

# Stabilität und Validität latenter Klassen im Beck-Depressionsinventar (BDI)

F. Keller, W. Kempf, Ravensburg-Weissenau/Ulm und Konstanz

## Zusammenfassung

Die Replikation einer Rasch-Analyse des BDI, bei der drei homogene Subskalen identifiziert wurden, erbrachte für die deutsche Version zwar keine Homogenität, jedoch insgesamt zufriedenstellende psychometrische Eigenschaften. In diesem Beitrag wird anhand von  $n = 714$  Fragebogen depressiver Patienten untersucht, inwieweit die gefundenen Symptommuster (latente Klassen) replizierbar sind, wenn Geschlecht und Testzeitpunkt als Kovariate hinzugenommen werden. Es zeigte sich eine fast völlige Unabhängigkeit vom Geschlecht; der Einfluß des Testzeitpunkts auf die Replizierbarkeit der Symptommuster war ebenfalls minimal. Angesichts der guten Ergebnisse für die Subskalen sollten diese Verwendung finden, zumal an der Unidimensionalität der Gesamtskala gezweifelt werden muß.

## Stability and Validity of Latent Classes in the Beck Depression Inventory (BDI)

Previous analyses of the BDI revealed that the BDI does not, in terms of the Rasch measurement model, constitute a unidimensional scale for the measurement of depression in clinical samples. Three subscales, however, turned out to have sufficient psychometric properties, as examined by latent class analysis (LCA). The stability and validity of these LCA solutions according to gender and time of assessment (admission vs. discharge/follow-up) were analysed at  $n = 714$  BDIs of depressive inpatients. Results show that the identification of the latent classes is practically independent of gender and time of assessment. Thus, the use of the subscale scores may be considered helpful for improving discriminant validity and is recommended in further research studies.

## Einleitung

Unter den Selbstbeurteilungsskalen zur Erfassung depressiver Symptomatik ist das Beck-Depressionsinventar (BDI) eines der meistverwendeten und in Therapiestudien zu depressiven Erkrankungen ist es inzwischen nahezu unerlässlich (2). Entsprechend zahlreich wird es in der Fachliteratur zitiert, jedoch vornehmlich in

seiner praktischen Anwendung als Testinstrument. Hinsichtlich der psychometrischen Eigenschaften des BDI ist es dagegen weit weniger bekannt. Bislang wurde nur eine geringe Zahl klinischer Stichproben daraufhin untersucht und es stellt sich vor allem die Frage, ob das BDI als unidimensionale Skala aufgefaßt werden kann, die ein einheitliches Konzept von Depression erfaßt. Nur dann ist eine Verwendung des Gesamtwertes aus den 21 Einzelitems als Indikator für die Schwere einer Depression sinnvoll.

Das Rasch-Modell (RM) leistet die in methodologischer Hinsicht wohl stringenteste Formalisierung der Unidimensionalitätsannahme. Homogenität eines Tests im Sinne des RM ist gleichbedeutend damit, daß die einzelnen Itemantworten ohne statistischen Informationsverlust zu einem Gesamtscore addiert werden können, so daß alle Versuchspersonen unabhängig von Diagnose und Ausprägungsgrad der gemessenen Eigenschaft (hier: Depressivität) auf derselben latenten Skala miteinander verglichen werden können.

Eine weitere Eigenschaft des RM ist die Unabhängigkeit von Items und Personen. Abb. 1 enthält die Items des BDI (Traurigkeit = Nr. 1) sowie die Anzahl der Nennungen, mit der die vier Kategorien jedes Items angeführt wurden. Der relative Vergleich zeigt, daß z. B. Entschluß- und Arbeitsunfähigkeit bald angegeben werden („leichte“ Items), während Appetit- und Gewichtsverlust „schwierige“ Items sind. Bei Gültigkeit des RM ist der Schwierigkeitsgrad jedes Items für alle Versuchspersonen gleich.

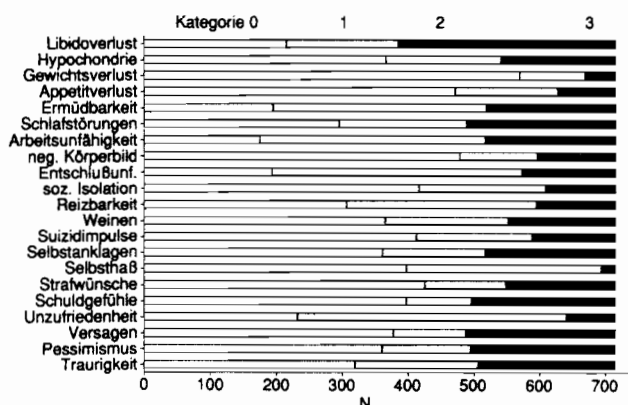


Abb. 1 Häufigkeiten in den Antwortkategorien der einzelnen BDI-Items für alle  $n = 714$  Fragebogen (im BDI ist jede Kategorie symptom-spezifisch ausformuliert, wobei 0 keine Symptomatik bedeutet).

Während faktorenanalytische Untersuchungen nahelegten, nur von einem generellen Faktor und damit einem eindimensionalen Konzept auszugehen (2) (die gefundenen Unstimmigkeiten werden auf die verschiedenen Auswertungsstrategien und Stichprobenunterschiede zurückgeführt), führte die von *Bouman* u. *Kok* (1) auf Grundlage der Daten von  $n = 389$  psychiatrischer Patienten vorgenommene Itemanalyse des BDI mittels des dichotomen Testmodells von *Rasch* (4) dagegen zu einer eindeutigen Zurückweisung der Unidimensionalitätsannahme. Danach kann im Falle des BDI nicht von einer homogenen Gesamtskala ausgegangen werden. Die Autoren identifizierten vielmehr drei verschiedene, in sich homogene Subskalen des BDI, die nur mäßig untereinander und nicht signifikant mit Alter, Geschlecht oder Bildung korrelierten.

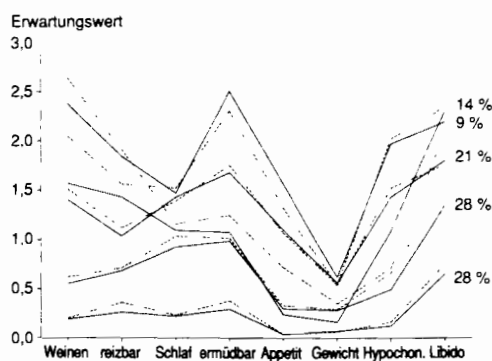
Eine Studie zur Replizierbarkeit dieser Ergebnisse für die deutsche Version des BDI sowie einige darüber hinausgehende Analysen (polytome Analyse und Latent-Class-Analyse [LCA]), die an  $n = 714$  Fragebogen depressiver Patienten durchgeführt wurde, bilden die Grundlage dieses Beitrags, in dem speziell auf die Validitätsaspekte eingegangen wird. Eine kurze Beschreibung der LCA findet sich im Methodenteil; zur vollständigen Darstellung der Modellgleichungen und aller Einzelergebnisse verweisen wir auf *Kempf* u. *Keller* (3).

Die Ergebnisse der Replikation lassen sich dahingehend zusammenfassen, daß jede der drei Subskalen bei dichotomierter Itemskalierung eine quantitative Differenzierung der Probanden in den jeweils erfaßten Symptombereichen zu leisten vermag. Alle drei Skalen erfüllen jedoch (im Unterschied zu [1]) nicht die strengen Homogenitätsvoraussetzungen des RM.

Die polytomen Antwortkategorien sind zwar durchaus geeignet, um eine gegenüber der dichotomen Auswertung verfeinerte quantitative Differenzierung zwischen den Probanden zu leisten. Andererseits kommen in den polytomen Antwortkategorien aber auch qualitative Unterschiede zwischen den Probanden zum Ausdruck, die dazu führen, daß die Rasch-Homogenität der Skalen verletzt ist.

Die Ergebnisse der LCA, die schwächere Modellannahmen macht (s.u.), zeigen, daß die Unterschiede zwischen den latenten Klassen bei den Subskalen G & F und M & I in erster Linie quantitativer Art sind (zur Zusammensetzung der Subskalen siehe Methodenteil). Bei der Skala SOM kommen qualitative Unterschiede zwischen den Probanden vermehrt zum Tragen, so daß eine Klasse, die immerhin 14 % der Stichprobe umfaßt, einen qualitativ verschiedenen Profilverlauf zeigt (vgl. Abb. 2).

Zusammenfassend kann dennoch gesagt werden, daß die psychometrischen Eigenschaften der Subskalen G & F, M & I und SOM des BDI zufriedenstellend sind. Obwohl die Homogenitätsvoraussetzungen des Rasch-Modells nicht erfüllt sind, leisten die Skalen doch vor allem eine quantitative Differenzierung zwischen den Probanden, so daß der mit der Scorebildung einhergehende Informationsverlust als geringfügig einzuschätzen ist.



**Abb. 2** Polytome LCA der Subskala SOM. Vergleich der Erwartungswerte der polytomen Antwortvariablen mit (---) und ohne (—) Geschlecht als Kovariater.

Im Anschluß an diese Untersuchungen zu psychometrischen Eigenschaften stellt sich die Frage, ob die identifizierten latenten Klassen durch die Hinzunahme von Kovariaten verändert werden.

### Methodik

Als Datengrundlage standen insgesamt  $n = 767$  Fragebogen zur Verfügung, von denen die  $n = 714$  vollständig ausgefüllten in die Analysen eingingen. Die Patienten wurden auf den Depressionsstationen des PLK Reichenau und Weissenau behandelt und wiesen als primäre psychiatrische Diagnose nach ICD-9 eine Depression auf. Die Aufteilung der Fragebogen nach Untersuchungszeitpunkt betrug  $n = 302$  bei Aufnahme (in der Regel innerhalb der ersten Woche),  $n = 279$  bei Entlassung und  $n = 133$  zum Katamnesezeitpunkt (ein Jahr nach Entlassung). Eine Aufteilung nach Geschlecht ergab, daß 70,6 % der Fragebogen von Frauen ausgefüllt waren, was in etwa dem oft gefundenen 2:1-Verhältnis entspricht.

Die von *Bouman* u. *Kok* (1) identifizierten homogenen Subskalen, die auch in dieser Arbeit verwendet wurden, sind:

1. "Guilt and Failure" (G & F): Items 3, 5, 7, 8
2. "Mood and Inhibition" (M & I): Items 1, 2, 9, 13, 15
3. "Somatic" (SOM): Items 10, 11, 16–21

Die Items 4, 6, 12, 14 werden keiner Subskala zugeordnet.

Die auf *Lazarsfeld* zurückgehende Latent-Class-Analyse (LCA) enthält eine Abschwächung der Annahmen des Rasch-Modells. Während im RM die Probanden auf einer latenten Dimension (z. B. „Depressivität“) quantitativ anordenbar sein sollen, wird in der LCA lediglich angenommen, daß jeder Proband einer latenten Klasse (Subpopulation) angehört, die durch die unterschiedlichen Wahrscheinlichkeiten ( $W_k$ ) charakterisiert sind, mit welchen die verschiedenen Items positiv beantwortet werden. In der LCA werden die Anzahl latenter Klassen und deren  $W_k$  geschätzt (mittels des Programms LACORD [5, 6]). Durch Verbinden der erwarteten Werte innerhalb jeder Klasse enthält man Symptommuster (z. B. eine sym-

ptomarme Klasse mit 28 % in Abb. 2). Sollen die Unterschiede zwischen den latenten Klassen quantifizierbar sein, so dürfen sich die Profillinien der Erwartungswerte nicht überschneiden. Andernfalls ist von qualitativen Unterschieden zwischen den Klassen auszugehen.

### Ergebnisse

Um Anhaltspunkte bezüglich der Validität der mittels polytomer LCA identifizierten Klassen zu gewinnen, wurde untersucht, ob sich diese latenten Klassen in gleicher Weise reproduzieren lassen, wenn das *Geschlecht* als zusätzliche Kovariate in die Berechnung der LCA einbezogen wird. Dabei interessiert sowohl, ob die Größe der einzelnen Klassen und die Profillinien stabil bleiben, als auch die Aufteilung von Männern und Frauen innerhalb der Klassen, die (bei Unabhängigkeit) in jeder Klasse gleich sein müßte wie das Geschlechterverhältnis insgesamt.

Tab. 1 zeigt, daß die ursprünglichen Klassengrößen bei Einbeziehung des Geschlechts als Kovariate bei den Subskalen G & F und M & I beinahe exakt reproduziert werden. Bei der Subskala SOM ist die Reproduktion der Klassengrößen immer noch gut, wobei allerdings die symptomarmen Klassen 1 und 2 hier stärker besetzt sind als in der LCA-Lösung ohne Kovariate.

Zugleich werden für die Subskalen G & F und M & I auch die Erwartungswerte der Antwortvariablen hervorragend reproduziert, so daß die Profile beinahe deckungsgleich sind. In der Subskala SOM sind die Profile ebenfalls sehr ähnlich, lediglich die qualitativ andere Klasse mit 14 % zeigt etwas höhere Differenzen (vgl. Abb. 2 mit dem Vergleich zwischen ursprünglicher LCA-Lösung [durchgezogene Linien] und der LCA-Lösung mit der Kovariaten Geschlecht [gestrichelt]). Die Vergleichbarkeit von männlichen und weiblichen Probanden auf Grundlage derselben latenten Klassen kann somit als sichergestellt gelten.

Wie aus Tab. 2 hervorgeht, sind die latenten Klassen bei der Subskala M & I weitgehend geschlechtsunabhängig, während Frauen bei der Subskala G & F in der symptomreichen Klasse fünf unterrepräsentiert und bei der Subskala SOM in der symptomreichen Klasse vier überrepräsentiert sind. In beiden Fällen handelt es sich jedoch um die mit einer Klassengröße von 2,6 % (G & F) bzw. 6,4 % (SOM) jeweils kleinsten Klassen (vgl. Tab. 1), so daß die Kovariation als geringfügig anzusehen ist.

Weiter interessiert, ob die latenten Klassen bezüglich des *Testzeitpunkts* invariant sind, so daß sie (unabhängig von der Symptomatik) eine Vergleichbarkeit von Fragebogenergebnissen erlauben, welche bei der Aufnahme, bei der Entlassung aus der Klinik oder zum Zeitpunkt der Katamnese (ein Jahr danach) erhoben wurden.

Zu diesem Zweck wurde eine nach Aufnahme versus Entlassung/Katamnese dichotomisierte Variable als Kovariate in die polytome LCA einbezogen. Die Ergebnisse, die ebenfalls in Tab. 1 aufgeführt sind, zeigen, daß die Klassengrößen bei Einbeziehung des so operatio-

**Tab. 1** Veränderung der Klassengrößen der 5-Klassen-Lösung der polytomen LCA bei Hinzunahme der Kovariaten Geschlecht und Zeitpunkt.

Skala	Kovariate	Klasse				
		1	2	3	4	5
G & F	ohne	47,2 %	19,8 %	13,8 %	16,0 %	3,2 %
	Geschlecht	48,5 %	19,1 %	13,1 %	16,7 %	2,6 %
	Zeitpunkt	44,6 %	25,1 %	10,3 %	16,3 %	3,7 %
M & I	ohne	25,0 %	24,2 %	22,0 %	16,7 %	12,1 %
	Geschlecht	24,5 %	25,4 %	21,1 %	16,5 %	12,6 %
	Zeitpunkt	34,8 %	33,9 %	16,6 %	10,7 %	4,0 %
SOM	ohne	27,9 %	28,7 %	20,8 %	8,8 %	13,8 %
	Geschlecht	31,4 %	35,0 %	15,6 %	6,4 %	11,6 %
	Zeitpunkt	26,3 %	29,0 %	22,2 %	7,0 %	15,6 %

**Tab. 2** Anteil der Frauen in den latenten Klassen (5-Klassen-Lösung mit Geschlecht als Kovariate). Anteil Frauen gesamt = 70,6 %.

Skala	Klasse				
	1	2	3	4	5
G & F	71,9 %	72,0 %	59,5 %	76,1 %	56,7 %
M & I	73,9 %	69,3 %	68,3 %	67,4 %	74,8 %
SOM	72,8 %	65,2 %	78,3 %	86,5 %	61,5 %

**Tab. 3** Anteil der Entlassungs- und Katamnesefragebogen in den latenten Klassen (5-Klassen-Lösung mit Testzeitpunkt als Kovariate). Anteil Entlassungs-/Katamnesebogen gesamt = 57,7 %.

Skala	Klasse				
	1	2	3	4	5
G & F	80,7 %	35,4 %	61,8 %	38,6 %	4,9 %
M & I	88,6 %	57,9 %	27,3 %	19,3 %	15,4 %
SOM	95,7 %	66,3 %	22,2 %	8,6 %	50,1 %

nalisierten Testzeitpunkts vor allem bei den Subskalen G & F sowie SOM sehr gut reproduziert werden. Bei der Subskala M & I gelingt die Reproduktion der Klassengrößen dagegen deutlich schlechter, aber doch hinreichend gut, um von einer Vergleichbarkeit der Testzeitpunkte mittels der latenten Klassen sprechen zu können, zumal auch die Erwartungswerte der Antwortvariablen recht gut reproduziert werden, wie aus der guten Übereinstimmung der Profile hervorgeht.

Von besonderem Interesse für die Validität der latenten Klassen ist deren Kovarianz mit dem Testzeitpunkt. Wie Tab. 3 zeigt, sind die Entlassungs- und Katamnesefragebogen bei der Subskala G & F in der symptomarmen Klasse 1 (wo sie 80,7 % aller Fragebogen dieser Klasse ausmachen) deutlich überrepräsentiert, während sie in der symptomreichen Klasse 5 (mit 4,9 % der Fragebogen) fast überhaupt nicht vertreten sind.

Noch deutlicher ist der Zusammenhang zwischen Symptomreichtum und Testzeitpunkt bei der Subskala M & I, wo die Entlassungs- und Katamnesefragebogen 88,6 % der Fragebogen ausmachen, welche in die symptomarme Klasse 1 fallen und mit zunehmendem

Symptomreichtum einer Klasse immer seltener vorkommen.

Dasselbe Bild zeigt sich auch für die Subskala SOM, wo die Entlassungs- und Katamnesefragebogen sogar 95,7% der Fragebogen ausmachen, welche in die symptomarme Klasse 1 fallen. In der symptomreichen Klasse 4 sind sie mit 8,6% der Fragebogen deutlich unterrepräsentiert. In der von den anderen Klassen qualitativ unterschiedenen Klasse 5 entspricht der Anteil der Entlassungs- und Katamnesefragebogen mit 50,1% in etwa deren Anteil an der Gesamtstichprobe (57,7%).

### Diskussion

Die mittels LCA identifizierten Symptommuster, die sich in erster Linie quantitativ bezüglich ihres Symptomreichtums unterscheiden (3), sind auch gegenüber dem Geschlecht der Probanden und dem Testzeitpunkt (Aufnahme vs. Entlassungs/Katamnese) invariant. Das Geschlechterverhältnis innerhalb jeder Klasse ist nahezu identisch und bezüglich des Testzeitpunkts ergibt sich (wie erwartet), daß Symptomreichtum einer Klasse und Aufnahmebogen sowie symptomarme Klassen und Entlassungs-/Katamnesefragebogen zusammenhängen.

Trotz dieser guten Ergebnisse für die Subskala bleibt freilich festzuhalten, daß die Unidimensionalität des gesamten BDI damit nicht gesichert ist. Die Rasch-Analyse und LCA einer Subskala aus sieben depressionstypischen Items, die den drei Subskalen entnommen sind, erbrachte im Gegenteil in ersten Auswertungen qualitativ verschiedene Symptommuster, was im Einklang steht mit den von (1) gefundenen geringen Interkorrelationen der Subskalen. Neben der Verwendung des Gesamtwertes sollten daher in Korrelations- und Verlaufsstudien auch die Werte der drei Subskalen Verwendung finden.

### Danksagung

Wir danken allen an der Datenerhebung Beteiligten, u. a. den Oberärzten der Depressionsstationen, Dr. H. Pecher (Reichenau) und PD Dr. M. Wolfersdorf (Weissenau) sowie besonders Frau Mica Endo für ihre wertvolle und angenehme Mithilfe.

### Literatur

- <sup>1</sup> Bouman, T. K., A. R. Kok: Homogeneity of Beck's Depression Inventory (BDI): Applying Rasch Analysis in Conceptual Exploration. *Acta Psychiatrica Scandinavica* 76 (1987) 568–573
- <sup>2</sup> Hautzinger, M.: Das Beck-Depressionsinventar (BDI) in der Klinik. *Nervenarzt* 62 (1991) 689–696
- <sup>3</sup> Kempf, W., F. Keller: Psychometrische Analysen des Beck-Depressionsinventars (BDI). Forschungsbericht PLK Weissenau, Ravensburg (1992)
- <sup>4</sup> Rasch, G.: Probabilistic Models for Some Intelligence and Attainment Tests. Danmarks Paedagogiske Institut, Copenhagen (1960)
- <sup>5</sup> Rost, J.: Quantitative und qualitative probabilistische Testtheorie. Huber, Bern (1988)
- <sup>6</sup> Rost, J.: LACORD. Latent class analysis for ordinal variables. A FORTRAN program. IPN-Institut, Kiel (1988)

*Dr. Ferdinand Keller, Dipl.-Psych.*

AG Verlaufsforschung  
Bereich Akutpsychiatrie II/Depression  
PLK Weissenau  
7980 Ravensburg

*Prof. Dr. W. Kempf-Palmbach, Dipl.-Psych.*

Universität Konstanz  
Fakultät für Wirtschaftswissenschaften und Statistik  
Postfach 5560  
7750 Konstanz