

TBS-TK
Rezension

Beck Depressions-Inventar (BDI-II). Revision

P. Y. HERZBERG & S. GOLDSCHMIDT,
Medizinische Psychologie der Universität Leipzig
N. HEINRICHS, Psychologisches Institut der Universität Bielefeld

Allgemeine Informationen über den Test, Beschreibung des Tests und seiner diagnostischen Zielsetzung

Das BDI-II dient der Erfassung der Schwere depressiver Beschwerden anhand von 21 Items, die als Selbstbeurteilungsinstrument zusammengestellt worden sind. Die Informationen zu den Altersgruppen sind in den Verfahrenshinweisen eingeschränkt. Als Anwendungsbereich werden Jugendliche ab 13 Jahren und Erwachsene angegeben. Bei Letzteren gibt es keine Information zu einer möglichen (oberen) Altersgrenze. Der Informationsgehalt der Darstellung von empirischen Untersuchungen ist insofern reduziert, als die 14 Evaluationsstichproben zwar kurz beschrieben werden, allerdings ohne Jahresangaben bzw. zeitliche Angaben der Erhebungen. Zudem finden sich dazu nur unvollständige deskriptive Stichprobenmerkmale. Die Durchführungszeit ist in den Verfahrenshinweisen angegeben, eine Angabe zur Auswertung fehlt zwar, kann aber aufgrund der Einfachheit vom Anwender abgeleitet werden. Bei dem berechneten reliablen Veränderungswert (unter Normwerte, S. 30) ist leider nicht nachvollziehbar, auf welche Stichprobe(n) sich dieser bezieht.

Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion

Die ursprüngliche Version des Tests beruhte auf klinischen Symptomen und war keiner Theorie der Depression verpflichtet. Als (theoretische) Grundlage werden in den Verfahrenshinweisen der Revision vor allem die Angleichung der Items an die Depressionskriterien des Diagnostischen und Statistischen Manuals in der vierten Revision (DSM-IV) benannt. Darüber hinaus sollen Items aktualisiert werden und auch solche Symptome mit einbezogen werden, die für eine schwerere Depression sprechen bzw. zu einer stationären Aufnahme führen können. Die Änderungen des Verfahrens werden diesbezüglich ausreichend beschrieben. In der Darstellung der Entstehungsgeschichte finden sich hingegen keine weiteren Hinweise auf die theoretischen Grundlagen der ursprünglichen Testkonstruktion (zum Beispiel, wie es zu der finalen Auswahl der 21 Items kam).

Objektivität

Die Objektivität ist hinsichtlich der Durchführung und Interpretation gegeben. Instruktionen werden standardisiert vorgegeben, und es liegen ausreichend detaillierte Informationen zur Interpretation in gut verständlicher Form vor. Es fehlen Angaben zur quantitativen Analyse der Objektivität. Einschränkungen bestehen hinsichtlich

der Auswertungsobjektivität. Die Verfahrenshinweise enthalten keine Informationen darüber, wie mit fehlenden, d. h. nicht bearbeiteten Items umgegangen werden soll. Es bleibt unklar, welcher Testwert für ein nicht bearbeitetes Item gegeben werden soll und bis zu welcher Anzahl von nicht bearbeiteten Items das Testergebnis noch interpretiert werden darf. Diese Informationen sind allgemein wichtig, besonders jedoch bei dem vorliegenden Testverfahren, das sich auf Personen bezieht, die möglicherweise an einer schweren Depression und damit verbundenen Konzentrationsschwierigkeiten leiden.

Normierung (Eichung)

Die Normen werden als Prozentränge für depressive und gesunde Personen angegeben. Beide Normstichproben setzen sich aus mehreren Stichproben von depressiven Patienten sowie aus mehreren Stichproben Gesunder zusammen. Unklar ist die Repräsentativität dieser Stichproben; der jeweilige Stichprobenumfang für die Normen muss von S. 19 f. erschlossen werden. Auf alters-, bildungs- und geschlechtsgetrennte Normen wird verzichtet, da für Alter und Bildung keine signifikanten Zusammenhänge mit dem Summenwert berichtet werden. Allerdings gibt es signifikante Zusammenhänge zwischen Geschlecht und BDI-II-Score bei den Gesunden, in beiden Stichproben weisen Frauen höhere Depressionswerte auf. Die Entscheidung, für die Stichprobe Gesunder keine geschlechtsgetrennten Normen zu berechnen, wäre besser nachzuvollziehen, wenn Effektstärken angegeben wären. Unsere Berechnung der Effektstärke ergibt $d = 0,15$, aufgrund der geringen praktischen Bedeutung kann auf geschlechtsgetrennte Normen verzichtet werden.

Zuverlässigkeit (Reliabilität, Messgenauigkeit)

Die Reliabilität des BDI-II wird in Form interner Konsistenzen (Cronbachs Alpha) berichtet und liegt je nach Stichprobe zwischen 0,89 und 0,93. Die auf Basis des Partial-Credit-Modells berechnete Reliabilität in der Stichprobe depressiver Patienten wird mit 0,92 angegeben. Für die Gesunden liegen die Reliabilitäten bei 0,80 bzw. 0,82. Für die IRT-Analysen fehlt die Angabe des Standardmessfehlers. Die Retestreliabilität beträgt für Gesunde über drei Wochen als auch über fünf Monate 0,78 und ist damit als hinreichend stabil anzusehen. Erwartungsgemäß liegt die Retestreliabilität für behandelte Patienten niedriger (0,46). Die Änderungssensitivität des BDI-II ist nicht hinreichend untersucht worden. Es wird der reliable Veränderungsindex angegeben, nach dem im Einzelfall ein Veränderungswert von 8 statistisch signifikant ist. Obgleich das BDI(-II) nicht für den Einsatz als

Beck Depressions- Inventar (BDI-II). Revision	Die TBS-TK-Anforderungen sind erfüllt			
	voll	weit- gehend	teil- weise	nicht
Allgemeine Informa- tionen, Beschreibung und diagnostische Zielsetzung		●		
Objektivität		●		
Zuverlässigkeit		●		
Validität	●			

Screeningverfahren, sondern zur Messung der Schwere einer Depression entwickelt wurde, wird es in Forschung und Praxis – entgegen dem von den Autoren empfohlenen Einsatzbereich – häufig als Screeninginstrument eingesetzt. Daher sollten auch für die deutsche Version im Hinblick auf eine kriteriumsorientierte Validierung Angaben zur Sensitivität und Spezifität in die Verfahrenshinweise aufgenommen werden.

Gültigkeit (Validität)

Die Inhaltsvalidität wird aus der Abdeckung der Kriterien des DSM-IV abgeleitet. Die Transitivität, d.h. die Ordnung der Antwortkategorien, ist bei 9 der 21 Items verletzt. Während die konvergente Validität mit anderen Selbstbeurteilungsverfahren gleichen Gültigkeitsanspruches belegt wird, ist die mangelnde Differenzierung zu Ängsten kritisch zu bewerten. Die Kriteriumsvalidität zeigt, wie schwierig eine deutliche Abgrenzung von solchen verwandten (aber konzeptuell dennoch anderen) Konstrukten, wie z.B. auch Lebensqualität mit Korrelationen zwischen 0,50 und 0,75, ist. Die den depressiven Symptomen zugrunde liegende negative Affektivität, die sich auch in den anderen Konstrukten wiederfindet, erklärt die zum Teil erheblichen Überlappungen von bis zu 50%. Besonders positiv hervorzuheben sind die überzeugende Differenzierungsfähigkeit des BDI-II und die international ähnliche Faktorenstruktur (faktorenanalytische Validität), auch wenn eine Arte-legis-Prüfung der Dimensionalität mittels Item-Response-Analyse oder konfirmatorischer Faktorenanalyse wünschenswert wäre.

Weitere Gütekriterien (Störanfälligkeit, Unverfälschbarkeit und Skalierung)

Wie alle Selbstbeurteilungsverfahren ist auch der BDI-II grundsätzlich verfälschbar. Studien zur Verfälschbarkeit werden in den Verfahrenshinweisen nicht berichtet. Es wird aber auf die potenzielle Möglichkeit von Simulations- oder auch Dissimulationstendenzen hingewiesen. Maßnahmen, diesen Tendenzen entgegenzuwirken, werden nicht aufgeführt (mit Ausnahme der Wieder-vorlage des ausgefüllten Fragebogens bei kontinuierlich extremen Antworten mit der Bitte, diese Antworten nochmals zu überdenken, da es unwahrscheinlich sei, dass alle Symptome in gleicher Stärke auftreten würden). Andere Gütekriterien, die neben den Hauptgütekriterien auch eine Rolle spielen können, werden nicht erwähnt (wie z.B. Fairness).

Abschlussbewertung/Empfehlung

Das BDI-II ist ein empfehlenswertes klinisches Testverfahren. Optimierbar im Manual wären (1) eine Angabe zu der Teilnahmequote bei den zugrunde liegenden empirischen Studien und damit verbunden Aussagen zur Repräsentativität der gezogenen Stichprobe für die Population, über die Aussagen gemacht werden sollen, (2) Hinweise auf die Übertragbarkeit der Normwerte von einer Vorgabeart zur anderen (z.B. von Papier/Bleistift zu computerisiert/Internet oder umgekehrt) und (3) die Angabe ROC-basierter Cut-off-Werte zur Einschätzung der Sensitivität und Spezifität für die deutsche Version des BDI-II. Bei einer Überarbeitung sollten diese und folgende Kritikpunkte berücksichtigt werden: Einige Items sind sprachlich unpräzise (z.B. Item 6, 16), sie weisen eine fehlende Transitivität und auch niedrige Itemtrennschärfen auf. In der Stichprobe der Patienten hatte Item 16 bereits im BDI eine sehr niedrige Trennschärfe, im BDI-II fällt sie ebenfalls gering aus, sodass die sprachliche Überarbeitung nicht erfolgreich war. Für das Rasch-Modell fehlen Angaben zur Anpassungsgüte, dadurch kann die Frage der Transitivität der Antwortkategorien und der Dimensionalität nicht beantwortet werden. Bei den Normwerten wäre eine T-Wert-Verteilung wünschenswert. Die Verfahrenshinweise sollten primär unter inhaltlichen, allerdings auch unter optischen Aspekten anwendungsfreundlicher gestaltet werden. Sie enthalten Flüchtigkeitsfehler wie fehlende Buchstaben und Wörter im Text sowie Ungenauigkeiten zwischen Textinhalten und Tabellen. Abbildung 5 ist nicht exakt dargestellt. Ein Veränderungswert von acht Punkten (Prä/Post) liegt innerhalb des schraffierten Bandes, was inhaltlich falsch ist und zu Fehlentscheidungen führen kann. Zudem stimmt die Bezeichnung dieser Abbildung nicht. Das BDI-II ist ein weitestgehend objektives, reliables und sehr valides Verfahren zur Erfassung depressiver Beschwerden, welches durch Berücksichtigung obiger Punkte noch verbessert werden könnte.

Diese Testrezension wurde im Auftrag des Testkuratoriums der Föderation deutscher Psychologengvereinigungen (DGPs und BDP) gemäß den TBS-TK-Richtlinien (Testkuratorium, 2006,2007) erstellt.

Testkuratorium (2006). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologengvereinigungen. Report Psychologie, 31, 492-499.

Testkuratorium (2007). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologengvereinigungen. Psychologische Rundschau, 58, 25-30.

TESTINFORMATIONEN

Hautzinger, M., Keller, F., & Kühner, C. (2006). Beck Depressions-Inventar (BDI-II). Revision. Frankfurt/Main: Harcourt Test Services. Deutsche Bearbeitung von Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck Depression Inventory-II (BDI-II). San Antonio, TX: Harcourt Assessment Inc.

Bezugsquelle: Harcourt Test Services GmbH, Baseler Str. 35-37, 60329 Frankfurt/Main
Manual 25 Euro, Fragebogen (50) 35 Euro.