	1,00	35	31	33
	7	hatte häufig flüchtige, nicht-wahnhafte paranoide Erlebnisse	1,42	21	17	63
	8	hatte wiederholte pseudo-psychotische Wahnvorstellungen oder Halluzinationen	0,65	65	6	29
Impulsbereich	**9**	betrieb einen ernsthaften Drogenmissbrauch	0,79	56	8	35
	10	hatte ein Muster sexuell abweichenden Verhaltens	0,17	88	8	4
	11	zeigte ein Muster von körperlicher Selbstbeschädigung	1,19	35	10	54
	12	zeigte ein Muster von manipulativen Selbstmorddrohungen, -gesten oder –versuchen	0,75	56	13	31
	13	zeigte ein anderes Muster impulsiven Verhaltens	1,00	42	17	42
Zwischenmenschliche Beziehungen	**14**	hat typischerweise versucht, das Alleinsein zu vermeiden oder fühlte sich extrem dysphorisch, wenn er allein war	1,58	15	13	73
	15	hat (wiederholt) Verlassenheits-, Verschlingungs- oder Vernichtungsängste erlebt	1,75	8	8	84
	16	hat seine Abhängigkeitswünsche stark abgewehrt oder befand sich in einem ernsten Konflikt zwischen Versorgen und Versorgtwerden	1,48	15	23	63
	17	neigte zu intensiven instabilen engen Beziehungen	0,63	63	13	25
	18	hatte in engen Beziehungen immer wieder Probleme mit Abhängigkeit oder Masochismus	1,44	23	10	67
	19	hatte in engen Beziehungen wiederkehrende Probleme mit Abwertung, Manipulation oder Sadismus	0,73	52	23	25
	20	hatte in engen Beziehungen immer wieder Probleme mit seiner Forderungs- oder Anspruchshaltung	0,67	54	25	21
	21	zeigte während der Therapie oder der psychiatrischen Hospitalisierung eine deutliche Regression	0,38	73	17	10
	22	hat auf der psychiatr. Station oder in einer Psychotherapie auffallende Gegenübertragungsreaktionen ausgelöst oder ist mit einem profess. Helfer eine ganz besondere Beziehung eingegangen	0,71	52	25	23

** Besonders invariant eingeschätzte Werte sind grau unterlegt dargestellt.*

Tab. 37: *Studenten-Rating; Reliabilitätsmaße für die Statements des DIB-R*

	S.	Der Patient ...	*ICC**	Finn's *r*	*Horst*
Affekte	**1**	litt an einer chronischen depressiven Verstimmung oder hatte eine oder mehrere Perioden von Major Depression	**0,39**	0,90	**0,37**
	2	hatte anhaltende Gefühle von Hilf-, Hoffnungs-, Wertlosigkeit oder Schuld	**0,20**	0,93	**0,17**
	3	hatte chronische Gefühle von Ärger, Wut oder verhielt sich häufig ärgerlich wütend	**0,25**	0,47	**0,23**
	4	hat sich chronisch sehr ängstlich gefühlt oder litt häufig unter körperlichen Angstsymptomen	**0,15**	0,78	**0,12**
	5	erlebte chronische Gefühle von Einsamkeit, Langeweile oder Leere	**0,17**	0,70	**0,15**
Kognition	**6**	neigte zu seltsamen Denken oder ungewöhnlichen Wahrnehmungserlebnissen	0,66	0,63	0,63
	7	hatte häufig flüchtige, nicht-wahnhafte paranoide Erlebnisse	0,80	0,79	0,79
	8	hatte wiederholte pseudo-psychotische Wahnvorstellungen oder Halluzinationen	0,79	0,72	0,77
Impulsbereich	**9**	betrieb einen ernsthaften Drogenmissbrauch	0,93	0,90	0,92
	10	hatte ein Muster sexuell abweichenden Verhaltens	**0,34**	0,77	**0,31**
	11	zeigte ein Muster von körperlicher Selbstbeschädigung	0,95	0,93	0,94
	12	zeigte ein Muster von manipulativen Selbstmorddrohungen, -gesten oder –versuchen	0,83	0,77	0,81
	13	zeigte ein anderes Muster impulsiven Verhaltens	0,67	0,56	0,65
Zwischenmenschliche Beziehungen	**14**	hat typischerweise versucht, das Alleinsein zu vermeiden oder fühlte sich extrem dysphorisch, wenn er allein war	0,73	0,77	0,71
	15	hat (wiederholt) Verlassenheits-, Verschlingungs- oder Vernichtungsängste erlebt	0,82	0,90	0,80
	16	hat seine Abhängigkeitswünsche stark abgewehrt oder befand sich in einem ernsten Konflikt zwischen Versorgen und Versorgtwerden	0,60	0,66	0,58
	17	neigte zu intensiven instabilen engen Beziehungen	0,57	0,50	0,55
	18	hatte in engen Beziehungen immer wieder Probleme mit Abhängigkeit oder Masochismus	0,62	0,57	0,60
	19	hatte in engen Beziehungen wiederkehrende Probleme mit Abwertung, Manipulation oder Sadismus	**0,32**	**0,26**	**0,29**
	20	hatte in engen Beziehungen immer wieder Probleme mit seiner Forderungs- oder Anspruchshaltung	0,57	0,56	0,54
	21	zeigte während der Therapie oder der psychiatrischen Hospitalisierung eine deutliche Regression	0,52	0,67	0,50
	22	hat auf der psychiatr. Station oder in einer Psychotherapie auffallende Gegenübertragungsreaktionen ausgelöst oder ist mit einem profess. Helfer eine ganz besondere Beziehung eingegangen	0,77	0,75	0,75

* Sämtliche ICC sind signifikant von Null verschieden. Kennwerte, die den angestrebten Wert ≥.70 für gute Reliabilitäten verfehlten, sind grau unterlegt dargestellt, ungenügende Reliabilitäten ≤.50 zusätzlich fett gedruckt.

Die nach dem ICC bisher ungenügenden Statements S. 1, S. 2, S. 4 und S. 5 aus dem Affektbereich liegen jetzt mit Werten über .70 durchweg im guten Bereich. Dies kann unter Berücksichtigung der beschriebenen extremen Werteverteilung als Zeichen dafür gewertet werden, dass die Varianz der Ratings zwischen den Fällen zu niedrig war, um trotz eigentlich akzeptabler Beurteilerübereinstimmung einen befriedigenden ICC zu ermöglichen.

Ein weiteres Statement hat sich verbessert von einem ICC=.34 zum befriedigenden Finn-Koeffizienten von .77. Es handelt sich um das oben beschriebene S. 10, bei welchem die Studenten in 88% der Ratings ein Nichtvorliegen sexueller Devianz einschätzten. Der Finn-Koeffizient spiegelt die zugrunde liegende Übereinstimmung der Studenten bei einförmiger Urteilstendenz wieder.

Zusammenfassend kann nach Analyse der Ergebnisse des Finn-Koeffizienten festgestellt werden, dass in der Studenten-Stichprobe die zwei Statements S. 3 und S. 19 ohne Einschränkung als unreliabel eingeschätzt gelten müssen. Bei den Statements S. 6, S. 13, S. 16, S. 17, S. 18, S. 20 und S. 21 besteht nach den vorliegenden Ergebnissen der studentischen Rater eine immerhin noch befriedigende gemeinsame Übereinstimmung.

Bei den Statements S. 1, S. 2, S. 4, S. 5 und S. 10 kann keine klare Aussage über die Reliabilität getroffen werden: Ein gutes Abschneiden im Finn-Koeffizienten bei gleichzeitig, vermutlich (!) auf sehr einseitiger Antworttendenz beruhenden, schlechtem Abschneiden im ICC erlaubt eben keine verlässliche Aussage. Bei einer größeren Variabilität im Merkmal hätte vielleicht durch die studentischen Rater ein guter ICC erzielt werden können – ebenso wäre auch ein schlechtes Abschneiden beider Koeffizienten möglich.

Lediglich die Statements S. 7, S. 8, S. 9, S. 11, S. 12, S. 14, S. 15 und S. 22 können eindeutig als reliabel eingeschätzt gelten.

6.2.3. Ergebnisse der Studenten zu DIB-R-Scores und Diagnosen

6.2.3.1. Rateranzahl, Mittelwerte und Verteilung

In Tab. 38 sind wieder die zur Einstufung erreichter Reliabilitäten notwendigen, zugrundeliegenden Verteilungen und Antworttendenzen der studentischen Rater angegeben.
Aus allen Bereichen herausragend ist der Affektbereich. Aus dem durchschnittlichen Summen-Score von annähernd 9 von 10 möglichen Punkten wird noch einmal klar, dass die Rater nahezu einförmig das Vorhandensein aller Symptome geratet haben. Bei den Skalierten Section-Scores

entzerrt sich das Verhältnis dann wieder etwas, was nur auf der Basis der verschiedenen Skalierungsregeln erfolgt sein kann. Dennoch wurde immer noch in fast der Hälfte der Ratings der maximale Skalierte Section-Score von "2" gegeben.

Tab. 38: *Studenten-Rating; Mittelwerte* der Skalierten Section-Scores, des Gesamt-Scores und der DIB-R-Diagnose, Nennungshäufigkeiten der Skalenstufen in Prozent*

		M	*Nennungsanteile der Skalierten Section-Scores in %*			
			"0"	*"1"*	*"2"*	*"3"*
Summen-Scores	**Affekte** *(Range 0-10)*	8,98				
	Kognition *(Range 0-6)*	3,08				
	Impulsivität *(Range 0-10)*	3,92				
	Zwischenmenschl. Beziehungen *(Range 0-18)*	9,29				
Skalierte Section-Scores	**Affekte** *(Range 0-2)*	1,17	27	29	44	-
	Kognition *(Range 0-2)*	1,00	31	38	31	-
	Impulsivität *(Range 0-3)*	1,40	48	-	17	35
	Zwischenmenschl. Beziehungen *(Range 0-3)*	2,23	19	-	21	60
Gesamt-Score	*(Range 0-10)*	5,77				
Diagnose	***Borderline-Persönlichkeitsstörung*** *(dichotom)*	33%				

**Mittelwerte über alle Ratings, d.h. jedes vorliegenden Raterurteils über alle Rater und Patienten.*
M für Diagnose wird als prozentualer Anteil der Borderline-Diagnosen angegeben.

Im Bereich Kognition ist die Verteilung offenbar recht gleichförmig erfolgt, sowohl im Summen- als im Skalierten Section-Score. Dieses ist auch im Impulsbereich der Fall, wenn auch hier eher die extremen Kategorien "0" und "3" gewählt wurden (der Wert "1" ist hier und im folgenden Bereich nicht vorgesehen). Im Bereich Zwischenmenschliche Beziehungen erfolgte die Beurteilung wieder einförmiger, in 60% der Ratings wurde eine "3" gegeben.

6.2.3.2. Reliabilitäts-Kennwerte nach ICC und Finn-Koeffizient

Die Ergebnisse zu den Reliabilitäten der DIB-R-Bereiche fallen auch in der Studenten-Stichprobe sehr heterogen aus und sind in Tab. 39 wiedergegeben.

6.2.3.2.1. Affektbereich

Im Affektbereich, in dem nach dem ICC sämtliche Statements unreliabel waren, ist der Summen-Score als reine Aufsummierung der Bereichs-Statements (also ohne Anwendung weiterer Transformationsregeln) mit einem ICC von 0,37 ebenfalls unreliabel. (Unterschiedlichkeiten in der Einschätzung mitteln sich also auch nicht etwa aus.) *Der Intraclass-Korrelationskoeffizient des Skalierten Section-Scores liegt mit .75 aber im guten Bereich.* Dieses zunächst überraschende Ergebnis liegt in der Anwendung der Transformations- bzw. Skalierungsregeln begründet. Lagen die Summen-Scores des Affektbereichs über alle Rater und Fälle in etwa gleicher Höhe, wurde insbesondere durch die recht übereinstimmende Anwendung der Nullsetzungsregel für den Skalierten Section-Score des Affektbereichs im Falle des Vorliegens eindeutiger manischer oder hypomanischer Episoden durch die Rater eine zum Erreichen hoher Werte nötige Varianz ins Spiel gebracht.

Nach dem Finn-Koeffizienten liegen mit .84 für den Summen-Score und .73 für den Skalierten Section-Score sämtliche Werte im guten Bereich, da dieser Koeffizient ja v.a. von der Übereinstimmung der Rater und nicht von der tatsächlichen Varianz der Urteile abhängig ist.

6.2.3.2.2. Kognitionsbereich

Die Reliabilität des Kognitionsbereichs erreicht für den Summen-Score den guten Reliabilitätskoeffizienten von ICC=.79. Der Skalierte Section-Score liegt jedoch nur bei knapp befriedigenden .58. Beim Finn-Koeffizienten ist das Ergebnis mit .75 für den Summen-Score und .58 für den Skalierten Section-Score praktisch identisch.

Auch dieses Ergebnis ist erklärungsbedürftig. Beim Summen-Score scheint die Übereinstimmung der Rater ausreichend. Der Summen-Score pendelt jedoch in vielen Fällen gerade in der für den Skalierten Section-Score kritischen Zone: Bei den Summen-Scores 2 und 3 ist ein Skalierter Section-Score von 1 zu geben, ab 4 aber ein Skalierter Section-Score von 2. Der Mittelwert liegt genau in dieser entscheidungsrelevanten Zone, so dass die Abweichungen der Rater zu unterschiedlichen Scores führen mussten.

Tab. 39: *Studenten-Rating; Reliabilitätsmaße der Summen-Scores, der Skalierten Section-Scores, des Gesamt-Scores und der DIB-R-Diagnose*

		ICC*	Finn's *r*	Horst
Summen-Scores	**Affekte**	**0,37**	0,84	**0,34**
	Kognition	0,79	0,75	0,77
	Impulsivität	0,91	0,93	0,91
	Beziehungen	0,80	0,89	0,78
Skalierte Section-Scores	**Affekte** *(Range 0-2)*	0,75	0,73	0,73
	Kognition *(Range 0-2)*	0,58	0,58	0,56
	Impulsivität *(Range 0-3)*	0,75	0,80	0,73
	Zwischenmenschl. Beziehungen *(Range 0-3)*	0,83	0,87	0,81
Gesamt-Score	*(Range 0-10)*	0,81	0,85	0,80
DIB-R Diagnose	*(dichotom)*	0,83	0,83	0,81

* Sämtliche ICC sind signifikant von Null verschieden. Kennwerte, die den angestrebten Wert ≥.70 für gute Reliabilitäten verfehlten, sind grau unterlegt dargestellt, ungenügende Reliabilitäten ≤.50 zusätzlich fett gedruckt.

6.2.3.2.3. Bereich Impulshandlungen

Der ICC des Summen-Scores des DIB-R-Bereichs Impulshandlungen, der weitgehend gute Statements aufwies, erreicht die sehr guten Werte von ICC=.91 bzw. *r*=.93.

Der Skalierte Section-Score des Bereichs ist mit ebenfalls guten ICC=.75 und *r*=.80 der höchste der Studenten-Stichprobe, so dass hier keine Widersprüche auftreten.

6.2.3.2.4. Bereich Zwischenmenschliche Beziehungen

Der Beziehungsbereich, in dem die meisten Statements eine unbefriedigende bzw. mäßige Reliabilität aufwiesen, ist in der Studenten-Stichprobe dennoch der zweitreliabelste DIB-R-Bereich.

Mit ICC=.80 bzw. *r*=.89 weist der Summen-Score eine gute Reliabilität auf. Offenbar mitteln sich die abweichenden Einschätzungen der Rater in den einzelnen Statements beim

Zusammenrechnen zu einem erheblichen Teil wieder heraus, was dann aber nicht für die Qualität bzw. Validität des erreichten Wertes spricht.

Entsprechend ist auch der Skalierte Section-Score des Bereichs Zwischenmenschliche Beziehungen mit ICC= .83 und *r*= .87 als reliabel zu bezeichnen.

6.2.3.2.5. Zusammenfassung zur Reliabilität der Bereichskennwerte

Nur die Reliabilitäten der Bereiche Impulshandlungen und Zwischenmenschliche Beziehungen sind uneingeschränkt als gut zu beurteilen.

Beim Bereich Affekte ist die Reliabilität insofern eingeschränkt, dass die Ratings sehr einförmig sind. Alle anderen ICC-Koeffizienten sind schlecht, die Finn-Koeffizienten hingegen liegen im guten Bereich. Diese Differenz zwischen dem ICC und dem Finn-Koeffizienten in diesem Bereich ist v.a. durch den unterschiedlichen Berechnungsmodus zu erklären.

(Die Einförmigkeit der Raterurteile lässt die – am vorliegenden Datenmaterial nicht überprüfbare – Vermutung zu, dass die studentischen Rater eher aufgrund von Antworttendenzen als aus einem wirklichen Verständnis der Statements heraus geurteilt haben könnten, was die Validität der Messergebnisse natürlich deutlich einschränken würde.)

Im Kognitionsbereich werden hingegen die Kriterien für reliable Skalierte Section-Scores nach ICC und Finn's *r* klar verfehlt, da die Raterurteile – wenn auch insgesamt relativ gut übereinstimmend – gerade im kritischen Bereich differieren.

6.2.3.3. Reliabilitäts-Kennwerte und Verteilung der DIB-R-Gesamtwerte und der Diagnosen

Die Reliabilität des Gesamt-Scores des DIB-R, der aus den vier Section-Scores aufsummiert wird, ist mit einem ICC von .81 gut. Die Reliabilität der aus dem Gesamt-Score anhand des Schwellenwertes (s.o.) gestellten Ein- bzw. Ausschlussdiagnosen "Borderline-Persönlichkeitsstörung" erreicht mit ICC=.83 ebenfalls einen guten Wert.

Es ist über alle 12 Videobänder eine weitgehende Übereinstimmung festzustellen: In nur zwei Fällen traten überhaupt unterschiedliche diagnostische Einschätzungen der Rater auf.

Nach der im Kapitel 2.4.6.5 zu den Grundlagen der Reliabilitätbestimmung geschilderten Rechenmethode wurde eine mittlere prozentuale Übereinstimmung von 92% bestimmt.

Zusammenfassend kann die Stellung der Ein- bzw. Ausschlussdiagnose Borderline-Persönlichkeitsstörung anhand des DIB-R durch die trainierten Studenten als reliabel gelten.

Die von den Ratern tatsächlich vergebenen Gesamt-Scores und Diagnosen sind zur besseren Vorstellbarkeit und Übersichtlichkeit in der folgenden Tabelle wiedergegeben.

Tab. 40: *Studenten-Rating; Raterurteile im Gesamt-Score und Diagnosen über alle Fälle*

Fälle*	Raterurteile								
	Gesamt-Scores				M	Diagnosen			
1	9	10	9	8	9,25	1	1	1	1
2	5	7	2	3	4,75	0	0	0	0
3	**10**	**9**	**6**	**8**	**9**	**1**	**1**	**0**	**1**
4	4	4	2	2	4	0	0	0	0
5	7	6	4	4	6,5	0	0	0	0
6	10	10	9	8	10,75	1	1	1	1
7	**6**	**8**	**6**	**6**	**8,25**	**0**	**1**	**0**	**0**
8	4	3	2	3	5	0	0	0	0
9	5	6	4	4	7	0	0	0	0
10	6	5	4	5	7,5	0	0	0	0
11	9	9	9	9	11,75	1	1	1	1
12	3	3	1	1	5	0	0	0	0

** Fälle mit Diagnoseabweichungen sind in Fettdruck dargestellt.*

Die Differenzen zwischen den Gesamt-Scores der Studenten sind insgesamt etwas größer als die der Experten: sie liegen zwischen einem und fünf Punkten. Die größten Differenzen sind hier nicht klar lokalisierbar: sie liegen bei den Mittelwerten 4,75 und 9 in den Fällen 2 und 3. Die weiteren Unterschiede verteilen sich über die ganze Bandbreite.

Die beiden Diagnoseabweichungen zwischen den Ratern in den Fällen 3 und 7 traten ähnlich der Experten-Stichprobe, wie zu erwarten nur dann auf, wenn der Gesamtwert in einem Bereich nahe des Cutoffs lag (hier 8,25 und 9). Bei allen anderen über die Rater gemittelten Gesamt-Scores gab es keine Diagnose-Abweichungen.

Zu beachten bleibt, dass die Werte der trainierten Studenten und der Experten vom Betrag her *nicht direkt vergleichbar* sind, da lediglich vier identische Fälle untersucht wurden.

Aufgrund dieser geringen übereinstimmenden Fallzahl wird eine Gegenüberstellung der Studenten- mit den Experten-Ratings nicht vorgenommen. Obwohl die Vermutung besteht, dass die Studenten z.T. wenig valide eingeschätzt haben, ist eine Prüfung dieser Frage aus o.g. Grund hier nicht möglich.

6.2.3.4. Zu den Ergebnissen der Horst-Koeffizienten

Auch bei den studentischen Ratern treten kaum Unterschiede zwischen dem Horst- und dem verwendeten Intraclass-Korrelationskoeffizienten auf. Eine eigene Interpretation der nur zu Vergleichszwecken erhobenen Horst-Ergebnisse erübrigt sich auch für die Studenten-Stichprobe. Von einer hohen Vergleichbarkeit kann ausgegangen werden.

6.3. TYPISCHE ANWENDUNGS- UND AUSWERTUNGSFEHLER DES DIB-R

Während des Einsatzes des beschriebenen Interviews in diagnostischen Gesprächen, des Gruppenratings von Interviews, die auf Videos aufgezeichnet wurden und beim Training von Studenten im Rahmen des Forschungsprojekts zur Evaluierung des DIB-R mussten immer wieder folgende Anwendungs- und Auswertungsfehler korrigiert werden, die die Reliabilität des DIB-R beeinträchtigen könnten.

So geben sich unerfahrene Interviewer gerne mit »Ja«- oder »Nein«-Antworten zufrieden, zum Beispiel auf die Frage: Waren Sie in den letzten zwei Jahren depressiv? Man sollte sich aber immer Beispiele, Häufigkeiten der Erlebnisse oder Verhaltensweisen schildern lassen und so lange nachfragen, bis man das Statement sicher einschätzen kann. In den Statements werden verschiedene Symptome, die erfragt werden, häufig durch »oder« verbunden. Dabei wird oft übersehen, dass dadurch das Zutreffen eines der genannten Phänomene für eine »2« im entsprechenden Statement genügt. Viele Beurteiler neigen dazu, die üblicherweise vorgesehene Wertung »l = wahrscheinlich« im Sinne einer abgestuften Ratingskala von »weniger stark vorhanden« oder »nicht ganz borderlinetypisch ausgeprägt« zu verwenden, was *nicht* vorgesehen ist.

Auch die teilweise unterschiedliche Belegung der Kodierung »(2, 1, 0)« führt regelmäßig zu Fehlern. So wird im Impulsbereich häufig vergessen, nach der *Häufigkeit* der Handlungen zu fragen, sodass nicht sichergestellt ist, ob wirklich ein Handlungs*muster* vorliegt. In den Fragen zu pseudopsychotischen versus echten psychotischen Erlebnissen führen jedoch nur »l«-Kodierungen der Fragen (pseudopsychotische Phänomene) zu einer »2«-Kodierung des zugehörigen Statements, während mit »2« eingeschätzte Antworten hingegen eine »0« im Statement zur Folge haben. Durch diese "Überkreuzung" sind Auswertungsfehler geradezu vorprogrammiert. Spezifische Anweisungen zur Berechnung des Skalierten Scores werden leider häufig überlesen.

Während sich die genannten Probleme überwiegend durch mangelnde Übung und Flüchtigkeitsfehler des Diagnostikers erklären lassen (insbesondere weniger erfahrene Rater sind davon betroffen), sind andere Unklarheiten auf die Anweisungen selbst zurückzuführen: Es ist eine Inkonsistenz im Interview, wenn der Patient ausschließlich nach seinem Erleben und Verhalten in den letzten beiden Jahren gefragt wird, die differentialdiagnostische Anweisung zum Affekt- und Kognitions-Section-Score aber fordert, manische und psychotische Erlebnisse zu berücksichtigen, die »jemals« aufgetreten sind – wonach man den Patienten ja unter Umständen gar nicht gefragt hat. Auch wenn die Fragen nach hypomanischen Episoden in den Revisionen des DIB ausdifferenziert wurden, sollte eine diagnostisch so schwerwiegende Entscheidung nur nach den DSM-IV-Kriterien vorgenommen werden.

Auch müsste unserer Einschätzung nach die Unterscheidung von pseudopsychotischen Erlebnissen und echt-psychotischen Phasen besser operationalisiert werden.

7. DISKUSSION DER ERGEBNISSE

In diesem Kapitel werden die in den Kapiteln 5 und 6 bereits beschriebenen, vorstrukturierten und inhaltlich eingeordneten Ergebnisse zusammenhängend diskutiert und interpretiert.

7.1. VALIDITÄT

7.1.1. Konvergente und divergente Validität bzgl. der DIB-R- und SKID-II-Diagnosen

7.1.1.1. Zusammenfassung der Ergebnisse

Die vorliegenden Ergebnisse sind durchgehend positiv. Die **konvergente Validität**, die im vorliegenden Anwendungsfall einer Übereinstimmungs-Validität entspricht, wurde an 100 psychiatrischen Patienten als Übereinstimmung der DIB-R-Diagnose Borderline-Persönlichkeitsstörung mit der entsprechenden Achse-II-Diagnose des SKID-II nach DSM-III-R bzw. -IV geprüft. Die absolute Übereinstimmung lag für dieses konzeptionell etwas abweichende Kriterium bei etwa 84%. Sie war mit einem Kappa-Koeffizienten von .59 signifikant überzufällig häufig, ein zufriedenstellender Validitätskoeffizient r_{tc}=.60 konnte ermittelt werden. Die hierzu formulierten Hypothesen 1.1. und 1.2. konnten beibehalten werden.

Im Sinne einer **diskriminanten Validität** wurde die Differenzierungsfähigkeit in Form des Fehlens substanzieller (univariater) Korrelationen der DIB-R-Borderline-Diagnose mit dem Vorliegen *anderer* Persönlichkeitsstörungen überprüft. Auch hier wurden gute Ergebnisse erzielt: Das Vorliegen einer Borderline-Persönlichkeitsstörung nach DIB-R korreliert gering oder gar nicht mit dem Vorliegen irgendeiner anderen Achse-II-Störung als der Borderline-Persönlichkeitsstörung.

Lediglich mit der *Schizotypischen Persönlichkeitsstörung* ergibt sich bzgl. deren Vorliegens eine signifikante Korrelation von φ=.28. Diese Korrelation ist aber wesentlich kleiner als jene mit der SKID-II-Borderline-Persönlichkeitsstörung. Zur Vorhersage der Borderline-Diagnose nach DIB-R aus der SKID-II-Diagnose *Schizotypische Persönlichkeitsstörung* oder umgekehrt würde der Zusammenhang bei einem Wert <.30 nach Lienert und Raatz (1989, s. Kap. 2.2.6) nahezu nutzlos sein.

Auch bei einer zur Sicherheit angewandten multivariaten Untersuchung, bei der *alle* SKID-II-Diagnosen simultan in eine Diskriminanzanalyse einbezogen wurden, konnte – außer der SKID-II-Borderline-Diagnose – keine SKID-II-Diagnose einen erheblichen Beitrag zur Vorhersage der

DIB-R-Borderline-Diagnose leisten. Lediglich die Diagnosen Schizotypische und Dependente Persönlichkeitsstörung deuten noch in geringem Umfang auf das Vorliegen einer DIB-R-Borderline-Diagnose hin.

7.1.1.2. Fazit

Insgesamt erreicht das DIB-R hinsichtlich der konvergenten und divergenten Validität ein voll befriedigendes Ergebnis. Hinsichtlich der Differenzierungsfähigkeit zu anderen Persönlichkeitsstörungen nach DSM-III-R bzw. -IV ist es ohne Einschränkungen als gut zu beurteilen.

7.1.2. Konvergente und divergente Validität bzgl. der DIB-R-Borderline-Diagnose und der DSM-Diagnosegruppen

7.1.2.1. Zusammenfassung der Ergebnisse

Die Verteilung der Patienten mit und ohne Borderline-Persönlichkeitsstörung ist signifikant unterschiedlich über die definierten DSM-Diagnosegruppen. Die häufigste Kombination ist dabei die der *Borderline-Persönlichkeitsstörung nach DIB-R und SKID-II*, die zweithäufigste *Borderline-Persönlichkeitsstörung nach DIB-R und Andere Persönlichkeitsstörungen*. Die hierzu formulierte Hypothese 2.1. konnte beibehalten werden.

Auch der DIB-R-Gesamtwert unterscheidet sich hochsignifikant über die gebildeten DSM-Diagnosegruppen. Die Gruppe der SKID-II-Borderlinepatienten weist dabei wie erwartet den höchsten DIB-R-Gesamt-Wert auf, welcher sich signifikant von dem aller anderen Diagnosegruppen unterscheidet. Die hierzu formulierten Hypothesen 2.2.a) und b) konnten beibehalten werden.

Bzgl. der entsprechenden Ergebnisse zu den Skalierten Section-Scores des DIB-R fällt das Ergebnis – wenn auch nicht durchgängig – insgesamt sehr positiv aus. In allen Skalierten Section-Scores unterscheiden sich die DSM-Diagnosegruppen signifikant voneinander, die entsprechende Hypothese 2.3.a) kann beibehalten werden.

Über alle Bereiche weist die SKID-II-Borderline-Gruppe dabei durchgehend den höchsten Wert auf. Dieser ist allerdings nur im Bereich *Zwischenmenschliche Beziehungen* signifikant von *allen* anderen DSM-Diagnosegruppen unterschiedlich. Die hierzu formulierte Hypothese 2.3.b) kann für den Bereich Zwischenmenschliche Beziehungen beibehalten werden. In den Bereichen *Kognition* und *Impulshandlungen* aber unterscheidet sich die SKID-II-Borderline-Gruppe von allen anderen DSM-Diagnosegruppen mit Ausnahme der Gruppe *Andere DSM-Störungen*. Da diese letzte Diagnosegruppe im Untersuchungsplan, also zum Zeitpunkt der Formulierung der Hypothesen, nicht vorgesehen war und nur eine Restkategorie darstellt, und außerdem das

Ausbleiben des signifikanten Mittelwertsunterschieds der sehr geringen Größe und nicht einem besonders kleinen Mittelwertsunterschied geschuldet sein dürfte, darf die Hypothese 2.3.b) auch für die Bereiche *Kognition* und *Impulshandlungen* dennoch als beizubehalten beurteilt werden.
Dieses gilt aber nicht für den Bereich *Affekte*. In diesem DIB-R-Bereich sind die Mittelwerte über alle DSM-Diagnosegruppen auffallend wenig differenziert. Trotz der o.g. *insgesamt* vorhandenen signifikanten Unterschiedlichkeit der Werte unterscheidet sich hier *keine* einzelne DSM-Diagnosegruppe signifikant von einer anderen. Für diesen Bereich ist daher die Hypothese 2.3.b) klar zurückzuweisen.
Dieses Ergebnis unterstreichen auch die berechneten Effektstärken für die oben genannten Varianzanalysen. Die Effektstärke für den Affektbereich ist klar die niedrigste. Wenngleich auch sie mit .40 einen gerade noch als *hoch* zu qualifizierenden Wert erreicht, liegen die der anderen Bereiche mit .75 – .79 fast in doppelter Höhe.

7.1.2.2. Fazit

Das DIB-R vermag insgesamt überzeugend Patienten mit Borderline-Persönlichkeitsstörung von Patienten mit anderen Diagnosen zu unterscheiden. Es demonstriert seine differentialdiagnostische Unterscheidungsfähigkeit sowohl hinsichtlich der Achse-I- als auch der Achse-II-Störungen nach DSM-III-R bzw. -IV.
Der Affektbereich ist hinsichtlich der Differenzierungsfähigkeit klar der schwächste DIB-R-Bereich. Er ist nur eingeschränkt in der Lage, Patienten mit Borderline-Persönlichkeitsstörung von denen mit anderen Störungsbildern zu differenzieren. Er erweist sich damit nach wie vor als relativ unspezifisch – ein Kritikpunkt, der bereits gegenüber dem Affektbereich der Ursprungsversion DIB geäußert worden war.

7.1.3. Variation der Cutoff-Werte von DIB-R und SKID-II

In der vorliegenden Studie waren zusätzlich die Cutoff-Werte des DIB-R einer überblicksweisen Überprüfung hinsichtlich ihrer Plausibilität unterzogen worden. Anhand des Vergleichs der DIB-R-Diagnosen unter veränderten Cutoff-Werten hinsichtlich der dann erhaltenen Prävalenz der Borderline-Persönlichkeitsstörung, der Diagnoseübereinstimmung mit dem SKID-II sowie der Parameter Sensitivität, Spezifität und prädiktiver Werte (PPW, NPW, GPW) erscheint der gegenwärtig gültige Cutoff-Wert von "8" als sinnvoll. Eine abschließende Beurteilung dieser Frage ist auf der Basis der vorliegenden Studie aber nicht möglich.
Ebenso brachte eine Berücksichtigung "unterschwelliger" Borderline-Diagnosen nach SKID-II keine verbesserte Übereinstimmung.

7.2. RELIABILITÄT

7.2.1. Interrater-Reliabilität

Die Überprüfung der Reliabilität war v.a. anhand einer Gruppe in der Anwendung des DIB-R trainierter Rater mit klinischer Erfahrung insbesondere im Bereich der Borderline-Persönlichkeitsstörung vorgenommen worden (Expertengruppe). Zu Vergleichszwecken war auch eine Gruppe von Studenten in der Anwendung des DIB-R trainiert und ebenfalls bzgl. der Interrater-Reliabilität überprüft worden. In beiden Gruppen waren unabhängige Ratings anhand von videographierten DIB-R-Interviews durchgeführt worden. Die Ergebnisse werden im Folgenden zusammengefasst und bewertet.

7.2.1.1. Zusammenfassung der Ergebnisse der Expertengruppe

Die Reliabilität des DIB-R in der Anwendung durch die Expertengruppe wurde mittels 19 Videoaufzeichnungen durchgeführter Interviews überprüft. Die Berechnung der Interrater-Reliabilität erfolgte mittels eines Intraclass-Korrelationskoeffizienten (ICC) sowie des Finn-Koeffizienten. Reliabilitäts-Koeffizienten zur individuellen Beurteilung wurden gemäß der Ausführungen in Kap. 2.4.7 (für beide Messgrößen) ab .50 als befriedigend, ab .70 als gut und ab .90 als sehr gut definiert.

7.2.1.1.1. Statements

Die Reliabilitäten der 22 Statements des DIB-R liegen im ICC breit gestreut zwischen .39 und .96. Hierbei liegen 13 Statements im Bereich ≥.70, sechs Statements immerhin noch ≥.50. Drei Statements schnitten im DIB-R nach dem ICC unbefriedigend ab.

Nach den Ergebnissen zum Finn-Koeffizienten waren 16 Statement-Reliabilitäten ≥.70, fünf ≥.50 und nur eine mit .49 nicht mehr ausreichend.

Wurden die Ergebnisse des ICC mit denen des Finn-Koeffizienten abgeglichen, wurden die Statements S. 3, S. 4, S. 6 und S. 22 als noch – wenn auch mit gewissen Schwierigkeiten – mit befriedigender Genauigkeit einschätzbar bezeichnet.

Die Statements S. 16 *«Der Patient hat seine Abhängigkeitswünsche stark abgewehrt oder befand sich in einem ernsten Konflikt zwischen Versorgen und Versorgtwerden.»* und S. 19 *«Der Patient hatte in engen Beziehungen wiederkehrende Probleme mit Abwertung, Manipulation oder Sadismus.»* wurden als nicht ausreichend reliabel einschätzbar eingestuft. Offenbar sind diese Statements heterogen und wenig eindeutig formuliert, was sich auch schon während der Durchführung der Ratings immer wieder gezeigt hatte. Beim Statement S. 19 war der Rater

erfahrungsgemäß stark von der Darstellungsweise des Patienten abhängig. Beim S. 16 handelt es sich nach unserer Einschätzung um das am wenigsten in Verhalten oder Erleben beobachtbare Merkmal des DIB-R – es muss weitgehend aus einem psychoanalytisch geprägten Hintergrund heraus *erschlossen* werden.

Verglichen mit den Ergebnissen aus der Reliabilitäts-Studie von Zanarini, Frankenburg und Vujanovic (2002), die auf der Statement-Ebene (allerdings mit Kappa-Koeffizienten) arbeitete, sind sehr wenige Gemeinsamkeiten erkennbar. Bis auf vier Koeffizienten lagen dort *sämtliche* über .80 – die schlechtesten vier Statements waren S. 6 (.74), S. 19 (.73), S. 22 (.73) und S. 21 (.55).

Es fällt zum Einen auf, dass auch hier das S. 19 (zu den Problemen mit Abwertung, Manipulation oder Sadismus) eines der am wenigsten reliablen war, zum Anderen, dass die Autoren offenbar keinerlei Schwierigkeiten hatten, die Abhängigkeitswünsche und die Konflikte bzgl. des Versorgungsthemas in S. 16 einzuschätzen. Dies könnte an einem hohen Maß der Vertrautheit mit den hinter der Statement-Formulierung stehenden psychoanalytischen Konzepten liegen – schließlich ist die Studie mit Zanarini und Frankenburg von zwei Co-Autoren des DIB-R veröffentlicht worden. Für eine allgemeine Verwendung in einem theoretisch nicht gebundenen Anwenderkreis scheint das Statement eher fragwürdig zu sein.

Auch insgesamt erscheint vor dem Hintergrund der vorliegenden Ergebnisse die Studie von Zanarini, Frankenburg und Vujanovic (2002) mit ihren sehr hohen Reliabilitäten schon auf Statement-Ebene (evtl. aus o.g. Gründen der enormen Vertrautheit mit den eigenen Konzepten) als die *Reliabilität* eines üblichen klinischen Anwenderkreises *überschätzend*.

Die Studie von Zanarini, Frankenburg und Vujanovic (2002) ist allerdings *nicht direkt* mit der vorliegenden Studie vergleichbar, da es sich ja um zwei sprachlich verschiedene Versionen, um eine andere Patienten-Stichprobe sowie um verschiedene Verfahren der Interrater-Reliabilitäts-Schätzung handelt und nicht nur um zwei verschiedene Ratergruppen.

7.2.1.1.2. Skalierte Section-Scores

Die Reliabilitäten der vier Skalierten Section-Scores fielen sehr unterschiedlich aus: Der ICC des *Affektbereichs*, in dem vier von fünf Statements nach dem ICC weniger reliabel (<.70) waren, lag bei nur .57. Die Reliabilität des *Kognitionsbereichs* mit zwei weniger reliablen und einem guten Statement erreichte im ICC den guten Bereich von .77. Der ICC des Bereichs *Impulshandlungen*, der durchgehend gute Statements aufwies, lag bei .80 und im Beziehungsbereich, in dem von neun Statements zwei eine weniger gute und eines eine schlechte Reliabilität aufwies, erreichte der ICC sogar .89.

Wurden die Skalierten Section-Scores mit dem robusteren, aber die Reliabilität gelegentlich überschätzenden Finn-Koeffizienten berechnet, bot sich hier ein absolut vergleichbares Bild. Lediglich im Affektbereich, der möglicherweise in der Variabilität der Raterurteile eingeschränkt ist, hob sich der Reliabilitäts-Koeffizient vom noch befriedigenden ICC mit .57 auf Finn's *r* von .67, was aber keinen grundlegenden Unterschied bedeutet.

Ein vergleichbares Bild ergab sich auch in den unskalierten Summen-Scores der Bereiche. Diese Werte waren auch von Zanarini, Frankenburg und Vujanovic (2002) erhoben worden. Auch hier ergibt sich wieder ein genau konträres Bild: Gerade der in der vorliegenden Untersuchung am schlechtesten eingeschätzte Affektbereich schnitt dort mit einem ICC von .99 deutlich am Besten ab. Dieser enorme Unterschied kann auf der vorliegenden Datenbasis zwar nicht erklärt werden, es muss aber vermutet werden, dass hier wieder die durch regelmäßige Absprachen noch erhöhte Vertrautheit zu diesen unwahrscheinlich hohen Übereinstimmungen geführt haben muss.

Im Vergleich mit den in der spanischen Studie von Szerman et al. (2005) erhobenen Skalierten Section-Scores, wobei hier allerdings nicht ICC- sondern Kappa-Koeffizienten berechnet worden waren, liegen hier alle Skalierten Section-Scores, außer dem des Kognitionsbereichs, über .90, während hier der Kognitionsscore nur .63 erreicht. Offenbar hängen die erzielten Interrater-Reliabilitäts-Koeffizienten, wie oben schon gesagt, in einem hier nicht genauer bestimmbaren Maße von einer ganzen Reihe von Faktoren ab, die die Ergebnisse erheblich beeinflussen.

Insgesamt ist das Ergebnis, das sich für die Reliabilitäten der Skalierten Section-Scores ergibt, sehr gut – lediglich für den Bereich Affekte ergibt sich offenkundig Verbesserungsbedarf. Während die Hypothesen 3.3.b)-d) beibehalten werden können, ist der Affektbereich der einzige, in dem die entsprechende Hypothese 3.3.a), in der ein ICC≥.70 gefordert wird, zurückzuweisen ist.

7.2.1.1.3. Gesamt-Score und Diagnose

Die Reliabilität des Gesamt-Scores des DIB-R, der aus den vier Section-Scores aufsummiert wird, ist mit .91 sehr reliabel. Die Reliabilität der aus dem Gesamt-Score anhand des Cutoff-Wertes "8" gestellten Diagnosen erreicht mit .75 immer noch einen guten Wert.

Die Reliabilität der Diagnosen *muss* niedriger als der Gesamt-Score, aus dem die Diagnose ermittelt wird, ausfallen, da sich im Bereich um den Cutoff herum die Möglichkeit ergibt, trotz einer eigentlich geringen Schwankung im Gesamt-Score – ein Punkt genügt – eine absolute Diagnosenabweichung zu erhalten. Dieses wurde anhand der Tab. 35, in der die absoluten Diagnosenunterschiede aufgezeigt sind, demonstriert.

Im Vergleich mit den Ergebnissen aus der spanischen Studie (Szerman et al., 2005) ergeben sich

vergleichbare Ergebnisse. Dort waren ein ICC von .90 für den Gesamt-Score und ein Kappa-Koeffizient von .78 für die Stellung der Diagnose (allerdings beim Cutoff-Wert ≥7) gefunden worden. In der Studie von Zanarini, Frankenburg und Vujanovic (2002) war der Gesamt-Score nicht überprüft worden – der dort allerdings für die Interrater-Übereinstimmung bei der Diagnosestellung ermittelte exzellente Kappa-Koeffizient von .94 liegt – im Vergleich zur vorliegenden Untersuchung – in konkurrenzloser Höhe.
Die entsprechenden Hypothesen 3.1. und 3.2. der vorliegenden Studie, in denen eine gute Reliabilität mit einem ICC≥.70 gefordert wird, können beibehalten werden.

7.2.1.2. Zusammenfassung der Ergebnisse der Studentengruppe

Leider sind die vorliegenden Ergebnisse nicht, wie ursprünglich geplant, direkt mit denen der Expertengruppe vergleichbar, da nur 4 der 12 untersuchten Patienten auch von den Experten gesehen worden waren. Der eigentlich angestrebte direkte Vergleich konnte in der Studie aus organisatorischen Gründen leider nicht ermöglicht werden.
Es besteht in einiger Hinsicht die – anhand der vorliegenden Daten nicht überprüfbare – Vermutung, dass das Rating der Studenten wenig valide ist und die Gruppe nicht ausreichend zu einer konzeptgemäßen Einschätzung der DIB-R-Statements in der Lage war.
Die eingehende Interpretation der auf dieser Basis erhaltenen Reliabilitäten würde zu wenig aussagekräftigen Ergebnissen führen.

7.2.1.2.1. Statements

Von den Studenten schnitten nach dem ICC nur 8 von 22 Statements mit einem guten Wert ≥.70 ab. Sieben Statements lagen noch im befriedigenden Bereich ≥.50, während weitere sieben Statements mit einem Wert <.50 als nicht reliabel eingeschätzt gelten müssen. Dazu gehören das S. 10 zu sexuell abweichendem Verhalten und das auch in der Expertengruppe mäßig eingeschätzte S. 19 zu Problemen mit Abwertung, Manipulation und Sadismus. Insbesondere aber wurde der *gesamte* Affektbereich mit allen Statements dem ICC zufolge unreliabel eingeschätzt. Berechnet mit dem Finn-Koeffizienten verschieben sich die Reliabilitäts-Koeffizienten für die meisten Statements im Affektbereich deutlich in die Höhe, was nahelegt, dass eine eingeschränkte Variabilität der Raterurteile eine der Ursachen der geringen Werte war. So berechnet erreichen 13 Statements ein gutes und sieben ein befriedigendes Niveau. Nur noch zwei statt vorher sieben Statements erreichen eine klar erkennbar ungenügende Reliabilität: Es handelt sich um das S. 3 aus dem Affektbereich zum Thema Ärger und Wut sowie um das bereits o.g. S. 19. Dass die Statements des Affektbereichs im Finn-Koeffizienten nun nicht mehr

eine allgemein schlechte Reliabilität aufweisen, legt in Verbindung mit den beinahe durchgehend hohen Ratings der Studenten in diesem Bereich (siehe Tab. 38) nahe, dass die Statements *inadäquat einförmig* eingeschätzt worden sind. Das scheint entsprechend auch für das S. 10 zu Mustern sexuell abweichenden Verhaltens aus dem Bereich Impulshandlungen zu gelten, in dem die Studenten beinahe durchgehend "0" geratet hatten.

7.2.1.2.2. Skalierte Section-Scores

In den Skalierten Section-Scores des DIB-R erreichen die Studenten, trotz der genannten Unzulänglichkeiten, in allen Bereichen brauchbare Reliabilitäten ≥.70 – lediglich im Kognitionsbereich wurde ein nur befriedigender ICC von .58 erzielt. Diese Werte erfahren durch eine Berechnung mittels des Finn-Koeffizienten *keine* Veränderung.

Für die Summen-Scores, die der als Summe aller Statements gebildete Rohwert zur Berechnung der Skalierten Section-Scores sind, stellen sich die Ergebnisse etwas anders dar. Im ICC erreicht der Bereich Affekte einen schlechten Wert von nur .37, alle anderen Summen-Scores liegen über .70, wobei der Kognitions-Score mit .79 der schlechteste der drei Werte ist.

Mittels des Finn-Koeffizienten berechnet bessert sich der Affekt-Summen-Score drastisch auf einen guten Wert von .84 – was abermals die eingeschränkte Urteilsvarianz aufzeigt. Im Finn-Koeffizienten ist, wie vorher bei den Skalierten Section-Scores, der Kognitions-Summen-Score mit .75 der schlechteste Wert.

7.2.1.2.3. Gesamt-Score und Diagnose

In der Studenten-Stichprobe erreichen sowohl der Gesamt-Score als auch die Diagnose Reliabilitätswerte ≥.70. Mit ICC-Werten von .81 für den Gesamt-Score und .83 für die Diagnose liegen sie für letztere sogar noch über den Werten der Expertengruppe, der Finn-Koeffizient bringt dabei nahezu identische Ergebnisse und bestätigt die gefundene Reliabilität.

7.2.1.3. Fazit

Zusammenfassend kann die Stellung der Ein- bzw. Ausschlussdiagnose Borderline-Persönlichkeitsstörung anhand des DIB-R als reliabel gelten.

Dies gilt in erster Linie für die von der Expertengruppe durchgeführte Reliabilitäts-Studie. Auch die Studentengruppe erreichte eine weitgehend gute Übereinstimmung. Leider kann nicht sichergestellt werden, dass die studentischen Rater auch konzeptgemäß gültige, valide Urteile fällten, sodass die gefunden Reliabilitäten nur mit dieser Einschränkung interpretierbar sind. Dennoch demonstrieren die Ergebnisse zur Reliabilität der Studentengruppe, dass auch Personen

mit einem geringeren klinischen Erfahrungshintergrund durch ein Anwendertraining zu Urteilen mit einer guten Übereinstimmung in der Lage sind.
In diesem Zusammenhang wäre eine breiter angelegte Studie, bei der eine größere Anzahl derselben Patienten bzw. Videoaufzeichnungen von je einer Gruppe von Studenten und Experten geratet wird, aufschlussreich, da sie einen direkten Vergleich gestatten und eine Überprüfung der Qualität der Studenten-Ratings gestatten würde.

Eine weitere, die Reliabilität möglicherweise negativ beeinflussende Quelle von Fehlern und Varianz wurde in der vorliegenden Studie nicht untersucht: Es handelt sich um die Vielzahl von Skalierungsregeln, die im DIB-R (und ähnlich schon in der Vorgängerversion DIB) Anwendung finden.
Ein Beispiel aus dem Affektbereich ist z.B. die "Nullsetzungsregel", die trotz eines möglicherweise erreichten hohen Bereichs-Summen-Scores wegen des Vorliegens häufiger eindeutiger manischer oder hypomanischer Episoden einen Skalierten Section-Score von "0" erzwingt. Oder im gleichen Bereich die Skalierungsregel, die nur bei Vorliegen von voll erfüllten Statements S. 3 *und* S. 5 einen Skalierten Affekt-Section-Score von "2" gestattet und ansonsten eine "1" vorschreibt. Derartige Regeln kommen in *jedem* DIB-R-Bereich vor.
Die Vielfalt dieser ineinander verstrickten, in Wechselwirkung stehenden Regeln in ihrer Auswirkung auf die Testergebnisse zu untersuchen wäre nur an einer wesentlich größeren Stichprobe möglich.

7.3. ZUSAMMENHÄNGENDE DISKUSSION VON RELIABILITÄT UND VALIDITÄT

Im Zusammenhang lassen die o.g. Ergebnisse zu Validität und Reliabilität den Schluss zu, dass die DIB-R-Bereiche Kognition, Impulshandlungen und Zwischenmenschliche Beziehungen ausreichend gut konstruiert sind, um eine qualitativ hochwertige Einschätzung von Patienten hinsichtlich ihrer borderlinetypischen Symptomatik zuzulassen. Das Gleiche gilt für die Gesamtkonstruktion des DIB-R in Bezug auf die Unterteilung in thematisch gegliederte Symptombereiche und ebenso auch für die Diagnosestellung anhand eines aus diesen Bereichswerten gebildeten Gesamtwertes bei einem Cutoff von "8".
Lediglich der DIB-R-Bereich Affekte scheint einer Überarbeitung und inhaltlichen Umgestaltung zu bedürfen. Er schnitt sowohl bei der Überprüfung von Validität und Reliabilität nur gerade noch ausreichend ab und hält hohen Maßstäben nicht stand. Er steht damit in einem

deutlichen Kontrast zu den übrigen Bereichen und trägt insgesamt wenig zur Qualität der Stellung der Diagnose einer Borderline-Persönlichkeitsstörung mittels des DIB-R bei.

7.4. VERSCHIEDENE ASPEKTE

Eine detaillierte Untersuchung der Ursachen für die Unterschiedlichkeit der Stellung einer Borderline-Diagnose zwischen DIB-R und SKID-II war anhand des vorliegenden Datenmaterials leider nicht möglich. Als Faktoren kamen z.B. die Unterschiede in der Konzeption beider Instrumente hinsichtlich Manie, Hypomanie und psychotischer Episoden in Betracht, die im DIB-R mittels der o.g. Skalierungsregeln systematisch durch Punkteabzug die Stellung einer Borderline-Diagnose erschweren oder beim Zusammentreffen mehrerer Symptome z.T. unmöglich machen, während dieses nach dem SKID-II vor dem Hintergrund des Komorbiditätskonzepts von DSM-III-R bzw. -IV praktisch unerheblich ist.

Auch könnte eine Untersuchung von weiteren Persönlichkeitsstörungs- und anderer Codiagnosen weiteren Aufschluss über eine Systematik in der abweichenden Diagnosestellung geben.

Leider ist die in dieser Studie vorhandene Fallzahl von nur 16 Patienten, die *entweder* nach dem einen *oder* dem anderen Instrument eine Borderline-Diagnose erhalten haben, hinsichtlich der Komplexität der beiden o.g. Fragestellungen zur genaueren Untersuchung zu klein.

7.5. MÖGLICHE FEHLERQUELLEN DER VORLIEGENDEN UNTERSUCHUNG

Gravierende Fehlerquellen in der Durchführung der Untersuchung konnten nicht ausgemacht werden.

Insgesamt wird vermutet, dass über das angewandte Rekrutierungsverfahren mit dem Angebot einer diagnostischen Dienstleistung für behandelnde Ärzte und Dipl.-Psychologen an der Klinik in hohem Maße diagnostisch schwierige und unklare Fällen Eingang in die Studie fanden. Dieses würde aber bedeuten, dass die erreichten Ergebnisse für Validität und Reliabilität insgesamt als eher konservativ zu bewerten wären, da sie unter relativ schwierigen Bedingungen zustande gekommen sind. In jedem Fall dürfte es sich nicht um eine für den Studienzweck ausgesucht günstige Stichprobe gehandelt haben – von einer Überschätzung der geschilderten Ergebnisse ist nicht auszugehen.

Lediglich das Verfahren einer Einschätzung der Reliabilität anhand von videographierten Interviews könnte zu einer gewissen, nicht quantifizierbaren Überschätzung der Reliabilitäten geführt haben, da alle Rater den exakt gleichen Interviewverlauf gesehen haben, was in der Praxis natürlich nicht vorkommt. Dieses Verfahren ist aber üblich und schränkt weder den

Vergleich der gefundenen Werte mit denen anderer Studien noch den Gebrauchswert der Ergebnisse ein.

7.6. EINORDNUNG IN DEN FORSCHUNGSSTAND

Die vorliegende Evaluations-Studie konnte zeigen, dass die deutsche Fassung des DIB-R ein praxistaugliches, valides und reliables Instrument zur Diagnose der Borderline-Persönlichkeitsstörung ist. Voraussetzung ist dabei eine fundierte psychiatrische oder psychologische Ausbildung, klinische Erfahrung im Umgang mit Borderline-Patienten und deren Symptomatik sowie eine allgemeine Erfahrung in der Stellung von Diagnosen, insbesondere unter Verwendung von strukturierten oder halbstrukturierten Verfahren.

Diese Erfahrung erscheint insbesondere für den Erhalt valider Einschätzungen notwendig zu sein. Eine ausreichende Interrater-Übereinstimmung im DIB-R lässt sich hingegen offenbar schon nach relativ kurzer Trainingszeit erreichen.

Verglichen mit dem am Meisten verbreiteten alternativen Instrument zur Stellung der Borderline-Diagnose, dem SKID-II, ergeben sich nur teilweise ähnliche Reliabilitäts-Werte. Nach der Untersuchung von Fydrich et al. (1996) war für die Borderline-Diagnose in der deutschen Fassung des SKID-II nach DSM-III-R ein Kappa-Koeffizient von .79 gefunden worden. Diese Größenordnung erreicht mit .74 auch das DIB-R. Eine italienische Studie mit der englischen Fassung des SKID-II nach DSM-IV (Maffei et al., 1997) erreichte allerdings mit einem Kappa-Koeffizienten von .91 einen weit darüber hinausgehenden – dem Autor allerdings kaum als verallgemeinerbar erscheinenden – Reliabilitäts-Koeffizienten.

Mit einem Validitätskoeffizienten von r_{tc}=.60 ergab sich eine hinreichende Übereinstimmung des DIB-R mit dem SKID-II-Bereich Borderline-Persönlichkeitsstörung.

Die Einführung des DIB-R als eines sehr ausführlichen, speziell für die differential-diagnostische Abklärung der Borderline-Persönlichkeitsstörung geschaffenen Interviews erscheint damit wohlbegründet.

Die Einführung der Revision des DIB, die aus dem Wunsch nach einer besseren differentialdiagnostischen Abgrenzung v.a. von anderen Persönlichkeitsstörungen erfolgte, kann mit der vorliegenden Evaluation der deutschen Fassung als eingelöst betrachtet werden. Lediglich die, durch die Revision des DIB angestrebte, bessere und auch trennschärfere Ausgestaltung des Bereichs *Affekte* erscheint auf der Basis der vorliegenden Ergebnisse als wenig geglückt.

8. RESUMEE

Manche interessante Fragestellungen, z.T. auch die Hauptanliegen der vorliegenden Studie zur DIB-R-Validierung betreffende, waren wegen des Fehlens eines letztlich zuverlässigen und gültigen Außenkriteriums nur schwierig zu bewerkstelligen gewesen. Ein Vergleich mit dem SKID-II als Referenzdiagnose war allein schon durch den unterschiedlichen konzeptuellen Hintergrund grundsätzlich erschwert. Ein direkter Vergleich und eine hohe Übereinstimmung beider Instrumente waren von vornherein nicht möglich.

Das DIB-R erschien dabei in vielerlei Hinsicht als ein dem SKID-II grundlegend verwandtes, aber v.a. wesentlich differenzierteres und elaborierteres Testverfahren.

Beim DIB-R zu hinterfragen ist die komplexe und verwirrend vielfältige, z.T. willkürlich anmutende Ausgestaltung der Skalierungsregeln und Bereichszuordnungen der Symptomfragen. In dieser Form werden Testverfahren nur selten konstruiert, da sie dann in dieser Hinsicht kaum überprüfbar sind und weniger auf einer empirischen als einer eher "intuitiven Erfahrungsbasis" beruhen, wenn auch der in der so erreichten Bildung und Bewertung von "Symptomzusammenhängen" und deren differentieller Bewertung bestehende Vorteil nicht bestritten wird.

Diese Regeln aber, so scheint es, sind die wesentliche Quelle der diagnostischen Differenzen zwischen der DIB-R- und der SKID-II-Borderline-Diagnose.

Es erscheint dem Autor der vorliegenden Studie als wahrscheinlich, dass sich, bei Zugrundelegung einer anderen Auswertungsweise, die meisten Kriterien des DIB-R ausgesprochen zuverlässig auch zu einer Vorhersage der Borderline-Persönlichkeitsstörung nach dem DSM-IV-Konzept nutzen lassen würden.

Die Entwicklung eines neuen Auswertungsmodus könnte dann mit dem Ziel der Vorhersage der SKID-II-Borderline-Diagnose auf der Basis eines multivariaten Verfahrens (wie der Diskriminanzanalyse) geschehen. Bei dieser würde eine Vorhersagefunktion der Gruppenzugehörigkeiten auf ausschließlich empirischer Basis erstellt werden. Auf diese Weise könnte eine höhere Zuordnungsgenauigkeit erreicht werden als dies heute der Fall ist. Die o.g. differentielle Bewertung von Symptomen und deren Zusammenhängen würde dabei rein rechnerisch erfolgen. Anhand einer Kreuzvalidierung der Vorhersageergebnisse an einer zweiten Stichprobe könnte dann die Genauigkeit sehr gut und einwandfrei überprüft werden.

(Selbstverständlich wäre alternativ dazu auch an eine Referenzdiagnose in Form einer größeren Anzahl von Expertenurteilen zu denken, deren Zuordnung dann vorhergesagt werden müsste. Auch dies wäre anhand der im DIB-R abgefragten Symptome vermutlich zuverlässig möglich.)

Die grundsätzliche Vereinheitlichung der ohnehin (u.a. durch ihre z.T. gemeinsame und sich in der Vergangenheit wechselseitig beeinflussende historische Entwicklung) relativ ähnlichen Konzepte der Borderline-Persönlichkeitsstörung hätte somit eine erhebliche Vereinfachung für die wissenschaftliche und diagnostische Arbeit zur Folge und wäre nach Ansicht des Autors wünschenswert.

Unabhängig davon erscheint bzgl. des DIB-R (und beispielsweise auch des SKID-II) die Durchführung weiterer Evaluations-Studien für die verschiedenen sprachlichen Versionen als absolut notwendig. Wahrscheinlich aufgrund des hohen Aufwands solcher Projekte werden aber nur wenige solcher Studien durchgeführt und veröffentlicht.
Der Sinn solcher Studien ist v.a. die *Generalisierbarkeit* der gefundenen Werte auf einen *breiten* allgemeinen Anwenderkreis. Gerade bei Studien zur Reliabilität aber scheint nach Meinung des Autors häufig nach dem Motto: "Reliabilitäten *von bis zu X* sind erreichbar" verfahren zu werden. Solche Werte dürften kaum auf die Population der praktischen Anwender verallgemeinerbar sein.
Eine dem Problem der allgemein zuverlässigen Anwendung solcher – für den Patienten und dessen Behandlung immens wichtiger – Diagnoseinstrumente angemessenere Losung wäre aber: "Klinische Sorgfalt und gute Schulung im jeweiligen Instrument vorausgesetzt, sind Reliabilitäten *von wenigstens X* durch die Anwender zuverlässig zu erzielen."
Der Autor hofft, dass die mit dieser Studie vorgelegten Ergebnisse dem o.g. Anspruch genügen können.

LITERATURVERZEICHNIS

Akiskal, H.S. (2000). Die Borderline-Persönlichkeit: affektive Grundlagen, Symptome und Syndrome. In: *Handbuch der Borderline-Störungen.* Kernberg, O.F., Dulz, B. & Sachsse, U. (Hrsg.). Stuttgart: Schattauer; 259-70.

American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders (DSM-III)* (3rd ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders DSM-III-R* (3rd ed., rev.). Deutsche Bearb.: Wittchen, H.-U., Saß, H., Zaudig, M. & Köhler, K. (1989). *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-III-R*. Weinheim: Beltz Verlag.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* (4th ed.). Washington, DC: American Psychiatric Association. Deutsche Bearb.: Saß, H., Wittchen, H.-U. & Zaudig, M. (1996). *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-IV*. Göttingen: Hogrefe.

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders – DSM-IV-TR* (4th ed., Text Revision). Washington, DC: American Psychiatric Association. Deutsche Bearb.: Saß, H., Wittchen, H.-U. & Zaudig, M. (2003). *Diagnostisches und Statistisches Manual Psychischer Störungen – Textrevision – DSM-IV-TR.* Göttingen: Hogrefe.

Armstrong, G.D. (1981). The intraclass correlation as a measure of interrater reliability of subjective judgments. *Nursing Research*, 30, 5, 314-5, 320A.

Asendorpf, J. & Wallbott, H.G. (1979). Maße der Beobachterübereinstimmung: Ein systematischer Vergleich. *Zeitschrift für Sozialpsychologie*, 10, 243-52.

Bartko, J.J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19, 3-11.

Bartko, J.J. & Carpenter, W.T. (1976). On the methods and theory of reliability. *Journal of Nervous and Mental Disease*, 163, 5, 307-17.

Bortz., J. (1993). *Statistik für Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.

Bortz, J. (2000). *Re: Höhe des Intraclass-Koeffizienten.* E-Mail: juergen.bortz@tu-berlin.de (2000-05-04).

Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2. Aufl.). Berlin: Springer.

Bronisch, T. (1999). Diagnostik von Persönlichkeitsstörungen. *Persönlichkeitsstörungen,* Sonderband 3, 5-15.

Buss A. & Plonin, R. (1975). *A temperament theory of personality development*. London: Wiley-Interscience.

Chaine, F., Guelfi, J.D., Monier, C., Brun, A. & Seunevel, F. (1995). Clinical diagnosis and standardized evaluation of borderline personality: preliminary report. *Encephale,* 21, 4, 247-56.

Clarkin, J.F. & Dammann, G. (2000). Psychometrische Verfahren zur Diagnostik und Therapie der Borderline-Störungen. In: *Handbuch der Borderline-Störungen.* Kernberg, O.F., Dulz, B. & Sachsse, U. (Hrsg.). Stuttgart: Schattauer; 125-48.

Cohen, J.A. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20, 37-46.

Davison, G.C. & Neal, J.M. (1996). *Klinische Psychologie* (4. Aufl.). Weinheim: Beltz.

Eckert, J., Biermann-Ratjen, E.-M. Papenhusen, R. Tönnies, S., Talmos-Gros, S., Seifert, R. & Spehr, W. (1987). Zur Diagnose von Borderline-Störungen: Überprüfung der Güte-Kriterien des "Diagnostischen Interview für Borderline-Störungen" (DIB). *Psychotherapie, Psychosomatik, Medizinische Psychologie,* 37, 2, 68-74.

Eckert, J., Papenhusen, R., Biermann-Ratjen, E.-M. & Wuchner, M. (1991). Untersuchung zur differentialdiagnostischen Abgrenzung von Boderline- gegenüber schizophrenen und neurotisch-depressiven Störungen. *Psychotherapie, Psychosomatik, medizinische Psychologie,* 41, 320-27.

Falret, J. (1854). De la folie circulaire. *Bulletin de l'Academie Medicinale*, 19, 382-94.

Finn, R.H. (1970). A note on estimating the reliability of categorical data. *Educational and Psychological Measurement*, 30, 1, 71-6.

Finn, R.H. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, 32, 2, 255-65.

First, M.B., Spitzer, R.L., Gibbon, M. & Williams, J.B.W. (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)*. Washington, DC: American Psychiatric Press.

First, M.B., Spitzer, R.L., Gibbon & M., Williams, J.B.W. (1997). *Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II).* Washington, DC: American Psychiatric Press.

Fleiss, J.L. (1981). The measurement of interrater agreement. In: *Statistical methods for rates and proportions* (2nd ed.). Fleiss, J.L. New York: John Wiley & Sons; 212-36.

Fleiss, J.L. & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613-19.

Fydrich, T., Schmitz, B., Hennch, C. & Bodem, M. (1996). Zuverlässigkeit und Gültigkeit diagnostischer Verfahren zur Erfassung von Persönlichkeitsstörungen. In: *Persönlichkeitsstörungen: Diagnostik und Psychotherapie.* Schmitz, B., Fydrich, T. & Limbacher, K. (Hrsg.). Weinheim: Psychologie Verlags Union.

Gagnon, J., Bouchard, M.-A., Rainville, C., Lecours, S. & St-Amand, J. (2006). Inhibition and object relations in borderline personality traits after traumatic brain injury. *Brain Injury*, 20, 1, 67-81.

Grinker, R.R. (1977). The Borderline Syndrome: A Phenomenological View. In: *Borderline Personality Disorders: The Concept, the Syndrome, the Patient.* Hartocollis, P. (Ed.). Madison, Conneticut: International Universities Press.

Grinker, R.R. (1979). Diagnosis of borderlines: a discussion. *Schizophr Bull* 5, 47-52.

Grinker, R.R. & Werble, B. (1977). *The Borderline Patient.* New York: Jason Aronson.

Grinker, R.R., Werble, B. & Drye, R.C. (1968). *The Borderline Syndrome: A Behavioural Study of Ego-Function.* New York: Basic Books.

Gunderson, J.G. (1985). *Diagnostisches Interview für das Borderline-Syndrom. Manual.* Dtsch. Bearb.: Pütterich, H. Weinheim: Beltz-Test.

Gunderson, J.G. (2005). *Borderline: Diagnostik, Therapie, Forschung.* Dilling, H. (Hrsg. d. dtsch. Ausg.). Bern: Huber.

Gunderson, J.G., Kolb, J.E. & Austin, V. (1981). The diagnostic interview for borderline patients. *Am J Psychiatry,* 138, 896-903.

Gunderson, J.G. & Sabo A.N. (1993). The phenomenological and conceptual interface between borderline personality disorder and PTSD. *Am J Psychiatry*, 150, 19-27.

Gunderson, J.G. & Singer, M.T. (1975). Defining borderline patients: an overview. *Am J Psychiatry*, 132, 1-10.

Gunderson, J.G. & Zanarini, M.C. (1987). Current overview of the borderline diagnosis. *J Clin Psychiatry*, 48 (Suppl.), 5-14.

Herpertz, S., Gretzer, A., Steinmeyer, E.M., Mühlbauer, V., Schürkens, A. & Saß., H. (1997). Affective instability and impulsivity in personality disorder: results of an experimental study. *J Affect Disord*, 44, 31-7.

Herpertz, S. & Saß, H. (1997). Impulsivität und Impulskontrolle – Zur psychologischen und psychopathologischen Konzeptionalisierung. *Nervenarzt*, 68, 171-83.

Herpertz, S. & Saß, H. (2000). Die Borderline-Persönlichkeitsstörung in der historischen und aktuellen psychiatrischen Klassifikation. In: *Handbuch der Borderline-Störungen.* Kernberg, O.F., Dulz, B. & Sachsse, U. (Hrsg.). Stuttgart: Schattauer; 115-23.

Hoch, P.H. & Polatin, P. (1949). Pseudoneurotic forms of schizophrenia. *Psychiatr Q*, 23, 248-76.

Hughes, C.H. (1884). Borderland psychiatric records – pro-dromal symptoms of psychical impairment. *Alienist and Neurologist*, 5, 85-91.

Kernberg, O.F. (1967). Borderline personality organization. *J Am Psychoanal Assoc*, 15, 641-85.

Kernberg, O.F. (1975). *Borderline Conditions and Pathological Narcissism.* New York: Jason Aronson.

Kernberg, O.F. (1975, 1990). *Borderline-Störungen und pathologischer Narzißmus.* Frankfurt/M.: Suhrkamp.

Kernberg, O.F. (1996). *Schwere Persönlichkeitsstörungen. Theorie, Diagnose, Behandlungsstrategien* (5. Aufl.). Stuttgart: Klett-Cotta

Kernberg, O.F., Goldstein, E.G., Carr, A.C., Hunt, H.F., Bauer, S.F. & Blumenthal, R. (1981). Diagnosing borderline personality. A pilot study using multiple diagnostic methods. *J Nerv Ment Dis*, 169, 4, 225-31.

Kraepelin, E. (1896). *Psychiatrie. Ein Lehrbuch für Studirende und Aerzte.* 5. Aufl. Leipzig: Barth.

Kraepelin, E. (1904). *Psychiatrie. Ein Lehrbuch für Studirende und Aerzte.* Bd. II, 7. Aufl. Leipzig: Barth.

Langer, I. & Schulz v. Thun, F. (1974). *Messung komplexer Merkmale in Psychologie und Pädagogik (Ratingverfahren).* München: Ernst Reinhardt Verlag.

Lawlis, G.F. & Lu, E. (1972). Judgement of counseling process: reliability, agreement and error. *Psychological Bullettin,* 78, 17-20.

Leichsenring, F. (2003). *Borderline-Stile: Denken, Fühlen, Abwehr und Objektbeziehungen; eine ganzheitliche Sichtweise* (2., vollst. überarb. und erw. Aufl.). Bern: Huber.

Lienert, G.A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Psychologie Verlags Union.

Loranger, A.W., Janca, A. & Sartorius, N. (1997). *Assessment and Diagnosis of Personality Disorders. The ICD-10 International Personality Disorder Examination (IPDE).* Cambridge: Cambridge University Press.

Loranger, A.W. & WHO (1996). *International Personality Disorder Examination (IPDE).* Deutschspr. Ausg.: Mombour, W., Zaudig, M., Berger, P., Gutierrez, K., Berner, W., Berger, K., v. Cranach, M., Giglhuber, O. & v. Bosse, M. Bern: Huber.

Ludwig-Mayerhofer, W. (2008). *ILMES - Internet-Lexikon der Methoden der empirischen Sozialforschung. Validität: Inhaltsvalidität*. WWW: http://www.lrz-muenchen.de/~wlm/ilm_v5.htm (2008-01-03).

Maffei, C., Fossati, A., Agostini, I., Barraco, A., Bagnato, M., Deborah, D., Namia, C., Novella, L. & Petrachi, M. (1997). Interrater reliability and internal consistency of the structured clinical interview for DSM-IV axis II personality disorders (SCID-II), version 2.0. *Journal of personality disorders*, 11, 3, 279-84.

Margraf, J., Schneider, S. & Ehlers, A. (1994). *Diagnostisches Interview bei psychischen Störungen, Handbuch* (2. Aufl.). Berlin: Springer.

Mellsop, G., Varghese, F., Joshua, S. & Hicks, A. (1982). The reliability of axis II of DSM-III. *Am J Psychiatry*, 139, 10, 1360-1.

Millon, T. (1995). *Disorders of Personality – DSM-IV and Beyond* (2nd. ed.). New York: Wiley.

Rae, G. (1988). The equivalence of multiple rater kappa statistics and intraclass correlation coefficients. *Educational and Psychological Measurement*, 48, 2, 367-74.

Rapaport, D., Gill, M. & Schafer, R. (1946). *Diagnostic Psychological Testing: The Theory, Statistical Evaluation and Diagnostic Evaluation of a Battery of Tests. Vol. 1 & 2.* Chicago, IL: Year Book Publishers.

Robinson, W.S. (1957). The statistical measurement of agreement. *American Sociological Review*, 22, 1, 17-25.

Rohde-Dachser, Chr. (1979). *Das Borderline-Syndrom.* Bern: Huber.

Rohde-Dachser, Chr. (1995). *Das Borderline-Syndrom* (5. überarb. u. erg. Aufl.). Bern: Huber.

Schneider, K. (1923). *Die psychopathischen Persönlichkeiten.* Leipzig: Thieme.

Schneider, S. & Margraf, J. (2006). *Diagnostisches Interview bei psychischen Störungen; Handbuch, Interviewleitfaden, Protokollbogen* (3., vollst. überarb. Aufl.). Heidelberg: Springer Medizin-Verlag.

Schödlbauer, M., Biermann-Ratjen, E.-M., Brodbeck, D., Ladendorf, R., Rohde-Dachser, Chr. & Eckert, J. (1997). Zur Revision des 'Diagnostischen Interviews für Borderlinepatienten' (DIB). *Persönlichkeitsstörungen*, 3, 148-52.

Serban, G., Conte, H.R. & Plutchik, R. (1987). Borderline and schizotypal personality disorders: mutually exclusive or overlapping? *J Pers Assess*, 51, 15-22.

Spitzer, R.L., Endicott, J. & Gibbon, M. (1979). Crossing the border into borderline personality and borderline schizophrenia: the development of criteria. *Arch Gen Psychiatry*, 36, 17-24.

Spitzer, R.L., Endicott, J. & Robins, E. (1984). *Forschungs-Diagnose Kriterien (RDC) für eine ausgewählte Gruppe von psychiatrischen Erkrankungen.* Dtsch. Bearb.: Klein, H.E. Weinheim: Beltz.

Spitzer, R.L., Williams, J.B.W., Gibbon, M. & First, M.B. (1990a). *Structured Clinical Interview for DSM-III-R, Patient Edition.* Washington, DC: American Psychiatric Press.

Spitzer, R.L., Williams, J.B.W., Gibbon, M. & First, M.B. (1990b). *Structured Clinical Interview for DSM-III-R Axis II Disorders (SCID-II).* Washington, DC: American Psychiatric Press.

Stern, A. (1938). Psychoanalytic investigation of and therapy in the borderline group of neuroses. *Psychoanalytic Quarterly*, 7, 467-89.

Strauß, B. & Schumacher, J. (Hrsg.) (2005). *Klinische Interviews und Ratingskalen. Reihe Diagnostik für Klinik und Praxis (Band 3)*. Göttingen: Hogrefe.

Szerman, N., Peris, M.D. , Ruiz, A., Ruiz, M. , Gunderson, J.G. & Rejas, J. (2005). Linguistic adaptation and validation into spanish of the Diagnostic Interview for Borderline Personality Disorders-Revised (DIB-R). *Current Medical Research and Opinion*, 21, 8, 1251-9.

Tinsley, H.E.A. & Weiss, D.J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22, 4, 358-76.

van der Kolk, B.A. (1999). Das Trauma in der Borderline-Persönlichkeit. *Persönlichkeitsstörungen*, 3, 21-9.

Wegner, R. (1976). Ratingmethoden. In: *Techniken der empirischen Sozialforschung (Band 5)*. van Koolwijk, J., Wieken-Mayser, M. (Hrsg.). München: Oldenbourg; 103-30.

Weltgesundheitsorganisation (1980). *Diagnosenschlüssel und Glossar psychiatrischer Krankheiten: deutsche Ausgabe der internationalen Klassifikation der Krankheiten der WHO, ICD, 9. Revision, Kapitel V*. (5. Aufl., korr. nach der 9. Rev. der ICD). Degkwitz, R. (Hrsg.). Berlin: Springer.

Weltgesundheitsorganisation (1993). *Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F); Klinisch-diagnostische Leitlinien.* Dilling, H., Mombour, W. & Schmidt, M.H. (Hrsg.). Bern: Huber.

Weltgesundheitsorganisation (2006). *Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F); diagnostische Kriterien für Forschung und Praxis.* (4. überarb. Aufl.) Dilling, H. (Hrsg.). Bern: Huber.

Widiger, T.A., Miele, G.M. & Tilly, S.M. (1992). Alternative perspectives on the diagnosis of borderline personality disorder. In: *Borderline Personality Disorder: Clinical and Empirical Perspectives.* Clarkin, J.F., Marziali, E. & Munroe-Blum, H. (Eds.). New York: Guilford, 89-115

Wittchen, H.-U., Schramm, E., Zaudig, M. & Unland, H. (1993). *Strukturiertes Klinisches Interview für DSM-III-R Achse II (Persönlichkeitsstörungen) – SKID II.* Weinheim: Beltz.

Wittchen, H.-U., Zaudig, M. & Fydrich, T. (1997). *SKID: Strukturiertes Klinisches Interview für DSM-IV; Achse I und II.* Göttingen: Hogrefe.

Wittchen, H.-U., Zaudig, M., Schramm, E., Spengler, P., Mombour, W., Klug, J. & Horn, R. (1991). *Strukturiertes Klinisches Interview für DSM-III-R (SKID).* Weinheim: Beltz Test Gesellschaft.

Wuchner, M. (1997). *Behandlungsergebnisse von Borderline-Patienten nach klientenzentrierter Psychotherapie.* Münster: Waxmann Verlag.

Zanarini, M.C. (2007). R*e: Reliability of the German Version of the DIB-R.* E-Mail: zanarini@mclean.harvard.edu (2007-08-24).

Zanarini, M.C. & Frankenburg, F.R. (2003). Omega-3 fatty acid treatment of women with borderline personality disorder: a double-blind, placebo-controlled pilot study. *Am J Psychiatry*, 160, 167–69.

Zanarini, M.C., Frankenburg, F.R., Hennen, J., Reich, D.B. & Silk, K.R. (2006). Prediction of the 10-year course of borderline personality disorder. *Am J Psychiatry*, 163, 827-32.

Zanarini, M.C., Frankenburg, F.R., Hennen, J. & Silk, K.R. (2003). The longitudinal course of borderline psychopathology: 6-year prospective follow-up of the phenomenology of borderline personality disorder. *Am J Psychiatry*, 160, 274–83.

Zanarini, M.C., Frankenburg, F.R. & Vujanovic, A.A. (2002). Inter-rater and test-retest reliability of the revised Diagnostic Interview for Borderlines. *Journal of Personality Disorders,* 16, 3, 270-6.

Zanarini, M.C., Frankenburg, F.R., Vujanovic, A.A., Hennen, J., Reich, D.B. & Silk, K.R. (2004). Axis II comorbidity of borderline personality disorder: description of 6-year course and prediction to time-to-remission. *Acta Psychiatrica Scandinavica*, 110, 416–20.

Zanarini, M.C., Gunderson, J.G. & Frankenburg, F.R. (1990). Cognitive features of the borderline personality disorder. *Am J Psychiatry*, 147, 1, 57-63.

Zanarini, M.C., Gunderson, J.G., Frankenburg, F.R. & Chauncey, D.L. (1989). The revised Diagnostic Interview for Borderlines: discriminating BPD from other Axis II disorders. *Journal of Personality Disorders*, 3, 10-8.

Zanarini, M.C., Williams, A.A., Lewis, R.E., Reich, R.B., Vera, S.C., Marino, M.F., Levin, A., Yong, L. & Frankenburg, F.R. (1997). Reported pathological childhood experiences associated with the development of borderline personality disorder. *Am J Psychiatry,* 154, 1101-6.

Zimmerman, M. & Coryell, W. (1989). DSM-III personality disorder diagnoses in a nonpatient sample. Demographic correlates and comorbidity. *Arch Gen Psychiatry,* 46, 8, 682-9.

ANHANG

Diagnostisches Interview für Borderline-Patienten DIB-R

(Projektversion mit Kommentar)

DIAGNOSTISCHES INTERVIEW FÜR BORDERLINE-PATIENTEN (DIB-R)

[Projektversion mit Kommentar]

(Gunderson und Zanarini 1983,
Modifikation 1992)

McLean Hospital
Harvard Medical School

Wegen weiterer Informationen wenden Sie sich bitte an Prof. Rohde-Dachser, Institut für Psychoanalyse der Universität Frankfurt, oder unmittelbar an die Autoren am McLean Hospital, Psychosocial Research Program, 115 Mill Street, Belmont, Massachusetts, 02178

Revidierte Fassung des "Diagnostischen Interviews für Borderline-Patienten (DIB-R)" von Gunderson und Zanarini (1983, Modifikation 1992)

Die revidierte Fassung des "Diagnostischen Interviews für Borderline-Patienten (DIB-R)" von Gunderson und Zanarini (1983) dient, anders als die ursprüngliche Form des "Diagnostischen Interviews für Borderline-Patienten (DIB)" von Gunderson und Kolb (1978), insbesondere der Abgrenzung der Borderline-Persönlichkeitsstörung von anderen Persönlichkeitsstörungen im Diagnostischen und Statistischen Manual Psychischer Störungen (DSM-III-R bzw. IV). Es stellt an die Diagnose *"Borderline-Persönlichkeitsstörung"* gleichzeitig strengere Maßstäbe als das ursprünglich von Gunderson und Kolb verfaßte "Diagnostische Interview für Borderline-Patienten". Aus diesem Grunde ist es insbesondere für Forschungen im Bereich der Persönlichkeitsstörungen geeignet. Im November 1992 wurde das Manual von Gunderson und Zanarini in einigen Punkten nochmals modifiziert. Auch diese Modifikationen sind in der hier abgedruckten Fassung des DIB-R enthalten.

BESCHREIBUNG:

Das revidierte DIB (im folgenden DIB-R genannt) ist ein halbstrukturiertes Interview, das Informationen aus vier Bereichen sammelt, die für die Borderline-Persönlichkeitsstörung von diagnostischer Bedeutung sind: Affektivität, Kognition, Impulshandlungen und Zwischenmenschliche Beziehungen.

Es mißt 97 Merkmale, die die Gefühle, die Gedanken und das Verhalten des Patienten in den *letzten zwei Jahren* betreffen. Für die Mehrzahl dieser Merkmale ist der Patient die einzige Informationsquelle; für einige wenige Merkmale können zusätzliche Informationsquellen herangezogen werden.

Das Interview ist in 24 Teilbereiche gegliedert; die Informationen aus 22 dieser Teilbereiche dienen der Einschätzung von 22 übergreifenden *Statements*, den sog. *Summary Statements*. Jedes dieser *Statements* repräsentiert ein wichtiges diagnostisches Kriterium der Borderline-Persönlichkeitsstörung und dient der Beurteilung über das Vorliegen oder Nicht-Vorliegen dieser Diagnose.

Die Informationen aus den anderen beiden Teilbereichen (Item 24 und 58) sprechen gegen eine Borderline-Diagnose und sollen bei der endgültigen Beurteilung des Affekt- und Kognitions-Bereichs mit herangezogen werden.

INTERVIEWER-ANLEITUNG:

1. Forschen Sie weiter nach, wenn der Patient eine Frage mißverstanden hat oder eine Antwort gegeben hat, die unvollständig, widersprüchlich oder unwahr erscheint. Forschen Sie auch weiter nach, wenn die Untersuchung zu bestimmten Fragen nicht genügend Informationen liefert, um ein *Summary Statement* vorzunehmen.
2. Kreisen Sie die Ziffer ein, die die beste Antwort für eine Frage oder ein *Summary Statement* liefert. Wenn nicht ausdrücklich anders vermerkt, werden alle Fragen und *Summary Statements* wie folgt bewertet: 2 = JA; 1 = WAHRSCHEINLICH, und 0 = NEIN. Wenn eine Frage nicht beantwortbar ist, schreiben Sie "n.b." rechts neben das Feld für das *Summary Statement*..
3. Addieren Sie für jeden Bereich die *Summary Statement Scores*, um einen sog. *SECTION SCORE* zu erhalten.
4. Transformieren Sie den *SECTION SCORE* in den *SKALIERTEN SECTION SCORE* von 0-2 oder 0-3, entsprechend den Richtlinien am Endes des betreffenden Bereichs.
5. Addieren Sie die *SKALIERTEN SECTION SCORES*, um den revidierten *DIB-R-Gesamt-Score* von 0-10 zu erhalten.
6. Für die Entscheidung über die Diagnose am Ende des Interviews gelten die folgenden Richtlinien: **Ein DIB-R-Score von 8-10 gilt als Indikator für eine Borderline-Persönlichkeitsstörung, während ein DIB-R-Score von 7 oder weniger als Indikator für ein anderes klinische Syndrom betrachtet wird.**

HINTERGRUNDINFORMATIONEN:

1. Patienten-Kennziffer: ☐
   (Name des Patienten: .....)
2. Klinischer Status ☐
   z.Z. des Interviews: 1. Hospitalisiert
   2. Ambulant
   3. Nicht in Behandlung
   (Datum des Interviews: .....)
   (Institution: .....)
   (Name des Interviewers: .....)
   (Name des Raters: .....)
3. Alter: ☐
4. Geschlecht: 1. männlich 2. weiblich ☐
5. Familienstand: 1. ledig ☐
   2. verheiratet oder geschieden
6. Nationalität: 1. deutsch ☐
   2. andere
7. Schulbildung: Schulabschluß ☐
8. Beruf: ☐
   01. Freier Beruf (Arzt, Anwalt, usw.), Leiter von Unternehmen
   02. Höherer Beamter, Leitender Angesteller
   03. Beamter, Angestellter
   04. Selbständiger Gewerbetreibender, Landwirt
   05. Facharbeiter
   06. Arbeiter
   07. Hausfrau/-mann, mithelfende/r Familienangehörige/r
   08. Schüler, Lehrling, Student
   09. Rentner, Pensionär
   10. ohne Beruf
   11. keiner der genannten Berufe, sondern: .....
   [Berufsbezeichnung: .....]
9. Schichtzugehörigkeit: ☐

In der Original-Version ist für die Einschätzung der Schichtzugehörigkeit die *Hollings- head-Redlich-Skala* vorgesehen, zusammen mit der Instruktion, bei Patienten, die nicht finanziell unabhängig sind, die Einschätzung der Schichtzugehörigkeit nach dem Bildungs- und beruflichen Niveau des Haushaltsvorstandes des Haushalts, in dem der Patient lebt, vorzunehmen. Auf den Abdruck dieser Skala wurde hier verzichtet.

Projekt: Evaluierung des DIB (2. Rev.): Kommentar

Dieser Fragebogen ist nur für die Mitarbeiter des Projekts bestimmt. Bis auf einzelne, eigens gekennzeichnete Stellen und sämtliche Anmerkungen, stimmt der Fragebogen mit der Übersetzung von Frau Rohde-Dachser überein.

Allgemeine Hinweise zur Durchführung:

Für den Zweck einer Evaluierung des DIB soll die Durchführung und Auswertung des DIB einheitlich gehandhabt werden. Auf der Grundlage der bisherigen Diskussion beim Raten von Modellbändern in Hamburg entstand folgender Kommentar. Er findet sich jeweils auf der Seite oberhalb der Fragen (Seite 1a - 22a).

- *Die Notierung der Antworten:*

Die Bewertung von Antworten und Statements ist mit "2" = Ja; "1" = Wahrscheinlich; "0" = Nein vereinbart. Einzelne Ausnahmen (Bereich *Impulshandlungen*) sind der Anweisung zu entnehmen.

Im Normalfall darf "1" und "2" nicht graduell verwendet werden. "1" gibt den Sicherheitsgrad der Einschätzung an.

-Wörtliche Durchführung:

Die Durchführung dieses halbstrukturierten Interviews ist sicher 'eleganter', wenn die Fragen nicht abgelesen werden. Viele Hinweise aus der direkten Interaktion gehen verloren, wenn man am Papier 'klebt'. Für eine 'freie' Durchführung spricht sicher auch, daß der Patient das Gespräch eher als auf ihn abgestimmt erlebt. Für den Zweck unserer Untersuchung ist es allerdings wichtig, daß **alle Fragen in der vom DIB vorgesehenen Form** gestellt und daß keine Unterfragen vergessen werden. Stellt man die Fragen nur teilweise, besteht die Gefahr, Borderline-Patienten nicht als solche zu erkennen (Falsch-Negative).

-Deskriptive Orientierung:

Wenn das DIB auch auf der Basis psychodynamischer Annahmen konzipiert wurde, verfolgen die Autoren (vgl. Gespräch zwischen M. Wuchner und M. Zanarini) das Ziel, die valide und reliable Anwendung des DIB auch klinisch wenig erfahrenen Interviewern zu ermöglichen. Auf Beurteilungen, die einen großen theoretischen Hintergrund erfordern, soll verzichtet werden. Der DIB hat den Anspruch, möglichst 'deskriptiv' das Borderline-Syndrom zu erfassen.

JA-Antworten auf Einzelfragen sind nur dann daraufhin zu überprüfen, ob das geschilderte Erleben oder Verhalten 'wirklich borderline-typisch' ist, wenn das DIB dies präzisiert.

Bsp.: Auf die Fragen Nr. 7: "Hatten Sie extreme Schuldgefühle?" ist das Ausmaß und die Häufigkeit dieser Gefühle relevant, nicht aber, ob diese Schuldgefühle sich darauf beziehen, daß sich der Pat. für böse hält, daß er versagt habe. . . Dies wären sicher wichtige Differenzierungen, die aber in der gegenwärtigen Fassung des DIB nicht vorgesehen sind.

- *Beispiele erfragen:*

Bei der Mehrzahl der Fragen genügt es zur Einschätzung nicht, daß der Patient nur 'Ja' oder 'Nein' antwortet. Man sollte sich im Zweifelsfalle (und immer, wenn es die Zeit erlaubt) Beispiele aus dem Leben des Patienten, die Häufigkeit etc. berichten lassen.

Viele Formulierungen der Fragen klingen 'niederschwellig' in dem Sinne, daß das Symptom in der Fragestellung als nicht allzu pathologisch erscheinen soll. Diese Fragen sollen zum Erzählen anregen.

Bsp.: Nr. 21: "Haben Sie sich gelangweilt?" Eine "Ja"-Antwort genügt sicher nicht, um eine "2" zu rechtfertigen. Wie Statement 5 präzisiert, muß es sich um ein chronisches Gefühl der Langeweile handeln.

Bsp.: Nr. 73: "Sind Sie auf Einkaufstouren gegangen, während derer Sie viel Geld für Dinge ausgegeben haben, die Sie nicht brauchten oder die Sie sich nicht leisten konnten?" Hier ist zu beurteilen, ob es sich um einen Konsumrausch i.S. eines Impulsdurchbruches handelt, der etwa zu einer Verschuldung geführt hat.

- *Videorating*

Unter Umständen kennt der Interviewer den Pat. aus Vorgesprächen. In diesem Fall sollten bekannte relevante Informationen kurz wiederholt werden ("Sie erzählten mir bereits, daß..."), denn ansonsten fehlen dem Video-Rater diese Hintergrundinformationen, was für die Bestimmung der Reliabilität sehr ungünstig wäre.

Fehlen bei einer Antwort genauere Angaben zu Art und Häufigkeit des Erfragten, so ist die 'Ja' oder 'Nein'-Antwort als "2" bzw. "0" zu raten, es sei denn, der Rater hat den Eindruck, diese Antwort stimme nicht (mit "0" zu kodieren) oder das Zutreffen sei nur wahrscheinlich (mit "1" zu kodieren).

*

Die Nummern des folgenden Kommentars beziehen sich auf die Fragen bzw. Statements des DIB's. Im Text weist ein hochgestellter Stern nach einer Nummer (z.B.: 5.*) auf einen Kommentar zu Frage 5. hin. Alle Anmerkungen, die sich nicht auf bestimmte Fragen, sondern z.B. auf Anweisungen beziehen, werden durch fortlaufende hochgestellte Buchstaben "a" gekennzeichnet.

Änderungen und Ergänzungen im DIB-Text sind in Klammern [. . .] gesetzt.

PATIENTEN-INSTRUKTION:

Bevor wir beginnen, möchte ich Sie darauf aufmerksam machen, daß sich der größte Teil der Fragen in dieser Untersuchung auf die letzten zwei Jahre Ihres Lebens bezieht, mit anderen Worten also auf die Zeit seit . . . . . *(entsprechendes Datum einsetzen)*. Dabei geht es besonders um die Gefühle, Gedanken und Verhaltensweisen, die in diesem Zeitraum für Sie *typisch* waren. Einige Fragen werden allerdings auch bestimmte Verhaltensweisen betreffen, die für Sie vielleicht nur dann zum Tragen kamen, wenn Sie in besonderer Weise aus dem Gleichgewicht geraten waren, oder wenn Sie sich in einer Krise befunden haben.

BEREICH "AFFEKTIVITÄT":

Depression

1. Gab es in den letzten 2 Jahren oft Zeiten, in denen Sie sich sehr niedergeschlagen oder deprimiert gefühlt haben? (2, 1, 0)

2. Gab es Perioden, in denen Sie über einen Zeitraum von 2 Wochen oder länger jeden Tag sehr deprimiert waren? (2, 1, 0)

3.* S. 1 DER PATIENT LITT AN EINER CHRONISCHEN, DEPRESSIVEN VERSTIMMUNG ODER HATTE EINE ODER MEHRERE PERIODEN VON MAJOR DEPRESSION 2, 1, 0

4. Haben Sie irgendwann in den letzten zwei Jahren tage- oder wochenlang unter einem Gefühl von Hilflosigkeit gelitten? (2, 1, 0)

5. Haben Sie sich hoffnungslos gefühlt? (2, 1, 0)

6. Wertlos? (2, 1, 0)

7.* Hatten Sie extreme Schuldgefühle? (2, 1, 0)

8. S. 2 DER PATIENT HATTE ANHALTENDE GEFÜHLE VON HILFLOSIGKEIT, HOFFNUNGSLOSIGKEIT, WERTLOSIGKEIT ODER SCHULD 2, 1, 0

Ärger / Wut

9.* Waren Sie während der letzten zwei Jahre häufig sehr ärgerlich? (2, 1, 0)

10. Oder wütend, oder zornig? (2, 1, 0)

11. Waren Sie oft sarkastisch? (2, 1, 0)

12. Oder besonders streitlustig? (2, 1, 0)

13. Oder leicht reizbar und aufbrausend? (2, 1, 0)

14. S. 3 DER PATIENT HATTE CHRONISCHE GEFÜHLE VON ÄRGER, WUT ODER VERHIELT SICH HÄUFIG ÄRGERLICH WÜTEND (WAR z.B. OFT SARKASTISCH, STREITLUSTIG ODER AUFBRAUSEND) 2, 1, 0

Angst

15. Waren Sie in den letzten zwei Jahren häufig sehr ängstlich? (2, 1, 0)

16. Hatten Sie in den letzten zwei Jahren häufig körperliche Symptome, die mit Spannungszuständen zusammenhängen, wie Kopfschmerzen, Herzklopfen oder starkes Schwitzen? (2, 1, 0)

17.* Wurden Sie in den letzten zwei Jahren oft von irgendwelchen irrationalen Befürchtungen oder Phobien beunruhigt? (2, 1, 0)

18. Hatten Sie in den letzten zwei Jahren irgendwelche Panikattacken (z.B. massive Angstanfälle, die Sie handlungsunfähig machten)? (2, 1, 0)

19. S. 4 DER PATIENT HAT SICH CHRONISCH SEHR ÄNGSTLICH GEFÜHLT ODER LITT HÄUFIG UNTER KÖRPERLICHEN ANGSTSYMPTOMEN 2, 1, 0

Andere dysphorische Affekte

20. Haben sie sich während der letzten zwei Jahre häufig sehr einsam gefühlt? (2, 1, 0)

21. Haben Sie sich gelangweilt? (2,1, 0)

22. Oder innerlich leer gefühlt? (2, 1, 0)

23. S. 5 DER PATIENT ERLEBTE CHRONISCHE GEFÜHLE VON EINSAMKEIT, LANGEWEILE ODER LEERE 2, 1, 0

3.* Bei allen "ODER"-Verbindungen in Statements genügt die Erfüllung eines Teilsatzes.

7.* Hier sind extreme Schuldgefühle jeder Art deskriptiv zu erfassen, ohne daß man sich von Annahmen über borderline-typische Schuldgefühle leiten läßt. Bitte notieren Sie die Art der Schuldgefühle:

.....

.....

9.* Es ist hier nicht erforderlich, daß der Ärger zu einer manifesten Auseinandersetzung führte oder von Dritten als Ärgerlichkeit wahrgenommen wurde. Es genügt das Gefühl der Ärgerlichkeit.

17.* Hier sind auch Ängste im Rahmen flüchtiger paranioder Ideen zu berücksichtigen.

Andere Merkmale

24.* [&58.]:

Gab es in den letzten zwei Jahren oft Tage oder Wochen, in denen Sie ohne ersichtlichen Grund übermäßig aufgedreht oder in Hochstimmung waren?

Oder ungewöhnlich gereizt, wenn Ihnen jemand über den Weg lief?

Glaubten Sie während dieser Zeit, daß Sie eine bedeutende Persönlichkeit sind, oder daß Sie besondere Fähigkeiten und Kräfte haben?

Schliefen Sie weniger als gewöhnlich ohne sich deshalb müde zu fühlen?

Waren Sie redseliger als gewöhnlich. . .
oder unfähig, mit dem Reden aufzuhören?

Hatten Sie das Gefühl, daß Ihre Gedanken sich überstürzten . . .
oder von einem Thema zum nächsten rasten?

Wurden Sie leichter abgelenkt als sonst? [Ließen Sie sich sehr leicht ablenken?]

Haben Sie sich in so viele (ungewöhnliche) Unternehmungen eingelassen, . . .

daß andere sich deshalb Sorgen machten?

Oder haben Sie sich körperlich unruhiger gefühlt als sonst?

Haben Sie unüberlegt viele spontane Dinge gemacht, die für Sie ganz untypisch sind? (z.B. Großeinkäufe gemacht, Affären gehabt, riskante Geschäfte getätigt)?

Haben andere Leute diese Veränderungen wahrgenommen? Was haben sie dazu gesagt?

Hat dieser Zustand Sie ernsthaft in Ihrer beruflichen Arbeit gestört? Wie wirkte er sich auf Ihr Leben zu Hause oder auf Ihre sozialen Kontakte aus? Mußten Sie wegen einer manischen Episode in eine psychiatrische Klinik?

(Beurteilen Sie, ob der Patient eine anhaltende Stimmungsschwankung hatte, . . .
während dieser Perioden sozial und beruflich ernsthaft beeinträchtigt war . . .
und zusätzlich drei der anderen sieben Kriterien einer hypomanischen, bzw. manischen Episode (nach DSM-III-R) erfüllte.)

(Manische Episoden, bzw. Hypomanische Episoden)

Hypomanische Episoden (Nr. 24) (2, 1, 0)

[Anzahl der hypomanischen Episoden in den letzten 2 Jahren: .....]

Manische Episoden [auch das Unterstrichene ist erfüllt] (Nr. 24 &58) (2, 1, 0)

[Anzahl der manischen Episoden in den letzten 2 Jahren: .....]

Anzahl der manischen Episoden vor mehr als 2 Jahren: .....]

25. AFFECT SECTION SCORE: _____

Skalierter Affect-Section Score:

2 wenn der Section Score 5 oder mehr beträgt und jeweils "2" in S. 3 und S. 5.

1 wenn der Section Score 3 oder 4 beträgt, oder bei jeder anderen Kombination von 5 oder mehr.

0 wenn der Section Score 2 oder weniger beträgt, oder wenn der Patient häufiger eindeutige [manische oder] hypomanische Episoden erlebt hat, die von anderen registriert wurden.

26. SKALIERTER AFFEKT-SECTION SCORE: _____

BEREICH "KOGNITION":

Dieser Bereich erfaßt Störungen des Denkens (seltsames Denken, ungewöhnliche Wahrnehmungen und nichtwahnhafte paranoide Erlebnisse), *pseudo*-psychotische und echte psychotische Denkstörungen. Als *pseudo*-psychotische Erlebnisse gelten Wahnvorstellungen und Halluzinationen, die flüchtig, umschrieben und atypisch für psychotische Störungen sind; als *echte* psychotische Erlebnisse gelten Wahnvorstellungen und Halluzinationen, die anhaltend sind, weit verzweigt und typisch für psychotische Störungen. **Außerdem beziehen sich alle *Summary Statements* und alle Items mit einer Ausnahme (Item 57) auf nicht von Substanzen induzierte Erlebnisse. Deshalb ist es wichtig, festzustellen, ob die vom Patienten beschriebenen Erlebnisse spontan oder unter dem Einfluß von Alkohol oder Drogen aufgetreten sind.**[a]

Seltsames Denken, ungewöhnliche Wahrnehmungen:

27. Waren Sie während der letzten zwei Jahre sehr abergläubisch (haben Sie z.B. oft auf Holz geklopft, Termine verschoben, weil Freitag, der 13. war, oder eine schwarze Katze als schlechtes Omen betrachtet)? (*Deutlich ausgeprägter Aberglaube*) (2, 1, 0)

28. Haben Sie oft geglaubt, daß Sie mit Ihren Gedanken, Worten oder Handlungen in einer besonderen oder magischen Weise Dinge verursachen oder verhindern könnten? (*Magisches Denken*) (2, 1, 0)

24.* & 58. Um eine ökonomischere Durchführung zu ermöglichen wurden die Texte der Fragen Nr. 24 (Hypomanische Episoden) und Nr. 58 (Manische Episoden) kombiniert. Was zusätzlich zu den Kriterien 'Hypomanie' für eine 'Manie' erfüllt sein muß, ist im Text unterstrichen. Mit der vorliegenden Fassung der Frage läßt sich bereits an dieser Stelle klären, ob bei dem Patienten hypo- oder manische Episoden in den letzten beiden Jahren aufgetreten sind.

Widersprüchlich ist, daß sich Nr. 24 und 58 auf den Zeitraum der letzten beiden Jahre beziehen, während die Anweisung (unter Nr. 59) lautet, daß der Kognitions-Score 0 beträgt, "wenn der Patient jemals [. . .] eine voll entwickelte manische Episode hatte"! Deshalb müssen auch manische Episoden erfragt werden, die mehr als zwei Jahre zurückliegen.

[a] Einschlägige Erlebnisse, die nur unter Drogeneinfluß aufgetreten sind, dürfen nicht mit einer "2" codiert werden. Wenn ein Erlebnis nur unter Drogen auftrat, kodieren Sie bitte "0" und vermerken Sie am Rand ein "D!" (für: drogeninduziert!). Vgl. auch Nr. 57.

29. Hatten Sie oft eine Art sechsten Sinn für Dinge, der über eine bloße [bzw. besondere] Sensibilität oder Wahrnehmungsfähigkeit für andere Menschen und deren Gefühle hinausging? (*Sechster Sinn*) (2, 1, 0)

30. Fühlten Sie sich oft in der Lage zu sagen, was andere Leute dachten oder fühlten, wenn Sie eine besondere oder magische Kraft benutzten, z.B. Telepathie? Haben Sie oft geglaubt, daß andere Leute wissen, was Sie denken oder fühlen, wenn Sie diese Art von Macht benutzten? (*Telepathie*) (2, 1, 0)

31. Hatten Sie oft hellseherische Erlebnisse, z.B. eine Vision von etwas, das an einem anderen Ort geschah? Konnten Sie häufig die Zukunft vorhersagen? (*Hellseherei*) (2, 1, 0)

32.* Hatten Sie irgendwelche Überzeugungen, die Sie nicht aufgeben konnten, obwohl Ihnen andere immer wieder sagten, daß Sie sich irrten (hielten Sie sich beispielsweise für dick, während Sie in Wirklichkeit untergewichtig waren)? (*Überwertige Ideen*) (2, 1, 0)

33. Haben Sie wiederholt die Anwesenheit einer Kraft oder einer Person gespürt, die nicht wirklich anwesend war? Haben Sie Dinge, die Sie gehört oder gesehen haben, oft falsch interpretiert (z.B. geglaubt, Sie hörten jemand Ihren Namen rufen, während es in Wirklichkeit ein anderes Geräusch war)? (*Wiederkehrende Sinnestäuschungen*) (2, 1, 0)

34.* Haben Sie sich wiederholt unwirklich gefühlt? So als ob Ihr Körper oder ein Teil von ihm fremd sei oder Größe und Form verändert habe? Als ob Sie körperlich von Ihren Gefühlen getrennt wären? Als ob Sie sich selbst aus einer Entfernung betrachten würden? (*Depersonalisation*) (2, 1, 0)

35.* Hatten Sie oft das Gefühl, als seien die Dinge um Sie herum unwirklich, so als ob sie fremd wären oder ihre Größe oder Form veränderten? Als ob Sie in einem Traum wären? Als ob etwas wie eine Glasscheibe zwischen Ihnen und der Welt wäre? (*Derealisation*) (2, 1, 0)

36. S. 6 DER PATIENT NEIGTE ZU SELTSAMEM DENKEN ODER UNGEWÖHNLICHEN WAHRNEHMUNGSERLEBNISSEN, (z.B. MAGISCHES DENKEN, WIEDERKEHRENDE SINNESTÄUSCHUNGEN, DEPERSONALISATION) 2, 1, 0

Nicht-wahnhafte paranoide Erlebnisse:

37. Waren Sie in den letzten zwei Jahren gegenüber anderen Leuten oft sehr mißtrauisch oder argwöhnisch? (*Übertriebenes Mißtrauen*) (2, 1, 0)

38. Haben Sie oft gedacht, daß andere Leute Sie anstarren? Hinter Ihrem Rücken über Sie reden? Über Sie lachten? (*Beziehungsideen*) (2, 1, 0)

39. Haben Sie oft gedacht, daß andere Leute darauf aus waren, Ihnen das Leben schwer zu machen oder sich mit Ihnen anzulegen? Haben Sie oft geglaubt, daß Sie übervorteilt wurden oder für etwas getadelt wurden, was nicht Ihr Fehler war? (*Andere paranoide Vorstellungen*) (2, 1, 0)

40.* S. 7 DER PATIENT HATTE HÄUFIG FLÜCHTIGE, NICHT-WAHNHAFTE PARANOIDE ERLEBNISSE (Z.B UNGERECHTFERTIGTES MISSTRAUEN, BEZIEHUNGSIDEEN, ANDERE PARANOIDE VORSTELLUNGEN). 2, 1, 0

Psychotische Erlebnisse:

Beurteilen Sie bitte jedes Erlebnis folgendermaßen:

2 = *echte* Wahnvorstellungen und Halluzinationen[b]
1 = *Pseudo*-Wahnvorstellungen und Halluzinationen
0 = *keine* Wahnvorstellungen und Halluzinationen.

41. Haben Sie geglaubt, daß Ihnen von irgendeiner äußeren Kraft Gedanken eingegeben werden? (*Gedankeneingebung*) (2, 1, 0)

42. . . ., daß Ihnen Gedanken gestohlen werden? (*Gedankenentzug*) (2, 1, 0)

43. . . ., daß Ihre Gedanken sich ausbreiten, so daß andere hören konnten, was Sie dachten? (*Gedankenlautwerden*) (2, 1, 0)

44. . . ., daß Ihre Gefühle, Gedanken und Handlungen von einer anderen Person oder einem Apparat kontrolliert werden? (*Beeinflussungswahn*) (2, 1, 0)

45. . . ., daß Sie tatsächlich hören können, was andere Menschen denken? Konnten andere buchstäblich Ihre Gedanken lesen, als wenn diese ein offenes Buch wären? (*Gedankenlesen*) (2, 1, 0)

46. . . ., daß andere sich in einer organisierten Weise gegen Sie verschworen haben? Daß sie Sie absichtlich verletzen oder bestrafen wollten? (*Verfolgungswahn*) (2, 1, 0)

32.* Überwertige Ideen sind abwegige, von den anderen nicht geteilte Überzeugungen, mit der sich die Person stark eins fühlt, und die keinen bizzarren oder echt wahnhaften Charakter haben (z.B. querulatorische Ideen, die noch nicht wahnhaft sind, oder eine eigentümliche Bedeutung religiöser Inhalte). Das Beispiel im DIB hat sich als eher ungünstig erwiesen.

34.* & 35* Bitte bei Interviews, die auf Video aufgezeichnet werden, diese beiden Fragen genau explorieren (Forschungsfragestellung).

40.* Unter *häufigen flüchtigen, nicht-wahnhaften paranoiden Erlebnissen* sind Vorstellungen zu verstehen, die nicht systematisiert sind. Ein systematisierter Wahn wäre Hinweis auf eine verzweigte psychotische Episode und ist unter Nr. 59 im Bereichswert 'Kognition' zu berücksichtigen.

[b] Abweichend von der üblichen Belegung von "2", "1", "0" spricht in diesem Bereich die Kodierung mit "1" (*Pseudo*-Wahnvorstellungen und [Pseudo-] Halluzinationen) für das Vorliegen eines Borderline-Syndroms, eine "2" jedoch dagegen.

47. . . ., daß andere Ihnen nachspionieren oder Ihnen folgen? Daß Dinge ganz speziell für Sie arrangiert wurden? Daß Ihnen besondere Botschaften durch das Radio oder Fernsehen gesendet wurden? (*Beziehungswahn*) (2, 1, 0)

48. . . ., daß Sie Strafe verdient haben, für etwas Schreckliches, das Sie getan haben? (*Versündigungswahn*) (2, 1, 0)

49. . . ., daß Sie eine außergewöhnlich bedeutende Person sind? Daß Sie besondere Fähigkeiten oder außergewöhnliche Kräfte haben? (*Größenwahn*) (2, 1, 0)

50. . . ., daß etwas Schreckliches geschehen ist oder in der Zukunft geschehen wird (z.B. daß morgen die Welt untergehen könnte oder Ihr Körper sich auflösen oder schmelzen könnte) ? (*Nihilistischer Wahn*) (2, 1, 0)

51. . . ., daß etwas mit Ihrem Körper nicht stimmt oder Sie eine ernsthafte Krankheit haben? (*Körperbezogener Wahn*) (2, 1, 0)

52. Gab es andere Überzeugungen, die von anderen als zweifelsfrei unwahr, fremdartig oder sogar bizarr angesehen wurden? (*Andere Wahnvorstellungen*) (2, 1, 0)

53. Hörten Sie Stimmen oder andere Geräusche, die niemand sonst hörte? (*Akustische Halluzinationen*) (2, 1, 0)

54. Hatten Sie irgendwelche Visionen oder sahen Sie andere Dinge, die niemand sonst sah? (*Optische Halluzinationen*) (2, 1, 0)

55. Hatten Sie andere Sinneswahrnehmungen, die von niemandem sonst geteilt wurden (z.B. die wiederholte Wahrnehmung eines Geruchs oder das Gefühl, daß etwas auf Ihrem Körper krabbelt, das nicht wirklich da war) (*Andere Halluzinationen*) (2, 1, 0)

56.* S. 8 DER PATIENT HATTE WIEDERHOLTE PSEUDO-PSYCHOTISCHE WAHNVORSTELLUNGEN ODER HALLUZINATIONEN 2, 1, 0

Andere Merkmale:

57. Welche von diesen Erlebnissen geschahen unter dem Einfluß von Alkohol oder Drogen? (*Substanzinduzierte psychotische Erlebnisse*) (2 = *echte* Erlebnisse, 1 = *pseudo*-psychotische Erlebnisse, 0 = *keine*) (2, 1, 0)

58. [s.o. Nr. 24]

[Übertrage von S. 7 (unten):]

Manische Episoden in den letzten zwei Jahren (2, 1, 0)
[Manische Episoden in früheren Jahren, nämlich: ..... (2, 1, 0)]

---

59. KOGNITIONS-SECTION SCORE: _____

Skalierter Kognitions-Section Score:

2 wenn der Section Score 4 oder mehr beträgt;
1 wenn der Section Score 2 oder 3 beträgt;
0 wenn der Section Score 0 oder 1 beträgt oder wenn der Patient jemals[c] entweder eine prolongierte verzweigte psychotische Episode oder eine voll entwickelte manische Episode hatte.

60. SKALIERTER KOGNITIONS-SECTION SCORE: _____

56.* "1"-Einschätzungen in den Items 41.-56. führen also zu einer "2" in S.8.

[c] Während in Nr. 58 nur nach Symptomen der Manie in den "letzten zwei Jahren" gefragt ist, soll nun im Kognitions-Section-Score berücksichtigt werden, ob irgendwann im Leben eine eindeutige Manie oder eine psychotische Störung aufgetreten ist!

BEREICH "IMPULSHANDLUNGEN"

Wenn die Antwort auf eine der folgenden Fragen "ja" lautet, fragen Sie bitte nach der Häufigkeit des betreffenden Verhaltens. Wenn es sich nicht um Substanzmißbrauch, Selbstbeschädigung und parasuizidales Verhalten handelt, beurteilen Sie jeden Typ von Impulsivität:

2 = 5mal oder mehr
1 = 3 - 4mal
0 = 2mal oder weniger[d].

Substanzmißbrauch

61. Haben Sie während der letzten zwei Jahre zu viel Alkohol getrunken oder waren Sie richtig betrunken? (*Alkohol-Abusus*) (2 = chronischer Abusus, 1 = episodischer Abusus, 0 = kein Abusus) (2, 1, 0)

62. Haben Sie während der letzten zwei Jahre Medikamente oder Drogen genommen, um sich in einen Rauschzustand zu versetzen? (*Substanz-Abusus*) (2 = chronischer Abusus, 1 = episodischer Abusus, 0 = kein Abusus) (2, 1, 0)

63. S. 9 DER PATIENT BETRIEB EINEN ERNSTHAFTEN DROGENMISSBRAUCH. 2, 1, 0

Sexuelle Abweichungen

64.* Haben Sie sich impulsiv mit x-beliebigen Menschen sexuell eingelassen oder hatten Sie sexuelle Affären? (*Promiskuität*) (2, 1, 0)

65. Gab es irgendwelche ungewöhnlichen sexuellen Praktiken (Haben Sie es z.B. genossen, beim Sex erniedrigt oder verletzt zu werden; anstatt selbst Sex zu haben, lieber andere dabei beobachtet)? (*Perversion*) (2, 1, 0)

65.1* [Hatten Sie sexuelle Kontakte mit irgendeinem Familienmitglied (außer Ihrem Ehemann bzw. Ihrer Ehefrau)? *(Inzest)* (2, 1, 0)]

66. S. 10 DER PATIENT HATTE EIN MUSTER SEXUELL ABWEICHENDEN VERHALTENS (z.B. PROMISKUITÄT ODER PERVERSION) 2, 1, 0

Selbstbeschädigung

67. Haben Sie sich absichtlich selbst verletzt, ohne daß Sie sich damit umbringen wollten (z.B. sich geschnitten, verbrannt oder geschlagen, ihre Hand durch die Scheibe gestoßen, gegen die Wand geschlagen, ihren Kopf angeschlagen)? (*Selbstbeschädigung*) (2 = 2x oder öfter, 1 = 1x, 0 = niemals) 2, 1, 0

68. S. 11 DER PATIENT ZEIGTE EIN MUSTER VON KÖRPERLICHER SELBSTBESCHÄDIGUNG 2, 1, 0

Suizidversuche

69. Haben Sie während der letzten zwei Jahre damit gedroht, sich umzubringen (*Suizidrohungen* (2 = 2x oder häufiger, 1 = 1x, 0 = niemals) (2, 1, 0)

70. Haben Sie irgendwelche Suizidversuche, auch solche leichterer Art, unternommen? (*Suizidgesten/Suizidversuche*) (2 = 2x oder öfter, 1 = 1x, 0 = keine) (2, 1, 0)

71. S. 12 DER PATIENT ZEIGTE EIN MUSTER VON MANIPULATIVEN SELBSTMORDDROHUNGEN, SELBSTMORDGESTEN ODER -VERSUCHEN (D.H. DIE PARASUIZIDALEN HANDLUNGEN WAREN HAUPTSÄCHLICH DARAUF ANGELEGT, HILFE ZU BEKOMMEN) 2, 1, 0

Andere impulsive Muster

72. Gab es Zeiten, wo Sie soviel gegessen haben, daß Sie starke Schmerzen bekamen oder sich übergeben mußten? (*Freßanfälle*) (2, 1, 0)

73.* Sind Sie auf Einkaufstouren gegangen, während derer Sie viel Geld für Dinge ausgegeben haben, die Sie nicht brauchten oder die Sie sich nicht leisten konnten? (*Einkaufstouren*) (2, 1, 0)

74. Oder auf Glücksspiel-Touren, wo Sie immer weitergesetzt bzw. gewettet haben, obwohl Sie laufend Geld verloren? (*Glücksspiel*) (2, 1, 0)

75. Haben Sie die Fassung verloren und irgendwen richtig angeschrien oder angebrüllt? (*Verbale Ausbrüche*) (2, 1, 0)

[d] Dies ist der einzige Bereich, in dem die Kodierung 2, 1, 0 eine graduelle Abstufung darstellt!

64.* alternativ kann gefragt werden:
"Haben Sie sich unüberlegt und ohne es eigentlich zu wollen mit x-beliebigen Menschen sexuell eingelassen oder hatten Sie sexuelle Affären? (*Promiskuität*)

65.1* Die in [Klammern] gesetzten Fragen sind zu Forschungszwecken aus einer früheren Version des DIB übernommen. Die Antwort geht nicht in Statement 10 ein.

73.* Ein Kriterium für eine "2" wäre z.B., daß sich der Patient durch die Einkaufstouren verschuldet hat.

76. Waren Sie in Schlägereien verwickelt? (*Körperliche Auseinandersetzungen*) (2, 1, 0)

77. Haben Sie jemandem körperliche Gewalt angedroht (z.B. jemandem gesagt, daß Sie ihn schlagen, niederstechen oder töten würden)? (*Androhung körperlicher Gewalt*) (2, 1,0)

78. Jemanden tätlich angegriffen oder mißhandelt (z.B. geohrfeigt, geschlagen oder getreten)? (*Tätliche Angriffe*) (2, 1, 0)

79.* Absichtlich Eigentum beschädigt (z.B. Geschirr zerschlagen, Möbel zertrümmert, ein fremdes Auto demoliert) (*Eigentumsbeschädigung*) (2, 1, 0)

80.* Sind Sie viel zu schnell gefahren? Geschah dies unter dem Einfluß von Alkohol oder Drogen? (*Rücksichtloses Fahren*) (2, 1, 0)

81. Haben Sie während der letzten zwei Jahre irgend etwas Ungesetzliches getan (z.B. Ladendiebstahl, Drogenhandel, Hehlerei)? (*Antisoziale Handlungen*) (2, 1, 0)

82.* S. 13 DER PATIENT ZEIGTE EIN ANDERES MUSTER IMPULSIVEN VERHALTENS 2, 1, 0

83.* IMPULSHANDLUNGEN-SECTION SCORE: _____

Skalierter Impulshandlungen-Section Score:

3 wenn der Section Score 6 oder mehr beträgt;
(2 entweder von S. 11 oder S. 12)

2 wenn der Section Score 4 oder 5 beträgt oder irgendeine andere Kombination von 6 oder mehr vorliegt;

0 wenn der Section Score 3 oder weniger ausmacht

84. SKALIERTER IMPULSHANDLUNGEN-SECTION SCORE: _____

BEREICH "ZWISCHENMENSCHLICHE BEZIEHUNGEN"

Unfähigkeit allein zu sein

85. Ist es Ihnen während der letzten zwei Jahre grundsätzlich sehr schwer gefallen, Ihre Zeit allein zu verbringen? (2, 1, 0)

86. Haben Sie oft verzweifelte Anstrengungen unternommen, um das Gefühl des Alleinseins zu vermeiden (z.B. stundenlange Telefongespräche geführt, ausgegangen, um einen Gesprächspartner zu finden)? (2, 1, 0)

87. Sich sehr deprimiert gefühlt, wenn Sie allein waren? (2, 1, 0)

88. Fühlten Sie sich dabei sehr ängstlich? Wütend? Leer? Schlecht? (2, 1, 0)

89. S. 14 DER PATIENT HAT TYPISCHERWEISE VERSUCHT, DAS ALLEINSEIN ZU VERMEIDEN ODER FÜHLTE SICH EXTREM DYSPHORISCH, WENN ER ALLEIN WAR 2, 1, 0

Verlassenheits- Verschlingungs- Vernichtungsängste

90. Hatten Sie während der letzten zwei Jahre immer wieder Angst, daß die Menschen, die Ihnen am nächsten stehen, Sie verlassen könnten? (2, 1, 0)

91. Hatten Sie immer wieder Angst, sich erdrückt zu fühlen oder Ihre Identität zu verlieren, wenn Sie anderen Menschen zu nahe kämen? (*Angst vor Verschlungenwerden*) (2, 1, 0)

92. Hatten sie immer wieder Angst, daß Sie buchstäblich zerfallen würden oder aufhören zu existieren, wenn Sie von jemandem verlassen würden, der Ihnen wichtig war? (*Angst vor Vernichtung*) (2, 1, 0)

93. S. 15 DER PATIENT HAT [WIEDERHOLT] VERLASSENHEITS-, VERSCHLINGUNGS- ODER VERNICHTUNGSÄNGSTE ERLEBT 2, 1, 0

Abwehr von Abhängigkeitswünschen

94. Waren Sie irgendwo beschäftigt, wo eine Ihrer Hauptaufgaben darin bestand, sich um andere Menschen oder um Tiere zu kümmern? (2, 1, 0)

79.* Die Frage betrifft auch die Beschädigung eigenen Eigentums und kann auch so gestellt werden:
"Haben Sie sich oder anderen etwas mit Absicht kaputtgemacht? (z.B. Geschirr zerschlagen, Möbel zertrümmert, ein fremdes Auto demoliert) (*Eigentumsbeschädigung*) (2, 1, 0)

80.* Rücksichtsloses Fahren, das nur in berauschtem Zustand vorkommt, kann mit einer "1" codiert werden.

82.* Es genügt also für eine "2" im Statement 13, daß eines der impulsiven Muster (Nr. 72-81) mit "2" erfüllt ist.

83.* In dem Impulshandlungen-Section Score sind nur die Einschätzungen "0, 2, 3" möglich. Es ist keine "1" vorgesehen.

95. Haben Sie Freunden, Verwandten oder Kollegen ständig Hilfe angeboten? (2, 1, 0)

96. Hat es Sie in den vergangenen zwei Jahren besonders gestört, wenn andere Menschen Ihnen helfen wollten oder sich um Sie zu kümmern versuchten? (2, 1,0)

97. Haben Sie auch dann nicht um Unterstützung oder Hilfe gebeten, wenn Sie das Gefühl hatten, wirklich Hilfe zu brauchen? (2, 1, 0)

98. Gab es in Ihrem Leben in den letzten zwei Jahren jemanden, von dem Sie das Gefühl hatten, Ihn wirklich zu brauchen? Hing Ihre Funktionstüchtigkeit von dieser Person ab? Hing Ihr Überleben von dieser Person ab? (2, 1, 0)

99. S. 16. DER PATIENT HAT SEINE ABHÄNGIGKEITSWÜNSCHE STARK ABGEWEHRT ODER BEFAND SICH IN EINEM ERNSTEN KONFLIKT ZWISCHEN VERSORGEN UND VERSORGTWERDEN 2, 1, 0

Instabile enge Beziehungen

100. Hatten Sie in den letzten zwei Jahren enge Beziehungen? Wie viele? Wie oft haben Sie diese Menschen gesehen? Wer war Ihnen darunter am wichtigsten?

(Wichtigste Beziehung: _____)

(2 = 4 oder mehr, 1 = 2-3, 0 = 1 oder weniger) (2, 1, 0)

101. War irgendeine dieser Beziehungen durch häufige intensive Streitigkeiten belastet? (2, 1, 0)

102. Gab es wiederholte Beziehungsabbrüche? (2, 1, 0)

103. S. 17 DER PATIENT NEIGTE ZU INTENSIVEN, INSTABILEN ENGEN BEZIEHUNGEN 2, 1, 0

Wiederkehrende Probleme in engen Beziehungen

104. Neigten Sie dazu, sich sehr abhängig von anderen zu fühlen? Brauchten Sie viel Unterstützung oder tatkräftige Hilfe, um zu funktionieren? Wurde Ihnen je gesagt, Sie seien zu abhängig? (*Abhängigkeit: Der Patient war wiederholt übermäßig abhängig von anderen*) (2, 1, 0)

105.* Haben Sie anderen wiederholt gestattet, Sie zu Handlungen zu zwingen, die Sie nicht tun wollten oder Sie grausam zu behandeln? Wurde Ihnen je gesagt, daß Sie anderen erlauben, Sie zum Opfer zu machen oder Sie zu mißbrauchen? (*Masochismus: Der Patient hat anderen wiederholt gestattet, ihn zu nötigen oder ihn zu verletzen*) (2, 1, 0)

106. S. 18 DER PATIENT HATTE IN ENGEN BEZIEHUNGEN IMMER WIEDER PROBLEME MIT ABHÄNGIGKEIT ODER MASOCHISMUS 2, 1, 0

107. Haben Sie häufiger die guten Eigenschaften von Menschen nicht wahrgenommen und nur ihre Fehler gesehen? Hat man Ihnen je gesagt, Sie seien eine sehr kritische oder abwertende Person? (*Abwertung: Der Patient hat wiederholt die Schwächen anderer übertrieben und ihre Stärken heruntergespielt*) (2, 1, 0)

108. Haben Sie wiederholt versucht, andere dazu zu bringen, zu tun, was Sie wollten, ohne sie wirklich zu bitten oder sie dazu aufzufordern? Hat man Ihnen jemals gesagt, daß Sie Menschen zu manipulieren versuchten? (*Manipulation: Der Patient hat wiederholt indirekte Mittel verwendet, um zu bekommen, was er wollte*) (2, 1, 0)

109. Haben Sie wiederholt versucht, andere zu zwingen, Dinge zu tun, die diese nicht tun wollten oder andere grausam behandelt? Hat man Ihnen jemals gesagt, daß Sie herrschsüchtig oder gemein seien? (*Sadismus: Der Patient hat wiederholt versucht andere zu nötigen oder zu verletzen*) (2, 1, 0)

110. S. 19 DER PATIENT HATTE IN ENGEN BEZIEHUNGEN WIEDERKEHRENDE PROBLEME MIT ABWERTUNG, MANIPULATION ODER SADISMUS 2, 1, 0

111. Haben Sie wiederholt Leute um Dinge gebeten, die diese Ihnen nicht geben konnten oder sollten? Viel von Ihrer Zeit und Aufmerksamkeit gefordert? Hat man Ihnen je gesagt, Sie seinen ein sehr fordernder Mensch? (*Forderungshaltung: Der Patient hat wiederholt unangemessene Forderungen gestellt*) (2, 1, 0)

112. Haben Sie sich wiederholt so verhalten, als hätten Sie ein Recht auf eine besondere Behandlung? Als ob Ihnen andere auf Grund dessen, was Sie durchgemacht haben, etwas schuldig seien? Ist Ihnen je gesagt worden, daß Sie sich so verhalten, als hätten Sie Anspruch auf besondere Fürsorge oder Rücksichtnahme? (*Anspruchshaltung: Der Patient hat wiederholt unrealistische Erwartungen gezeigt*) (2, 1, 0)

105.* Die Frage kann auch so gestellt werden:
"Wurde Ihnen je gesagt, daß Sie anderen erlauben, Sie zum Opfer zu machen oder Sie zu mißbrauchen oder hatten Sie den Eindruck, daß Sie das öfter tun? Lassen Sie es öfters zu, daß man Sie zu Handlungen zwingt, die Sie nicht tun wollten. Lassen Sie es zu, daß man Sie grausam behandelt?"

Hier ist eine Unterscheidung vom herkömmlichen Masochismus zu treffen, bei dem eine Verletzung, Erniedrigung etc. als lustvoll erlebt wird oder Bedingung der Lust ist. Sexueller Masochismus ist nicht hier, sondern unter Item Nr. 65 (Perversion) zu werten.

113. S. 20 DER PATIENT HATTE IN ENGEN BEZIEHUNGEN IMMER WIEDER PROBLEME MIT SEINER FORDERUNGS- ODER ANSPRUCHSHALTUNG 2, 1, 0

Schwierige therapeutische Beziehungen

114. Waren Sie in den letzten zwei Jahren in einer anderen Einzeltherapie? In wievielen? (*Anzahl der Einzeltherapien*) 2=2 oder mehr, 1=1, 0=keine). (2, 1, 0)

115. Wieviele Monate der letzten zwei Jahre sind Sie in Einzeltherapie gewesen? (*Zeit, die in Einzeltherapie verbracht wurde*) (2=12 Monate oder länger, 1=1-11 Monate, 0=keine Einzeltherapie) (2, 1, 0)

116. Ging es Ihnen als Ergebnis dieser anderen Einzeltherapie(n) schlechter als zuvor? Inwiefern? (*Regression in Einzeltherapie*) (2, 1, 0)

117. Waren Sie schon früher einmal in stationärer oder psychiatrischer Behandlung? Wie oft? (*Anzahl der psychiatrischen Hospitalisierungen*) (2=2 und mehr, 1=1 und 0=keine) (2, 1, 0)

118. Wieviele Monate sind Sie in den letzten zwei Jahren in stationärer psychiatrischer Behandlung gewesen? (Zeit, die in psychiatrischen Kliniken verbracht wurde) (2=12 Monate oder länger, 1=1-11 Monate, 0=keine stationäre Behandlung) (2, 1, 0)

119. Ging es Ihnen als Ergebnis eines solchen Klinikaufenthaltes sehr viel schlechter als zuvor? Inwiefern? (*Regression*) (2, 1, 0)

120. S. 21 DER PATIENT ZEIGTE WÄHREND DER THERAPIE ODER DER PSYCHIATRISCHEN HOSPITALISIERUNG EINE DEUTLICHE REGRESSION 2, 1, 0

121. Waren Sie auf einer psychiatrischen Station schon einmal Mittelpunkt von Teamkonflikten oder -problemen? (*Beurteilen Sie, ob das Stationspersonal a.f den Patienten mit einer auffallenden Gegenübertragungsreaktion reagierte. Andere Quellen sollten, soweit verfügbar, für diese Beurteilung herangezogen werden*) (2, 1, 0)

122. Hatten Sie einen Therapeuten, der sehr wütend auf Sie wurde? Hat er Sie aufgefordert, die Behandlung zu beenden? Hat er sich sehr viel stärker um Sie gekümmert als andere Therapeuten? (Diese Fragen gelten auch für Therapeutinnen) (*Urteilen Sie, ob der Patient Gegenstand einer auffallenden Gegenübertragungsreaktion des Therapeuten oder der Therapeutin war. Andere Quellen sollten, soweit verfügbar, für diese Beurteilung mit herangezogen werden.)* (2, 1, 0)

123. Haben sie eine enge Freundschaft oder eine Liebesaffäre mit einem Mitglied des Stationspersonals entwickelt? (2, 1, 0)

124. Oder mit einem Therapeuten (einer Therapeutin)? (2, 1, 0)

125. S. 22 DER PATIENT HAT AUF DER PSYCHIATRISCHEN STATION ODER IN EINER PSYCHOTHERAPIE, AUFFALLENDE GEGENÜBERTRAGUNGSREAKTIONEN AUSGELÖST ODER IST MIT EINEM PROFESSIONELLEN HELFER EINE GANZ BESONDERE BEZIEHUNG EINGEGANGEN 2, 1, 0

Andere Merkmale

125.1* Neigten Sie dazu, Leute auf eine Podest zu stellen und nur ihre guten Seiten zu sehen? Betrachteten Sie sie als ungewöhnlich gut? Perfekt? Unzerstörbar? Hat man Ihnen jemals gesagt, daß Sie es offenbar schwer haben, anderer Leute Fehler zu sehen? (*Idealisierung: Der Patient hat wiederholt die Stärken anderer übertrieben und ihre Schwächen heruntergespielt*) (2, 1, 0)

125.2.* Wechselten Sie Ihre Meinung über Menschen oft von einem Tag zum anderen (so daß Sie sie z.B. an einem Tag mochten und am nächsten Tag nicht ausstehen konnten)? (*Auffallende Einstellungsverschiebungen*) (2, 1, 0)

125.3.* Waren Sie die meiste Zeit über außergewöhnlich empfindlich gegenüber Kritik? Haben Sie sich oft kritisiert gefühlt, wenn dies wahrscheinlich gar nicht der Fall war? (*Überempfindlichkeit gegenüber Kritik*) (2, 1, 0)

125.4.* Waren Sie oft unsicher, wer Sie sind oder wie Sie wirklich sind? Hatten Sie oft das Gefühl, keine Identität zu haben? Waren Sie sich oft über Ihre Werte oder Ziele im unklaren? Haben Sie Ihre Werte oder Ziele häufig gewechselt? Waren Sie sich oft unsicher, aus welchen Menschen Sie sich wirklich etwas machen oder wem gegenüber Sie loyal sein sollten? Wechselten Sie öfters von einer Gruppe von Freunden zur nächsten? Waren Sie sich Ihrer sexuellen Identität oft unsicher? Gab es Zeiten, wo Sie ein Mann sein wollten, und Zeiten, in denen Sie lieber eine Frau gewesen wären? (*Ernsthafte Identitätsstörungen in zwei oder mehr Bereichen*) (2, 1, 0)

125.5.* Stellten Sie in den letzten zwei Jahren oft fest, daß Ihre Stimmung innerhalb nur weniger Stunden oder Tage von Depression über Ärger zu Angst wechselte? Waren Sie ein sehr launischer Mensch? (Affektive Instabilität) (2, 1, 0)

125.1* - 125.4* Die in [Klammern] gesetzten Fragen sind zu Forschungszwecken aus einer früheren Version des DIB übernommen. Die Antworten gehten nicht in die Wertung des DIB's ein.

126.* BEZIEHUNGS-SECTION SCORE: _____

Skalierter Beziehungs-Section Score:

3 wenn der Section Score 9 oder mehr beträgt
2 wenn der Section Score 6 bis 8 beträgt
0 wenn der Section Score 5 oder weniger ausmacht,
oder wenn der Patient ein absonderlicher, sozial isolierter Einzelgänger war

127. SKALIERTER BEZIEHUNGS-SECTION SCORE: _____

ERGEBNIS:

BEREICH "AFFEKTIVITÄT":

1. Affekt-Section Score: 0 - 10 _____
2. Skalierter Affekt-Section Score: 0 - 2 _____

BEREICH "KOGNITION":

3. Kognitions-Section Score: 0 - 6 _____
4. Skalierter Kognitions-Section Score: 0 - 2 _____

BEREICH "IMPULSIVE HANDLUNGSMUSTER":

5. Impulshandlungs-Section Score: 0 - 10 _____
6. Skalierter Impulshandlungs-Section Score: 0 - 3 _____

BEREICH "INTERPERSONALE BEZIEHUNGEN":

7. Beziehungs-Section Score: 0 - 18 _____
8. Skalierter Beziehungs-Section Score: 0 - 3 _____

9. **DIB-R-GESAMT-SCORE:** -------

Sicherheit der Einschätzung:

1 = sehr sicher
2 = ziemlich sicher
3 = ziemlich unsicher
4 = sehr unsicher

Vollständigkeit der Information, bzw. Güte der Exploration

1 = sehr gut
2 = gut
3 = befriedigend
4 = mangelhaft

Sonstige Bemerkungen:

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

126.* In dem Beziehungs-Section Score sind nur die Einschätzungen "0, 2, 3" möglich. Es ist keine "1" vorgesehen.