



# Rezension von *Occupational Burnout Inventory (OBOI)*

R. Hossiep und R. Schulz (2024).

Occupational Burnout Inventory (OBOI).

Göttingen: Hogrefe, kompletter Test € 158,36, Best.-Nr. 01 605 01

Gyde Wartenberg

Institut für Erziehungswissenschaften, Kultur-, Sozial- und Bildungswissenschaftliche Fakultät, Humboldt-Universität zu Berlin, Deutschland

## **Beschreibung des Tests und seiner diagnostischen Zielsetzung**

Das Occupational Burnout Inventory (OBOI; Hossiep & Schulz, 2024) stellt ein Screening-Verfahren zur systematischen Erfassung der berufsbedingten Burnout-Gefährdung von Individuen im Alter von 16 bis 67 Jahren dar. In einem deduktiven Verfahren wurden die fünf Skalen *Berufliches Leistungsdefizit*, *Körperliche Beschwerden*, *Intrapersonale Anzeichen*, *Sozialer Rückzug* und *Erschöpfung* mit jeweils neun Items entwickelt, die im Selbstbericht auf einer sechsstufigen Antwortskala beurteilt werden. Neben den Skalen veranschaulicht der OBOI-Gesamtwert das individuelle Gefährdungspotenzial. Das Verfahren dient explizit nicht einer klinischen Diagnostik, sondern vielmehr der frühzeitigen Erkennung von gefährdeten Individuen, um präventiv in den Symptomverlauf durch personalpsychologische Maßnahmen eingreifen zu können. Es ist sowohl eine individuelle Ergebnisrückmeldung (d.h. Einzelfalldiagnostik), als auch eine gruppenspezifische Auswertung (z.B. Teamanalyse) möglich. Dadurch weist das Instrument in der Anwendung eine hohe praktische und wissenschaftliche Relevanz auf. Anwendungsbereiche umfassen somit nicht nur die individuelle Risikoeinschätzung und Verlaufsdiagnostik (u.a. in betrieblichem Gesundheitsmanagement, Coaching oder Psychotherapie) oder das betriebliche Eingliederungsmanagement nach vorangegangener Burnout-Erkrankung, sondern auch den Einsatz in der Forschung, wie beispielsweise die Evaluation von Interventionsansätzen.

Bei der Testkonstruktion wurde darauf geachtet, dass eine generalisierbare Anwendung bei Personen unterschiedlicher Berufsgruppen möglich ist. Die Testdurchführung (als Papier-Bleistift oder digitale Version) ist mit fünf bis sieben Minuten angegeben und kann in Einzel- oder Gruppentestung stattfinden. Bei der Testauswertung werden sowohl die Skalenmittelwerte als auch der Gesamtmittelwert in Normwerte transformiert und in einem individuellen Gefährdungsprofil abgebildet, wobei positiv gepolte Items umkodiert werden, sodass höhere Werte eine höhere Burnout-Gefährdung verdeutlichen. Bei der computergestützten Bearbeitung geschieht die Auswertung vollständig automatisiert. Die Auswertung der Papier-Bleistift-Version wird nach der Eingabe der Daten in das System ebenfalls computergestützt ausgeführt. Ab einem Sten-Wert<sup>1</sup> von acht wird von einem erhöhten bzw. sehr hohen Burnout-Risiko gesprochen.

## **Bewertung des Informationsgehalts der Verfahrenshinweise**

Insgesamt wird eine strukturierte und prägnante Beschreibung der Test- und Itementwicklung, der Evaluation, sowie der Durchführung, Auswertung und Interpretation gegeben. Positiv hervorzuheben sind die expliziten Empfehlungen zur Wahrung der Anonymität und zum Umgang mit Ergebnissen beim organisationsinternen Einsatz. Zudem liefert eine im Manual enthaltene FAQs-

<sup>1</sup> Sten-Werte zählen zu den etablierten Normwertskalen (auch Standard-Ten-Skala) und umfassen einen Wertebereich von 1–10 mit einem Mittelwert von 5.5 und einer Standardabweichung von 2 (Fröhlich, Pieter & Kemmler, 2024).

Liste Antworten auf potenzielle Rückfragen und Unsicherheiten.

## Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion

Das Verfahren beruht auf einer etablierten Definition des Burnout-Syndroms nach ICD-11 (World Health Organization, 2023), die sich an den einschlägigen Arbeiten von Maslach et al. (2001) und Schaufeli und Enzmann (2020) orientiert. Die Items beschreiben empirisch abgeleitete Burnout-Symptome, die sich den Phasen des Burnout-Prozesses (Burisch, 2014) zuordnen lassen und so das Fortschreiten des Krankheitsbildes abbilden sollen.

## Objektivität

Ein Instrument gilt als objektiv, wenn die Durchführung auf Basis der Testmaterialien, -durchführung, -auswertung und -interpretation unabhängig von Ort, Zeit und Testleitung ohne Verzerrung der Ergebnisse durchgeführt werden kann (Moosbrugger & Kelava, 2020). Die Objektivität des OBOI ist hinsichtlich der Durchführung, Auswertung und Interpretation gewährleistet. Die Bearbeitung des Fragebogens wird durch detaillierte schriftliche Instruktionen eingeleitet und Testleiterinnen und Testleitern liegen ausreichend standardisierte Informationen zur Beantwortung möglicher Rückfragen vor. Sowohl die digitale Durchführung als auch die Papier-Bleistift-Version werden computergestützt vollständig standardisiert ausgewertet, wodurch Testleitungseffekte weitestgehend ausgeschlossen werden können. Im Manual wird darauf hingewiesen, dass eine Auswertung bei fehlenden Werten, d.h. bei nicht bearbeiteten Items, nicht valide möglich ist. Dieses Problem kann allerdings nur in der Papier-Bleistift-Version auftreten. Eine fundierte Interpretation wird durch detaillierte Skalenbeschreibungen, Normtabellen, Fallbeispiele und ein visualisiertes Ergebnisprofil ermöglicht.

## Normierung

Es liegt eine Tabelle mit Normen für die deutsche Gesamtbevölkerung vor. Die Normen werden als Sten-Werte angegeben. Es wird plausibel begründet, dass auf gruppenspezifische Normen verzichtet wird, weil die Gefähr-

dungsbeurteilung explizit im Vergleich zur Gesamtbevölkerung erwerbstätiger Personen in Deutschland erstellt werden soll. Beispielsweise soll das Testergebnis nicht verdeutlichen, ob eine individuelle Lehrkraft ein höheres Burnout-Risiko im Vergleich zu anderen Lehrkräften hat, sondern eine generelle Risikoeinschätzung im Vergleich zur erwerbstätigen Gesamtbevölkerung darstellen. Die Daten der Normstichprobe sind seit 2013 im Rahmen unterschiedlicher Online-Befragungen verschiedener Kooperationsprojekte erhoben worden. Mit einem  $N$  von 16 168 Teilnehmerinnen und Teilnehmern basieren die Normen des OBOI im Vergleich zu anderen Testverfahren zur Erfassung von Burnout oder Depression auf einer deutlich größeren Stichprobe. Mit einem Anteil von knapp 50 % ist der Anteil an Frauen in der Normstichprobe vergleichbar zur erwerbstätigen Bevölkerung in Deutschland. Auch hinsichtlich der Altersverteilung ist die Normstichprobe überwiegend repräsentativ für die Gruppe erwerbstätiger Personen in Deutschland ( $M = 42.57$ ,  $SD = 11.94$ ). Die Verteilung des Bildungsniveaus, der hierarchischen Position, des Beschäftigungsverhältnisses sowie des Arbeitsbereichs und der Branche werden ebenfalls für die Normstichprobe berichtet. Allerdings liegen hier keine Einschätzungen dazu vor, ob die Normstichprobe in diesen Merkmalen repräsentativ für die Gesamtbevölkerung ist. Es zeigten sich kleine Effekte des Geschlechts, des Bildungsniveaus und der hierarchischen Position hinsichtlich der OBOI-Skalen, mit einem geringfügig niedrigeren Burnout-Risiko für Männer, Personen mit einem höheren Bildungsabschluss und Personen mit Führungsverantwortung. Da jedoch alle Effektstärken bei  $d < 0.20$  liegen, sind aufgrund der geringen praktischen Bedeutung keine merkmalsgetrennten Normen notwendig. Für das Beschäftigungsverhältnis, den Arbeitsbereich und die Branche zeigten sich ebenfalls nur für vereinzelte OBOI-Skalen geringfügige Effekte. Die Frage, ob sich die Normwerte, die auf der Online-Bearbeitung des OBOI basieren, auch auf die Papier-Bleistift-Version übertragen lassen, wird im Manual nicht adressiert.

Insgesamt liegen repräsentative Normen, die auf einer sehr großen Stichprobe basieren, für die OBOI-Skalen vor. Die Verfügbarkeit von Sten-Werten erleichtert die Vergleichbarkeit mit anderen Testergebnissen, wobei diese Werte in der Interpretation aber eher grobe Normwerte darstellen.

## Zuverlässigkeit

Die interne Konsistenz wurde anhand der Normstichprobe ( $N = 16 168$ ) ermittelt und wird für die fünf Skalen mit  $\alpha = .84$  (Erschöpfung) bis  $\alpha = .90$  (Intrapersonale Anzei-

chen) mit *gut bis sehr gut* angegeben. Die interne Konsistenz für den Gesamtwert liegt mit  $\alpha = .96$  ebenfalls im sehr guten Bereich.

Die Retest-Reliabilität wurde über eine wiederholte Testung zu insgesamt vier Messzeitpunkten bestimmt. Die Retest-Reliabilität liegt bei 2 260 Personen im Alter zwischen 17 bis 73 Jahren ( $M = 47.63$ ;  $SD = 8.97$ ; 55.66 % weiblich) bei einer erneuten Testung nach 8 bis 14 Wochen für die Skalen zwischen  $r_{tt} = .80$  (*Sozialer Rückzug*) und  $r_{tt} = .83$  (*Körperliche Beschwerden* und *Intrapersonale Anzeichen*) und für den Gesamtwert bei  $r_{tt} = .85$ . Bei einer erneuten Testung ( $N = 1\,029$ ;  $M_{Alter} = 48.33$ ;  $SD_{Alter} = 8.61$ ; 56 % weiblich) nach 33 bis 41 Wochen ergab sich eine Retest-Reliabilität für die Skalen zwischen  $r_{tt} = .72$  (*Erschöpfung*) und  $r_{tt} = .77$  (*Intrapersonale Anzeichen*) und für den Gesamtwert von  $r_{tt} = .78$ . Die Retest-Reliabilität zwischen dem ersten und vierten Messzeitpunkt (nach 58 bis 68 Wochen) wurde anhand von 302 Personen im Alter von 26 bis 72 Jahren ( $M = 49.47$ ;  $SD = 7.70$ ; 57.95 % weiblich) ermittelt und lag für die Skalen zwischen  $r_{tt} = .74$  (*Berufliches Leistungsdefizit* und *Erschöpfung*) und  $r_{tt} = .76$  (*Körperliche Beschwerden*) und für den Gesamtwert bei  $r_{tt} = .79$ . Damit zeigt sich für unterschiedliche Zeitintervalle eine hohe Retest-Reliabilität, die neben der individuellen Standortbestimmung auch eine verlässliche Verlaufsdiagnostik ermöglicht. So kann das Verfahren beispielsweise auch effektiv zur Evaluation von Interventionen genutzt werden.

## Gültigkeit

Zur Itemmetrischen Analyse wurden für die insgesamt 45 Items Item-Mittelwerte und -Standardabweichungen sowie Trennschärfen berechnet, die zwischen .41 und .80 im mittleren bis hohen Bereich liegen. Konfirmatorische Faktorenanalysen wurden nicht berichtet, sodass eine statistische Überprüfung der deduktiv konstruierten Skalen nicht gegeben ist. Die Berechnung des Gesamtscores wird über die hohe Skaleninterkorrelation von .61 bis .84 gerechtfertigt, aber ebenfalls nicht faktorenanalytisch bestätigt.

Die Inhaltsvalidität wird im Manual nicht diskutiert. Aufgrund der Herleitung der Items auf Basis einer umfangreichen Literaturrecherche und empirisch belastbarer Burnout-Symptome kann jedoch angenommen werden, dass die Auswahl der Items repräsentativ für das intendierte Konstrukt ist.

Die Konstruktvalidität wird über den Zusammenhang mit einer deutschen Version des Maslach-Burnout Inventory – General Survey (MBI-GS; Büsing & Glaser, 1998) und den Burnout-Screening Skalen (BOSS I; Geuenich &

Hagemann, 2014) weitestgehend überzeugend überprüft (konvergente Validität). Die Gesamtwerte korrelieren erwartungskonform hoch miteinander ( $r_{MBI} = .81$ ;  $r_{BOSS} = .85$ ). Überraschenderweise hängt die OBOI-Skala *Berufliches Leistungsdefizit* stärker mit den MBI-Skalen *Erschöpfung* ( $r = .75$ ) und *Zynismus* ( $r = .75$ ) zusammen als mit der MBI-Skala *Berufliche Effektivität* ( $r = .54$ ). Der sehr hohe Zusammenhang mit der Screening-Skala aus dem Trierer Inventar zum chronischen Stress (TICS; Schulz, Schlotz & Becker, 2004) und die hohen Korrelationen mit der Allgemeinen Depressionsskala (ADS-L; Hautzinger, Bailer, Hofmeister & Keller, 2012) deuten darauf hin, dass eine Differenzierung zu verwandten Konstrukten wie dem allgemeinen chronischen Stresserleben ( $r = .72$  bis  $r = .88$ ) und Depression ( $r = .66$  bis  $r = .84$ ) nicht klar ist. Explizite Angaben zur divergenten Validität werden allerdings nicht gemacht.

Die Kriteriumsvalidität wurde anhand der Korrelation mit einer Reihe von Außenkriterien, die jeweils im Selbstbericht mit einem Item erfragt wurden, bestimmt. Dabei zeigten sich erwartungskonform negative Zusammenhänge der OBOI-Skalen mit der subjektiven Einschätzung des Berufserfolges ( $r = -.13$  bis  $r = -.46$ ), dem Organisationsklima ( $r = -.25$  bis  $r = -.46$ ) und der subjektiven Arbeitszufriedenheit ( $r = -.33$  bis  $r = -.68$ ) mit einer Varianzaufklärung von jeweils 25 %, 21 % und 47 %. Interessanterweise wies die OBOI-Skala *Berufliches Leistungsdefizit* dabei jeweils die höchste und die OBOI-Skala *Erschöpfung* die niedrigste Korrelation mit den drei Außenkriterien auf. Für die Außenkriterien des aktuellen Belastungserlebens und der subjektiven Einschätzung des aktuellen Burnout-Risikos ergaben sich erwartungskonform positive Zusammenhänge mit den OBOI-Skalen mit einer Varianzaufklärung von jeweils 31 % bzw. 45 %, wobei die OBOI-Skala *Erschöpfung* am höchsten mit diesen beiden Außenkriterien korrelierte ( $r_{Belastung} = .55$ ;  $r_{Risiko} = .60$ ). Angaben zur inkrementellen Validität werden nicht gemacht, sodass keine Einschätzung darüber erfolgen kann, welchen zusätzlichen Vorhersagewert das OBOI im Vergleich zu bereits etablierten diagnostischen Verfahren hat.

Im Manual werden Sten-Werte zwischen acht und neun als erhöhtes Burnout-Risiko angegeben, während bei einem Wert von zehn von einer sehr hohen Gefährdung gesprochen wird. Vor dem Hintergrund, dass das Instrument unter anderem zur Einzelfalldiagnostik vorgesehen ist, ist kritisch anzumerken, dass keine Angaben zur Sensitivität oder Spezifität der Einstufung der Testkennwerte vorliegen.

Insgesamt sprechen diese Ergebnisse für die Validität des OBOI, auch wenn eine Faktorenanalytische Prüfung der Skalenstruktur fehlt und die inhaltliche Zuordnung der Items zu den unterschiedlichen Skalen nicht in allen

Teilen überzeugend ist. Darüber hinaus wäre eine explizite Prüfung der divergenten Validität, die Verwendung etablierter Skalen zur Erfassung der Außenkriterien, eine breitere Auswahl praxisrelevanter Außenkriterien, die Prüfung der inkrementellen Validität und Angaben zur Spezifität und Sensitivität von Cut-Off Werten wünschenswert.

## Weitere Gütekriterien

Die Ökonomie ist durch die kurze Testdauer (fünf bis sieben Minuten) gegeben. Wird eine große Anzahl an Fragebögen benötigt, fallen die Anschaffungskosten jedoch vergleichsweise hoch aus. Zumutbarkeit und Akzeptanz können durch die kurze Testdauer und Art der Itemformulierung als gegeben angesehen werden und auch die Nützlichkeit des Testverfahrens ist nachvollziehbar. Im Manual wird eine Empfehlung für den kombinierten Einsatz des OBOI mit persönlichkeitsorientierten Verfahren gegeben, um das Gefährdungspotenzial bestimmter Persönlichkeitsprofile abzuschätzen. Wie bei allen Selbstbeurteilungsverfahren kann auch bei dem OBOI die Gefahr der Verfälschbarkeit nicht ausgeschlossen werden. Hinweise zum Umgang oder Studien bezüglich der Verfälschbarkeit werden in den Verfahrenshinweisen nicht diskutiert. In den FAQs ist jedoch ein kurzer Hinweis enthalten, welche vorbeugenden Maßnahmen ergriffen werden können. In der Anlage wird in einer Übersicht detailliert und leicht nachvollziehbar dargestellt, inwiefern das vorliegende Testverfahren die DIN-Standards für berufsbezogene Eignungsdiagnostik erfüllt (DIN 33430) und an welchen Stellen des Manuals die zur Einschätzung relevanten Informationen zu finden sind.

## Abschlussbewertung

Mit dem OBOI liegt ein ökonomisches, reliables und in Teilen valides Instrument zur frühzeitigen Einschätzung der Burnoutgefährdung vor, das eine hohe praktische Relevanz aufweist. Insbesondere die Verständlichkeit und die transparenten Durchführungshinweise im Manual, die berufsübergreifende Anwendung, die Möglichkeiten zur Erstellung individueller und gruppenspezifischer Risiko-profile und die konkreten Handlungsoptionen zum Umgang mit einer hohen Burnout-Gefährdung können vor dem Hintergrund der praktischen Relevanz positiv betont werden. Kritisch anzumerken sind lediglich die mangelnden Hinweise zur Anwendbarkeit der computergestützten Normen auf die Papier-Bleistift-Version, die fehlende

empirische Validierung der deduktiven Skalenkonzeption durch konfirmatorische Faktorenanalysen und die mangelnde Einschätzung der Spezifität und Sensitivität des Instruments hinsichtlich der Einzelfalldiagnostik. Dennoch stellt das OBOI ein empfehlenswertes Testverfahren dar, das durch die kurze Bearbeitungsdauer nicht nur für die individuelle Testdiagnostik in der Praxis, sondern auch für den wissenschaftlichen Einsatz praktikabel erscheint.

## Formalisierte Bewertung

**Tabelle 1.** Testbeurteilungskriterien (TBS-TK)

| Occupational Burnout Inventory<br>(OBOI)  | Die TBS-TK-Anforderungen sind<br>erfüllt |            |           |       |
|---|--|------------|-----------|-------|
|   | Voll                                     | Weitgehend | Teilweise | Nicht |
| Allg. Informationen, Beschreibung und diagnostische Zielsetzung   | X  |            |           |       |
| Objektivität  | X  |            |           |       |
| Zuverlässigkeit   | X  |            |           |       |
| Validität   |  | X          |           |       |
|   | Ja                                       | Nein       |           |       |
| In den Verfahrenshinweisen ist verzeichnet, wo die nach den DIN 33430 notwendigen Informationen zu finden sind. | X  |            |           |       |

## Informationen

Diese Testrezension wurde im Auftrag der Diagnostica, Zeitschrift für Psychologische Diagnostik und Differentielle Psychologie erstellt. Bei der Erstellung wurden die TBS-TK-Richtlinien (Diagnostik- und Testkuratorium, 2023) berücksichtigt.

## Literatur

- Burisch, M. (2014). *Das Burnout-Syndrom. Theorie der inneren Er schöpfung* (5. Aufl.). Berlin: Springer.  
 Büssing, A. & Glaser, J. (1998). *Managerial stress and burnout. A collaborative international study (CISMS)*. Die deutsche Untersuchung (Bericht Nr. 44 aus dem Lehrstuhl für Psychologie). München: Technische Universität.  
 Diagnostik- und Testkuratorium (2023). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Vierte, revidierte Fassung vom 31. Juli 2023. *Report Psychologie*, 48(11+12),

- 18–27. Verfügbar unter [https://www.bdp-verband.de/fileadmin/user\\_upload/BDP/website/dokumente/PDF/Profession/Testrezensionen/reportpsychologie\\_11\\_12-2023\\_TBS-DTK.pdf](https://www.bdp-verband.de/fileadmin/user_upload/BDP/website/dokumente/PDF/Profession/Testrezensionen/reportpsychologie_11_12-2023_TBS-DTK.pdf)
- Fröhlich, M., Pieter, A. & Kemmler, W. (2024). Normen, Benchmark und Ranking. In M. Fröhlich, A. Pieter & W. Kemmler (Hrsg.), *Normbezüge und alternative statistische Verfahren im Kontext Sport und Gesundheit* (S. 1–30). Berlin: Springer Spektrum. [https://doi.org/10.1007/978-3-662-70085-3\\_1](https://doi.org/10.1007/978-3-662-70085-3_1)
- Geuenich, K. & Hagemann, W. (2014). *Burnout-Screening-Skalen (BOSS)* (2. Aufl.). Göttingen: Hogrefe.
- Hautzinger, M., Bailer, M., Hofmeister, D. & Keller, F. (2012). *Allgemeine Depressionsskala (ADS)* (2. Aufl.). Göttingen: Hogrefe.
- Hossiep, R. & Schulz, R. (2024). *Occupational Burnout Inventory (OBI)*: Ein personalpsychologischer Fragebogen zur Burnout-Prävention. Göttingen: Hogrefe.
- Maslach, C., Schaufeli, W. B. & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, 52, 397–422.
- Moosbrugger, H. & Kelava, A. (2020). Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien“). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 13–38). Berlin: Springer.
- Schaufeli, W. & Enzmann, D. (2020). *The burnout companion to study and practice: A critical analysis*. Boca Raton, FL: CRC press.
- Schulz, P., Schlotz, W. & Becker, P. (2004). *Trierer Inventar zum chronischen Stress (TICS)*. Göttingen: Hogrefe.
- World Health Organization (WHO). (2023). *International Classification of Diseases 11<sup>th</sup> Revision (ICD-11)*. Verfügbar unter <https://icd.who.int/en>
- Onlineveröffentlichung: 21.10.2025

**Dr. Gyde Wartenberg**

Institut für Erziehungswissenschaften  
Kultur-, Sozial- und Bildungswissenschaftliche Fakultät  
Humboldt-Universität zu Berlin  
Geschwister-Scholl-Straße 7  
10117 Berlin  
Deutschland  
[gyde.wartenberg@hu-berlin.de](mailto:gyde.wartenberg@hu-berlin.de)

<https://doi.org/10.1026/0012-1924/a000355>