



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Vergleich der psychometrischen Güte zweier Verfahren
zur Erfassung des Burnout-Syndroms bei berufstätigen
Erwachsenen“

verfasst von / submitted by

Nicole Perak, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the
degree of

Master of Science (MSc)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 840

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Psychologie UG2002

Betreut von / Supervisor:

Mag. Dr. Reinhold Jagsch

Inhaltsverzeichnis

ABSTRACT Deutsch	4
ABSTRACT English	5
Theoretischer Hintergrund	6
Einleitung.....	6
Anfänge der Burnout-Forschung	6
Entwicklung erster Erhebungsinstrumente	8
Weitere Instrumente	10
Definitionsprobleme	10
Burnout als Diagnose.....	11
Maslach Burnout Inventory	12
Reliabilität.....	13
Faktorielle Validität	14
Kriteriumsvalidität und Diskriminationsfähigkeit	16
Burnout-Screening-Skalen.....	18
Reliabilität.....	19
Faktorielle Validität	20
Kriteriumsvalidität und Diskriminationsfähigkeit	21
Fazit.....	23
Zielsetzung	23
Methodik	24
Untersuchungsdurchführung	24
Studiendesign.....	24
Untersuchungsinstrumente	25
Soziodemographischer Fragebogen	25
Maslach Burnout Inventory	25
Burnout-Screening-Skalen I und II	25
Perceived Stress Questionnaire (PSQ).....	25

Fragestellungen und Hypothesen.....	26
Fragestellung 1: Vollständigkeit der Daten	26
Fragestellung 2: Häufigkeitsverteilung der Antwortkategorien	26
Fragestellung 3: Reliabilität.....	26
Fragestellung 4: Faktorielle Validität	27
Fragestellung 5: Diskriminationsfähigkeit.....	27
Fragestellung 6: Relative Validität	28
Fragestellung 7: Vergleich der Verfahren.....	28
Datenaufbereitung und statistische Auswertung	28
Bildung der Gruppen auf Basis des subjektiven Stresserlebens	28
Datenvollständigkeitsanalyse.....	29
Deskriptive Datenanalyse	29
Reliabilitätsanalyse	29
Trennschärfeanalyse	30
Hauptachsenanalyse	30
Receiver-Operating-Characteristics-Kurven (ROC-Kurven)	30
Ergebnisdarstellung.....	31
Stichprobenbeschreibung.....	31
Nationalität.....	31
Beziehungsstatus.....	31
Beschäftigungsart.....	31
Durchschnittliche Arbeitsstunden pro Woche.	32
Beruf.	32
Berufserfahrung.	33
Höchste abgeschlossene Ausbildung.	33
Krankenstandstage.	33
Gruppeneinteilung A.....	33
Gruppeneinteilung B.....	33

Untersuchung der Fragestellungen	34
Datenvollständigkeitsanalyse.....	34
Deskriptive Analyse.....	35
Reliabilitätsanalyse	38
Faktorenanalyse.	38
ROC-Analyse.....	47
Analyse der relativen Validität.	47
Diskussion	48
Datenvollständigkeit.....	49
Häufigkeitsverteilung der Antwortkategorien	50
Reliabilität	52
Faktorielle Validität.....	54
Diskriminationsfähigkeit und relative Validität	56
Vergleich der Verfahren	57
Limitationen und Ausblick	58
Literaturverzeichnis.....	59
Tabellenverzeichnis.....	67
Anhang A: Anteil an fehlenden Werten pro Item	68
Anhang B: Verteilung der Skalenwerte in den Stresserlebensgruppen.....	71
Anhang C: Trennschärfeanalyse	75
Anhang D: Faktorenladungen mit postulierter Faktorenanzahl	78

ABSTRACT Deutsch

Hintergrund. Berufsbedingtes Burnout ist ein populäres Thema in den Medien und der Forschung. Nur ein sehr geringer Anteil der Burnout-Forschung beschäftigt sich mit den psychometrischen Eigenschaften der Erhebungsinstrumente. Das Maslach Burnout Inventory (MBI) gilt als „Goldstandard“ zur Erfassung des Burnout-Syndroms. Bisherige Untersuchungen zur Testgüte des MBI zeichnen jedoch kein eindeutiges Bild, und seine Anwendbarkeit in der klinischen Praxis ist fragwürdig. Die Burnout-Screening-Skalen (BOSS) I und II könnten möglicherweise eine vielversprechende Alternative darstellen. Eine von den Testautor_innen unabhängige psychometrische Begutachtung der BOSS fehlt bisher.

Methodik. Ziel dieser Studie war es, die beiden Instrumente anhand einer Auswahl von verschiedenen Indikatoren zur Beurteilung der psychometrischen Güte zu untersuchen. Dazu wurden die Verfahren einer Stichprobe deutschsprachiger berufstätiger Erwachsener ($N = 392$) vorgegeben. Die Teilnehmer_innen wurden auf Basis ihres subjektiven Stresserlebens in zwei bzw. vier Gruppen eingeteilt.

Ergebnisse. Die erhobenen Daten waren in hohem Maße vollständig. Bei beiden Instrumenten traten Bodeneffekte auf. Die interne Konsistenz der Subskalen des MBI war akzeptabel bis ausgezeichnet und jene der BOSS gut bis ausgezeichnet. Die Skalenstruktur der Instrumente konnte faktorenanalytisch nicht repliziert werden. Die Diskriminationsfähigkeit der Skalen wurde mittels ROC-Analyse untersucht, wobei das subjektive Stresserleben als Außenkriterium herangezogen wurde. Darüber hinaus wurde die relative Validität der Skalen berechnet. Alle Subskalen des MBI und der BOSS I und II konnten signifikant zwischen den Gruppen differenzieren. Die Diskriminationsfähigkeit der BOSS war insgesamt jener des MBI überlegen.

Schlussfolgerung. Die Ergebnisse dieser Studie sprechen dafür, dass sich die BOSS I und II besser zur Erfassung des Burnout-Syndroms bei berufstätigen Erwachsenen eignen als das MBI.

ABSTRACT English

Background. Occupational burnout is a popular subject in research and media. Only a small share of burnout research concerns the psychometric properties of the instruments used for its assessment. The Maslach Burnout Inventory (MBI) is considered the “gold standard” of burnout measurement. Previous findings regarding the psychometric quality of the MBI are inconclusive and its use in clinical practice is questionable. The Burnout-Screening-Scales (BOSS) I and II might represent a promising alternative. No independent psychometric evaluations of the BOSS have been carried out thus far. **Methods.** The aim of this study was to examine the psychometric quality of these instruments using a selection of various indicators. For this purpose the measures were administered to a sample of German-speaking working adults ($N = 392$). The participants were then sorted into two and four groups on the basis of their subjective experienced stress. **Results.** The sampled data was highly complete. Floor effects occurred on both instruments. The resulting internal consistency of the instrument subscales can be rated reasonable to good for the MBI and good to excellent for the BOSS. The instruments’ scale structure could not be replicated using factor analysis. To test the discriminating power of the scales, a ROC analysis was computed using subjective experienced stress as an external criterion. Furthermore the relative validity of the scales was determined. All subscales of the MBI and BOSS I and II were able to significantly distinguish between groups. Overall the discriminating power of the BOSS was superior to the MBI’s. **Conclusion.** The results of this study suggest that the BOSS I and II are more expedient for the assessment of burnout in working adults than the MBI.

Theoretischer Hintergrund

Einleitung

Berufsbedingtes Burnout ist seit Jahren ein beliebtes Thema in den Populärmedien, der Selbsthilfeliteratur und auch in der psychologischen Forschung. Ungefähr 13000 Einträge finden sich derzeit zu dem Thema in der Datenbank PsycINFO der American Psychological Association. Besonders in den letzten 15 Jahren ist ein deutlicher Anstieg an wissenschaftlichen Veröffentlichungen zum Thema Burnout zu verzeichnen (Heinemann & Heinemann, 2017). Eine Analyse der englischsprachigen Publikationen von 1978 bis 2012 zeigt, dass sich der Großteil der Studien mit der Erforschung von Ursachen, zusammenhängenden Faktoren und der Prävalenz des Syndroms beschäftigt (Heinemann & Heinemann, 2017). Überschaubar ist hingegen die Forschung zu Interventionen, Symptomatik, Psychometrie und Biomarkern. Ein genauerer Blick auf die Literatur zeigt ein kontrovers diskutiertes Forschungsfeld, das definitorisch, diagnostisch und psychometrisch durchaus auf wackeligen Beinen zu stehen scheint.

Anfänge der Burnout-Forschung

Um zu verstehen, was Burnout ist, lohnt sich ein Blick auf die historische Entwicklung der Burnout-Forschung. Das Burnout-Syndrom wurde erstmals 1974 von dem Psychologen Herbert J. Freudenberger beschrieben. Sein Artikel handelt von seinen Beobachtungen an sich selbst und anderen Helfer_innen, die in alternativen „Free Clinics“ für Drogenabhängige in den USA arbeiteten. Für Freudenberger sei Burnout eine Art von Überarbeitung, die sich in somatischen und verhaltensbezogenen Symptomen äußere. Körperlich zeige sich Burnout in einem Gefühl von Erschöpfung oder Ermüdung, verringerter Widerstandsfähigkeit gegenüber Krankheiten, häufigen Kopfschmerzen, Magen-Darm-Problemen, Kurzatmigkeit, Schlafstörungen und Depressivität. Betroffene haben Schwierigkeiten, ihre Gefühle zurückzuhalten. Sie seien frustriert, reizbar, werden schnell laut und brechen leichter in Tränen aus. Kleine Belastungen überfordern sie bereits. Des Weiteren entwickeln sie eine negative, zynische, verbitterte, resignierte und misstrauische bis paranoide Haltung. In ihrer Arbeit mit Klient_innen gehen sie teilweise große Risiken ein. Sie leisten Widerstand gegen jede Art von Veränderung und üben viel Kritik. Auf andere wirken sie deprimiert, nach innen gekehrt und zurückgezogen. Einige greifen auch zu Beruhigungsmitteln. Freudenberger beschreibt anekdotisch fünf verschiedene Etiketten von Persönlichkeiten, die zu Burnout neigen (Freudenberger, 1975). Gemeinsam sei ihnen allen ihr übertrieben aufopferungsvolles

Engagement. Sie schaffen es nicht, ihr Privatleben von ihrem Beruf abzugrenzen, arbeiten zu viel, zu lang und zu intensiv. Die Persönlichkeitsetiketten unterscheiden sich vor allem in ihren Positionen und den Motiven für ihre übermäßige Hingabe. Für manche sei der Wunsch, anderen zu helfen, und ihr aufopferndes Selbstbild eine treibende Kraft. Andere haben ein unbefriedigendes Privatleben, weshalb sie es vorziehen, ihre Identität vollständig über ihren Beruf zu definieren. Die Unfähigkeit oder der Widerwille, Aufgaben zu delegieren, stelle laut Freudenberger einen weiteren Grund für das übertriebene Engagement dar. Aber nicht nur Persönlichkeitsfaktoren, sondern auch Umweltfaktoren spielen für Freudenberger eine Rolle in der Entstehung von Burnout. Eine hohe Anzahl an Arbeitsstunden, Langeweile durch eintönige Aufgaben, fehlende Anerkennung, Leistungsdruck durch Vorgesetzte, Überidentifizierung mit den Klient_innen sowie der Druck durch deren Bedürftigkeit werden unter anderem als Ursachen des Ausbrennens angeführt (Freudenberger, 1975). Es war allerdings nicht Freudenberger, der den Begriff Burnout prägte, da die Helfer_innen ihn selbst bereits untereinander nutzten.

Etwa zur selben Zeit (1976) befasste sich auch die Sozialpsychologin Christina Maslach mit dem Phänomen Burnout bei Personen in helfenden Berufen. Bei ihrer Exploration des Syndroms ging sie systematischer vor als Freudenberger. Gemeinsam mit Kolleg_innen der University of California beobachtete sie mehr als 200 Helfer_innen bei ihrer Arbeit, führte Interviews durch und gab Fragebögen vor. Die Teilnehmer_innen arbeiteten als Armen-Anwält_innen, Ärzt_innen, Gefängnisangestellte, Sozialarbeiter_innen, klinische Psycholog_innen und Psychiater_innen. In diesem ersten Artikel beschreibt Burnout hauptsächlich eine Art von emotionaler Abstumpfung gegenüber Patient_innen oder Klient_innen. Diese Abstumpfung sei ein Versuch der Bewältigung des emotionalen Stresses, den die Arbeit als Helfer_in mit sich bringe. Sie entwickeln negative oder zynische Gefühle ihren Klient_innen gegenüber, kümmern sich weniger um sie, versuchen den direkten Kontakt mit ihnen zu limitieren und betrachten sie oft nicht mehr als menschliche Individuen (Dehumanisierung). Darüber hinaus entwickeln viele Betroffene psychosomatische Symptome wie beispielsweise Muskelverspannungen. Manche Helfer_innen greifen zu Alkohol, Beruhigungsmitteln oder Drogen. Der Stress werde auch oft von der Arbeit mit nachhause genommen und sorge für Probleme im familiären Umfeld. Burnout führe auf institutioneller Ebene zu schlechterer Versorgung der Klient_innen oder Patient_innen, niedriger Arbeitsmoral, mehr Krankenstandstagen und häufigeren Kündigungen. Maslach identifizierte eine zu hohe Anzahl an Patient_innen oder Klient_innen, zu viel direkten Kontakt mit ihnen ohne Rückzugsmöglichkeiten und einen Mangel an Training im Umgang

mit emotionalem Stress als mit Burnout in Zusammenhang stehende Faktoren (Maslach, 1976). In weiteren Arbeiten beschrieb Maslach (1978, 1981) auch eine emotionale Erschöpfung, die sich zusätzlich zur körperlichen Erschöpfung bei betroffenen Helfer_innen zeige.

Spätestens Anfang der 1980er war Burnout ein sehr populäres Thema. Eine Vielzahl an Veröffentlichungen über Burnout erschien in wissenschaftlichen Journals und den Massenmedien. Es wurden Selbsttests angeboten, Beratungsliteratur geschrieben und Symposien abgehalten (Maslach & Jackson, 1984). Zu dieser Zeit entstanden viele Theorien über Burnout und dessen Entstehungsursachen, die aber kaum empirisch geprüft wurden (Maslach, 1982).

Entwicklung erster Erhebungsinstrumente

Gemeinsam mit ihrer Kollegin Ayala Pines und der Studentin Susan E. Jackson untersuchte Maslach in den darauffolgenden Jahren das Phänomen bei verschiedenen helfenden Berufen näher (Maslach & Jackson, 1979; Maslach & Pines, 1977; Pines & Maslach, 1978, 1980). Maslach und Jackson (1981a) entwickelten im Zuge dessen einen Selbsteinschätzungsfragebogen zur Erhebung des Burnout-Syndroms bei Personen in helfenden Berufen, das Maslach Burnout Inventory (MBI). Sie beschreiben Burnout nun als ein Syndrom mit drei Aspekten: (1) emotionale Erschöpfung, (2) negative, zynische Gefühle gegenüber den Klient_innen oder Patient_innen und (3) eine Tendenz zur negativen Selbstbewertung, insbesondere in Bezug auf die eigene Arbeit. Das Syndrom trete nur bei Personen auf, die mit Menschen arbeiten. Ihr Instrument spiegle diese drei Aspekte in je einer Skala wider. Optional ist eine weitere Skala, die eine Art von emotionaler Verwicklung mit anderen Menschen abbilde (Maslach & Jackson, 1981b).

Während sich Maslach auf Burnout konzentrierte, begann Pines, gemeinsam mit einer weiteren Kollegin, Ditsa Kafry, ihren Fokus auf ein Phänomen zu legen, das sie *Tedium* nannten (Kafry & Pines, 1980; Pines & Kafry, 1978, 1981). Im Deutschen wird Tedium als Überdross bezeichnet. Überdross sei das Erleben von körperlicher, emotionaler und mentaler Ermüdung (Kafry & Pines, 1980). Betroffene fühlen sich belastet und ausgebrannt. Sie weisen negative Einstellungen sich selbst, ihrer Umwelt und ihrem Leben gegenüber auf. Überdross in seiner schlimmsten Form sei ein Punkt, an dem die Bewältigungsstrategien eines Individuums nicht mehr greifen. An diesem Punkt überschneide es sich mit dem Störungsbild der Depression. Die Ursachen für Überdross können etwa traumatische Lebensereignisse oder chronischer Stress sein. Chronischer Stress trete wiederum dann auf, wenn die Anforderungen

die eigenen wahrgenommenen Ressourcen unter- oder überschreiten, oder in Situationen, in denen die Belohnungen nicht den eigenen Erwartungen entsprechen. Überdruss sei eng verbunden mit der Stressforschung und dem Konzept der Arbeitszufriedenheit (Kafry & Pines, 1980). Zur Erhebung von Überdruss entwickelten sie ein Selbsteinschätzungsinstrument, die Überdrussskala (Tedium Measure, TM, Aronson, Pines, & Kafry, 1983). Der Unterschied zwischen Überdruss und Burnout liegt für die Testautor_innen darin, dass Burnout nur bei Personen in helfenden Berufen auftrete, während Überdruss bei allen Professionen auftreten könne. Bereits einige Jahre später trafen Pines und Aronson (1988) diese Unterscheidung allerdings nicht mehr, und die Überdrussskala wurde von Tedium Measure in Burnout Measure (BM) umbenannt, wobei Burnout alle Berufsgruppen betreffe.

Mit der zweiten Auflage des MBI wurde zusätzlich eine alternative Version für Lehrende veröffentlicht, in der sich die Items nicht mehr auf Empfänger_innen der Hilfsleistungen von Helfer_innen, sondern auf Schüler_innen beziehen (Maslach & Jackson, 1986). In der dritten Auflage des MBI wurde das originale Instrument in MBI-HSS (MBI Human Service Survey, Maslach & Jackson, 1996) und die Version für Lehrende in MBI-ES (MBI Educators Survey, Maslach, Jackson, & Schwab, 1996) umbenannt. Zusätzlich wurde mit dem MBI-GS (MBI General Survey, Schaufeli, Leiter, Maslach, & Jackson, 1996) eine weitere alternative Form veröffentlicht. Sie dient zur Erfassung von Burnout bei Personen, die nicht in helfenden Berufen arbeiten (Maslach, Jackson, & Leiter, 1997). Die Arbeitsdefinition des MBI-GS unterscheidet sich von Maslachs und Jacksons ursprünglicher Definition, laut der Burnout ausschließlich bei Personen, die mit Menschen arbeiten, auftrete und sie negative und zynische Gefühle gegenüber ebendiesen entwickeln. Laut dem MBI-GS richten sich diese Gefühle bei Personen, die nicht oder nur oberflächlich mit Menschen arbeiten, gegen die Arbeit an sich. Die Beziehung zu ihrer Arbeit stelle sich über ein Kontinuum von Engagement bis Burnout dar (Maslach et al., 1997). Mit diesem Schritt weitete also auch Maslach ihre Burnout-Definition auf alle Professionen aus. Doch Maslach (2000) betont, dass sich ihre Burnout-Definition immer noch von jener von Aronson et al. (1983) unterscheide. Während letztere sich ausschließlich auf die Dimension der emotionalen Erschöpfung konzentriere, erweitere Maslachs Definition diese um zwei weitere Dimensionen, negative, zynische Gefühle und geringe erlebte Kompetenz und Produktivität.

Weitere Instrumente

Das große Interesse der Forschung führte in weiterer Folge zur Entwicklung einer Vielzahl von weiteren Messinstrumenten (Burisch, 2014; Schaufeli, Leiter, & Maslach, 2009), wie etwa jene von Melamed, Kushnir und Shirom (1992), Demerouti, Bakker, Vardakou und Kantas (2003) und Kristensen, Borritz, Villadsen und Christensen (2005). Melamed et al. (1992) definieren Burnout als eine durch körperliche, emotionale und kognitive Erschöpfung gekennzeichnete Konstellation. Zu dessen Erhebung fertigten sie das Selbstbeschreibungsinstrument Shirom-Melamed Burnout Measure (S-MBM) an.

Ein weiteres Instrument, das den Aspekt der Erschöpfung in den Mittelpunkt seiner Operationalisierung stellt, ist das Copenhagen Burnout Inventory (CBI, Kristensen et al., 2005). Das CBI enthält drei verschiedene Teile, die sich auf Burnout in verschiedenen Bereichen beziehen: (1) persönliches Burnout, (2) arbeitsbezogenes Burnout und (3) Burnout in Bezug auf Klient_innen. Der erste Teil versuche die Ausprägung von körperlicher und psychischer Erschöpfung im Allgemeinen zu erfassen. Die anderen beiden Teile beziehen sich auf die körperliche und psychische Erschöpfung in Bezug auf die Arbeit und andere Personen. Diese Personen können in allen möglichen Beziehungen zu den Teilnehmer_innen stehen, zum Beispiel Kolleg_innen, Nachbarn, Kinder, Schüler_innen, und müssen nicht tatsächliche Klient_innen sein.

Demerouti et al. (2003) entwickelten in Deutschland einen weiteren Fragebogen zur Selbstbeschreibung, das Oldenburg Burnout Inventory (OLBI). Das Instrument besteht aus zwei Skalen: *Erschöpfung* und *Engagement*. Erschöpfung bezieht sich, wie bei Aronson et al. (1983) und Melamed et al. (1992), auf körperliche, affektive und kognitive Aspekte, während sich Engagement auf emotionales Distanzieren von der Arbeit und negative Einstellung gegenüber der Arbeit bezieht, ähnlich wie bei Maslach und Jackson (1981b).

Die Burnout-Screening-Skalen (BOSS, Hagemann & Geuenich, 2009) wurden ebenfalls in Deutschland entwickelt. Das Verfahren besteht aus zwei Teilen: Die Burnout-Screening-Skala I (BOSS I) konzentriert sich auf Beschwerden im Beruf und sozialen Umfeld sowie psychische Beschwerden. Bei der Burnout-Screening-Skala II (BOSS II) steht die psychosomatische Symptomatik auf körperlicher, emotionaler und kognitiver Ebene im Vordergrund.

Definitionsprobleme

Die erwähnten Burnout-Definitionen verschiedener Forschungsgruppen unterscheiden sich in gewissen Aspekten voneinander. Grob lassen sie sich in zwei Kategorien einteilen:

eindimensionale und mehrdimensionale Definitionen (Schaufeli et al., 2009). Erstere setzen Burnout im Grunde mit Erschöpfung gleich (Aronson et al., 1983; Hagemann & Geuenich, 2009; Kristensen et al., 2005; Melamed et al., 1992). Diese Erschöpfung wird oft noch näher als eine Kombination aus körperlicher, psychischer und manchmal auch kognitiver Erschöpfung beschrieben. Mehrdimensionale Definitionen sehen Burnout als ein durch Erschöpfung und weitere Aspekte gekennzeichnetes Syndrom (wie etwa Zynismus, Demerouti et al., 2003; Maslach & Jackson, 1981b).

Bereits 1982 bemängelt Maslach das Fehlen einer einheitlich akzeptierten Definition von Burnout. Forscher_innen würden auch zukünftig aneinander vorbeireden, schafften sie es nicht, sich auf eine solche zu einigen. Maslach (1982) kritisiert auch den Begriff „Burnout“ an sich. Er habe zwar den Vorteil, dass sich unter dem Begriff jeder sofort etwas vorstellen könne, die Bilder, die dabei aufkommen können, seien allerdings sehr unterschiedlich. So könne Burnout als eine erloschene Flamme gesehen werden, also als etwas, das einmal war und jetzt nicht mehr ist. Ein anderes Bild sei jenes eines Feuers, das eine Person von innen heraus ausbrennt. Dieses Feuer könne übermäßige interne Motivation darstellen, die zur Selbstverausgabung führt, oder eine externe Situation, die so stressig ist, dass man sie nicht länger bewältigen kann. Ein fehlender wissenschaftlicher Konsens bei der Definition könne gerade bei einem so bildlichen Wort dazu führen, dass jeder etwas anderes darunter verstehe (Maslach, 1982). Abgesehen davon meinen Maslach und Jackson (1984), dass die bildhafte Eigenschaft des Begriffs zu seiner Popularität beitrage. Allerdings führe die aufgekommene Popularität beim gleichzeitigen Fehlen einer einheitlichen Definition dazu, dass das Konzept des Burnouts von vielen Wissenschaftlern als „Pop-Psychologie“ bezeichnet und nicht ernst genommen werde. Obwohl Burnout nach nunmehr 45 Jahren Forschung immer noch ein populäres Thema in den Medien ist und Schaufeli et al. die Anzahl der wissenschaftlichen Publikationen zu dem Thema schon im Jahre 2009 auf mehr als 6000 schätzten, ist es bis heute nicht gelungen, einen Konsens bezüglich der Definition zu erzielen (Burisch, 2014; Heinemann & Heinemann, 2017; Schaufeli & Enzmann, 1998).

Burnout als Diagnose

Durch die Popularität des Themas wurde auch das Interesse der klinischen Psychologie und Psychiatrie geweckt, doch das Definitionsproblem verhindert die Bestimmung von einheitlichen Diagnosekriterien. Darüber hinaus gibt es symptomatisch große Überschneidungen mit bereits bestehenden Diagnosen wie depressiven Störungen oder Angststörungen (Bianchi, Schonfeld, & Laurent, 2015a, 2015b, 2019; Koutsimani,

Montgomery, & Georganta, 2019; Schonfeld, Bianchi, & Palazzi, 2018). Aus diesen Gründen wird Burnout in den meisten Ländern nicht als offizielle Diagnose von den Gesundheitssystemen anerkannt (Heinemann & Heinemann, 2017). Ausnahmen stellen lediglich Schweden und die Niederlande dar, wo Burnout eine offizielle medizinische Diagnose darstellt (Burisch, 2014; Friberg, 2009; Heinemann & Heinemann, 2017). In der 5. Auflage des *Diagnostic and statistical manual of mental disorders* (DSM, American Psychiatric Association, 2013) ist Burnout nicht enthalten. Im aktuell gültigen Klassifikationssystem medizinischer Diagnosen der WHO, der ICD-10, ist Burnout in Kapitel XXI (Z 73.0) knapp als „Zustand der totalen Erschöpfung“ erwähnt (World Health Organization, 2016). Dieses Kapitel hat keine Krankheiten oder Störungsbilder zum Inhalt, sondern andere Faktoren, die den Gesundheitszustand beeinflussen und dazu führen können, dass Personen das Gesundheitssystem beanspruchen. Seit Mai 2019 ist Burnout in der neuesten Version der ICD-11, die ab 2022 in Kraft treten soll, in Kapitel 24 „Factors influencing health status or contact with health services“ unter dem Code QD85 erwähnt (World Health Organization, 2019). Burnout ist der ICD-11 zufolge daher immer noch nicht als psychische Störung klassifiziert, enthält aber eine detailliertere Beschreibung als in der ICD-10. Es wird beschrieben als ein Syndrom, das aus chronischem, unbewältigtem Stress am Arbeitsplatz hervorgeht und durch drei Dimensionen gekennzeichnet ist: (1) Gefühle von Erschöpfung, (2) erhöhte psychische Distanz von der Arbeit oder negative Gefühle oder Zynismus gegenüber der Arbeit und (3) reduzierter professioneller Wirksamkeit. Darüber hinaus wird darauf hingewiesen, dass sich Burnout nur auf den Arbeitskontext und nicht auf Erfahrungen in anderen Lebensbereichen beziehe. Diese Definition entspricht jener von Maslach, Schaufeli und Leiter (2001).

Maslach Burnout Inventory

Das Maslach Burnout Inventory ist ein Selbstbeschreibungsinstrument zur Erfassung des Burnout-Syndroms (Maslach & Jackson, 1981a). Es gilt als „Goldstandard“ der Burnout-Forschung und kam in mehr als 90% aller Studien zum Einsatz (Schaufeli & Enzmann, 1998; Schaufeli et al., 2009). Das Instrument wurde faktorenanalytisch konstruiert und beinhaltet in seiner Erstauflage insgesamt 25 Items und vier Skalen (Maslach & Jackson, 1981a). Des Weiteren verfügt es über ein siebenstufiges Antwortformat, getrennt nach Häufigkeit, von 0 (*nie*) bis 6 (*jeden Tag*), und Intensität, von 1 (*sehr schwach, kaum wahrnehmbar*) bis 7 (*bedeutend, sehr stark*). In der zweiten Auflage des Instruments wurden die vierte Skala *Betroffenheit* und die Intensitätsmessung schließlich entfernt (Maslach & Jackson, 1986). Die

verbleibenden 22 Items gliedern sich in drei Skalen (Maslach & Jackson, 1986). Die Skala *Emotionale Erschöpfung* (EE, 9 Items) beschreibt Gefühle von emotionaler Überlastung und Ermüdung durch die Arbeit, die Skala *Depersonalisation* (DP, 5 Items) Gefühllosigkeit und Distanziertheit gegenüber Klient_innen oder Patient_innen und die Skala *Persönliche Erfüllung* (PE, 8 Items) ein Gefühl von Kompetenz und Erfolg in der Arbeit mit Patient_innen oder Klient_innen (Maslach & Jackson, 1981b). Burnout ist durch hohe Ausprägungen in den Skalen EE und DP und niedrige Ausprägungen in der Skala PE charakterisiert. Die Berechnung eines Gesamtwerts über alle drei Skalen ist nicht vorgesehen (Maslach & Jackson, 1981a). Der Fragebogen existiert in mehreren deutschen Übersetzungen (Barth, 1985; Burisch, 1984; Büssing & Perrar, 1992; Enzmann & Kleiber, 1989). Die Übersetzung von Büssing und Perrar aus dem Jahr 1992 ist bis heute die einzige deutsche Version, die durch Maslach autorisiert ist (Burisch, 2014).

Eine bedeutende Einschränkung des MBI für die Burnout-Diagnostik stellt seine fehlende Normierung dar. Das Testmanual gibt zwar Cut-off-Werte an, jedoch betonen die Autor_innen, dass diese nicht für die klinische Diagnostik geeignet seien und dass das MBI ein reines Forschungsinstrument sei (Maslach & Jackson, 1986; Maslach et al., 2001). Die Cut-offs wurden lediglich durch das Dritteln der Stichprobe in niedrige, mittlere und hohe Ausprägungen bestimmt. Versuche, universelle Cut-offs zu etablieren, sind jedoch auch kritisch zu sehen, da sie nicht für jedes Land und jede Stichprobe gleich gut geeignet sind und oftmals aus Daten von gesunden, arbeitenden Stichproben abgeleitet werden, anstatt von Daten aus klinischen Stichproben (Schaufeli, Bakker, Hoogduin, Schaap, & Kladler, 2001).

Reliabilität. Maslach und Jackson (1981b) berichten akzeptable Reliabilitätskoeffizienten. Sie fanden Cronbachs-Alpha-Werte von .89 für die Skala EE, .77 für die Skala DP und .74 für die Skala PE ($N = 420$). Die Retest-Reliabilität bei einer Stichprobe von 53 Personen nach zwei bis vier Wochen betrug .82 für die Skala EE, .60 für die Skala DP und .80 für die Skala PE. Eine jüngere Abschätzung der Reliabilität findet sich bei Riley, Mohr und Waddimba (2018). Bei der Anwendung an einer Stichprobe von 308 Ärzt_innen in den USA zeigte sich ebenfalls eine hohe interne Konsistenz der drei Skalen ($\alpha \geq .87$).

Eine Metaanalyse von Wheeler, Vassar, Worley und Barnes (2011) verglich Daten aus 84 internationalen Studien und berechnete durchschnittliche Cronbachs-Alpha-Werte von .87 für EE, .71 für DP und .76 für PE. Während die Skala EE in allen Studien eine zufriedenstellende interne Konsistenz aufwies ($\alpha_{\min} = .72$, $\alpha_{\max} = .95$), war dies bei den Skalen DP ($\alpha_{\min} = .50$, $\alpha_{\max} = .91$) und PE ($\alpha_{\min} = .52$, $\alpha_{\max} = .87$) nicht immer der Fall. Eine zweite

Metaanalyse von Aguayo, Vargas, de la Fuente und Lozano (2011) mit Daten aus 45 internationalen Studien berechnete beinahe identische durchschnittliche Alpha-Werte in der Höhe von .88 für EE ($\alpha_{\min} = .66$, $\alpha_{\max} = .95$), .71 für DP ($\alpha_{\min} = .43$, $\alpha_{\max} = .83$) und .78 für PE ($\alpha_{\min} = .49$, $\alpha_{\max} = .94$). Aguayo et al. (2011) betonen, dass Reliabilitätswerte aufgrund der gefundenen Heterogenität nicht auf verschiedene Stichproben und Kontexte generalisiert werden dürfen. Da nur die Skala EE konsistent zufriedenstellende interne Konsistenzen aufweise, sollten die anderen Skalen nicht angewendet werden, um wichtige klinische Entscheidungen zu treffen (Wheeler et al., 2011). Die Sprache des MBI erklärte in beiden Metaanalysen einen signifikanten Anteil der Variabilität der internen Konsistenz der Skalen EE und DP. Die Alpha-Koeffizienten, besonders jene der Skala DP, waren in den Studien niedriger, die eine fremdsprachige Übersetzung des englischen Originals einsetzten (Aguayo et al., 2011; Wheeler et al., 2011).

Dieses Bild zeigt sich auch bei der Reliabilität der deutschen Übersetzungen des MBI. Büssing und Perrar (1992) berichten für ihre Version Alpha-Koeffizienten von .88 für EE, .58 für DP und .61 für PE ($N = 449$). Eine weitere Studie, die dieselbe Übersetzung verwendete ($N = 362$), erzielte Cronbachs-Alpha-Werte von .85 für EE, .42 für DP und .71 für PE (Neubach & Schmidt, 2000). Enzmann und Kleiber (1989) übersetzten den Fragebogen ebenfalls ins Deutsche und fanden interne Konsistenzen von $\alpha = .82$ für EE, $\alpha = .67$ für DP und $\alpha = .75$ für PE ($N = 129$). Eine weitere Übersetzung durch Burisch (1984) berichtete Alpha-Werte von .81 für EE, .60 für DP und .67 für PE ($N = 51$). Einige Items wiesen in den Untersuchungen der deutschen Übersetzungen eine geringe Trennschärfe ($< .40$) auf: Item 7 und 22 (Burisch, 1984; Büssing & Perrar, 1992; Enzmann & Kleiber, 1989), Item 4, 5, 9, 10, 18 und 21 (Burisch, 1984; Büssing & Perrar, 1992), Item 14 und 16 (Burisch, 1984) sowie Item 15 und 17 (Büssing & Perrar, 1992).

Es zeigt sich, dass die Skala DP in jeder der deutschsprachigen Stichproben eine unzufriedenstellende interne Konsistenz aufwies, unabhängig von der verwendeten Übersetzung. Die Skala EE wies auch in den deutschsprachigen Studien gute Cronbachs-Alpha-Werte auf. Die Ergebnisse für die Skala PE schwankten zwischen akzeptablen und unzufriedenstellenden Reliabilitätswerten.

Faktorielle Validität. Die Testautor_innen unterzogen die 25 Items der ersten MBI-Version einer Hauptachsenanalyse mit Varimax-Rotation (Maslach & Jackson, 1981b). Sie fanden eine vierfaktorielle Lösung ohne Fehlladungen ($< .40$) und Doppelladungen. Der erste Faktor entsprach der Skala EE und wies Ladungen im Bereich von .54 bis .84 auf. Die Items der Skala PE luden in einer Höhe von .43 bis .59 auf dem zweiten Faktor. Der dritte Faktor

verfügte über Ladungen zwischen .41 und .67 und ergab die Skala DP. Der vierte Faktor stellte die Skala Betroffenheit dar, die in der ersten Version des MBI (Maslach & Jackson, 1981a) zur optionalen Vorgabe inkludiert war. Die Skala wurde nicht zu den Hauptskalen gezählt, da der Eigenwert ihres Faktors kleiner als 1 war. In der zweiten Auflage wurden die drei Items, aus denen die Skala bestand, aus dem Instrument entfernt (Maslach & Jackson, 1986).

Büssing und Perrar (1992) unterzogen die 25 Items ihrer deutschen Übersetzung einer Hauptkomponentenanalyse mit Varimax-Rotation. Wie Maslach und Jackson (1981b) berichteten sie eine vierfaktorielle Struktur, wobei allerdings nur der erste Faktor EE befriedigende Ladungen zwischen .61 und .77 ohne Doppelladungen aufwies. Beim zweiten Faktor PE kam es zu einer Fehlladung und einer Doppelladung. Die Ladungen lagen zwischen .18 und .62. Der dritte Faktor stellte nur eingeschränkt die Skala DP dar. Bei den fünf Items der DP, die Ladungen von .00 bis .56 aufwiesen, kam es zu zwei Fehlladungen und einer Doppelladung. Drei der Items luden stärker auf dem ersten Faktor als auf dem vorgesehenen dritten Faktor. Der vierte Faktor, der die drei Items der optionalen Skala Betroffenheit beinhalten sollte, wies ausschließlich Fehlladungen auf.

Enzmann und Kleiber (1989) führten eine Hauptachsenanalyse mit Varimax-Rotation mit den 25 Items ihrer Übersetzung durch. Sie fanden ebenfalls vier Faktoren. Sie berichteten von Ladungen zwischen .32 und .73 mit drei Fehlladungen und einer Doppelladung für den Faktor EE. Der zweite Faktor PE wies eine befriedigendere Struktur mit Ladungen zwischen .43 und .63 ohne Doppelladungen auf. Der dritte Faktor DP verfügte ebenfalls über keine Doppelladungen und wies eine Fehlladung auf. Die Ladungen lagen zwischen .39 und .79. Der vierte Faktor sollte die Skala Betroffenheit darstellen und wies eine Fehlladung auf.

Eine Metaanalyse der faktoriellen Validität des MBI verglich 32 internationale, hauptsächlich englischsprachige Studien, die explorative Faktorenanalysen anwandten (Worley, Vassar, Wheeler, & Barnes, 2008). Die in den inkludierten Studien vorgegebenen Versionen des MBI waren sehr heterogen. Zur Analyse wurden entweder orthogonale oder oblique Hauptachsenanalysen oder Hauptkomponentenanalysen angewandt. 24 der Studien fanden eine dreifaktorielle Struktur. Der Rest fand entweder vier oder zwei Faktoren. Obwohl die meisten Studien drei Faktoren erhielten, berichtete der Großteil von Fehl- oder Doppelladungen. Insgesamt konnten lediglich sieben von 32 Studien den Faktor EE exakt replizieren, weil alle Items der Skala auf ihrer vorgesehenen Dimension geladen hatten. Die Skala DP konnte von neun und die Skala PE von zehn Studien exakt repliziert werden. In der Metaanalyse wurden darüber hinaus 21 Studien verglichen, die konfirmatorische

Faktorenanalysen mit der 22-Item-Version des MBI durchführten. 18 der 21 Studien fanden den besten Fit mit einem korrelierten dreifaktoriellen Modell. Worley et al. (2008) betonen, dass die dreifaktorielle Struktur durch die Ergebnisse überwiegend unterstützt werde, die Faktoren allerdings nicht unabhängig seien. Die Art des Zusammenhangs zwischen den Faktoren variere zudem stark zwischen den Studien.

Einen anderen Ansatz zur Überprüfung der faktoriellen Validität des MBI wählten Jordan, Steingen, Terschüren und Harth (2018). Sie führten eine psychometrische Analyse auf Basis der Item- Response-Theorie durch. Zur Erhebung wurde eine deutsche Übersetzung von Barth (1985) mit 22 Items verwendet ($N = 994$). Die Ergebnisse konnten das dreifaktorielle Modell des Instruments nicht unterstützen und werfen Zweifel an der Eindimensionalität der Messung durch die Skalen auf.

Es gibt einige Berichte über problematische Items. So wurde oft beobachtet, dass Item 12 der Skala PE, das inhaltlich einen hohen Energielevel beschreibt, negative Kreuzladungen auf der Skala EE aufwies (Belcastro, Gold, & Hays, 1983; Byrne, 1993; Schaufeli & van Dierendonck, 1993). Auch Item 16 der Skala EE zeigte in einigen Studien bedeutsame Kreuzladungen auf der Skala DP (Byrne, 1993; Enzmann & Kleiber, 1989; Schaufeli & van Dierendonck, 1993). Inhaltlich geht es bei dem Item um Stress durch direkten Kontakt mit anderen Menschen während der Arbeit. Schaufeli und van Dierendonck (1993) empfehlen, die Items aus dem MBI zu streichen.

Worley et al. (2008) berichten in ihrer Metaanalyse von sieben von 21 Studien, die konfirmatorische Faktorenanalysen durchführten und durch die Löschung der Items 12 und 16 einen besseren Fit erzielen konnten. Auch Item 6 der Skala EE zeigte sich in mehreren in der Metaanalyse inkludierten Studien als problematisch, da es nicht auf seiner Skala lud. Item 22 der Skala DP wies häufig generelle Fehlladungen auf (Worley et al., 2008). Inhaltlich beschreibt Item 6 die Anstrengung durch das Arbeiten mit Menschen und Item 22 eine Schuldzuweisung durch Klient_innen.

Kriteriumsvalidität und Diskriminationsfähigkeit. Da es keine einheitlichen diagnostischen Kriterien für Burnout gibt, kann die Fähigkeit eines Instruments, zwischen Burnout-Fällen und Nicht-Burnout-Fällen zu unterscheiden, nur über Umwege untersucht werden. Dafür werden in der Regel stellvertretende Außenkriterien herangezogen. Die Auswahl eines solchen Kriteriums unterliegt jedoch einer gewissen Willkürlichkeit.

So untersuchten Kleijweg, Verbraak und van Dijk (2013), wie gut die Skalen der dänischen Übersetzung des MBI-GS zwischen Personen mit und ohne der DSM-IV-Diagnose „undifferenzierte Somatisierungsstörung“ (American Psychiatric Association, 1994)

differenzieren können. Die Diagnose sollte dabei stellvertretend für klinisches Burnout stehen und wurde von Psycholog_innen anhand des Ergebnisses eines strukturierten Interviews vergeben. Die Autor_innen argumentieren, dass die Diagnose der wesentlichen Symptomatik des Burnout-Syndroms entspreche (Kleijweg et al., 2013). Die Stichprobe bestand aus 419 ambulanten Patient_innen eines dänischen Therapiezentrums für arbeitsbezogene psychische Probleme. Eine Receiver-Operating-Characteristics-Analyse (ROC-Analyse) ergab einen signifikanten Area-Under-The-Curve-Wert (AUC-Wert) von .64 für die Skala EE, was für eine dürftige Diskriminationsfähigkeit spricht. Die beiden anderen Skalen konnten nicht signifikant zwischen Personen mit oder ohne undifferenzierter Somatisierungsstörung differenzieren. Kleijweg et al. (2013) empfehlen deshalb, das MBI-GS nicht als einziges Instrument in der Diagnostik von klinischem Burnout zu verwenden.

Die Diskriminationsfähigkeit der singhalesischen Version des MBI-ES wurde in einem ähnlichen Ansatz von de Silva, Hewage und Fonseka (2013) untersucht. Als Außenkriterium wurde die Diagnose Burnout durch Psychiater_innen herangezogen. Genaue Diagnosekriterien werden von den Autor_innen nicht genannt. Die Stichprobe bestand aus 123 weiblichen Grundschullehrkräften aus Sri Lanka. Der AUC-Wert für das gesamte MBI-ES lag bei .94. Dies spricht für eine exzellente Diskriminationsfähigkeit des Instruments.

Roelen et al. (2015) führten eine ROC-Analyse der niederländischen Version des MBI-GS mit dem Risiko für „long-term sickness absence“ (LTSA), also Langzeitkrankenstand, als Außenkriterium durch. Sie definierten LTSA als durchgehenden Krankenstand im Umfang von mindestens 42 Tagen. Ein Jahr nach Vorgabe des MBI wurden die Werte der 4921 Teilnehmer_innen mit dem Register des niederländischen nationalen Arbeitsgesundheitservice abgeglichen. Dieses Register enthielt Informationen über die Langzeitkrankenstände der Teilnehmer_innen des letzten Jahres. So konnten die Teilnehmer_innen den Gruppen mit oder ohne Risiko für Langzeitkrankenstand zugeordnet werden. Der errechnete AUC-Wert von .68 spricht dafür, dass das MBI-GS nur dürftig zwischen Personen mit oder ohne Risiko für Langzeitkrankenstand unterscheiden kann.

Eine weitere niederländische Studie untersuchte den Zusammenhang zwischen den Skalenwerten des MBI-GS und der für Burnout stellvertretenden ICD-10-Diagnose „Neurasthenie“ (Roelofs, Verbraak, Keijsers, de Bruin, & Schmidt, 2005). Neurasthenie sei Burnout, den Autor_innen zufolge, im Kern sehr ähnlich. Die Diagnose wurde durch Interviews mit klinischen Psycholog_innen vergeben. Zusätzlich wurde die Symptom-Checkliste SCL-90 (Derogatis, 1977) vorgegeben, um Depressivität und allgemeine psychische Belastung zu erheben. Die Stichprobe setzte sich aus 168 Personen zusammen, die

eine Überweisung in ein spezielles Diagnostik- und Therapiezentrum für arbeitsbezogene psychische Probleme erhalten hatten. Die Forscher_innen fanden eine Korrelation der stellvertretenden Burnout-Diagnose von .48 mit der Skala EE, .15 mit der Skala DP und -.03 (n.s.) mit der Skala PE. Partielle Korrelationen mit Depressivität als Kontrollvariable lagen bei .53 für die Skala EE, .22 für die Skala DP und -.10 (n.s.) für die Skala PE. Wurde die allgemeine psychische Belastung (SCL-90-Gesamtwert) berücksichtigt, fand man einen Zusammenhang von .53 mit der Skala EE, .20 mit der Skala DP und -.07 (n.s.) mit der Skala PE. Die Skala EE zeigte demnach die größten Zusammenhänge mit der stellvertretenden Diagnose (Roelofs et al., 2005).

Fazit. Zusammenfassend kann man sagen, dass die Studienergebnisse zur Reliabilität, Validität und Diskriminationsfähigkeit des MBI kein eindeutiges Bild zeichnen und die Anwendung des Instruments in der klinischen Praxis fragwürdig ist. Trotz der widersprüchlichen Forschungslage zu seiner psychometrischen Güte gilt das MBI weiterhin als „Goldstandard“ in der Burnout-Forschung (Burisch, 2014; Dreher, Theune, Kersting, Geiser, & Weltermann, 2019; Schaufeli et al., 2009). Schaufeli et al. (2009) gehen sogar so weit, das gesamte Konzept des Burnout-Syndroms mit dem MBI gleichzusetzen: Burnout ist, was das MBI misst und umgekehrt (S. 211). Allerdings gibt es auch Kritik an der weitverbreiteten Verwendung des Instruments, nicht zuletzt aufgrund der dreidimensionalen Burnout-Definition, auf der das MBI basiert (Burisch, 2014; Kristensen et al., 2005; Shirom, 2005).

Ein relativ junges Instrument, das einen ganz anderen Weg zur Erhebung des Burnout-Syndroms einschlägt, ist die Burnout-Screening-Skalen-Batterie von Hagemann und Geuenich (2009).

Burnout-Screening-Skalen

Bei den Burnout-Screening-Skalen I und II (BOSS I und II) handelt es sich um zwei Selbstbeschreibungsfragebögen zur Erfassung von aktuellen psychischen, körperlichen und psychosozialen Beschwerden, wie sie in der Regel beim Burnout-Syndrom auftreten. Die Items wurden praxisorientiert aus Expertenmeinungen von therapeutischem Personal, Forschungsliteratur und der therapeutischen Erfahrung der Autor_innen abgeleitet (Geuenich & Hagemann, 2014).

Geuenich und Hagemann (2014) beschreiben Burnout als einen Zustand psychischer und mentaler Erschöpfung. Die theoretische Grundlage der BOSS I und II bilden systemische, psychosomatische und Stress-Coping-Modelle. Die Items beziehen sich auf klinische

Symptome und Beschwerden im zwischenmenschlichen Bereich. Im systemischen Sinne liege die Ursache von Burnout in einem aus dem Gleichgewicht geratenen Beziehungssystem. Aus psychosomatischer Sicht sei Burnout auf die psychophysische Verarbeitung von Stressoren sowie inneren und äußeren Konflikten zurückzuführen. Stresstheoretisch sei Burnout das Resultat eines Stressors, der übermäßig intensiv und/oder lange auf das Individuum einwirke. Ziel des Instruments sei es, eine ganzheitliche Perspektive auf das individuelle Beschwerdebild zu schaffen (Geuenich & Hagemann, 2014).

Die BOSS I besteht aus 30 Items. Sie erfasst bereichsspezifische Beschwerden und Belastungen in den vier Subskalen *Beruf* (B, 10 Items), *Eigene Person* (EP, 10 Items), *Familie* (FA, 5 Items) und *Freunde* (FR, 5 Items). Die Antworten werden auf einer sechsstufigen Likert-Skala von 0 (*trifft nicht zu*) bis 5 (*trifft stark zu*) erfasst. Der Beurteilungszeitraum beträgt drei Wochen. Hohe Werte in den einzelnen Skalen stehen für eine hohe Ausprägung an Belastungen und Beschwerden in dem jeweiligen Bereich (Geuenich & Hagemann, 2014).

Die BOSS II besteht ebenfalls aus 30 Items und erfasst mit drei Subskalen *Körperliche Beschwerden* (KOEB), *Kognitive Beschwerden* (KOB) und *Emotionale Beschwerden* (EB) zu jeweils zehn Items. Das Antwortformat gleicht dem der BOSS I. Der Beurteilungszeitraum beträgt sieben Tage. Hohe Werte in den einzelnen Skalen stehen für eine hohe Ausprägung an Beschwerden in dem jeweiligen Bereich (Geuenich & Hagemann, 2014). Das Verfahren verfügt über eine für Deutschland repräsentative Normstichprobe aus dem Jahr 2008 ($N = 300$). Die Rohwerte können damit in T-Werte umgerechnet werden. Für BOSS I und II können drei verschiedene Globalwerte über alle Skalen hinweg bestimmt werden: Gesamt-, Intensitäts- und Breitenwerte. Ein hoher Gesamtwert steht für eine starke Belastung, ein hoher Intensitätswert für eine hohe Intensität der Beschwerden, und ein hoher Breitenwert deutet darauf hin, dass die Beschwerden in vielen Bereichen auftreten. Für die Verdachtsdiagnose berufsbedingtes Burnout-Syndrom ist hauptsächlich die BOSS I relevant, wobei die BOSS II die Diagnose unterstützen kann, indem sie das Gesamtbild vervollständigt. Ein T-Wert > 59 in einem oder mehreren Globalwerten der BOSS I gilt als Cut-off für die Verdachtsdiagnose. Je stärker der Wert vom Mittelwert ($T = 50$) abweicht und je mehr Kennwerte erhöht sind, desto wahrscheinlicher ist die Diagnose Burnout-Syndrom (Geuenich & Hagemann, 2014).

Reliabilität. Die Cronbachs-Alpha-Werte der BOSS I, ermittelt anhand der Eichstichprobe, lagen bei .78 für die Skala Beruf, .87 für die Skala EP, .83 für die Skala FA und .75 für die Skala FR. Für die BOSS II fand man für die Skala KOEB ein Alpha von .80 und für die Skalen KOB und EB ein Alpha von .91. Die Retest-Reliabilität ($N = 91$) nach

zwei Tagen lag für die BOSS I zwischen .84 und .91 und für die BOSS II zwischen .90 und .91. Darüber hinaus wurde die interne Konsistenz für drei weitere Stichproben berechnet: eine Internetstichprobe ($N = 2634$ für BOSS I, $N = 631$ für BOSS II), Ärzt_innen ($N = 74$) und Psychotherapiepatient_innen ($N = 377$). Die gemittelten Cronbachs-Alpha-Werte der BOSS I über alle vier Stichproben hinweg betrugen .86 für die Skala Beruf, .90 für die Skala EP, .87 für die Skala FA und .80 für die Skala FR. Für die BOSS II lagen die gemittelten Alpha-Werte bei .82 für die Skala KOEB, .90 für die Skala KOB und .90 für die Skala EB (Geuenich & Hagemann, 2014).

Bis auf wenige Ausnahmen war die korrigierte Trennschärfe der Items der BOSS I in der Eichstichprobe gut. Items mit einer unzureichenden Trennschärfe waren Item 3 („Um Fehler zu vermeiden, arbeite ich oft bis spät in die Nacht.“) und 9 („Um beruflichen Stress abbauen zu können, greife ich auf Alkohol und/oder Medikamente zurück.“) der Skala B sowie Item 30 („Es fällt mir schwer, mich abzugrenzen, nein zu sagen.“) der Skala FR. Aus inhaltlichen Gründen wurden die Items 3 und 9 von den Autor_innen beibehalten. Item 30 wurde ebenfalls beibehalten, weil es in der Internetstichprobe eine zufriedenstellende Itemtrennschärfe aufwies. Die Trennschärfe der Items der BOSS II waren ebenfalls überwiegend gut, bis auf Item 1 („Mich quälen heftige Kopfschmerzen.“) und 10 („Meine Blutdruckwerte sind zu hoch/schwanken stark.“) der Skala KOEB, die unter .40 lagen. Die Items wurden aus inhaltlichen Gründen und wegen ihrer zufriedenstellenden Kennwerte in der Internetstichprobe beibehalten (Geuenich & Hagemann, 2014).

Da es bisher nur sehr wenige Studien gibt, die die BOSS I und II angewendet haben, und der Großteil davon keine internen Konsistenzen berichtet, gibt es nur sehr wenig Information zur Reliabilität des Instruments, die nicht von den Testautor_innen stammt. Die BOSS I und II wurde in einer Studie von Steinlin et al. (2016) bei einer Stichprobe von pädagogischen Mitarbeiter_innen ($N = 319$) eingesetzt und lieferte Cronbachs-Alpha-Werte zwischen .77 und .90. Im Rahmen einer Dissertation wurden zwei Skalen der BOSS I und BOSS II Flugbegleiter_innen ($N = 420$) vorgegeben (Kinschel, 2013). Dabei ergaben sich Alpha-Koeffizienten von .85 für die Skala Beruf und .89 für die Skala KOEB. Zusammengefasst scheinen BOSS I und II laut den Testautor_innen über eine akzeptable bis gute und BOSS II über eine gute bis ausgezeichnete interne Konsistenz zu verfügen. Die wenigen Ergebnisse anderer Studien sind mit diesen Angaben im Einklang.

Faktorielle Validität. Die Testautor_innen führten mit den Daten der Eichstichprobe eine Hauptkomponentenanalyse mit Varimax-Rotation durch, um die faktorielle Struktur des Fragebogens zu untersuchen (Geuenich & Hagemann, 2014). Es wurden mehrere Lösungen

mit jeweils drei bis sieben Faktoren für die BOSS I berechnet. Die theoretisch festgelegte vierfaktorielle Struktur des Fragebogens konnte nicht gefunden werden. Die fünffaktorielle Struktur passte am besten im Sinne der Einfachstruktur. Die einzelnen Faktoren setzten sich dabei jeweils aus Items unterschiedlicher Skalen zusammen, mit Ausnahme des dritten Faktors, der erwartungsgemäß die Skala FA abbildete. Inhaltlich beschrieben die Faktoren laut den Autor_innen (1) psychosoziale und psychosomatische Belastungen, (2) berufliche Belastungen mit depressiv-defizitorientierten Beschwerden, (3) familiäre und partnerschaftliche Belastungen, (4) einen Prozess der Dekompensation oder inneren Kündigung und (5) angstbesetzte Überforderung und Unzufriedenheit im Beruf. Man sollte allerdings erwähnen, dass die von den Testautor_innen beschriebene Lösung einige Doppelladungen aufwies. Für die BOSS II wurden ebenfalls mehrere Lösungen berechnet. Hier zeigte sich, dass eine einfaktorielle Lösung das Kriterium der Einfachstruktur am ehesten erfüllte, jedoch deutlich weniger Varianz aufklärte als eine alternative Struktur mit fünf Faktoren. Da sich die letztere jedoch schlecht interpretieren ließ, befanden die Autor_innen eine vierfaktorielle Lösung für am besten geeignet. Die theoretisch festgelegte dreifaktorielle Struktur der BOSS II konnte ebenfalls nicht gestützt werden. Inhaltlich bildeten die Faktoren laut den Autor_innen (1) die eigenen Handlungsfreiräume und Kontrollmöglichkeiten, (2) depressive Symptome, (3) psychosomatische und funktionale Beschwerden außerhalb des Herz-Kreislauf-Systems und (4) Beschwerden des Herz-Kreislauf- und Atmungssystems ab. Auch bei dieser Lösung kam es zu Doppelladungen, und teilweise wurden Items nicht den Faktoren zugeordnet, auf denen sie am höchsten luden (Geuenich & Hagemann, 2014).

Bisher wurde die faktorielle Validität der BOSS I und II noch nicht von unabhängigen Autor_innen untersucht. Durch die Ergebnisse der Faktorenanalyse der Testautor_innen lässt sich die Skalenstruktur der Instrumente nicht stützen. Geuenich und Hagemann (2014) argumentieren, dass die faktorenanalytisch ermittelten Lösungen keine Vorteile für die Interpretation durch die Anwender_innen bringen und daher die theoretische Skalenzuteilung beibehalten wurde. Die theoretische Skalenstruktur der BOSS I und II habe den Vorteil, Belastungen in spezifischen Lebens- und Symptombereichen aufzuzeigen. Da die empirisch ermittelten Faktoren zum Großteil Items aus unterschiedlichen Bereichen beinhalten, ginge die bereichsspezifische Information bei einer Anpassung verloren. Des Weiteren wird bemerkt, dass die theoretisch formulierten Skalen intern konsistent seien (Geuenich & Hagemann, 2014).

Kriteriumsvalidität und Diskriminationsfähigkeit. Die Testautor_innen untersuchten den Zusammenhang zwischen BOSS I und II und dem „Goldstandard“ der

Burnouterhebung, dem MBI (Geuenich & Hagemann, 2014). Die Instrumente wurden dazu einer Stichprobe von Berufsschullehrkräften ($N = 237$) vorgegeben. Es zeigten sich mittlere bis hohe Korrelationen der Skalen der BOSS I und II mit dem Gesamtwert des MBI ($r_{\min} = .40$, $r_{\max} = .64$). Durchwegs hohe Zusammenhänge wiesen die Skalen der BOSS I und II mit der Skala EE des MBI auf ($r_{\min} = .53$, $r_{\max} = .75$). Kleine bis mittlere Zusammenhänge zeigten sich zwischen den Skalen der BOSS I und der MBI-Skala DP ($r_{\min} = .15$, $r_{\max} = .40$). Die umgepolte Skala PE des MBI korrelierte lediglich mit den Skalen B, EP, FR, KOB und EB signifikant niedrig bis mittel ($r_{\min} = .13$, $r_{\max} = .30$). Die Skalen FR und KOEB korrelierten nicht signifikant mit der MBI-Skala PE. Die Autor_innen interpretieren die Zusammenhänge mit dem MBI-Gesamtwert als zufriedenstellend. Sie weisen aber auch auf die Unterschiede der beiden Verfahren hin, die sich in den niedrigeren Korrelationen mit den MBI-Skalen DP und PE abzeichnen. BOSS I und II bilden allerdings den Aspekt der emotionalen Erschöpfung von Burnout in ausreichendem Maße ab (Geuenich & Hagemann, 2014).

Zur Untermauerung der konvergenten Validität untersuchten Geuenich und Hagemann (2014) den Zusammenhang der BOSS I und II mit dem Trier Inventar zum chronischen Stress (TICS, Schulz, Schlotz, & Becker, 2004) an einer Stichprobe von Mediziner_innen ($N = 59$). Die Skalen der BOSS I und II korrelierten allesamt hoch mit den TICS-Skalen *soziale Isolation* ($r_{\min} = .50$, $r_{\max} = .64$) und *chronische Besorgnis* ($r_{\min} = .50$, $r_{\max} = .78$) und mittel bis hoch mit den TICS-Skalen *Arbeitsunzufriedenheit* ($r_{\min} = .43$, $r_{\max} = .66$), *mangelnde soziale Anerkennung* ($r_{\min} = .39$, $r_{\max} = .58$) und *soziale Spannungen* ($r_{\min} = .34$, $r_{\max} = .59$). Darüber hinaus korrelierten, mit Ausnahme der Skala KOEB, alle Skalen der BOSS I und II hoch mit den TICS-Skalen *Arbeitsüberlastung* ($r_{\min} = .28$, $r_{\max} = .61$) und *Arbeitsüberforderung* ($r_{\min} = .29$, $r_{\max} = .67$). Die TICS-Skala *soziale Überforderung* korrelierte nur mit den Skalen B, EP, FR und KOB signifikant ($r_{\min} = .31$, $r_{\max} = .33$). Die TICS-Skala *Erfolgsdruck* wies keine signifikanten Korrelationen auf.

Die BOSS I wurde mit derselben Fragestellung einer Stichprobe von Psychotherapiepatient_innen ($N = 74$), von denen die Hälfte die Diagnose Burnout erhalten hatten, vorgelegt (Geuenich & Hagemann, 2014). Die Korrelationen zwischen den Skalen der BOSS I und den Skalen des TICS fielen in dieser Stichprobe, im Vergleich zu den Mediziner_innen, niedriger aus. Die Skala B zeigte signifikante Korrelationen, überwiegend im mittleren Bereich, mit allen neun TICS-Skalen außer der Skala soziale Spannungen ($r_{\min} = .27$, $r_{\max} = .65$). Kleine bis mittlere Zusammenhänge zeigten sich zwischen der Skala EP und den TICS-Skalen Arbeitsunzufriedenheit, Arbeitsüberforderung und chronische Besorgnis ($r_{\min} = .27$, $r_{\max} = .47$). Die Skala FA korrelierte niedrig bis mittel mit den TICS-Skalen soziale

Spannungen, soziale Isolation und chronische Besorgnis ($r_{\min} = .23$, $r_{\max} = .45$). Weitere niedrige bis mittlere Zusammenhänge fanden sich zwischen den Skalen FR und den TICS-Skalen Arbeitsunzufriedenheit, Arbeitsüberforderung, mangelnde soziale Anerkennung, soziale Isolation, und chronische Besorgnis ($r_{\min} = .24$, $r_{\max} = .48$). Darüber hinaus fanden sich keine weiteren signifikanten Korrelationen.

Der Zusammenhang zwischen BOSS I und TICS wurde auch anhand einer Stichprobe von Psychologiestudierenden ($N = 58$) untersucht (Geuenich & Hagemann, 2014). Dabei zeigten sich mittlere bis hohe Korrelationen zwischen allen Skalen der BOSS I und TICS ($r_{\min} = .32$, $r_{\max} = .64$).

Die Testautor_innen gaben die BOSS I und II einer Stichprobe von 38 deutschen Psychotherapiepatient_innen vor, die Burnout entweder als Erstdiagnose ($n = 9$) oder als Zusatzdiagnose ($n = 29$) erhalten hatten (Geuenich & Hagemann, 2014). Ziel der Untersuchung war es, den Anteil an Personen zu ermitteln, die erwartungsgemäß die Cut-off-Werte der einzelnen Skalen überschritten. Bei der BOSS I erzielten 82% signifikant erhöhte Gesamtwerte in der Skala B, 95% in der Skala EP, 50% in der Skala FA und 92% in der Skala FR. Bis auf das Ergebnis der Skala FA deutet dies auf eine relativ gute Übereinstimmung hin. Die Skalenwerte der BOSS II waren in 55% der Fälle in der Skala KOEB, in 82% der Fälle in der Skala KOB und in 84% in der Skala EB erhöht. Die Skalen KOB und EB weisen demnach eine relativ hohe Diskriminationsfähigkeit auf. Weitere Ergebnisse zur Diskriminationsfähigkeit der BOSS I und II finden sich in der Literatur nicht.

Fazit. Zusammenfassend kann festgestellt werden, dass die Ergebnisse der Testautor_innen für eine hohe Reliabilität der BOSS I und II sowie eine niedrige faktorielle Validität sprechen. Die Belege der Kriteriumsvalidität schwanken je nach Skala, Stichprobe und Kriterium deutlich. Die Diskriminationsfähigkeit der meisten Skalen scheint gut zu sein. Die Skalen FA und KOEB scheinen allerdings kaum besser als ein Münzwurf zwischen Personen mit oder ohne Burnout unterscheiden zu können. Eine unabhängige umfangreiche Begutachtung der Testgüte der BOSS I und II gibt es bis dato noch nicht.

Zielsetzung

In Anbetracht der widersprüchlichen Ergebnisse zur Testgüte des MBI, insbesondere der deutschen Übersetzung, ist das Hinterfragen seines Prädikats als „Goldstandard“ zur Erfassung des Burnout-Syndroms angebracht. Die Burnout-Screening-Skalen I und II unterscheiden sich in ihrem Ansatz vom MBI und könnten möglicherweise eine vielversprechende Alternative darstellen. Allerdings gibt es noch kaum Literatur, die sich

eingehend mit dem Instrument beschäftigt hat. Ziel dieser Untersuchung soll daher sein, die Testgüte des MBI und der BOSS I und II an einer deutschsprachigen berufstätigen Erwachsenenstichprobe zu untersuchen und abschließend miteinander zu vergleichen. Dafür soll eine Auswahl von verschiedenen Indikatoren zur Beurteilung der psychometrischen Güte herangezogen werden.

Methodik

Untersuchungsdurchführung

Für die Stichprobe wurden berufstätige Personen ab 18 Jahren aus dem deutschsprachigen Raum rekrutiert, die seit mindestens drei Jahren ihre aktuelle Berufstätigkeit ausübten. Von der Teilnahme ausgeschlossen waren Personen, die sich vorrangig als Studierende bezeichneten.

Die Daten wurden im Rahmen des Seminars Theorie und Empirie wissenschaftlichen Arbeitens (Klinische Psychologie und Gesundheitspsychologie) unter der Leitung von Mag. Dr. Reinhold Jagsch erhoben. Jede_r Teilnehmer_in sollte mindestens 16 Personendaten erheben. Insgesamt wurde eine Stichprobe im Umfang von mindestens 304 Personen angestrebt. Um eine möglichst ausgewogene Stichprobe zu erhalten, wurde eine Stratifizierung auf Basis des Geschlechts (weiblich vs. männlich), Bildungsgrads (Akademiker_in vs. Nicht-Akademiker_in) und Alters (unter 40 Jahren vs. über 40 Jahren) vorgenommen. Jede_r Seminarteilnehmer_in sollte mindestens zwei Personen aus jeder der acht Kategorien, die sich aus den möglichen Kombinationen der Faktoren ergeben, rekrutieren. Die Personen wurden mittels persönlichen Kontakts angeworben. Die Erhebung erfolgte von Mai 2018 bis Jänner 2019.

Die Testbatterie wurde im Paper-Pencil-Format vorgegeben und erhob soziodemographische Daten, Burnout, subjektives Stresserleben, Copingstile, die subjektive Beeinträchtigung durch körperliche und psychische Symptome, Alkohol- und Nikotinkonsum sowie gesundheitsbezogene Lebensqualität. Im Folgenden sollen jedoch nur die für diese Arbeit interessierenden Merkmale näher beschrieben werden.

Studiendesign

Bei der vorliegenden Studie handelt es sich um eine quantitative Querschnittsuntersuchung.

Untersuchungsinstrumente

Soziodemographischer Fragebogen. Es werden folgende soziodemographischen Variablen erfasst: Geschlecht, Alter, Körpergröße, Körpergewicht, Nationalität, Beziehungsstatus, höchste abgeschlossene Ausbildung, Beruf, derzeitiger Beschäftigungsstatus, Anzahl der durchschnittlichen Arbeitsstunden pro Woche, Berufserfahrung insgesamt und an der aktuellen Arbeitsstelle (in Jahren), die Anzahl der Krankenstandstage in den letzten drei Monaten, kritische Lebensereignisse innerhalb des letzten Jahres und schwere Erkrankungen oder Unfälle innerhalb des letzten Jahres.

Maslach Burnout Inventory. Das Verfahren wurde im Theorieteil dieser Arbeit bereits umfangreich beschrieben. Für die Erhebung wurde eine deutsche Übersetzung des MBI mit 22 Items herangezogen („Der Maslach Burnout Inventory-Test,” n.d.). Die Formulierungen „Kollegen“, „Kollegen/Vorgesetzten“, „Klienten/Kollegen“ und „Kollegen/Mitarbeitenden“ ersetzen in dieser Version das Wort „Patienten“.

Burnout-Screening-Skalen I und II. Das Instrument wurde im Theorieteil dieser Arbeit bereits umfangreich beschrieben.

Perceived Stress Questionnaire (PSQ). Das Instrument von Fliege, Rose, Arck, Levenstein und Klapp (2001) eignet sich zur Erhebung des subjektiven Stresserlebens. Es handelt sich dabei um eine deutsche Übersetzung und faktorenanalytische Restrukturierung des anglo-amerikanischen Originals von Levenstein et al. (1993). Die Originalversion besteht aus 30 Items, die zu insgesamt sieben Skalen zusammengefasst werden. Die deutsche Version umfasst 20 Items und vier Skalen. Der Fragebogen weist ein vierstufiges Antwortformat von 1 (*fast nie*) bis 4 (*meistens*) auf. Die Items beziehen sich auf die letzten vier Wochen. Die drei Subskalen *Sorgen*, *Anspannung* und *Freude* bilden die interne Stressreaktion der Personen ab, während die Skala *Anforderungen* die externen Stressoren berücksichtigt (Fliege et al., 2001). Der Wertebereich der Skalen liegt zwischen 0 und 1, wobei ein hoher Wert für eine hohe Ausprägung der Dimension steht. Darüber hinaus kann ein Gesamtwert (ebenfalls von 0 bis 1) berechnet werden. Ein hoher Gesamtwert spricht für ein hohes Ausmaß an Stresserleben. Die durchschnittliche Bearbeitungszeit liegt bei ungefähr fünf Minuten (Fliege et al., 2005). Eine Reliabilitätsanalyse der deutschsprachigen 20-Item-Version durch die Autor_innen zeigte sehr gute Cronbachs-Alpha-Werte zwischen .90 und .94 für den Gesamtwert und akzeptable bis gute Cronbachs-Alpha-Werte zwischen .76 und .87 für die vier Skalen (Fliege et al., 2005).

Fragestellungen und Hypothesen

Fragestellung 1: Vollständigkeit der Daten. Wie vollständig sind die erhobenen Daten des MBI, BOSS I und BOSS II auf Itemebene und auf Skalenebene?

Kommt es innerhalb einer Skala zu einem hohen Anteil an fehlenden Werten, kann dies zu Problemen beim Aufsummieren der Skalenwerte führen. Die Summenscores können mit unvollständigen Daten nicht mit derselben Zuverlässigkeit berechnet werden. Die Analyse der Vollständigkeit auf Item- und Skalenebene kann Hinweise auf etwaige Probleme mit der Formulierung der Items oder dem Antwortformat liefern (Ware & Gandek, 1998).

Fragestellung 2: Häufigkeitsverteilung der Antwortkategorien. Wie gestaltet sich die Häufigkeitsverteilung der Antwortkategorien auf Itemebene und auf Skalenebene in der Gesamtstichprobe und in den vier Gruppen mit unterschiedlichem Stresserleben (niedrig, eher niedrig, eher hoch, hoch)? Gibt es Boden- oder Deckeneffekte?

Die Analyse der Häufigkeiten soll sich auf die folgenden Kriterien konzentrieren: Wurden alle Antwortkategorien genutzt? Ist die Verteilung schief? Gibt es bedeutende Boden- oder Deckeneffekte? Stellt sich die Verteilung verschieden über die vier Gruppen mit unterschiedlichem subjektiven Stresserleben (gemessen mit dem PSQ) hinweg dar? Eine schiefe Verteilung könnte problematisch für die Nutzung von Pearson-Korrelationskoeffizienten in der weiterführenden Analyse sein. Gibt es in manchen Gruppen Boden- oder Deckeneffekte und in anderen nicht, so könnte das ein Hinweis darauf sein, dass das Instrument für bestimmte Populationen besser geeignet ist.

Fragestellung 3: Reliabilität. Wie gestaltet sich die interne Konsistenz der Skalen und die Trennschärfe der Items des MBI, BOSS I und BOSS II?

Diese Fragestellung soll sich mit der Messgenauigkeit der Instrumente auseinandersetzen. Cronbachs Alpha ist einer der gebräuchlichsten Reliabilitätskoeffizienten und gibt Aufschluss über die interne Konsistenz einer Skala (Bortz & Döring, 2006). Je höher die Items im Durchschnitt miteinander korrelieren, umso höher ist die interne Konsistenz einer Skala (Schermelleh-Engel & Werner, 2012). Die Itemtrennschärfe beschreibt, wie gut der Punktwert eines einzelnen Items den Gesamtscore seiner Skala repräsentiert. Fällt die Trennschärfe eines Items zu niedrig aus, kann es als ein schlechter Indikator für das Konstrukt, das die Skala zu erheben versucht, gesehen werden (Bortz & Döring, 2006). Eine akzeptable Skala sollte mindestens ein Cronbachs Alpha von .70 (Hussy, Schreier, &

Echterhoff, 2013) und Itemtrennschärfen größer oder gleich .40 (Kelava & Moosbrugger, 2012) aufweisen.

Fragestellung 4: Faktorielle Validität. Kann die durch die Testautor_innen festgelegte Faktorenstruktur des MBI, BOSS I und BOSS II repliziert werden?

Exemplarisch werden hier lediglich die Hypothesen zu einem der Verfahren angeführt.

H_0 (4.1): Die von den Testautor_innen festgelegte dreifaktorielle Faktorenstruktur des MBI kann nicht repliziert werden. Die Items laden nicht eindeutig auf ihren Skalen.

H_1 (4.1): Die von den Testautor_innen festgelegte dreifaktorielle Faktorenstruktur des MBI kann repliziert werden. Die Items laden eindeutig auf ihren Skalen.

Analog zu H_0 bzw. H_1 (4.1) sind die Hypothesen H_0 bzw. H_1 (4.2.) und (4.3) zu der vierfaktoriellen Faktorenstruktur der BOSS I und der dreifaktoriellen Faktorenstruktur der BOSS II formuliert.

Die jeweiligen Items sollten bedeutsam ($\geq |.40|$, Pituch & Stevens, 2015) auf der ihnen zugeschriebenen Skala laden und keine Kreuzladungen aufweisen. Eine Kreuzladung tritt dann auf, wenn ein Item bedeutsam ($\geq |.40|$) auf einer anderen Skala lädt.

Fragestellung 5: Diskriminationsfähigkeit. Differenzieren die einzelnen Skalen der Verfahren MBI, BOSS I und BOSS II signifikant zwischen Personen mit niedrigem und hohem subjektiven Stresserleben?

Exemplarisch werden hier lediglich die Hypothesen zu einer der Skalen des MBI, BOSS I und BOSS II angeführt.

H_0 (5.1): Es gibt keinen Unterschied zwischen Personen mit niedrigem und hohem subjektiven Stresserleben in der Skala Emotionale Erschöpfung des MBI.

H_1 (5.1): Es gibt einen Unterschied zwischen Personen mit niedrigem und hohem subjektiven Stresserleben in der Skala Emotionale Erschöpfung des MBI.

Analog zu H_0 bzw. H_1 (5.1) sind die Hypothesen H_0 bzw. H_1 (5.2.) bis (5.10) zu den weiteren beiden Skalen des MBI, den vier Skalen der BOSS I und den drei Skalen der BOSS II formuliert.

Diese Fragestellung soll untersuchen, ob und wie gut die Instrumente zwischen Personen mit hohem und niedrigem Stresserleben unterscheiden können. Laut den meisten Definitionen entsteht Burnout aufgrund von chronischem Stress (Geuenich & Hagemann, 2014; Maslach et al., 2001; Pines & Aronson, 1988; Schaufeli & Enzmann, 1998; Shirom, 2005). Geht man davon aus, dass Burnout eine Reaktion auf chronische Stressoren ist, dann

sollte ein Instrument zur Messung des Burnout-Syndroms zwischen Menschen auf Basis der Ausprägung ihres subjektiven Stresserlebens differenzieren können.

Fragestellung 6: Relative Validität. Wie stellt sich die relative Validität der einzelnen Skalen des MBI, BOSS I und BOSS II dar?

Diese Fragestellung soll die Diskriminationsfähigkeit der Skalen relativ zueinander betrachten. Auch hier soll das subjektive Stresserleben als Außenkriterium herangezogen werden. Es soll untersucht werden, welche Skalen am besten zwischen den Gruppen mit niedrigem, eher niedrigem, eher hohem und hohem Stresserleben differenzieren können.

Fragestellung 7: Vergleich der Verfahren. Welches Verfahren (MBI oder BOSS I und BOSS II) eignet sich aufgrund seiner psychometrischen Güte besser zur Erfassung des Burnout-Syndroms bei berufstätigen Erwachsenen?

Abschließend soll ein Resümee gezogen werden in Bezug auf die Ergebnisse der vorangehenden Fragestellungen. Im Fokus soll der Vergleich des MBI mit BOSS I und II stehen.

Datenaufbereitung und statistische Auswertung

Die Datenaufbereitung und die statistische Auswertung erfolgten mithilfe der Programme IBM SPSS Statistics 25, Microsoft Excel 2010 und MATLAB R2016. Die Daten wurden überprüft, fehlerhafte Werte entfernt und fehlende Werte einzelner Items interpoliert, sofern mindestens 80% der übrigen Werte der jeweiligen Skala vorhanden waren.

Bildung der Gruppen auf Basis des subjektiven Stresserlebens. Der PSQ-Gesamtwert wurde zur besseren Anschaulichkeit auf einen Wertebereich von 0 bis 100 transformiert. Es wurden zwei unterschiedliche Gruppeneinteilungen auf Basis des PSQ-Gesamtwerts vorgenommen. Einteilung A enthält vier Gruppen: (1) niedriges, (2) eher niedriges, (3) hohes und (4) eher hohes subjektives Stresserleben. Ursprünglich war die Einteilung in vier Perzentilklassen des Wertebereichs des PSQ vorgesehen, was allerdings dazu führte, dass die Fallzahlen in den beiden Gruppen mit eher hohem und hohem subjektiven Stresserleben zu gering ausfielen (weniger als 30 Fälle pro Gruppe) und die vier Gruppen ungleich groß waren. Deshalb wurde die Stichprobe anhand der empirisch ermittelten Quartile in die vier Gruppen aufgeteilt. Die binäre Gruppeneinteilung B in eine Gruppe mit niedrigem und eine mit hohem subjektiven Stresserleben erfolgte durch die Trennung der Stichprobe am empirischen Median. Auch hier war die geplante Aufteilung

anhand des Wertebereichs des PSQ nicht sinnvoll, da bei diesem Vorgehen die Gruppe mit niedrigem Stresserleben mehr als 79% der Fälle enthielt und die Gruppengrößen demnach sehr ungleich waren.

Datenvollständigkeitsanalyse. Der prozentuale Anteil an fehlenden Werten für jedes Item und für jede Skala wurde berechnet. Des Weiteren wurde für jede Skala untersucht, für wieviel Prozent der Stichprobe Summenwerte aufgrund der Vollständigkeit der Items berechnet werden konnten (weniger als 20% der Items einer Skala fehlend).

Deskriptive Datenanalyse. Die Häufigkeitsverteilung der Antwortkategorien wurde auf Item- und Skalenebene für die gesamte Stichprobe sowie für die vier Gruppen der Einteilung A mit niedrigem, eher niedrigem, eher hohem und hohem subjektiven Stresserleben betrachtet. Mittelwerte, Standardabweichung, Schiefe und Kurtosis wurden berechnet. Zur besseren Vergleichbarkeit wurden die Skalenrohwerte des MBI, BOSS I und II transformiert, sodass sie einen Wertebereich von 0 bis 100 abbildeten (Mancuso, Peterson, & Charlson, 2001). Dazu wurde der niedrigstmögliche Rohwert der jeweiligen Skala vom tatsächlichen Rohwert abgezogen und im Anschluss durch den Rohwertebereich der Skala dividiert. Das Ergebnis wurde mit 100 multipliziert. Als Richtwert für die Beurteilung der Schiefe wurde eine z-Standardisierung derselben vorgenommen. Dazu wurde die Schiefe durch ihren Standardfehler geteilt. Ist ein dadurch erhaltener Wert größer als $z = \pm 1.96$, so weicht der Wert der Schiefe mit einer Sicherheit von 95% bedeutsam von 0 (symmetrisch) ab. Analog erfolgte die Berechnung der z-standardisierten Kurtosis. Darüber hinaus wurden die transformierten Skalenwerte der Skala PE des MBI umgepolt, um sie besser vergleichbar zu machen. Um mögliche Boden- oder Deckeneffekte zu untersuchen, wurde der Prozentsatz von Personen berechnet, die den jeweils niedrigstmöglichen (Bodeneffekt) oder höchstmöglichen (Deckeneffekt) Summenwert einer Skala erreichten.

Reliabilitätsanalyse. Zur Bestimmung der internen Konsistenz wurde der Reliabilitätskoeffizient Alpha (Cronbach, 1951) berechnet. Im Zuge der deskriptiven Datenanalyse stellte sich heraus, dass die Verteilungen auf Skalen- und auf Itemebene (mit Ausnahme von zwei Items des MBI) deutlich schief waren ($z > 1.96$). Deshalb wurde bei der Berechnung von Cronbachs Alpha anstelle einer Pearson-Korrelationsmatrix auf eine polychorische Korrelationsmatrix zurückgegriffen, um einen ordinalen Alpha-Koeffizienten zu berechnen (Gadermann, Guhn, & Zumbo, 2012). Ordinale Cronbachs-Alpha-Werte sind bessere Reliabilitätsschätzer bei Daten, die mit Likert-Skalen erhoben wurden und eine deutlich schiefe Verteilung und/oder wenige Antwortkategorien aufweisen (Gadermann et al.,

2012). Cronbachs Alpha tendiert in solchen Fällen dazu, die interne Konsistenz zu unterschätzen.

Die Berechnung der polychorischen Korrelationsmatrizen erfolgte mittels Two-Step-Methode durch das SPSS-Plugin Essentials for R und dessen Erweiterung SPSSINC_HETCOR. Die Berechnung der ordinalen Alpha-Werte erfolgte, mit Unterstützung von Microsoft Excel, durch die Anwendung einer Formel für Cronbachs Alpha, die auf einem Ein-Faktor-Modell basiert (McDonald, 1985, zitiert nach Gadermann et al., 2012). Dafür wurde für jede Skala eine unrotierte Hauptachsenanalyse mit einem einzigen Faktor auf Basis der polychorischen Korrelationsmatrizen in SPSS berechnet.

Trennschärfeanalyse. Die Itemtrennschärfe (korrigierte Item-Skala-Korrelation) wurde berechnet. Auch hier wurden aufgrund der schiefen Verteilungen auf Skalen- und Itemebene polychorische Korrelationen anstelle von Pearson-Korrelationen verwendet.

Hauptachsenanalyse. Zur Untersuchung der faktoriellen Validität wurde eine Hauptachsenanalyse mit Varimax-Rotation in SPSS durchgeführt. Aufgrund der vorliegenden schiefen Verteilungen der Daten wurde auf polychorische Korrelationsmatrizen zurückgegriffen. Die Abbruchkriterien für die Faktorenextraktion wurden mittels Parallelanalyse ermittelt (Horn, 1965). Diese basierte ebenfalls auf den zuvor berechneten polychorischen Korrelationsmatrizen. Sie wurde nach der von Garrido, Abad und Ponsoda (2013) beschriebenen Methode berechnet. Dazu wurden die von Garrido (2017) zur Verfügung gestellten Syntax-Files für MATLAB verwendet.

Receiver-Operating-Characteristics-Kurven (ROC-Kurven). Hierbei handelt es sich um eine statistische Analysemethode, die Aufschluss über die Diskriminationsfähigkeit eines Messinstruments geben kann. Konkret ist eine ROC-Kurve die grafische Darstellung der Sensitivität und Spezifität. Für diese Analyse ist ein binäres Außenkriterium notwendig. In dieser Untersuchung wurde der Gesamtwert der subjektiven Stressbelastung (erhoben mittels PSQ) als Außenkriterium herangezogen. Die zuvor beschriebene binäre Gruppeneinteilung B, (1) niedriges und (2) hohes subjektives Stresserleben, wurde verwendet. Mit den ROC-Kurven lässt sich graphisch darstellen, wie gut die jeweilige Skala zwischen Personen mit hoher und niedriger Stressbelastung differenzieren kann. Die im Zuge der Analyse berechneten Area-Under-The-Curve-Werte (AUC) haben einen Wertebereich von 0 bis 1 und können als Maß für die Diskriminationsfähigkeit herangezogen werden. Das Konfidenzniveau wurde bei 95% festgelegt. Ein AUC-Wert von .50 entspricht einer zufälligen Diskriminationsfähigkeit. Ein AUC-Wert über .60 kann als dürftig, ein Wert über .70 kann als mittelmäßig, ein Wert über .80 als gut und ein Wert über .90 als exzellent interpretiert

werden. Signifikanztests geben darüber hinaus Auskunft darüber, ob ein AUC-Wert signifikant von .50 abweicht. Der Vorteil dieser Methode ist, dass man die Diskriminationsfähigkeit verschiedener Skalen direkt miteinander vergleichen kann. Die Items der Skala PE des MBI wurden für diese Analyse umgepolt.

Relative Validität. Während die ROC-Kurven nur Aussagen über die Diskriminationsfähigkeit in Bezug auf ein binäres Außenkriterium erlauben, können mit der Analyse der relativen Validität mehrere Abstufungen des Außenkriteriums getätigt werden (Mancuso et al., 2001). Hierzu wurde die zuvor beschriebene Einteilung A in vier Gruppen herangezogen: (1) niedriges, (2) eher niedriges, (3) eher hohes und (4) hohes subjektives Stresserleben. Für die Berechnung der relativen Validität wurden zuerst einfaktorielle Varianzanalysen für jede Skala über die vier Gruppen des Stresserlebens berechnet. Da die Voraussetzung der Varianzhomogenität für alle Skalen mit Ausnahme der Skala PE des MBI verletzt war, wurde für die betroffenen Skalen auf den Welch-Test zurückgegriffen. Die dabei ausgegebenen *F*-Werte wurden danach standardisiert, indem sie durch den niedrigsten *F*-Wert geteilt wurden. Die *F*-Ratios der beiden Fragebögen können direkt miteinander als Maße für die relative Validität verglichen werden.

Ergebnisdarstellung

Stichprobenbeschreibung

Insgesamt wurden Daten von 392 Personen erhoben. 54.6% der Stichprobe waren weiblich, 45.4% männlich. Die Personen waren zwischen 18 und 67 Jahren alt, wobei der Mittelwert bei 41.15 Jahren ($SD = 12.21$) lag.

Nationalität. 67.1% der Teilnehmer_innen gaben als Herkunftsland Österreich, 24.2% Deutschland und 6.6% Italien an. 1.8% nannten andere Länder wie Kroatien, Rumänien, Serbien, Slowenien und Spanien. Eine Person (0.3%) machte keine Angabe zu ihrer Nationalität.

Beziehungsstatus. 75.5% der Teilnehmer_innen gaben an, in einer Beziehung zu leben (34.2%) oder verheiratet zu sein (41.3%). 19.6% beschrieben sich als Single und 4.1% als getrennt oder geschieden. Zwei Personen (0.5%) wählten die Kategorie „anderes“, und eine Person (0.3%) machte keine Angabe zu ihrem Beziehungsstatus.

Beschäftigungsart. Die meisten Teilnehmer_innen (68.6%) erklärten, einer Vollzeitbeschäftigung nachzugehen. 16.1% gaben an, Teilzeit angestellt zu sein. Weitere 12.3% beschrieben sich als selbstständig tätig. Einer Kombination aus Anstellung und

Selbstständigkeit gingen laut eigener Angabe 1.5% der Stichprobe nach, und weitere 1.5% gaben an, im Rahmen eines professionellen Sportförderprogrammes beschäftigt zu sein.

Durchschnittliche Arbeitsstunden pro Woche. Die Angaben lagen zwischen 10 und 98 Stunden mit einem Mittelwert von 39.59 Stunden ($SD = 9.79$).

Beruf. Zur Berufsklassifikation wurde das Alphabetikum der ÖISCO-08 (Statistik Austria, 2011) herangezogen. Dabei handelt es sich um die österreichische Version der internationalen Standardklassifikation der Berufe (ISCO-08, International Labour Organization, 2012). Die im offenen Antwortformat erhobenen Berufe wurden, sofern eindeutig möglich, einer der zehn Berufshauptgruppen der ÖISCO-08 zugeordnet. Ein großer Teil der Nennungen (48.5%) konnte dabei der Gruppe *akademische Berufe* zugeordnet werden. Im Allgemeinen ist für Berufe dieser Kategorie ein Studienabschluss notwendig. Bei einigen Berufen kann die formale Ausbildung jedoch durch On-the-job-Training oder praktische Erfahrung ersetzt werden (Statistik Austria, 2011). Der meistgenannte Beruf dieser Hauptgruppe war Lehrer_in (11.9% der Gesamtstichprobe), der auch insgesamt über alle Kategorien hinweg die häufigste Nennung darstellte. Weitere mehrfach genannte Berufe dieser Kategorie waren diplomierte_r Gesundheits- und Krankenpfleger_in (12.1%), Pädagogin/Pädagoge (6.3%), Ärztin/Arzt (5.8%), Jurist_in (4.7%), Psychologin/Psychologe (2.3%), Software-Entwickler_in (2.3%), Bibliothekar_in (2.0%), Journalist_in (1.5%), Jugendberater_in (1.5%) und Sozialarbeiter_in (1.5%). 11.5% der Stichprobe wurden der Berufshauptgruppe *Techniker_innen und gleichrangige nichttechnische Berufe* zugeordnet. Die meistgenannten Berufe dieser Kategorie waren Sozialpädagogin/Sozialpädagoge (1.5% der Gesamtstichprobe), Sportler_in (1.5%), Jugendarbeiter_in (0.8%) und Bautechniker_in (0.5%). 9.2% der Stichprobe nannten Berufe, die der Hauptgruppe der *Führungskräfte* zugeordnet wurden. Die häufigsten Nennungen in dieser Gruppe waren Geschäftsführer_in (2.8% der Gesamtstichprobe), Manager_in (1.3%), Unternehmer_in (1.3%) und Direktor_in (0.8%). Der Berufshauptgruppe *Dienstleistungsberufe und Verkäufer_innen* konnten 5.6% der Stichprobe zugeordnet werden. Am häufigsten genannt wurden in dieser Gruppe die Berufe Servicekraft in der Gastronomie (1.3%), Pflegeassistent_in (1.0%) und Verkäufer_in (1.0%). 4.1% der Nennungen wurden der Berufshauptgruppe *Bürokräfte und verwandte Berufe* zugeordnet. Weitere 1.8% wurden der Gruppe *Handwerks- und verwandte Berufe*, 0.8% der Kategorie *Bediener_innen von Anlagen und Maschinen und Montageberufe*, 0.5% der Gruppe der *Fachkräfte in Land- und Forstwirtschaft und Fischerei* und jeweils 0.2% den Gruppen *Angehörige der regulären Streitkräfte* und der *Hilfsarbeitskräfte* zugeordnet. 15.8% der Nennungen konnten aufgrund von mangelhafter Berufsbeschreibung, wie etwa Angestellte_r

(8.4% der Stichprobe), keiner Kategorie eindeutig zugeordnet werden. 1.8% der Teilnehmer_innen machten keine Angaben zum Beruf.

Berufserfahrung. Die gesamte Berufserfahrung der Teilnehmer_innen ($N = 392$) bewegte sich zwischen drei und 46 Jahren. Im Schnitt waren es 18.72 Jahre ($SD = 12.00$). An der aktuellen Arbeitsstelle wiesen sie ($N = 391$) eine Erfahrung von mindestens einem und maximal 45 Jahren auf, mit einem Durchschnitt von 12.06 Jahren ($SD = 10.02$)

Höchste abgeschlossene Ausbildung. Die Stichprobe setzte sich aus 52.3% Nicht-Akademiker_innen und 47.7% Akademiker_innen zusammen. Für 6.6% war der höchste Abschluss jener der Pflichtschule. 8.7% verfügten über einen Lehrabschluss, und 1.0% hatten eine Meisterprüfung abgelegt. 6.9% der Stichprobe hatten eine berufsbildende mittlere Schule abgeschlossen. Über die Hochschulreife verfügten 29.1% der Teilnehmer_innen. 14.0% hatten ein Bachelorstudium abgeschlossen und 28.8% entweder ein Masterstudium oder ein gleichwertiges Studium absolviert. 4.9% der Stichprobe wies einen PhD oder ein Doktorat auf.

Krankenstandstage. In den drei Monaten vor der Erhebung wiesen die meisten Teilnehmer_innen (69.4%) laut eigener Angabe keine Krankenstandstage auf. Im Schnitt waren es 1.39 Tage ($SD = 3.43$, $N = 387$). Das Maximum lag bei 31 Tagen.

Gruppeneinteilung A. Die drei Trennwerte, die zur Bildung der vier Gruppen herangezogen wurden, entsprachen den empirischen Quartilen. Diese lagen bei PSQ-Gesamtwerten von 21.67, 35 und 48.33. Die Gruppe niedriges Stresserleben setzte sich aus 104 (26.5%) Personen zusammen. Ihr PSQ-Mittelwert lag bei 14.62 ($SD = 5.68$). Der Gruppe eher niedriges Stresserleben wurden 103 (26.3%) der Teilnehmer_innen zugeordnet. Sie wies einen PSQ-Mittelwert von 29.68 ($SD = 3.76$) auf. 86 Fälle (21.9%) bildeten die Gruppe eher hohes Stresserleben. Ihr PSQ-Mittelwert lag bei 41.10 ($SD = 3.40$). Die Gruppe hohes Stresserleben bestand aus 25.1% der Stichprobe ($n = 98$) und wies einen mittleren PSQ-Wert von 59.73 ($SD = 9.79$) auf. Eine Person (0.3%) konnte keiner Gruppe zugeordnet werden, da aufgrund von fehlenden Werten kein Gesamtwert des PSQ berechnet werden konnte.

Gruppeneinteilung B. Die Stichprobentrennung erfolgte durch Mediansplit. Der Median lag bei einem PSQ-Gesamtwert von 35. 207 Personen (52.8%) wiesen nach dieser Einteilung niedriges Stresserleben mit einem PSQ-Mittelwert von 22.11 ($SD = 8.95$) auf. Die Gruppe hohes Stresserleben setzte sich aus 184 Teilnehmer_innen (46.9%) zusammen und wies im Mittel einen PSQ-Wert von 51.02 ($SD = 11.96$) auf. Eine Person (0.3%) konnte aufgrund des fehlenden PSQ-Gesamtwerts keiner Gruppe zugeordnet werden.

Untersuchung der Fragestellungen

Datenvollständigkeitsanalyse. 0.5% der Werte des MBI, 0.4% der Werte der BOSS I und 0.1% der Werte der BOSS II fehlten. Tabelle 1 zeigt den prozentuellen Anteil an fehlenden Werten pro Skala für MBI, BOSS I und BOSS II. Der im Vergleich zu den anderen Skalen höchste Anteil an fehlenden Werten (1.3%) zeigte sich in der Skala FA der BOSS I. Auf Itemebene wiesen vor allem die Items 21 bis 24 der Skala FA höhere Anteile an fehlenden Werten auf (1.3% – 2.0% fehlende Antworten). Mehr als 1.0% fehlende Antworten fanden sich darüber hinaus noch bei Item 10 der Skala DP und Item 18 der Skala PE des MBI. Der Anteil an fehlenden Werten für jedes Item des MBI, der BOSS I und BOSS II ist in den Tabellen A1 bis A3 im Anhang angegeben.

Tabelle 1

Deskriptive Skalenanalyse der Instrumente MBI, BOSS I und BOSS II

Skala	N	M (SD)	Schiefe	Kurtosis	Fehlende Werte (%)
MBI					
EE	391	27.53 (16.79)	1.04***	1.19***	0.2
DP	390	16.29 (16.19)	1.41***	2.02***	0.7
PE ^a	390	35.19 (16.39)	0.39**	-0.28	0.6
BOSS I					
B	392	14.97 (13.03)	1.37***	2.77***	0.2
EP	392	15.07 (15.45)	1.59***	2.82***	0.1
FA	387	21.78 (19.16)	0.96***	0.49*	1.3
FR	392	15.79 (17.18)	1.61***	3.05***	0.2
BOSS II					
KOEB	392	12.53 (14.35)	1.91***	4.11***	0.2
KOB	392	17.23 (16.75)	1.40***	2.22***	0.2
EB	392	13.64 (15.55)	1.84***	4.28***	0.1

Anmerkung. SE der Schiefe = 0.12; SE der Kurtosis = 0.25. EE = Emotionale Beschwerden; DP = Depersonalisation; PE = Persönliche Erfüllung; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden.

^aumgepolt.

* $\alpha = 5.0\%$. ** $\alpha = 1.0\%$. *** $\alpha = 0.1\%$.

Aufgrund von mehr als 20% fehlenden Werten innerhalb einer Skala konnte für 0.3% aller Fälle beim MBI kein Summenwert der Skala EE berechnet werden. Bei den Skalen DP und PE traf dies in jeweils 0.5% aller Fälle zu. Bei der BOSS I gab es keine Verluste von Summenwerten bei den Skalen B, EP und FR. In der Skala FA konnten für 1.3% der Fälle keine Summenwerte berechnet werden. Bei der BOSS II konnten für alle Fälle Summenwerte bestimmt werden.

Deskriptive Analyse. Die Mittelwerte, Standardabweichung, Schiefe und Kurtosis für jede Skala des MBI, der BOSS I und BOSS II sind ebenfalls in Tabelle 1 angeführt. Die Verteilungen aller Skalen der drei Instrumente waren signifikant rechtsschief. Mit Ausnahme der Skala PE wiesen alle Skalen eine signifikant spitzgipflige Wölbung auf. Die Mittelwerte lagen je nach Skala zwischen 13.64 und 35.19.

Die Angaben zu Mittelwert, Standardabweichung, Schiefe und Kurtosis der Skalen des MBI, der BOSS I und BOSS II sind, für jede der vier Stresserlebensgruppen getrennt, in den Tabellen B1 bis B4 im Anhang zu finden. Die Mittelwerte der Skalen in der Gruppe eher niedriges Stresserleben schwankten zwischen 3.42 und 28.08. Die Skalenmittelwerte der Gruppe mit eher niedrigem Stresserleben lagen zwischen 8.93 und 32.34. Die Gruppe eher hohes Stresserleben wies Skalenmittelwerte zwischen 12.37 und 35.90 auf. Die Mittelwerte der Skalen der Gruppe hohes Stresserleben lagen im Bereich von 23.84 und 44.90.

Die Verteilungen der gewählten Antwortoptionen in den vier Stresserlebensgruppen für die Skalen des MBI sind in Tabelle 2 angeführt. Alle sieben Antwortoptionen wurden in allen Gruppen bei allen Skalen des MBI mindestens einmal gewählt. Bodeneffekte zeigten sich vor allem in der Skala DP. 27.9% der Personen mit niedrigem Stresserleben erreichten hier den niedrigstmöglichen Summenscore. Die prozentuellen Bodeneffekte in dieser Skala für die Gruppen eher niedriges, eher hohes und hohes Stresserleben lagen bei 24.5%, 7.0% und 4.1%. Ein weiterer Bodeneffekt von 2.9% zeichnete sich in der Gruppe mit niedrigem Stresserleben in der Skala EE ab. Es gab keine Deckeneffekte zu verzeichnen.

Tabelle 3 zeigt die Verteilung der gewählten Antwortoptionen der BOSS I in allen vier Stresserlebensgruppen. Mit Ausnahme der Antwortoption 5 der Skala FR, die in der Gruppe mit niedrigem Stresserleben unbenutzt blieb, wurden die sechs Antwortoptionen mindestens einmal bei jeder Skala und jeder Gruppe verwendet. Es zeigten sich Bodeneffekte in allen Skalen bei den Gruppen niedriges, eher niedriges und eher hohes Stresserleben. Das Ausmaß der Bodeneffekte in der Gruppe niedriges Stresserleben schwankte zwischen 26.9% (B) und 46.2% (FR). Die Gruppe eher niedriges Stresserleben wies Bodeneffekte zwischen 7.8% (B) und 28.2% (FR) auf. Zwischen 1.2% (B) und 14.0% (FR) der Teilnehmer_innen der Gruppe

eher hohes Stresserleben erreichten den niedrigstmöglichen Summenscore der Skalen. In der Gruppe hohes Stresserleben zeichnete sich lediglich in der Skala FA ein Bodeneffekt von 4.1% ab. Es gab keine Deckeneffekte.

Tabelle 2

Verteilung der gewählten Antwortoptionen der Skalen des MBI in der Gesamtstichprobe und den vier Stresserlebensgruppen

Gruppe Stresserleben		Antwortoption							Boden- effekt
		0	1	2	3	4	5	6	
EE									
niedrig		42.1	40.7	7.6	6.8	1.8	0.7	0.3	2.9
eher niedrig		25.4	40.5	16.9	10.3	3.8	3.0	0.1	0.0
eher hoch		15.1	38.5	18.6	17.2	6.0	3.6	1.0	0.0
hoch		8.2	21.3	16.9	23.0	14.1	12.8	3.7	0.0
gesamt		23.2	35.2	15.0	14.0	6.3	5.0	1.3	0.8
DP									
niedrig		67.1	19.6	3.9	4.4	2.5	1.7	0.8	27.9
eher niedrig		60.0	24.4	7.8	4.1	1.7	1.0	1.0	24.5
eher hoch		43.9	30.2	11.9	7.2	2.6	2.6	1.6	7.0
hoch		35.1	27.7	11.1	11.3	7.2	5.1	2.5	4.1
gesamt		52.0	25.3	8.5	6.7	3.5	2.6	1.4	16.4
PE									
niedrig		2.9	6.3	4.7	13.3	13.9	33.3	25.6	0.0
eher niedrig		5.2	11.0	10.5	14.3	15.7	29.3	14.0	0.0
eher hoch		1.6	6.4	12.4	18.5	20.5	28.0	12.6	0.0
hoch		3.7	12.4	17.1	20.9	18.5	17.2	10.2	0.0
gesamt		2.5	7.7	10.7	16.8	17.9	28.4	16.0	0.0

Anmerkung. 0 = nie; 1 = ein paar Mal im Jahr; 2 = monatlich; 3 = einige Male pro Monat; 4 = wöchentlich; 5 = mehrmals pro Woche; 6 = täglich; EE = Emotionale Erschöpfung; DP = Depersonalisation; PE = Persönliche Erfüllung; gesamt = Gesamtstichprobe. Alle Angaben in Prozent.

Die Anteile der gewählten Antwortoptionen der BOSS II in den vier Stresserlebensgruppen sind in Tabelle 4 angeführt. Es zeigten sich unbenutzte Antwortoptionen in der Gruppe mit niedrigem Stresserleben. Keine_r der Teilnehmer_innen

dieser Gruppe wählte die Antwortoption 5 der Skala KOB oder die Optionen 4 und 5 der Skala EB.

Tabelle 3

Verteilung der gewählten Antwortoptionen der Skalen der BOSS I in der Gesamtstichprobe und den vier Stresserlebensgruppen

Gruppe Stresserleben		Antwortoption						Boden- effekt
		0	1	2	3	4	5	
B								
	niedrig	80.8	14.0	3.8	0.7	0.5	0.2	26.9
	eher niedrig	60.7	25.9	10.8	1.3	0.7	0.6	7.8
	eher hoch	49.3	32.2	14.6	1.9	1.5	0.5	1.2
	hoch	30.3	27.7	21.9	10.2	7.0	2.9	0.0
	gesamt	56.0	24.3	12.6	3.5	2.4	1.0	9.4
EP								
	niedrig	83.2	11.2	5.0	0.3	0.2	0.1	34.6
	eher niedrig	66.2	23.4	8.2	1.1	0.8	0.3	9.7
	eher hoch	49.1	30.2	14.5	3.6	1.9	0.7	3.5
	hoch	29.1	26.7	20.2	11.8	8.2	4.0	0.0
	gesamt	57.7	22.5	11.7	4.1	2.7	1.3	12.5
FA								
	niedrig	68.2	19.4	8.7	2.5	0.6	0.6	36.9
	eher niedrig	43.4	30.3	19.4	5.5	1.2	0.2	10.9
	eher hoch	30.4	31.1	25.9	5.5	4.7	2.4	7.1
	hoch	22.5	26.3	27.3	9.0	10.6	4.3	4.1
	gesamt	42.0	26.5	19.9	5.6	4.2	1.8	15.5
FR								
	niedrig	80.0	14.8	4.0	0.6	0.6	0.0	46.2
	eher niedrig	63.7	22.1	11.3	2.1	0.6	0.2	28.2
	eher hoch	50.7	28.6	11.4	5.4	2.3	1.6	14.0
	hoch	29.4	26.9	21.0	7.4	12.0	3.3	0.0
	gesamt	56.7	22.8	11.8	3.7	3.8	1.2	24.0

Anmerkung. 0 = trifft nicht zu; 1 = trifft kaum zu; 2 = trifft teilweise zu; 3 = trifft überwiegend zu; 4 = trifft zu; 5 = trifft stark zu; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde; gesamt = Gesamtstichprobe. Alle Angaben in Prozent.

Es ließen sich Bodeneffekte in allen Gruppen und allen Skalen verzeichnen. Diese schwankten zwischen 29.8% (KOEB, KOB) und 48.1% (EB) in der Gruppe mit niedrigem Stresserleben. Von den Teilnehmer_innen mit eher niedrigem Stresserleben erreichten zwischen 11.7% (KOB) und 19.4% (KOEB) die niedrigstmöglichen Summenwerte der jeweiligen Skalen. Die Bodeneffekte in der Gruppe mit eher hohem Stresserleben lagen zwischen 2.3% (KOB) und 10.5% (KOEB). In der Gruppe mit hohem Stresserleben schwankten die Bodeneffekte zwischen 2.0% (KOB, EB) und 5.1% (KOEB). Es traten keine Deckeneffekte auf.

Über alle Gruppen hinweg zeigte die Skala PE des MBI als einzige keine Bodeneffekte. Die Skala EE des MBI wies insgesamt den geringsten Anteil an Bodeneffekten auf (0.8% der Gesamtstichprobe), gefolgt von der Skala B der BOSS I (9.4%) und der Skala KOB der BOSS II (12.0%). Weitere Bodeneffekte in aufsteigender Reihenfolge verzeichneten insgesamt die Skalen FA der BOSS I (15.5%), DP des MBI (16.4%), KOEB (16.6%) und EB der BOSS II (18.9%). Der größte Bodeneffekt fand sich bei der Skala FR der BOSS I (24.0%).

Reliabilitätsanalyse. Tabelle 5 zeigt die Ergebnisse der Reliabilitätsanalyse der Instrumente MBI, BOSS I und BOSS II. Detaillierte Ergebnisse der Trennschärfenanalyse für jedes einzelne Item des MBI, BOSS I und BOSS II sind in den Tabellen C1 bis C3 im Anhang angeführt. Die Skalen des MBI wiesen Cronbachs-Alpha-Werte von .90 für die Skala EE und .79 für die Skalen PE und DP auf. Auf Itemebene zeigte Item 22 der Skala DP eine unzureichende korrigierte Trennschärfe von .36. In der Skala PE wiesen zwei Items unzureichende Werte von jeweils .37 auf: Item 4 und Item 18. Das Entfernen dieser Items würde die interne Konsistenz jedoch nicht steigern. Die übrigen Items des MBI wiesen eine akzeptable Trennschärfe auf.

Die Cronbachs-Alphawerte der Skalen der BOSS I betrugen .88 für die Skala B, .93 für die Skala EP, .91 für die Skala FA und .89 für die Skala FR. Item 9 der Skala B verfügte über eine unzureichende Trennschärfe von .37. Die übrigen Items des Instruments wiesen akzeptable Trennschärfewerte auf.

Die interne Konsistenz der Skalen der BOSS II lag bei .91 für die Skala KOEB und .95 für die Skalen KOB und EB. Die Trennschärfe aller Items war akzeptabel.

Faktorenanalyse. Das Ergebnis der durchgeführten Parallelanalyse deutete beim MBI auf eine dreifaktorielle Lösung hin (siehe Abbildung 1). Die drei rotierten Faktoren erklärten

gemeinsam 46.2% der Gesamtvarianz. Eine Auflistung der Faktorladungen der rotierten Lösung für jedes Item des MBI findet sich in Tabelle 6.

Tabelle 4

Verteilung der gewählten Antwortoptionen der Skalen der BOSS II in der Gesamtstichprobe und den vier Stresserlebensgruppen

		Antwortoption					Boden- effekt
		0	1	2	3	4	
KOEB							
niedrig	83.7	9.3	5.3	0.7	0.9	0.1	29.8
eher niedrig	70.7	18.4	7.0	2.0	1.3	0.6	19.4
eher hoch	67.2	14.9	12.1	1.8	2.8	1.2	10.5
hoch	51.2	16.8	12.7	6.6	6.5	6.2	5.1
gesamt	68.5	14.8	9.1	2.8	2.7	2.1	16.6
KOB							
niedrig	76.9	18.6	4.0	0.2	0.3	0.0	29.8
eher niedrig	55.7	30.3	11.0	1.6	0.9	0.5	11.7
eher hoch	42.0	36.2	16.7	2.4	1.9	0.8	2.3
hoch	18.8	30.6	26.9	11.6	9.1	3.0	2.0
gesamt	49.2	28.5	14.4	3.9	3.0	1.0	12.0
EB							
niedrig	86.0	11.0	2.9	0.1	0.0	0.0	48.1
eher niedrig	67.9	22.5	7.7	1.0	0.5	0.4	16.5
eher hoch	53.0	27.7	14.5	3.4	1.0	0.4	5.8
hoch	29.7	29.5	21.4	9.4	6.9	3.1	2.0
gesamt	59.9	22.4	11.3	3.4	2.1	0.9	18.9

Anmerkung. 0 = trifft nicht zu; 1 = trifft kaum zu; 2 = trifft teilweise zu; 3 = trifft überwiegend zu; 4 = trifft zu; 5 = trifft stark zu; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden; gesamt = Gesamtstichprobe. Alle Angaben in Prozent.

Tabelle 5*Reliabilitätsanalyse der Instrumente MBI, BOSS I und BOSS II*

Skala	N	ordinale Itemtrennschärfe		ordinales α
		min. – max.	Ø	
MBI				
EE	391	.56 – .81	.65	.90
DP	390	.36 – .69	.56	.79
PE	390	.37 – .58	.49	.79
BOSS I				
B	392	.37 – .68	.59	.88
EP	392	.53 – .85	.71	.93
FA	387	.70 – .85	.75	.91
FR	392	.61 – .78	.72	.89
BOSS II				
KOEB	392	.48 – .77	.64	.91
KOB	392	.64 – .84	.77	.95
EB	392	.70 – .86	.78	.95

Anmerkung. Ordinale Itemtrennschärfe = ordinale, korrigierte Item-Skala-Korrelation; α = Cronbachs Alpha; EE = Emotionale Erschöpfung; DP = Depersonalisation; PE = Persönliche Erfüllung; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden.

Der erste Faktor, der 21.3% der Varianz erklärte, entsprach im Wesentlichen der Skala EE. Die bedeutsamen Ladungen der neun Items lagen zwischen .52 und .82. Item 16 zeigte allerdings zusätzlich zu seiner Ladung von .52 auf dem ersten Faktor eine bedeutsame Kreuzladung von .42 auf dem zweiten Faktor.

Ebenfalls auf dem zweiten Faktor luden vier der fünf Items der Skala DP bedeutsam hoch (.57 bis .75). Item 22 der Skala zeigte jedoch weder eine bedeutsame Ladung auf diesem Faktor (.36) noch auf einem der anderen beiden. Die durch den zweiten Faktor erklärte Varianz betrug 12.5%.

Der dritte Faktor erklärte weitere 12.4% der Varianz. Die acht Items, die bedeutsame Ladungen auf diesem Faktor zeigten (.43 bis .67), konnten alle der Skala PE zugeordnet werden. Item 12 der Skala zeigte jedoch zusätzlich zu seiner Ladung von .53 auf dem dritten Faktor eine Kreuzladung von -.52 auf dem ersten Faktor.

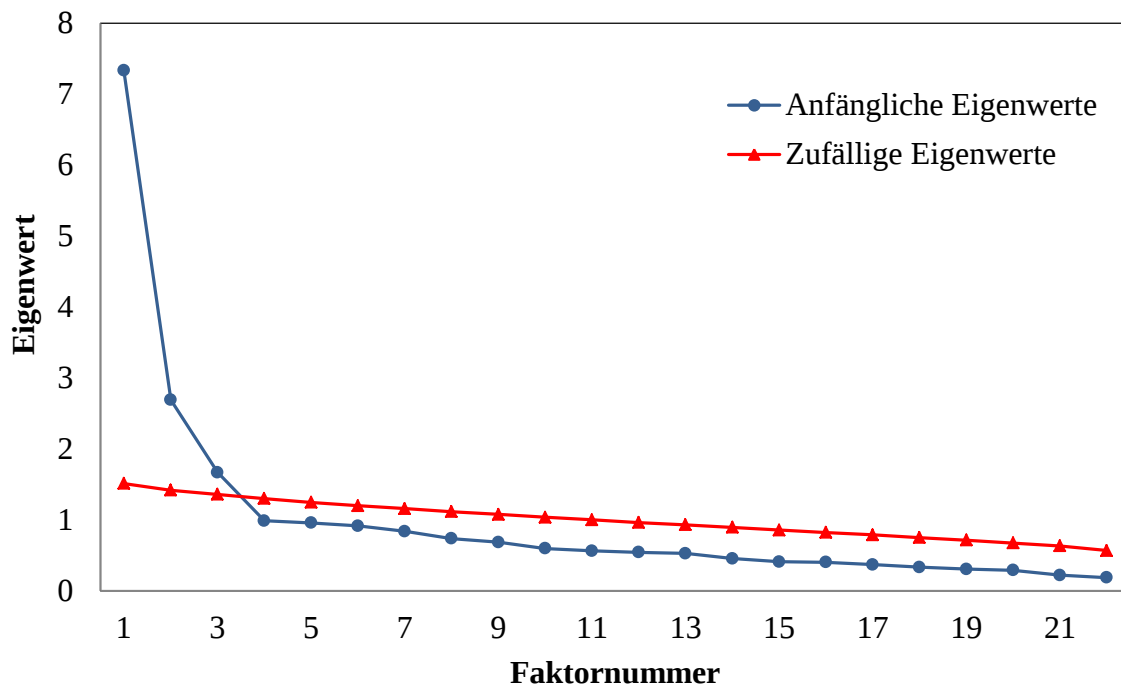


Abbildung 1. Verlauf der anfänglichen Eigenwerte der Faktorenanalyse des MBI und der zufälligen Eigenwerte der Parallelanalyse.

Für die BOSS I ergab sich eine zweifaktorielle Lösung durch die Parallelanalyse (siehe Abbildung 2). Die beiden rotierten Faktoren erklärten gemeinsam 53.6% der Gesamtvarianz. Die Faktorladungen der rotierten Lösung sind in Tabelle 7 aufgelistet.

Auf dem ersten Faktor, der einen Anteil von 34.0% der Varianz erklärte, zeigten 22 der 30 Items der BOSS I ihre höchsten bedeutsamen Ladungen (.47 bis .85). Es handelte sich dabei um alle zehn Items der Skala EP, acht von zehn Items der Skala B und die Items 27 bis 30 der Skala FR. Item 20 der Skala EP sowie die Items 27 bis 30 der Skala FR zeigten zwar ihre höchsten Ladungen auf dem ersten Faktor, wiesen allerdings auch bedeutsame Kreuzladungen zwischen .44 und .52 auf dem zweiten Faktor auf. Die fünf Items der Skala FA luden ausschließlich auf dem zweiten Faktor bedeutsam hoch (.69 bis .85). Das Item 26 der Skala FR wies ebenfalls eine bedeutsame Ladung von .60 auf dem zweiten Faktor auf, zeigte allerdings auch eine Kreuzladung von .42 auf dem ersten Faktor. Der zweite Faktor erklärte 19.6% der Gesamtvarianz. Die Items 3 und 9 der Skala B wiesen weder auf dem ersten noch auf dem zweiten Faktor bedeutsame Ladungen auf.

Da die durch die Parallelanalyse gefundene zweifaktorielle Struktur nicht der postulierten Struktur der BOSS I mit vier Skalen entsprach, wurde zusätzlich eine Hauptachsenanalyse mit vier Faktoren und Varimaxrotation berechnet. Die Faktoren erklärten

gemeinsam 58.8% der Gesamtvarianz. Die Faktorladungen sind im Anhang in Tabelle D1 angeführt.

Tabelle 6

Faktorladungen der Items des MBI

Nr.	Item	Skala	Faktor		
			1	2	3
1	durch Arbeit emotional erschöpft	EE	.82	.11	-.07
2	am Ende des Arbeitstages verbraucht	EE	.79	.03	.06
3	morgens vor Arbeitstag ermüdet	EE	.63	.16	-.11
6	Anstrengung, mit Menschen zu arbeiten	EE	.52	.33	-.03
8	ausgebrannt durch die Arbeit	EE	.80	.29	-.18
13	frustriert durch Arbeit	EE	.69	.36	-.17
14	zu verbissen arbeiten	EE	.56	.32	.03
16	direkter Kontakt stresst zu sehr	EE	.52	.42	-.19
20	am Ende der Weisheit zu sein	EE	.52	.34	-.14
5	K./V. wie Objekte behandeln	DP	.30	.57	-.19
10	gleichgültiger gegenüber Menschen	DP	.26	.75	-.20
11	emotional verhärtet durch Arbeit	DP	.31	.72	-.14
15	kein Interesse an manchen K.	DP	.06	.58	-.18
22	K./M. geben mir Schuld für Probleme	DP	.26	.36	.00
4	nachvollziehen können was in K./V. vorgeht	PE	.11	-.18	.47
7	guter Umgang mit Problemen anderer	PE	.00	.00	.66
9	andere positiv beeinflusst durch Arbeit	PE	-.08	-.11	.67
12	voller Energie	PE	-.52	-.07	.53
17	leichtes Schaffen entspannter Atmosphäre	PE	-.30	-.22	.57
18	angeregt, wenn eng mit K. gearbeitet	PE	.03	-.12	.43
19	viele lohnende Ziele bei Arbeit erreicht	PE	-.09	-.12	.57
21	gelassener Umgang mit emot. Problemen	PE	-.20	.01	.49

Anmerkung. Nr. = Itemnummerierung des MBI; K. = Kolleg_innen; V. = Vorgesetzte; M. = Mitarbeitende; EE = Emotionale Erschöpfung; DP = Depersonalisation; PE = Persönliche Erfüllung. Hauptachsenanalyse mit Varimax-Rotation, basierend auf polychorischen Korrelationen. Abbruchkriterium: polychorische Parallelanalyse. $N = 390$. Bedeutsame Ladungen ($\geq |.40|$) sind fettgedruckt.

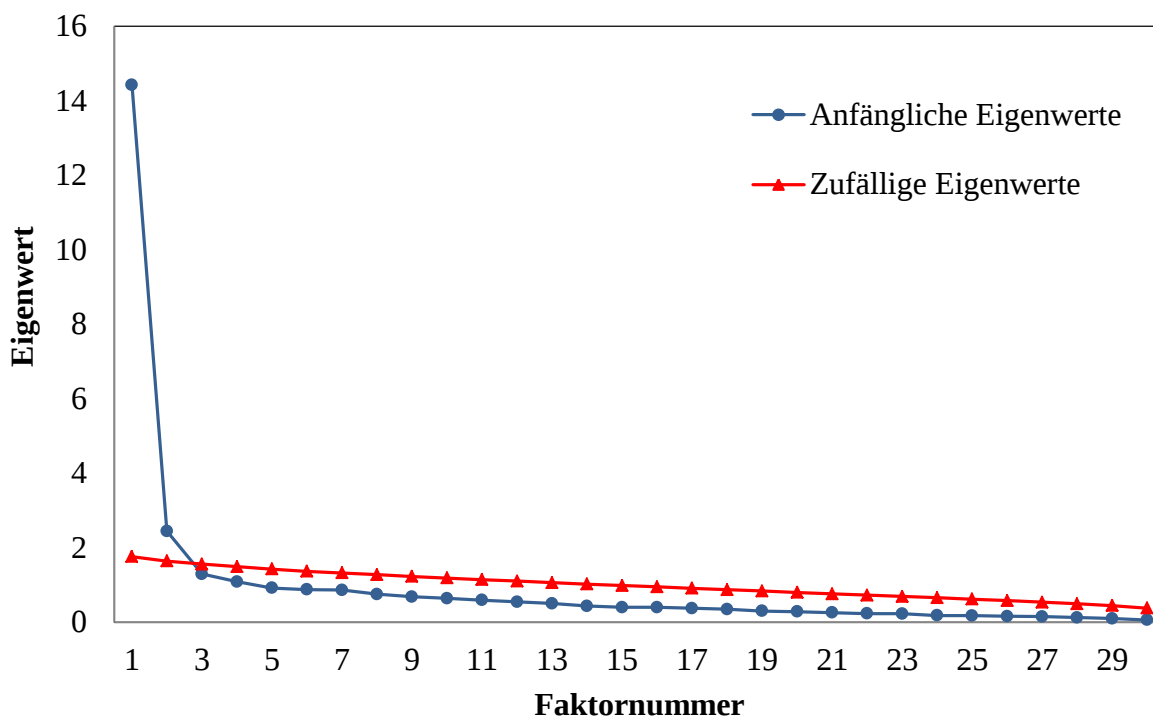


Abbildung 2. Verlauf der anfänglichen Eigenwerte der Faktorenanalyse der BOSS I und der zufälligen Eigenwerte der Parallelanalyse.

Auf dem zweiten Faktor der Lösung, der 13.8% der Varianz erklärte, luden ausschließlich die Items der Skala FA und das Item 26 der Skala FR bedeutsam hoch. Letzteres lud darüber hinaus auch auf dem vierten Faktor. Der vierte Faktor, der 10.2% der Varianz erklärte, wies hauptsächlich bedeutsame Ladungen der fünf Items der Skala FR auf, die allerdings alle auch auf einem der anderen drei Faktoren bedeutsam luden. Der erste und dritte Faktor konnten nicht eindeutig einer der beiden Skalen B und EP zugeordnet werden. Die Überschneidungen der Items, die auf diesen beiden Faktoren bedeutsam luden, waren sehr groß. Sie erklärten jeweils 22.8% und 12.0% der Varianz.

Bei der BOSS II waren zwei der anfänglichen Eigenwerte größer als die zufälligen Eigenwerte der Parallelanalyse (siehe Abbildung 3). Die daraus resultierende rotierte zweifaktorielle Lösung erklärte 61.0% der Gesamtvarianz. Die rotierten Faktorladungen jedes Items der BOSS II sind in Tabelle 8 aufgelistet.

Auf dem ersten Faktor luden alle 20 Items der Skalen KOB und EB bedeutsam hoch (.52 bis .84). 38.4% der Varianz wurden durch diesen Faktor erklärt. Die zehn Items der Skala KOEB luden ausschließlich auf dem zweiten Faktor in bedeutsamer Höhe (.48 bis .84). Das Item 11 der Skala KOB („Ich kann mich schlecht konzentrieren.“) wies neben seiner Ladung von .52 auf dem ersten Faktor auch eine bedeutsame Ladung in der Höhe von .50 auf dem zweiten Faktor auf.

Tabelle 7*Faktorladungen der Items der BOSS I*

Nr.	Item	Skala	Faktor	
			1	2
1	Sorge, Job nicht zu bewältigen	B	.66	.26
2	vermehrtes Kontrollieren, misstrauischer	B	.56	.28
3	Arbeiten bis spät in die Nacht	B	.38	.37
4	keine nachhaltige Freude über Erfolge	B	.67	.24
5	Fassung verlieren, wenn nicht nach Plan	B	.60	.19
6	Dauerstress durch lange Arbeitswege	B	.50	.21
7	Verlust der Freude an der Arbeit	B	.78	.02
8	wenig treffsicher in berufl. Entscheidungen	B	.73	.15
9	Alkohol/Medikamente zum Stressabbau	B	.31	.31
10	gefährdetes positives Image im Job	B	.69	.25
11	sich selbst fremd geworden sein	EP	.74	.30
12	keine Lebensfreude mehr empfinden können	EP	.70	.31
13	herabgesetzte Frustrationstoleranz	EP	.74	.27
14	zunehmende Sorgen um Gesundheit	EP	.59	.37
15	häufige banale Infekte	EP	.47	.26
16	kleine Anforderungen bereits belastend	EP	.80	.27
17	schlechter Schlaf	EP	.59	.32
18	Verlust der Routine, alles schwerer	EP	.85	.28
19	zunehmende Konzentrationsschwierigkeiten	EP	.68	.33
20	Schwierigkeiten beim Entspannen	EP	.64	.44
21	Unstimmigkeiten in Familie/Partnerschaft	FA	.27	.69
22	zu wenig Zeit für Gemeinsamkeiten	FA	.18	.78
23	drohende Entfremdung	FA	.19	.85
24	Sexualität/Erotik kommt zu kurz	FA	.09	.81
25	weniger Anteil an Familie/Partnerschaft	FA	.36	.74
26	besorgte Rückmeldungen von Freunden	FR	.42	.60
27	häufiger Rückzug aus sozialen Kontakten	FR	.57	.49
28	Belastungen sind einziges Gesprächsthema	FR	.63	.45
29	weniger Anteil an Freundschaften	FR	.60	.52
30	Schwierigkeiten, sich abzugrenzen	FR	.51	.44

Anmerkung. Nr. = Itemnummerierung der BOSS I; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde. Hauptachsenanalyse mit Varimax-Rotation, basierend auf polychorischen Korrelationen. Abbruchkriterium: polychorische Parallelanalyse. $N = 387$. Bedeutsame Ladungen ($\geq |.40|$) sind fettgedruckt.

Die Items 24, 26 und 27 der Skala EB luden ebenfalls auf beiden Faktoren bedeutsam, wobei ihre Ladungen auf dem ersten Faktor höher sind. Der zweite Faktor erklärte 22.6% der Varianz.

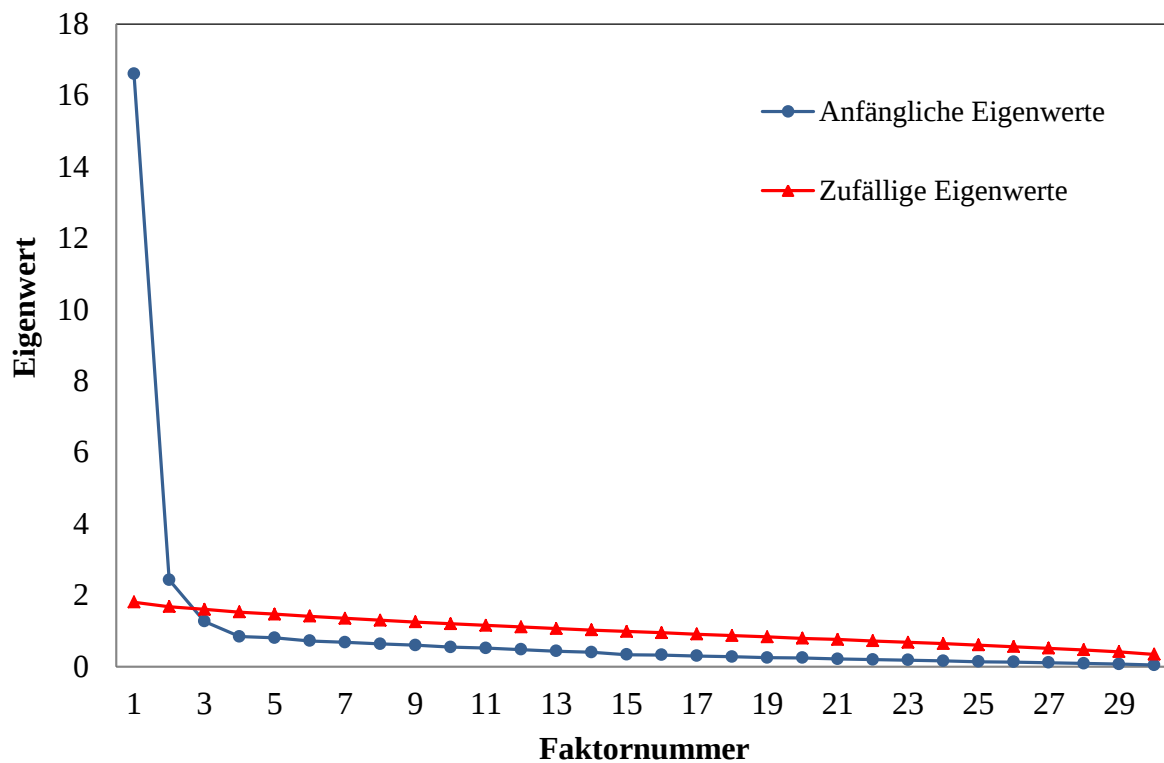


Abbildung 3. Verlauf der anfänglichen Eigenwerte der Faktorenanalyse der BOSS II und der zufälligen Eigenwerte der Parallelanalyse.

Da die Autor_innen eine dreifaktorielle Struktur der BOSS II postulieren, wurde ebenfalls eine Hauptachsensanalyse mit drei Faktoren und Varimaxrotation berechnet. Die drei Faktoren dieser Lösung erklärten gemeinsam 64.0% der Gesamtvarianz. Die rotierten Faktorladungen sind in Tabelle D2 im Anhang aufgelistet. Der erste Faktor, der 37.4% der Varianz erklärte, entspricht bei dieser Lösung dem ersten Faktor der zweifaktoriellen Lösung und enthält bedeutsame Ladungen der 20 Items der Skalen KOB und EB. Auf dem zweiten Faktor, der 17.7% der Varianz erklärte, luden die Items 3 bis 10 der Skala KOEB. Das Item 11 der Skala KOB (Konzentrationsbeschwerden) lud, wie in der zweifaktoriellen Lösung, auf den ersten beiden Faktoren bedeutsam hoch. Auf dem dritten Faktor luden ausschließlich die Items 1 bis 3 der Skala KOEB bedeutsam hoch. Das dritte Item, das bedeutsame Ladungen auf dem zweiten und dritten Faktor aufwies, zeigte seine höchste Ladung auf dem dritten Faktor. Der dritte Faktor erklärte 8.9% der Varianz.

Tabelle 8*Faktorladungen der Items der BOSS II*

Nr.	Item	Skala	Faktor	
			1	2
1	heftige Kopfschmerzen	KOEB	.27	.49
2	häufig wiederkehrende Infekte	KOEB	.31	.48
3	ständige Muskelverspannungen	KOEB	.33	.68
4	beklemmende Enge über der Brust	KOEB	.34	.75
5	Atembeschwerden	KOEB	.23	.84
6	Schmerzen	KOEB	.23	.68
7	Herzbeschwerden	KOEB	.31	.76
8	Magen-Darmbeschwerden	KOEB	.22	.64
9	Schlafstörungen	KOEB	.37	.58
10	hohe oder schwankende Blutdruckwerte	KOEB	.19	.57
11	Konzentrationsschwierigkeiten	KOB	.52	.50
12	zu negative Bewertungen	KOB	.71	.38
13	Verlust der Entschlussfreudigkeit	KOB	.77	.33
14	vermehrte Fehlerneigung	KOB	.70	.38
15	nur noch Reagieren statt Agieren	KOB	.78	.35
16	Verlust der Kreativität	KOB	.75	.34
17	wiederkehrender, quälender Gedanke	KOB	.70	.39
18	verstärkte Selbstkontrolle	KOB	.59	.30
19	Verlust der Souveränität	KOB	.79	.33
20	verminderte Durchsetzungsfähigkeit	KOB	.81	.30
21	Zukunftsängste	EB	.71	.26
22	eingeschränkte Selbstwahrnehmung	EB	.79	.33
23	sozialer Rückzug	EB	.74	.31
24	gedrückte Stimmung	EB	.77	.43
25	Antriebslosigkeit	EB	.84	.27
26	Gereiztheit	EB	.67	.47
27	erhöhtes Misstrauen	EB	.64	.40
28	Scham	EB	.81	.14
29	erhöhte Ängstlichkeit	EB	.83	.29
30	Geduldverlust bei Misslingen	EB	.61	.35

Anmerkung. Nr. = Itemnummerierung der BOSS II; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden. Hauptachsenanalyse mit Varimax-Rotation, basierend auf polychorischen Korrelationen. Abbruchkriterium: polychorische Parallelanalyse. $N = 392$. Bedeutsame Ladungen ($\geq |.40|$) sind fettgedruckt.

ROC-Analyse. Die erhaltenen ROC-Kurven sind in Abbildung 4 abgebildet. Je weiter die Kurve einer Skala am linken oberen Rand liegt und je weiter sie von der Bezugslinie entfernt ist, desto besser ist ihre Diskriminationsfähigkeit zwischen Personen, die ein hohes Stresserleben aufweisen, und jenen mit niedrigem Stresserleben. Die Bezugslinie stellt die zufällige Diskriminationsfähigkeit mit einer Fläche unter der Kurve (AUC) von .50 dar. Die Area-Under-The-Curve-Werte jeder Skala sind in Tabelle 9 aufgelistet. Alle Skalen konnten signifikant zwischen den Zustandsausprägungen diskriminieren. Die höchsten AUC-Werte erzielten die Skalen EP der BOSS I (.85), EE des MBI (.84) und EB der BOSS II (.84). Den niedrigsten AUC-Wert von .69 wies die Skala PE des MBI auf. Betrachtet man die gemittelten AUC-Werte der drei Instrumente, so zeigte die BOSS I den höchsten Wert (.81), gefolgt von der BOSS II (.79) und dem MBI (.74)

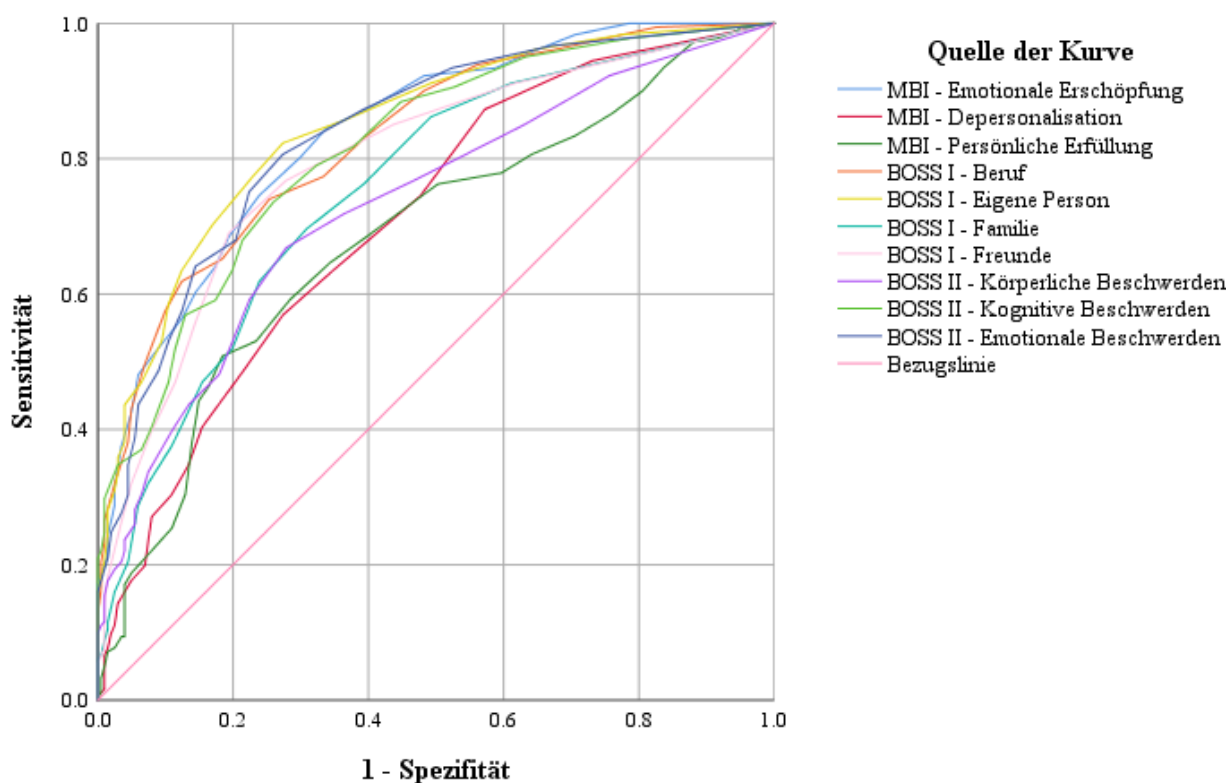


Abbildung 4. ROC-Kurven der Skalen des MBI, der BOSS I und der BOSS II. Zustandsvariable: hohes subjektives Stresserleben. Basisrate = 47.38%.

Analyse der relativen Validität. Die Skalenmittelwerte aller Skalen der drei Instrumente über die vier Gruppen des Stresserlebens hinweg sowie die F -Werte und relative Validität der Skalen sind in Tabelle 10 angeführt. Zur Berechnung der relativen Validität wurden die in der Varianzanalyse erhaltenen F -Werte der einzelnen Skalen durch den niedrigsten F -Wert von 16.806 dividiert. Das ist der F -Wert der Skala DP des MBI. Die

höchsten Werte relativer Validität erreichten die Skalen B der BOSS I (5.2), EE des MBI (5.0) und EP der BOSS I (4.6).

Tabelle 9

AUC-Werte der Skalen des MBI, der BOSS I und BOSS II

Skala	AUC	SE	p^a	95 % KI ^a
BOSS I - Eigene Person	.85	0.02	< .001	.81 - .89
MBI - Emotionale Erschöpfung	.84	0.02	< .001	.80 - .88
BOSS II - Emotionale Beschwerden	.84	0.02	< .001	.80 - .88
BOSS I - Beruf	.83	0.02	< .001	.79 - .87
BOSS II - Kognitive Beschwerden	.82	0.02	< .001	.77 - .86
BOSS I - Freunde	.80	0.02	< .001	.75 - .84
BOSS I - Familie	.76	0.03	< .001	.71 - .80
BOSS II - Körperliche Beschwerden	.73	0.03	< .001	.68 - .78
MBI - Depersonalisation	.71	0.03	< .001	.66 - .76
MBI - Persönliche Erfüllung	.69	0.03	< .001	.63 - .74

Anmerkung. AUC = Fläche unter der Kurve. $N = 382$. Zustandsvariable: hohes subjektives Stresserleben. Basisrate = 47.38%. Anordnung absteigend basierend auf AUC-Werten.

^aasymptotische Signifikanz. Nullhypothese: Wahrheitsfläche = .50.

Diskussion

Ziel dieser Untersuchung war es, die Testgüte des MBI, der BOSS I und BOSS II anhand von verschiedenen psychometrischen Indikatoren zu untersuchen und abschließend miteinander zu vergleichen. Dazu wurden die Fragebögen einer deutschsprachigen berufstätigen Erwachsenenstichprobe vorgelegt. Im Fokus stand die Verteilung und die Vollständigkeit der erhobenen Daten sowie die Reliabilität, faktorielle Validität, Diskriminationsfähigkeit und relative Validität der Instrumente.

Tabelle 10

Skalenmittelwerte der Instrumente MBI, BOSS I und BOSS II nach Ausprägung des Stresserlebens

Skala	Stresserleben				F-Wert ^a	relative ^b Validität
	niedrig (n = 103)	eher niedrig (n = 98)	eher hoch (n = 84)	hoch (n = 97)		
MBI						
EE	14.92	22.58	29.41	44.71	83.848	5.0
DP	10.71	11.73	18.06	25.46	16.806	1.0
PE ^c	28.03	32.31	35.96	44.85	21.599 ^d	1.3
BOSS I						
B	5.38	11.41	15.14	29.01	87.620	5.2
EP	4.72	9.57	16.05	31.11	76.909	4.6
FA	9.94	18.37	26.10	34.39	38.289	2.3
FR	5.40	10.78	17.00	31.09	50.589	3.0
BOSS II						
KOEB	5.22	9.20	12.24	23.92	31.156	1.9
KOB	5.61	12.51	17.45	34.06	75.157	4.5
EB	3.46	8.80	14.48	28.70	67.595	4.0

Anmerkung. EE = Emotionale Erschöpfung; DP = Depersonalisation; PE = Persönliche Erfüllung; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden. $p < .001$ bei den Welch-Tests aller Skalen und der ANOVA für die Skala PE.

^aWelch-korrigiert aufgrund von Varianzheterogenität.

^bder kleinste F-Wert von 16.806 wurde als Divisor zur Berechnung der relativen Validität verwendet.

^cumgepolt.

^dnicht Welch-korrigiert.

Datenvollständigkeit

Die erste Fragestellung beschäftigte sich mit der Datenvollständigkeit der Instrumente auf Item- und Skalenebene. Die erhobenen Daten waren in hohem Ausmaß vollständig. Tendenziell wies die BOSS II die geringste Anzahl an fehlenden Werten auf. Bei keinem Instrument gab es jedoch mehr als 0.5% fehlende Werte. Die Unterschiede zwischen den Instrumenten in Bezug auf dieses Untersuchungsmerkmal sind demnach äußerst gering. Der höchste Anteil an fehlenden Werten zeigte sich in der Skala FA der BOSS I. Dieser Umstand

könnte auf den Inhalt der Items 21 bis 24 dieser Skala zurückzuführen sein. Verglichen mit allen anderen Items der drei Instrumente wiesen diese vier Items den höchsten Anteil an fehlenden Werten auf. Es handelt sich dabei um sehr private und intime Fragen zum Familienleben und zur Sexualität. Die Skalen DP und PE des MBI wiesen ebenfalls vergleichsweise größere Anteile an fehlenden Werten auf. Auf Itemebene fallen hier besonders Item 10 der Skala DP („Seit ich diese Arbeit ausübe, bin ich Menschen gegenüber gleichgültiger geworden.“) und Item 18 der Skala PE („Ich fühle mich angeregt, wenn ich eng mit meinen Kollegen zusammengearbeitet habe.“) auf. Möglicherweise empfanden einige Teilnehmer_innen Item 10 als irritierend, da es eine einmalige, nachhaltige Veränderung in der Vergangenheit („seit ich diese Arbeit ausübe“) beschreibt, das Antwortformat aber Häufigkeitsangaben von „nie“ bis „jeden Tag“ verlangt. Bei Item 18 wäre es denkbar, dass manche Teilnehmer_innen die Antwort verweigerten, da sie entweder keine Kolleg_innen haben oder nie eng mit ihnen zusammenarbeiten.

Die meisten fehlenden Werte konnten interpoliert werden, da nur in wenigen Fällen mehr als 20% der Werte einer Skala fehlten. Bei der BOSS II kam es so zu keinem Verlust an Skalengesamtwerten. Die BOSS I wies nur in der Skala FA Verluste auf, hatte dafür aber den höchsten Anteil im Vergleich zu allen anderen untersuchten Skalen (1.3%). Das MBI wies geringfügige Verluste in allen Skalen von weniger als einem halben Prozent auf. An dieser Stelle sollte betont werden, dass der Datenverlust insgesamt bei allen Instrumenten äußerst gering war. Die Ergebnisse der Datenvollständigkeitsanalyse sprechen für eine hohe Akzeptanz der drei Verfahren durch die Teilnehmer_innen.

Häufigkeitsverteilung der Antwortkategorien

Die nächste Fragestellung beschäftigte sich mit der Verteilung der gewählten Antwortmöglichkeiten in der Gesamtstichprobe und in den vier Stresserlebensgruppen. Darüber hinaus sollten etwaige Decken- oder Bodeneffekte untersucht werden. Die deskriptive Analyse der Ergebnisse der Gesamtstichprobe konnte zeigen, dass die Werte aller Skalen eine signifikant rechtsschiefe und spitzgipflige Verteilung aufwiesen. In Anbetracht dessen, dass es sich bei der vorliegenden Stichprobe um keine klinische Stichprobe handelt, war dies zu erwarten. Die relative Gesundheit der Stichprobe spiegelte sich auch in den Skalenmittelwerten aller Instrumente wider, die zwischen 13.64 und 35.19 lagen, bei einem Wertebereich von 0 bis 100. Burisch (2014) und Schaufeli et al. (2001) weisen darauf hin, dass ein Großteil der Burnout-Forschung an überwiegend gesunden Stichproben geschehe. In solchen Stichproben findet sich vermutlich nur ein sehr kleiner Anteil an Burnout-

Betroffenen. Man kann also erwarten, dass die Burnout-Scores der Instrumente in überwiegend gesunden Stichproben schief verteilt sind. Ein Problem, das eine schiefe Verteilung mit sich bringt, ist die potentielle Verzerrung der Ergebnisse von Methoden, die auf dem Pearson-Korrelationskoeffizienten basieren (Gadermann et al., 2012; Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010). Bisherige Untersuchungen der Reliabilität und faktoriellen Validität des MBI, der BOSS I und II ergriffen keine Maßnahmen, um einer solche Verzerrung vorzubeugen. Aufgrund der vorliegenden signifikant schiefen Verteilungen wurden polychorische Korrelationsmatrizen als Basis zur Berechnung der Reliabilitätsmaße und Hauptachsenanalyse herangezogen, die gegenüber schief verteilten Daten robust sind (Gadermann et al., 2012; Holgado-Tello et al., 2010). Hierin liegt eine klare Stärke der vorliegenden Untersuchung.

Die Daten aller Skalen zeigten keinerlei Deckeneffekte. Während dies für die meisten Skalen aufgrund der Stichprobe zu erwarten war, ist es ein gutes Ergebnis für die Skala PE des MBI, die umgekehrt gepolt ist. Bodeneffekte traten in allen Skalen der Instrumente mit Ausnahme der Skala PE auf. Der gefundene Bodeneffekt von 0.8% in der Skala EE des MBI ist vernachlässigbar. Die Bodeneffekte der übrigen Skalen lagen zwischen 9.4% und 24.0%. Da es sich bei der vorliegenden Untersuchung um eine relativ gesunde Stichprobe mit einem vermeintlich geringen Anteil an Burnout-Betroffenen handelte, sind Bodeneffekte in diesem Ausmaß noch kein Indiz für eine schlechte Testgüte.

Die Verteilungen der vier Stresserlebensgruppen unterschieden sich deutlich voneinander. Die Höhe des Schwankungsbereichs aller Skalenmittelwerte stieg über die vier Gruppen hinweg an. Die Ausprägung der Bodeneffekte in den Skalen sank über die vier Gruppen hinweg. In der Gruppe mit niedrigem Stresserleben waren in allen Skalen, außer der Skala EE (2.9%) und PE (0.0%) des MBI, große Bodeneffekte zwischen 26.9% und 48.1% zu verzeichnen. In der Gruppe mit hohem Stresserleben hingegen gab es nur noch in den Skalen DP des MBI, FA der BOSS I und KOEB, KOB und EB der BOSS II gering ausgeprägte Bodeneffekte zwischen 2.0% und 5.1%. Die Ergebnisse zeigen, dass alle drei Instrumente für Populationen mit höherer Stressbelastung besser geeignet sind, da sie deren Burnout-Ausprägungen differenzierter abbilden können. Es wäre demnach denkbar, dass sie für die Anwendung an ihrer Zielpopulation, also Personen mit Burnout, gut geeignet sind. Diese Hypothese könnte allerdings nur mit einer Stichprobe von Burnout-Betroffenen geprüft werden.

Reliabilität

Die Reliabilität der drei untersuchten Instrumente stand im Fokus der dritten Fragestellung. Dafür wurde Cronbachs Alpha und die korrigierte Itemtrennschärfe berechnet. Als Mindestmaß wurde ein Cronbachs Alpha von .70 und eine Itemtrennschärfe von .40 festgelegt. Die Skalen DP und PE des MBI wiesen eine akzeptable (.79) und die Skala EE eine ausgezeichnete (.90) interne Konsistenz auf. Diese Ergebnisse sind höher als jene von den Testautor_innen (Maslach & Jackson, 1981b) und höher als die durchschnittlichen Reliabilitätskoeffizienten der beiden beschriebenen Metaanalysen (Aguayo et al., 2011; Wheeler et al., 2011). Dieses Ergebnis lässt sich zum Teil durch die Methodik der vorliegenden Untersuchung erklären. Durch die Berechnung von ordinalen Cronbachs-Alpha-Werten, basierend auf polychorischen Korrelationen, konnte eine Unterschätzung der Reliabilität durch eine zugrundeliegende schiefe Verteilung verhindert werden. Berechnet man die „normalen“ Cronbachs-Alpha-Werte für den MBI in der vorliegenden Stichprobe, erhält man Werte, die näher an den Ergebnissen der Metaanalysen liegen: .88 für die Skala EE, .74 für die Skala DP und .77 für die Skala PE.

Im Vergleich zu den Ergebnissen der Reliabilitätsstudien der deutschen Übersetzungen des MBI ist die in dieser Studie vorliegende akzeptable innere Konsistenz der Skala DP überraschend. Diese lag in den bisherigen Untersuchungen lediglich zwischen .42 und .67 (Burisch, 1984; Büssing & Perrar, 1992; Enzmann & Kleiber, 1989; Neubach & Schmidt, 2000). Möglicherweise ist dieser Unterschied auf die Stichprobenzusammensetzung zurückzuführen. Die bisherigen deutschen Studien untersuchten ausschließlich Personen in helfenden Berufen. Die vorliegende Stichprobe deckt auch andere Professionen ab.

Die ermittelte Trennschärfe der Items der Skala EE war hoch. Die Skala DP enthielt ein Item mit unzureichender Trennschärfe (.36). Es handelt sich dabei um Item 22 („Ich habe das Gefühl, dass mir meine Kollegen/Mitarbeitenden für manche ihrer Probleme die Schuld geben.“). Durch die Entfernung dieses Items könnte die interne Konsistenz der Skala von .79 auf .81 angehoben werden. Item 22 wies auch in den Untersuchungen der deutschen Übersetzungen von Burisch (1984), Büssing und Perrar (1992) und Enzmann und Kleiber (1989) sehr geringe Itemtrennschärfen auf. Die Ursache der Problematik dieses Items könnte in seiner inhaltlichen Ambiguität zu finden sein. Einerseits könnte der Eindruck, von anderen beschuldigt zu werden, ein negatives Gefühl gegenüber anderen sein, das objektiv nicht unbedingt den Tatsachen entsprechen muss, ganz im Sinne der Depersonalisation. Andererseits könnte es aber auch eine korrekte Einschätzung der Tatsache sein, dass man für die Probleme der anderen beschuldigt wird. Item 4 („Ich kann es leicht nachvollziehen, was in

meinen Kollegen/Vorgesetzten vorgeht.“) und Item 18 („Ich fühle mich angeregt, wenn ich eng mit meinen Kollegen zusammengearbeitet habe.“) der Skala PE wiesen ebenfalls eine unzureichende Trennschärfe (.37) auf. Das Entfernen der Items würde die interne Konsistenz der Skala allerdings nicht weiter erhöhen und im Falle von Item 4 sogar leicht absenken. Auch in den Studien von Burisch (1984) und Büssing und Perrar (1992) fielen diese Items aufgrund ihrer geringen Trennschärfe auf. Bei Enzmann und Kleiber (1989) zeigten sich mit diesen Items allerdings keine Probleme.

Die Skalen B (.88) und FR (.89) der BOSS I zeigten eine gute interne Konsistenz. Die Skalen EP (.93) und FA (.91) wiesen ausgezeichnete Cronbachs-Alpha-Werte auf. Die Ergebnisse sind höher als die von den Autor_innen berichteten (Geuenich & Hagemann, 2014). Die Skalenmittelwerte, die Geuenich und Hagemann (2014) für ihre Stichproben angeben, liegen deutlich im unteren Drittel des Wertebereichs. Dies weist darauf hin, dass auch dort schiefe Verteilungen vorlagen. Da von den Autor_innen keine ordinalen Reliabilitätskoeffizienten berechnet wurden, dürften ihre Angaben daher einer gewissen Unterschätzung unterliegen. Wie zuvor beschrieben, ist die höhere Reliabilitätsschätzung in der vorliegenden Studie deshalb vermutlich auf die Verwendung von ordinalen Koeffizienten zurückzuführen.

Die Trennschärfe der Items der BOSS I war, mit Ausnahme eines Items, zufriedenstellend. Item 9 der Skala B („Um beruflichen Stress abbauen zu können, greife ich auf Alkohol und/oder Medikamente zurück.“) wies eine unzureichende Trennschärfe (.37) auf. Die interne Konsistenz der Skala kann durch das Entfernen des Items allerdings nicht erhöht werden. Dieses Item wies auch in der Untersuchung von Geuenich und Hagemann (2014) eine niedrige Trennschärfe (.18) auf. Die Autor_innen argumentierten allerdings, dass eine Frage nach Substanzkonsum in einem Burnout-Instrument nicht fehlen dürfe, weshalb es beibehalten wurde. Aufgrund der guten internen Konsistenz der Skala und weil das Löschen dieses Items keine Verbesserung mit sich bringen würde, spricht auch mit den vorliegenden Daten nichts gegen das Beibehalten des Items. Die Trennschärfe zweier weiterer Items, die in der Eichstichprobe der Autor_innen zu niedrig ausfielen, wiesen in dieser Untersuchung zufriedenstellende Trennschärfen auf.

Die Skalen der BOSS II wiesen ausgezeichnete Reliabilitätskoeffizienten auf. Der Alphawert der Skala KOEB lag bei .91, jener der Skalen KOB und EB bei .95. Die Trennschärfe aller Items war hoch. Die Ergebnisse der Reliabilitätsanalyse der BOSS II sind ebenfalls besser als jene der Testautor_innen (Geuenich & Hagemann, 2014). Auch hier erklärt die unterschiedliche Methodik vermutlich den Großteil der Diskrepanz. Berechnet man

nicht-ordinale Cronbachs-Alpha-Werte mit den vorliegenden Daten, so sind die Ergebnisse jener der Gesamtstichprobe der Autor_innen sehr ähnlich.

Faktorielle Validität

Die vierte Fragestellung befasste sich mit der faktoriellen Validität der drei Instrumente und ob die durch die Testautor_innen festgelegte Faktorenstruktur repliziert werden konnte. Die Hauptachsenanalyse des MBI ergab eine dreifaktorielle Struktur mit zwei Kreuzladungen und einer Fehlladung. Die Items der Skala EE zeigten hauptsächlich auf dem ersten Faktor bedeutsame Ladungen, jene der Skala DP hauptsächlich auf dem zweiten Faktor und jene der Skala PE hauptsächlich auf dem dritten Faktor. Item 16 der Skala EE („Bei der Arbeit in direktem Kontakt zu Menschen zu stehen, stresst mich zu sehr.“) lud sowohl auf dem ersten (.52) als auch auf dem zweiten Faktor (.52) bedeutsam hoch. Eine Kreuzladung des Items 16 auf der Skala DP wurde auch schon von Byrne (1993), Enzmann und Kleiber (1989) und Schaufeli und van Dierendonck (1993) berichtet. Inhaltlich gesehen steht die Skala DP für negative Reaktionen, Gefühle oder Einstellungen gegenüber den Menschen, mit denen man arbeitet, sowie Distanzierung von ebendiesen. Die Skala EE soll emotionale Überlastung und Übermüdung bei der Arbeit darstellen. Es scheint, als würden sich diese beiden Dimensionen in Item 16 überschneiden. Die Formulierung „stresst mich zu sehr“ beschreibt ein Gefühl von Überforderung, könnte aber gewissermaßen auch eine negative Bewertung des direkten Kontakts und einen Wunsch nach Rückzug implizieren. Item 12 der Skala PE („Ich fühle mich voller Energie.“) zeigte ebenfalls eine Kreuzladung auf dem ersten (-.52) und dem dritten Faktor (.53). Auch dieses Item wies bereits in anderen Studien Kreuzladungen auf der Skala EE auf (Belcastro et al., 1983; Byrne, 1993; Schaufeli & van Dierendonck, 1993). Inhaltlich ist es nicht überraschend, dass ein Item, das einen hohen Energielevel beschreibt, eine hohe negative Ladung auf einer Skala aufweist, die Erschöpfung darstellt. Item 22 der Skala DP („Ich habe das Gefühl, dass mir meine Kollegen/Mitarbeitenden für manche ihrer Probleme die Schuld geben.“) lud als einziges Item des MBI auf keiner der Skalen bedeutsam hoch. Fehlladungen dieses Items traten auch in einigen in der Metaanalyse von Worley et al. (2008) inkludierten Studien sowie bei Enzmann und Kleiber (1989) auf. Darüber hinaus zeigte es eine unzureichende Trennschärfe in der vorliegenden und anderen Untersuchungen der deutschen MBI-Versionen (Burisch, 1984; Büssing & Perrar, 1992; Enzmann & Kleiber, 1989). Die Ursache für diese Ergebnisse könnte in der zuvor erklärten Ambiguität des Items liegen. Eine Entfernung des Items würde in den vorliegenden Daten zu einem stabilen Faktor führen, der eindeutig der Skala DP entspricht,

sowie die interne Konsistenz der Skala verbessern. Allerdings bestünde die Skala dann nur noch aus vier Items.

Aufgrund der Ergebnisse muss die Alternativhypothese H_1 (4.1) verworfen werden. Die von den Testautor_innen festgelegte dreifaktorielle Faktorenstruktur des MBI konnte zwar repliziert werden, die Items luden jedoch nicht eindeutig auf ihren Skalen.

Die Hauptachsenanalyse der BOSS I ergab eine zweifaktorielle Lösung. Der erste Faktor enthielt bedeutsame Ladungen aller Items der Skala EP, FR und B, bis auf Item 3 und 9, die auf keinem Faktor bedeutsame Ladungen aufwiesen. Item 9 der Skala B zeigte in dieser Untersuchung eine unzureichende Itemtrennschärfe. Der zweite Faktor enthielt bedeutsame Ladungen aller Items der Skala FA und Kreuzladungen aller Items der Skala FR und des Items 20 der Skala EP. In Bezug auf den Inhalt der Items könnte man sagen, dass der erste Faktor eher innerpsychische Beschwerden abbildete und der zweite Faktor eher zwischenmenschliche Probleme. Eine zusätzlich berechnete Hauptachsenanalyse mit vier Faktoren brachte eine schlechtere Lösung im Sinne der Einfachstruktur hervor. Obwohl Geuenich und Hagemann (2014) die faktorielle Validität der BOSS I ebenfalls nicht stützen konnten, unterscheidet sich ihre gefundene Lösung stark von der in dieser Studie vorliegenden. Die vorliegende zweifaktorielle Lösung ähnelt der tatsächlichen Skalenstruktur etwas besser, da die Items einer Skala nicht kreuz und quer auf vielen verschiedenen Faktoren bedeutsam luden, wie in der Lösung der Testautor_innen, sondern tendenziell einheitlich auf demselben Faktor oder auf beiden Faktoren. Die H_1 (4.2) muss verworfen werden. Die von den Testautor_innen festgelegte vierfaktorielle Faktorenstruktur der BOSS I konnte nicht repliziert werden. Die Items luden nicht eindeutig auf ihren Skalen.

Auch für die BOSS II wurde in der vorliegenden Untersuchung eine zweifaktorielle Lösung gefunden. Auf dem ersten Faktor luden alle Items der Skala KOEB und auf dem zweiten Faktor alle Items der Skalen KOB und EB in bedeutsamer Höhe. Item 11 der Skala KOB und Items 24, 26 und 27 der Skala EB, in denen es inhaltlich um Konzentrationsschwierigkeiten, gedrückte Stimmung, Gereiztheit und erhöhtes Misstrauen geht, luden auf beiden Faktoren bedeutsam. Der erste Faktor scheint inhaltlich körperliche Beschwerden und der zweite Faktor kognitiv-emotionale Beschwerden abzubilden. Darüber hinaus wurde eine dreifaktorielle Lösung berechnet, die aber im Sinne der Einfachstruktur nicht besser war. Die in dieser Untersuchung gefundene zweifaktorielle Lösung wies starke Unterschiede zu der Lösung von Geuenich und Hagemann (2014) auf. Legt man die Skalen KOB und EB zusammen, entspricht die zweifaktorielle Lösung, bis auf einige Kreuzladungen, den Skalen der BOSS II, was bei der Lösung der Testautor_innen kaum der

Fall war. Die H_1 (4.3) muss aufgrund der vorliegenden Ergebnisse trotzdem verworfen werden. Die von den Autor_innen postulierte dreifaktorielle Struktur der BOSS II konnte nicht repliziert werden. Die Items luden nicht eindeutig auf ihren Skalen.

Diskriminationsfähigkeit und relative Validität

Die nächsten beiden Fragestellungen beschäftigten sich mit der Diskriminationsfähigkeit der drei Instrumente. Zuerst wurde eine ROC-Analyse mit allen Skalen durchgeführt. Die Ergebnisse zeigten, dass alle Skalen signifikant zwischen Personen mit niedrigem und hohem Stresserleben unterscheiden konnten. Die Alternativhypothesen (5.1) bis (5.10) werden demnach angenommen: Es gab einen Unterschied zwischen Personen mit niedrigem und hohem subjektiven Stresserleben in allen Skalen des MBI, BOSS I und BOSS II. Die Diskriminationsfähigkeit unterschied sich aber deutlich zwischen den Skalen. Für das MBI zeigte sich, dass die Skala EE über eine gute, die Skala DP über eine mittelmäßige und die Skala PE über eine dürftige Diskriminationsfähigkeit verfügte. Die Skalen EP, B und FR der BOSS I konnten gut und die Skala FA mittelmäßig zwischen den Gruppen unterscheiden. Die Diskriminationsfähigkeit der Skalen EB und KOB war ebenfalls als gut zu beurteilen, jene der Skala KOEB als mittelmäßig. Gemittelt wies die BOSS I die besten AUC-Werte auf (.81), gefolgt von der BOSS II (.79) und dem MBI (.74). Vergleicht man die AUC-Werte der einzelnen Skalen über alle Instrumente hinweg, so zeigt sich, dass die Skala EE insgesamt den zweithöchsten Wert erreichte, während die anderen beiden Skalen des MBI die schlechtesten Werte hatten. Offenbar gibt es also innerhalb des MBI starke Unterschiede in der Diskriminationsfähigkeit bezüglich des Außenkriteriums. Auch bei der BOSS II gab es einen relativ großen Unterschied zwischen den AUC-Werten der Skalen EB und KOP einerseits und der Skala KOEB andererseits.

Darüber hinaus wurde die relative Validität der Skalen berechnet. Insgesamt erreichte die Skala B der BOSS I die höchste relative Validität, gefolgt von der Skala EE des MBI, der Skala EP der BOSS I und den Skalen KOB und EB der BOSS II. Die Skalen FA und FR der BOSS I sowie die Skala KOEB der BOSS II lagen im Mittelfeld. Die Skalen PE und DP konnten am schlechtesten zwischen den vier Stresserlebensgruppen unterscheiden. Auch bei diesen Ergebnissen wird der Unterschied zwischen den Skalen des MBI deutlich: die Skala EE zeigte eine fünfmal höhere relative Validität als die Skala DP.

Vergleich der Verfahren

Die abschließende Fragestellung lautete: Welches Verfahren (MBI oder BOSS I und BOSS II) eignet sich aufgrund seiner psychometrischen Güte besser zur Erfassung des Burnout-Syndroms bei berufstätigen Erwachsenen? Die Daten der beiden Verfahren waren in hohem Maße vollständig. Die BOSS I und II zeigten stärkere Bodeneffekte als das MBI, aber in der Gruppe mit hohem subjektiven Stresserleben wiesen beide Verfahren lediglich minimale Bodeneffekte auf. In Bezug auf die Reliabilität schnitten die BOSS I und II besser als das MBI ab. Während die Skala EE des MBI eine mit der BOSS I und II vergleichbare interne Konsistenz aufwies, blieb jene der anderen beiden MBI-Skalen deutlich zurück. Das Ergebnis der Faktorenanalyse des MBI lag relativ nahe an der postulierten Struktur. Es zeigt sich jedoch Verbesserungsbedarf bei den Skalen PE und DP des MBI. Einige problematische Items dieser Skalen sind schon seit 25 bis 30 Jahren bekannt und tauchen immer wieder in Validitätsstudien auf, wurden aber auch nach mehreren Neuauflagen von den Testautor_innen nicht verbessert, ersetzt oder entfernt. Im Gegensatz zum MBI wurden die BOSS I und II nicht faktorenanalytisch konstruiert, weshalb es nicht überraschend ist, dass sich ihre Skalenstruktur faktorenanalytisch nicht replizieren ließ. Die in dieser Untersuchung gefundene Faktorenstruktur der BOSS I und II war allerdings nicht so weit von der tatsächlichen Skalenstruktur entfernt wie jene von den Testautor_innen selbst gefundene. Aus dieser Sicht legen die vorliegenden Daten eine Zusammenlegung der Skalen B und EB sowie FA und FR der BOSS I und KOB und EB der BOSS II nahe. Da sich weder die Struktur des MBI noch jene der BOSS I und II eindeutig replizieren ließ, trägt dieses Kriterium wenig zur Beantwortung der Fragestellung bei, zeigt aber Ansatzpunkte zur Verbesserung der untersuchten Verfahren auf. In Bezug auf die Diskriminationsfähigkeit zeigte sich, dass BOSS I und II insgesamt besser zwischen den zwei und vier Stresserlebensgruppen differenzieren konnten. Auch die Skala EE konnte im Vergleich gut differenzieren. Die beiden anderen Skalen des MBI waren allerdings sowohl in der ROC-Analyse als auch in Bezug auf die relative Validität die jeweils am schlechtesten abschneidenden Skalen. Zusammenfassend kann man also sagen, dass die Ergebnisse dieser Studie für eine bessere Eignung der BOSS I und II sprechen.

Unabhängig von den untersuchten Kriterien haben BOSS I und II aber einige Nachteile. So ist ihre Itemanzahl höher (60 Items im Vergleich zu 22 Items beim MBI) und damit die Bearbeitungszeit länger. Allerdings betonen die Autoren, dass für die Verdachtsdiagnose Burnout vor allem die BOSS I relevant ist und BOSS II lediglich Zusatzinformationen liefere. Setzt man also nur die BOSS I ein, sind es im Vergleich zum

MBI nur acht Items mehr. Darüber hinaus ist die Vergleichbarkeit mit anderen Studien zum Thema Burnout, die meist den MBI einsetzen, eingeschränkt. Über die theoretische Basis der Instrumente soll und kann hier kein Urteil gefällt werden, da es keine einheitlich akzeptierte Burnout-Definition und keine eindeutigen Diagnosekriterien gibt, woran sich ein solches Urteil festmachen lassen könnte.

Limitationen und Ausblick

Die vorliegende Studie hat zwei wichtige Einschränkungen. Die erhobene Stichprobe erlebte insgesamt relativ wenig Stress und Burnout-Symptomatik im Sinne des MBI und der BOSS I und II. Die Ergebnisse lassen sich daher nicht uneingeschränkt auf eine „belastetere“ Population generalisieren.

Darüber hinaus ist das zur Beurteilung der Diskriminationsfähigkeit herangezogene Außenkriterium des subjektiven Stresserlebens keine optimale Operationalisierung von Burnout, sondern lediglich ein Versuch der Annäherung an das Konstrukt. Die Ergebnisse haben diesbezüglich eingeschränkte Aussagekraft. Solange es keinen Konsens in Bezug auf die Definition von Burnout und auf deren Basis entwickelte Diagnosekriterien gibt, ist keine optimale Operationalisierung möglich. Die Burnout-Forschung sollte sich daher in Zukunft unbedingt stärker mit ihrem wackeligen Fundament auseinandersetzen und eine gemeinsame Basis finden. Erst dann wird eine systematische Untersuchung des Syndroms möglich sein. Bis dahin sollten die momentan eingesetzten Erhebungsverfahren besser durchleuchtet werden. Es gibt große Versäumnisse in der psychometrischen Untersuchung und der Verbesserung gebräuchlicher Messinstrumente.

Literaturverzeichnis

- Aguayo, R., Vargas, C., de la Fuente, E. I., & Lozano, L. M. (2011). A meta-analytic reliability generalization study of the Maslach Burnout Inventory. *International Journal of Clinical and Health Psychology*, 11, 343–361.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: APA.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: APA. doi:10.1176/appi.books.9780890425596
- Aronson, E., Pines, A., & Kafry, D. (1983). *Ausgebrannt: Vom Überdruß zur Selbstentfaltung*. Stuttgart: Klett-Cotta.
- Barth, A. (1985). *Erste Untersuchung mit einer deutschen Übersetzung des Maslach Burnout Inventory bei klientenzentrierten Gesprächstherapeuten und Hochschullehrern* (Unveröffentlichte Zulassungsarbeit zur staatlichen Ergänzungsprüfung im Fach Psychologie). Universität Nürnberg.
- Belcastro, P. A., Gold, R. S., & Hays, L. C. (1983). Maslach Burnout Inventory: Factor structures for samples of teachers. *Psychological Reports*, 53, 364–366. doi:10.2466/pr0.1983.53.2.364
- Bianchi, R., Schonfeld, I. S., & Laurent, E. (2015a). Burnout-depression overlap: A review. *Clinical Psychology Review*, 36, 28–41. doi:10.1016/J.CPR.2015.01.004
- Bianchi, R., Schonfeld, I. S., & Laurent, E. (2015b). Is it time to consider the “burnout syndrome” a distinct illness? *Frontiers in Public Health*, 3, 158. doi:10.3389/fpubh.2015.00158
- Bianchi, R., Schonfeld, I. S., & Laurent, E. (2019). Burnout: Moving beyond the status quo. *International Journal of Stress Management*, 26, 36–45. doi:10.1037/str0000088
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer. doi:10.1007/978-3-540-33306-7
- Burisch, M. (1984). *The Maslach Burnout Inventory and the Tedium Measure: Reliability and validity in a German sample*. Unveröffentlichtes Manuskript.
- Burisch, M. (2014). *Das Burnout-Syndrom: Theorie der inneren Erschöpfung* (5. Aufl.). Berlin: Springer. doi:10.1007/978-3-642-36255-2
- Büssing, A., & Perrar, K.-M. (1992). Die Messung von Burnout: Untersuchung einer deutschen Fassung des Maslach Burnout Inventory (MBI-D). *Diagnostica*, 38, 328–353.
- Byrne, B. M. (1993). The Maslach Burnout Inventory: Testing for factorial validity and invariance across elementary, intermediate and secondary teachers. *Journal of*

- Occupational and Organizational Psychology*, 66, 197–212. doi:10.1111/j.2044-8325.1993.tb00532.x
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. doi:10.1007/BF02310555
- de Silva, P. V., Hewage, C. G., & Fonseka, P. (2013). Measurement of burnout: Validation of the Sinhala translation of Maslach Burnout Inventory – Educators Survey among female primary school teachers in Sri Lanka. *Galle Medical Journal*, 18, 8–11. doi:10.4038/gmj.v18i1.5512
- Demerouti, E., Bakker, A. B., Vardakou, I., & Kantas, A. (2003). The convergent validity of two burnout instruments: A multitrait-multimethod analysis. *European Journal of Psychological Assessment*, 18, 296–307. doi:10.1027//1015-5759.19.1.12
- Der Maslach Burnout Inventory-Test. (n.d.). Abgerufen von [https://staedteverband.ch/cmsfiles/test_de_maslach_\(burnout\)_-_d.pdf](https://staedteverband.ch/cmsfiles/test_de_maslach_(burnout)_-_d.pdf)
- Derogatis, L. R. (1977). *SCL-90-R: Administration, scoring & procedures manual-I for the r(evised) version*. Baltimore, MD: Johns Hopkins University School of Medicine.
- Dreher, A., Theune, M., Kersting, C., Geiser, F., & Weltermann, B. (2019). Prevalence of burnout among German general practitioners: Comparison of physicians working in solo and group practices. *PLoS ONE*, 14(2):e0211223 doi:10.1371/journal.pone.0211223
- Enzmann, D., & Kleiber, D. (1989). *Helfer-Leiden: Streß und Burnout in Psychosozialen Berufen*. Heidelberg: Asanger.
- Fliege, H., Rose, M., Arck, P., Levenstein, S., & Klapp, B. F. (2001). Validierung des “Perceived Stress Questionnaire” (PSQ) an einer deutschen Stichprobe. *Diagnostica*, 47, 142–152. doi:10.1026//0012-1924.47.3.142
- Fliege, H., Rose, M., Arck, P., Walter, O. B., Kocalevent, R. D., Weber, C., & Klapp, B. F. (2005). The Perceived Stress Questionnaire (PSQ) reconsidered: Validation and reference values from different clinical and healthy adult samples. *Psychosomatic Medicine*, 67, 78–88. doi:10.1097/01.psy.0000151491.80178.78
- Freudenberger, H. J. (1974). Staff burn-out. *Journal of Social Issues*, 30, 159–165. doi:10.1111/j.1540-4560.1974.tb00706.x
- Freudenberger, H. J. (1975). The staff burn-out syndrome in alternative institutions. *Psychotherapy: Theory, Research & Practice*, 12, 73–82. doi:10.1037/h0086411
- Friberg, T. (2009). Burnout: From popular culture to psychiatric diagnosis in Sweden. *Culture, Medicine, and Psychiatry*, 33, 538–558. doi:10.1007/s11013-009-9149-z
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for

- Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17(3), 1–13.
- Garrido, L. E. (2017). Horn's parallel analysis method with polychoric correlations [MATLAB Syntax Files]. Abgerufen von <https://www.mathworks.com/matlabcentral/fileexchange/64213-horn-s-parallel-analysis-method-with-polychoric-correlations>
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods*, 18, 454–474. doi:10.1037/a0030005
- Geuenich, K., & Hagemann, W. (2014). *BOSS. Burnout Screening Skalen* (2. Aufl.). Göttingen: Hogrefe.
- Hagemann, W., & Geuenich, K. (2009). *BOSS. Burnout-Screening-Skalen*. Göttingen: Hogrefe.
- Heinemann, L. V., & Heinemann, T. (2017). Burnout research: Emergence and scientific investigation of a contested diagnosis. *SAGE Open*, 7:215824401769715. doi:10.1177/2158244017697154
- Holgado-Tello, F. P., Chacón-MoscOSO, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity*, 44, 153–166. doi:10.1007/s11135-008-9190-y
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185. doi:10.1007/BF02289447
- Hussy, W., Schreier, M., & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (2. Aufl.). Berlin: Springer. doi:10.1007/978-3-642-34362-9
- International Labour Organization. (2012). *ISCO-08 – International Standard Classification of Occupations*. Genf: ILO.
- Jordan, P., Steingen, U., Terschüren, C., & Harth, V. (2018). The Maslach Burnout Inventory: A test dimensionality assessment via item response theory. *Testing, Psychometrics, Methodology in Applied Psychology*, 25, 101–120. doi:10.4473/TPM25.1.6
- Kafry, D., & Pines, A. (1980). The experience of tedium in life and work. *Human Relation*, 33, 477–503.
- Kelava, A., & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 75–102). Berlin: Springer. doi:10.1007/978-3-642-20072-4_4

- Kinschel, M. (2013). *Über den Zusammenhang der Mitarbeiterzufriedenheit auf die verschiedenen Facetten von Burnout bei Flugbegleiterinnen und Flugbegleitern* (Dissertation). Universitaet Koblenz-Landau.
- Kleijweg, J. H. M., Verbraak, M. J. P. M., & van Dijk, M. K. (2013). Comparing the clinical utility of the Maslach Burnout Inventory in a clinical population. *Psychological Assessment*, 25, 435–441. doi:10.1037/a0031334
- Koutsimani, P., Montgomery, A., & Georganta, K. (2019). The relationship between burnout, depression, and anxiety: A systematic review and meta-analysis. *Frontiers in Psychology*, 10:284. doi:10.3389/fpsyg.2019.00284
- Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. *Work & Stress*, 19, 192–207. doi:10.1080/02678370500297720
- Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., Berto, E., Luzi, C., & Andreoli, A. (1993). Development of the Perceived Stress Questionnaire: A new tool for psychosomatic research. *Journal of Psychosomatic Research*, 37, 19–32. doi:10.1016/0022-3999(93)90120-5
- Mancuso, C. A., Peterson, M. G. E., & Charlson, M. E. (2001). Comparing discriminative validity between a disease-specific and a general health scale in patients with moderate asthma. *Journal of Clinical Epidemiology*, 54, 263–274. doi:10.1016/S0895-4356(00)00307-3
- Maslach, C. (1976). Burned-out. *Human Behavior*, 9, 16–22.
- Maslach, C. (1978). The client role in staff burn-out. *Journal of Social Issues*, 34, 111–124.
- Maslach, C. (1981). Burnout: a social psychological analysis. In J. W. Jones (Ed.), *The burnout syndrome: Current research, theory and interventions* (pp. 30–53). Park Ridge, IL: London House Press.
- Maslach, C. (1982). Understanding burnout: Definitional issues in analyzing a complex phenomenon. In W. S. Paige (Ed.), *Job stress and burnout: Research, theory, and intervention perspectives* (pp. 29–40). Beverly Hills, CA: Sage.
- Maslach, C. (2000). A multidimensional theory of burnout. In C. L. Cooper (Ed.), *Theories of organizational stress* (pp. 68–85). Oxford: University Press.
- Maslach, C., & Jackson, S. E. (1979). Burned-out cops and their families. *Psychology Today*, 12, 59–62.
- Maslach, C., & Jackson, S. E. (1981a). *The Maslach Burnout Inventory*. Palo Alto, CA: Consulting Psychologists Press.

- Maslach, C., & Jackson, S. E. (1981b). The measurement of experienced burnout. *Journal of Occupational Behaviour*, 2, 99–113. doi:10.1002/job.4030020205
- Maslach, C., & Jackson, S. E. (1984). Burnout in organization settings. *Applied Social Psychology Annual*, 5, 133–153.
- Maslach, C., & Jackson, S. E. (1986). *The Maslach Burnout Inventory* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Maslach, C., & Jackson, S. E. (1996). Maslach Burnout Inventory – Human Services Survey (MBI-HSS). In C. Maslach, S. E. Jackson, & M. P. Leiter (Eds.), *MBI manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Maslach, C., Jackson, S. E., & Leiter, M. P. (1997). Maslach Burnout Inventory. In C. P. Zalaquett & R. J. Wood (Eds.), *Evaluating stress: A book of resources* (pp. 191–218). Lanham, MD: The Scarecrow Press.
- Maslach, C., Jackson, S. E., & Schwab, R. L. (1996). Maslach Burnout Inventory – Educators Survey (MBI-ES). In C. Maslach, S. E. Jackson, & M. P. Leiter (Eds.), *MBI manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Maslach, C., & Pines, A. (1977). The burn-out syndrome in the day care setting. *Child Care Quarterly*, 6, 100–113. doi:10.1007/BF01554696
- Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, 52, 397–422. doi: 10.1146/annurev.psych.52.1.397
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.
- Melamed, S., Kushnir, T., & Shirom, A. (1992). Burnout and risk factors for cardiovascular diseases. *Behavioral Medicine*, 18, 53–60. doi:10.1080/08964289.1992.9935172
- Neubach, B., & Schmidt, K. H. (2000). Gütekriterien einer deutschen Fassung des Maslach Burnout Inventory (MBI-D) – Eine Replikationsstudie bei Altenpflegekräften. *Zeitschrift für Arbeits- und Organisationspsychologie A&O*, 44, 140–144. doi:10.1026//0932-4089.44.3.140
- Pines, A., & Aronson, E. (1988). *Career burnout: Causes and cures*. New York, NY: Free Press.
- Pines, A., & Kafry, D. (1978). Occupational tedium in the social services. *Social Work*, 23, 499–507. doi:10.1093/sw/23.6.499
- Pines, A., & Kafry, D. (1981). Tedium in the life and work of professional women as compared with men. *Sex Roles*, 7, 963–977. doi:10.1007/BF00288498
- Pines, A., & Maslach, C. (1978). Characteristics of staff burn-out in mental health settings. *Psychiatric Services*, 29, 233–237. doi:10.1176/ps.29.4.233

- Pines, A., & Maslach, C. (1980). Combatting staff burn-out in a day care center: A case study. *Child Care Quarterly*, 9, 5–16. doi:10.1007/BF01555032
- Pituch, K. A., & Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS* (6th ed.). New York, NY: Routledge. doi:10.4324/9781315814919
- Riley, M. R., Mohr, D. C., & Waddimba, A. C. (2018). The reliability and validity of three-item screening measures for burnout: Evidence from group-employed health care practitioners in upstate New York. *Stress and Health*, 34, 187–193. doi:10.1002/smi.2762
- Roelen, C. A. M., van Hoffen, M. F. A., Groothoff, J. W., de Bruin, J., Schaufeli, W. B., & van Rhenen, W. (2015). Can the Maslach Burnout Inventory and Utrecht Work Engagement Scale be used to screen for risk of long-term sickness absence? *International Archives of Occupational and Environmental Health*, 88, 467–475. doi:10.1007/s00420-014-0981-2
- Roelofs, J., Verbraak, M., Keijsers, G. P. J., de Bruin, M. B. N., & Schmidt, A. J. M. (2005). Psychometric properties of a Dutch version of the Maslach Burnout Inventory General Survey (MBI-DV) in individuals with and without clinical burnout. *Stress and Health*, 21(1), 17–25. doi:10.1002/smi.1032
- Schaufeli, W. B., Bakker, A. B., Hoogduin, K., Schaap, C., & Kladler, A. (2001). On the clinical validity of the Maslach Burnout Inventory and the Burnout Measure. *Psychology and Health*, 16, 565–582. doi:10.1080/08870440108405527
- Schaufeli, W. B., & Enzmann, D. (1998). *The burnout companion to study & practice: A critical analysis*. London: Taylor & Francis.
- Schaufeli, W. B., Leiter, M. P., & Maslach, C. (2009). Burnout: 35 years of research and practice. *Career Development International*, 14, 204–220. doi:10.1108/13620430910966406
- Schaufeli, W. B., Leiter, M. P., Maslach, C., & Jackson, S. E. (1996). Maslach Burnout Inventory – General Survey (MBI-GS). In C. Maslach, S. E. Jackson, & M. P. Leiter (Eds.), *MBI manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Schaufeli, W. B., & van Dierendonck, D. (1993). The construct validity of two burnout measures. *Journal of Organizational Behavior*, 14, 631–647. doi:10.1002/job.4030140703
- Schermelleh-Engel, K., & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S.

- 119–141). Berlin: Springer. doi:10.1007/978-3-642-20072-4_6
- Schonfeld, I. S., Bianchi, R., & Palazzi, S. (2018). What is the difference between depression and burnout? An ongoing debate. *Rivista di Psichiatria*, 53, 218–219.
- Schulz, P., Schlotz, W., & Becker, P. (2004). *TICS. Trier Inventar zum chronischen Stress*. Göttingen: Hogrefe.
- Shirom, A. (2005). Reflections on the study of burnout. *Work & Stress*, 19, 263–270. doi:10.1080/02678370500376649
- Statistik Austria. (2011). *Systematik der Berufe: ÖISCO-08*. Wien: Autor.
- Steinlin, C., Dölitzsch, C., Fischer, S., Schmeck, K., Fegert, J. M., & Schmid, M. (2016). Der Zusammenhang zwischen Burnout-Symptomatik und Arbeitszufriedenheit bei pädagogischen Mitarbeitenden in der stationären Kinder- und Jugendhilfe. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 65, 162–180. doi:10.13109/prkk.2016.65.3.162
- Ware, J. E., & Gandek, B. (1998). Methods for testing data quality, scaling assumptions, and reliability: The IQOLA Project approach. *Journal of Clinical Epidemiology*, 51, 945–952. doi:10.1016/S0895-4356(98)00085-7
- Wheeler, D. L., Vassar, M., Worley, J. A., & Barnes, L. L. B. (2011). A reliability generalization meta-analysis of coefficient alpha for the Maslach Burnout Inventory. *Educational and Psychological Measurement*, 71, 231–244. doi:0.1177/0013164410391579
- World Health Organization. (2016). *International Statistical Classification of Diseases and Related Health Problems* (10th ed.). Abgerufen von <https://icd.who.int/browse10/2016/en>
- World Health Organization. (2019). *ICD-11 for mortality and morbidity statistics* (Version: 04 / 2019). Abgerufen von <https://icd.who.int/browse11/l-m/en>
- Worley, J. A., Vassar, M., Wheeler, D. L., & Barnes, L. L. B. (2008). Factor structure of scores from the Maslach Burnout Inventory: A review and meta-analysis of 45 exploratory and confirmatory factor-analytic studies. *Educational and Psychological Measurement*, 68, 797–823. doi:10.1177/0013164408315268

Abbildungsverzeichnis

Abbildung 1. Verlauf der anfänglichen Eigenwerte der Faktorenanalyse des MBI und der zufälligen Eigenwerte de Parallelanalyse.....	41
Abbildung 2. Verlauf der anfänglichen Eigenwerte der Faktorenanalyse der BOSS I und der zufälligen Eigenwerte der Parallelanalyse.....	43
Abbildung 3. Verlauf der anfänglichen Eigenwerte der Faktorenanalyse der BOSS II und der zufälligen Eigenwerte der Parallelanalyse.....	45
Abbildung 4. ROC-Kurven der Skalen des MBI, der BOSS I und der BOSS II.....	47

Tabellenverzeichnis

Tabelle 1. Deskriptive Skalenanalyse der Instrumente MBI, BOSS I und BOSS II.....	34
Tabelle 2. Verteilung der gewählten Antwortoptionen der Skalen des MBI in der Gesamtstichprobe und den vier Stresserlebensgruppen.....	36
Tabelle 3. Verteilung der gewählten Antwortoptionen der Skalen der BOSS I in der Gesamtstichprobe und den vier Stresserlebensgruppen.....	37
Tabelle 4. Verteilung der gewählten Antwortoptionen der Skalen der BOSS II in der Gesamtstichprobe und den vier Stresserlebensgruppen.....	39
Tabelle 5. Reliabilitätsanalyse der Instrumente MBI, BOSS I und BOSS I.....	40
Tabelle 6. Faktorladungen der Items des MBI.....	42
Tabelle 7. Faktorladungen der Items der BOSS I.....	44
Tabelle 8. Faktorladungen der Items der BOSS II.....	46
Tabelle 9. AUC-Werte der Skalen des MBI, der BOSS I und BOSS II.....	48
Tabelle 10. Skalenmittelwerte der Instrumente MBI, BOSS I und BOSS II nach Ausprägung des Stresserlebens.....	49
Tabelle A1. Anteil an fehlenden Werten der Items des MBI.....	68
Tabelle A2. Anteil an fehlenden Werten der Items der BOSS I.....	69
Tabelle A3. Anteil an fehlenden Werten der Items des BOSS II.....	70
Tabelle B1. Verteilung der Skalenwerte des MBI, BOSS I und BOSS II in der Gruppe mit niedrigem Stresserleben.....	71
Tabelle B2. Verteilung der Skalenwerte des MBI, BOSS I und BOSS II in der Gruppe mit eher niedrigem Stresserleben.....	72
Tabelle B3. Verteilung der Skalenwerte des MBI, BOSS I und BOSS II in der Gruppe mit eher hohem Stresserleben.....	73
Tabelle B4. Verteilung der Skalenwerte des MBI, BOSS I und BOSS II in der Gruppe mit hohem Stresserleben.....	74
Tabelle C1. Itemanalyse des MBI.....	75
Tabelle C2. Itemanalyse der BOSS I.....	76
Tabelle C3. Itemanalyse der BOSS II.....	77
Tabelle D1. Faktorladungen der Items der BOSS I mit 4 Faktoren.....	78
Tabelle D2. Faktorladungen der Items der BOSS II mit 4 Faktoren.....	79

Anhang A: Anteil an fehlenden Werten pro Item

Tabelle A1

Anteil an fehlenden Werten der Items des MBI

Item	Skala	Fehlende Werte	
		absolut	prozentuell
1. durch Arbeit emotional erschöpft	EE	3	0.5
2. am Ende des Arbeitstages verbraucht	EE	1	0.3
3. morgens vor Arbeitstag ermüdet	EE	1	0.3
4. nachvollziehen können was in K./V. vorgeht	PE	3	0.8
5. K./V. wie Objekte behandeln	DP	2	0.5
6. Anstrengung, mit Menschen zu arbeiten	EE	2	0.5
7. guter Umgang mit Problemen anderer	PE	2	0.5
8. ausgebrannt durch die Arbeit	EE	3	0.8
9. andere positiv beeinflusst durch Arbeit	PE	2	0.5
10. gleichgültiger gegenüber Menschen	DP	5	1.3
11. emotional verhärtet durch Arbeit	DP	2	0.5
12. voller Energie	PE	1	0.3
13. frustriert durch Arbeit	EE	1	0.3
14. zu verbissen arbeiten	EE	1	0.3
15. kein Interesse an manchen K.	DP	2	0.5
16. direkter Kontakt stresst zu sehr	EE	2	0.5
17. leichtes Schaffen entspannter Atmosphäre	PE	1	0.3
18. angeregt, wenn eng mit K. gearbeitet	PE	4	1.0
19. viele lohnende Ziele bei Arbeit erreicht	PE	3	0.8
20. am Ende der Weisheit zu sein	EE	3	0.8
21. gelassener Umgang mit emot. Problemen	PE	3	0.8
22. K./M. geben mir Schuld für Probleme	DP	3	0.8

Anmerkung. K. = Kolleg_innen; V. = Vorgesetzte; M. = Mitarbeitende; EE = Emotionale Erschöpfung; PE = Persönliche Erfüllung; DP = Depersonalisation. *N* = 392.

Tabelle A2*Anteil an fehlenden Werten der Items der BOSS I*

Item	Skala	Fehlende Werte	
		absolut	prozentuell
1. Sorge, Job nicht zu bewältigen	B	0	0.0
2. vermehrtes Kontrollieren, misstrauischer	B	0	0.0
3. Arbeiten bis spät in die Nacht	B	2	0.5
4. keine nachhaltige Freude über Erfolge	B	2	0.5
5. Fassung verlieren, wenn nicht nach Plan	B	0	0.0
6. Dauerstress durch lange Arbeitswege	B	2	0.5
7. Verlust der Freude an der Arbeit	B	0	0.0
8. wenig treffsicher in berufl. Entscheidungen	B	0	0.0
9. Alkohol/Medikamente zum Stressabbau	B	1	0.3
10. gefährdetes positives Image im Job	B	2	0.5
11. sich selbst fremd geworden sein	EP	0	0.0
12. keine Lebensfreude mehr empfinden können	EP	0	0.0
13. herabgesetzte Frustrationstoleranz	EP	1	0.3
14. zunehmende Sorgen um Gesundheit	EP	0	0.0
15. häufige banale Infekte	EP	0	0.0
16. kleine Anforderungen bereits belastend	EP	0	0.0
17. schlechter Schlaf	EP	0	0.0
18. Verlust der Routine, alles schwerer	EP	2	0.5
19. zunehmende Konzentrationsschwierigkeiten	EP	0	0.0
20. Schwierigkeiten beim Entspannen	EP	1	0.3
21. Unstimmigkeiten in Familie/Partnerschaft	FA	3	0.8
22. zu wenig Zeit für Gemeinsamkeiten	FA	6	1.5
23. drohende Entfremdung	FA	5	1.3
24. Sexualität/Erotik kommt zu kurz	FA	8	2.0
25. weniger Anteil an Familie/Partnerschaft	FA	3	0.8
26. besorgte Rückmeldungen von Freunden	FR	1	0.3
27. häufiger Rückzug aus sozialen Kontakten	FR	2	0.5
28. Belastungen sind einziges Gesprächsthema	FR	0	0.0
29. weniger Anteil an Freundschaften	FR	0	0.0
30. Schwierigkeiten, sich abzugrenzen	FR	0	0.0

Anmerkung. B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde. *N* = 392.

Tabelle A3*Anteil an fehlenden Werten der Items der BOSS II*

Item	Skala	Fehlende Werte	
		absolut	prozentuell
1. heftige Kopfschmerzen	KOEB	1	0.3
2. häufig wiederkehrende Infekte	KOEB	1	0.3
3. ständige Muskelverspannungen	KOEB	1	0.3
4. beklemmende Enge über der Brust	KOEB	0	0.0
5. Atembeschwerden	KOEB	1	0.3
6. Schmerzen	KOEB	0	0.0
7. Herzbeschwerden	KOEB	0	0.0
8. Magen-Darmbeschwerden	KOEB	1	0.3
9. Schlafstörungen	KOEB	0	0.0
10. hohe oder schwankende Blutdruckwerte	KOEB	1	0.3
11. Konzentrationsschwierigkeiten	KOB	1	0.3
12. zu negative Bewertungen	KOB	0	0.0
13. Verlust der Entschlussfreudigkeit	KOB	1	0.3
14. vermehrte Fehlerneigung	KOB	0	0.0
15. nur noch Reagieren statt Agieren	KOB	0	0.0
16. Verlust der Kreativität	KOB	1	0.3
17. wiederkehrender, quälender Gedanke	KOB	0	0.0
18. verstärkte Selbstkontrolle	KOB	1	0.3
19. Verlust der Souveränität	KOB	2	0.5
20. verminderte Durchsetzungsfähigkeit	KOB	0	0.0
21. Zukunftsängste	EB	0	0.0
22. eingeschränkte Selbstwahrnehmung	EB	0	0.0
23. sozialer Rückzug	EB	0	0.0
24. gedrückte Stimmung	EB	2	0.5
25. Antriebslosigkeit	EB	0	0.0
26. Gereiztheit	EB	1	0.3
27. erhöhtes Misstrauen	EB	0	0.0
28. Scham	EB	0	0.0
29. erhöhte Ängstlichkeit	EB	1	0.3
30. Geduldverlust bei Misslingen	EB	0	0.0

Anmerkung. KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden. *N* = 392.

Anhang B: Verteilung der Skalenwerte in den Stresserlebensgruppen

Tabelle B1

Verteilung der Skalenwerte des MBI, BOSS I und BOSS II in der Gruppe mit niedrigem Stresserleben

Skala	N	M (SD)	Schiefe	Kurtosis
MBI				
EE	104	14.81 (9.11)	1.08***	2.50***
DP	104	10.64 (12.39)	1.45***	1.70***
PE ^a	104	28.08 (15.32)	0.72**	0.36
BOSS I				
B	104	5.33 (5.74)	1.38***	1.85***
EP	104	4.67 (5.82)	1.94***	4.54***
FA	103	9.94 (12.78)	1.64***	2.45***
FR	104	5.38 (8.36)	3.13***	13.84***
BOSS II				
KOEB	104	5.17 (6.73)	2.80***	11.75***
KOB	104	5.67 (6.87)	1.63***	2.53***
EB	104	3.42 (5.47)	2.50***	7.14***

Anmerkung. SE der Schiefe = 0.24; SE der Kurtosis = 0.47. EE = Emotionale Beschwerden; DP = Depersonalisation; PE = Persönliche Erfüllung; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden.

^aumgepolt.

** $\alpha = 1.0\%$. *** $\alpha = 0.1\%$.

Tabelle B2

Verteilung der Skalenwerte des MBI, BOSS I und BOSS II in der Gruppe mit eher niedrigem Stresserleben

Skala	N	M (SD)	Schiefe	Kurtosis
MBI				
EE	102	22.69 (10.66)	0.46	0.09
DP	102	11.67 (13.56)	2.23***	6.72***
PE ^a	101	32.34 (13.63)	0.65**	0.98*
BOSS I				
B	103	11.44 (7.99)	0.60*	0.06
EP	103	9.53 (8.51)	1.48***	2.24***
FA	101	18.26 (15.06)	1.03***	0.85
FR	103	10.87 (11.65)	1.37***	2.18***
BOSS II				
KOEB	103	9.32 (9.32)	1.37***	1.59***
KOB	103	12.62 (10.37)	0.62**	-0.73
EB	103	8.93 (9.04)	1.32***	1.09*

Anmerkung. SE der Schiefe = 0.24; SE der Kurtosis = 0.47. EE = Emotionale Beschwerden; DP = Depersonalisation; PE = Persönliche Erfüllung; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden.

^aumgepolt.

* $\alpha = 5.0\%$. ** $\alpha = 1.0\%$. *** $\alpha = 0.1\%$.

Tabelle B3

Verteilung der Skalenwerte des MBI, BOSS I und BOSS II in der Gruppe mit eher hohem Stresserleben

Skala	N	M (SD)	Schiefe	Kurtosis
MBI				
EE	86	29.22 (11.87)	0.98***	0.83
DP	86	18.06 (14.95)	1.18***	0.90
PE ^a	86	35.90 (15.00)	0.23	-0.60
BOSS I				
B	86	15.09 (8.17)	0.30	-0.45
EP	86	16.21 (11.10)	0.63*	-0.47
FA	84	26.10 (17.35)	0.47	-0.35
FR	86	16.98 (13.90)	1.55***	5.05***
BOSS II				
KOEB	86	12.37 (11.24)	1.34***	1.70***
KOB	86	17.70 (12.79)	0.85**	0.30
EB	86	14.56 (11.65)	0.95***	0.22

Anmerkung. SE der Schiefe = 0.26; SE der Kurtosis = 0.51. EE = Emotionale Beschwerden; DP = Depersonalisation; PE = Persönliche Erfüllung; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden.

^aumgepolt.

* $\alpha = 5.0\%$. ** $\alpha = 1.0\%$. *** $\alpha = 0.1\%$.

Tabelle B4

Verteilung der Skalenwerte des MBI, BOSS I und BOSS II in der Gruppe mit hohem Stresserleben

Skala	N	M (SD)	Schiefe	Kurtosis
MBI				
EE	98	44.48 (17.67)	0.52 [*]	0.02
DP	97	25.46 (18.91)	0.91 ^{***}	0.63
PE ^a	98	44.90 (16.69)	-0.05	-0.37
BOSS I				
B	98	28.92 (14.60)	0.82 ^{***}	1.36 ^{**}
EP	98	31.04 (18.25)	0.80 ^{***}	0.28
FA	98	34.37 (21.32)	0.59 [*]	-0.31
FR	98	31.10 (20.51)	0.86 ^{***}	0.53
BOSS II				
KOEB	98	23.84 (19.54)	0.96 ^{***}	0.28
KOB	98	34.10 (19.19)	0.72 ^{**}	0.35
EB	98	28.71 (19.26)	1.13 ^{***}	1.23 [*]

Anmerkung. SE der Schiefe = 0.26; SE der Kurtosis = 0.48. EE = Emotionale Beschwerden; DP = Depersonalisation; PE = Persönliche Erfüllung; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden.

^aumgepolt.

* $\alpha = 5.0\%$. ** $\alpha = 1.0\%$. *** $\alpha = 0.1\%$.

Anhang C: Trennschärfeanalyse

Tabelle C1

Itemanalyse des MBI

Item	Skala	M (SD)	r_{IT}	α ohne Item
1. durch Arbeit emotional erschöpft	EE	1.94 (1.36)	.73	.88
2. am Ende des Arbeitstages verbraucht	EE	2.75 (1.52)	.68	.89
3. morgens vor Arbeitstag ermüdet	EE	1.67 (1.55)	.58	.89
4. nachvollziehen können was in K./V. vorgeht	PE	1.83 (1.58)	.37	.78
5. K./V. wie Objekte behandeln	DP	0.67 (1.15)	.60	.74
6. Anstrengung, mit Menschen zu arbeiten	EE	1.97 (1.48)	.57	.89
7. guter Umgang mit Problemen anderer	PE	1.71 (1.45)	.54	.76
8. ausgebrannt durch die Arbeit	EE	1.45 (1.41)	.81	.87
9. andere positiv beeinflusst durch Arbeit	PE	1.71 (1.71)	.56	.75
10. gleichgültiger gegenüber Menschen	DP	1.05 (1.51)	.69	.70
11. emotional verhärtet durch Arbeit	DP	1.22 (1.60)	.66	.71
12. voller Energie	PE	2.16 (1.48)	.52	.76
13. frustriert durch Arbeit	EE	1.46 (1.32)	.71	.88
14. zu verbissen arbeiten	EE	1.54 (1.51)	.56	.89
15. kein Interesse an manchen K.	DP	1.01 (1.42)	.40	.77
16. direkter Kontakt stresst zu sehr	EE	1.09 (1.21)	.62	.89
17. leichtes Schaffen entspannter Atmosphäre	PE	1.92 (1.53)	.58	.76
18. angeregt, wenn eng mit K. gearbeitet	PE	2.31 (1.67)	.37	.79
19. viele lohnende Ziele bei Arbeit erreicht	PE	2.39 (1.60)	.50	.77
20. am Ende der Weisheit zu sein	EE	1.00 (1.18)	.57	.89
21. gelassener Umgang mit emot. Problemen	PE	2.58 (1.69)	.45	.79
22. K./M. geben mir Schuld für Probleme	DP	0.94 (1.23)	.36	.81

Anmerkung. r_{IT} = ordinale, korrigierte Item-Skala-Korrelation; α = ordinale Cronbachs Alpha; K. = Kolleg_innen; V. = Vorgesetzte; M. = Mitarbeitende; EE = Emotionale Erschöpfung; PE = Persönliche Erfüllung; DP = Depersonalisation. Skala EE: $N = 391$, Skala DP, PE: $N = 390$. Skala PE wurde umgepolt. Fettgedruckte Trennschärfe-Werte unterschreiten Anforderungen von $\geq .40$. Fettgedruckte Alphawerte sind höher als das Gesamtalpha der Skala.

Tabelle C2*Itemanalyse der BOSS I*

Item	Skala	<i>M</i> (<i>SD</i>)	<i>r</i> _{IT}	α ohne Item
1. Sorge, Job nicht zu bewältigen	B	0.77 (1.03)	.66	.86
2. vermehrtes Kontrollieren, misstrauischer	B	0.97 (1.09)	.61	.86
3. Arbeiten bis spät in die Nacht	B	0.68 (1.09)	.44	.88
4. keine nachhaltige Freude über Erfolge	B	1.01 (1.25)	.68	.86
5. Fassung verlieren, wenn nicht nach Plan	B	1.01 (1.03)	.59	.87
6. Dauerstress durch lange Arbeitswege	B	0.70 (1.13)	.52	.87
7. Verlust der Freude an der Arbeit	B	1.08 (1.26)	.65	.86
8. wenig treffsicher in berufl. Entscheidungen	B	0.63 (0.86)	.68	.86
9. Alkohol/Medikamente zum Stressabbau	B	0.26 (0.72)	.37	.88
10. gefährdetes positives Image im Job	B	0.38 (0.75)	.65	.86
11. sich selbst fremd geworden sein	EP	0.35 (0.75)	.72	.92
12. keine Lebensfreude mehr empfinden können	EP	0.44 (0.81)	.70	.92
13. herabgesetzte Frustrationstoleranz	EP	0.92 (1.14)	.71	.92
14. zunehmende Sorgen um Gesundheit	EP	1.03 (1.15)	.66	.93
15. häufige banale Infekte	EP	0.45 (0.93)	.53	.92
16. kleine Anforderungen bereits belastend	EP	0.56 (0.91)	.78	.92
17. schlechter Schlaf	EP	1.14 (1.37)	.67	.93
18. Verlust der Routine, alles schwerer	EP	0.56 (1.02)	.85	.92
19. zunehmende Konzentrationsschwierigkeiten	EP	0.84 (1.12)	.74	.92
20. Schwierigkeiten beim Entspannen	EP	1.24 (1.34)	.71	.93
21. Unstimmigkeiten in Familie/Partnerschaft	FA	0.89 (0.99)	.70	.89
22. zu wenig Zeit für Gemeinsamkeiten	FA	1.57 (1.37)	.73	.89
23. drohende Entfremdung	FA	0.79 (1.07)	.85	.87
24. Sexualität/Erotik kommt zu kurz	FA	1.33 (1.39)	.76	.88
25. weniger Anteil an Familie/Partnerschaft	FA	0.87 (1.06)	.72	.89
26. besorgte Rückmeldungen von Freunden	FR	0.44 (0.90)	.69	.88
27. häufiger Rückzug aus sozialen Kontakten	FR	0.86 (1.19)	.78	.85
28. Belastungen sind einziges Gesprächsthema	FR	0.46 (0.86)	.72	.87
29. weniger Anteil an Freundschaften	FR	0.88 (1.13)	.77	.85
30. Schwierigkeiten, sich abzugrenzen	FR	1.30 (1.39)	.61	.89

Anmerkung. *r*_{IT} = ordinale, korrigierte Item-Skala-Korrelation; α = ordinale Cronbachs Alpha; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde. Skala: B, EP, FR: *N* = 392. Skala FA: *N* = 387. Fettgedruckte Trennschärfe-Werte unterschreiten Anforderungen von $\geq .40$.

Tabelle C3*Itemanalyse der BOSS II*

Item	Skala	<i>M (SD)</i>	<i>r</i> _{IT}	α ohne Item
1. heftige Kopfschmerzen	KOEB	0.53 (1.07)	.55	.91
2. häufig wiederkehrende Infekte	KOEB	0.39 (0.91)	.53	.91
3. ständige Muskelverspannungen	KOEB	1.25 (1.46)	.73	.90
4. beklemmende Enge über der Brust	KOEB	0.36 (0.82)	.73	.89
5. Atembeschwerden	KOEB	0.27 (0.72)	.77	.89
6. Schmerzen	KOEB	1.09 (1.46)	.68	.90
7. Herzbeschwerden	KOEB	0.29 (0.79)	.71	.89
8. Magen-Darmbeschwerden	KOEB	0.68 (1.14)	.61	.89
9. Schlafstörungen	KOEB	0.99 (1.34)	.60	.90
10. hohe oder schwankende Blutdruckwerte	KOEB	0.42 (0.98)	.41	.90
11. Konzentrationsschwierigkeiten	KOB	1.04 (1.08)	.67	.95
12. zu negative Bewertungen	KOB	1.25 (1.21)	.76	.94
13. Verlust der Entschlussfreudigkeit	KOB	0.90 (1.13)	.83	.95
14. vermehrte Fehlerneigung	KOB	0.73 (0.90)	.76	.94
15. nur noch Reagieren statt Agieren	KOB	0.80 (1.06)	.81	.94
16. Verlust der Kreativität	KOB	0.90 (1.15)	.78	.95
17. wiederkehrender, quälender Gedanke	KOB	0.92 (1.28)	.77	.95
18. verstärkte Selbstkontrolle	KOB	0.85 (1.16)	.64	.95
19. Verlust der Souveränität	KOB	0.60 (0.95)	.84	.94
20. verminderte Durchsetzungsfähigkeit	KOB	0.62 (0.86)	.83	.95
21. Zukunftsängste	EB	0.89 (1.18)	.71	.95
22. eingeschränkte Selbstwahrnehmung	EB	0.39 (0.80)	.83	.94
23. sozialer Rückzug	EB	0.62 (1.00)	.76	.94
24. gedrückte Stimmung	EB	0.54 (0.91)	.86	.94
25. Antriebslosigkeit	EB	0.67 (1.05)	.83	.94
26. Gereiztheit	EB	0.94 (1.12)	.78	.94
27. erhöhtes Misstrauen	EB	0.85 (1.11)	.70	.94
28. Scham	EB	0.34 (0.82)	.78	.95
29. erhöhte Ängstlichkeit	EB	0.49 (0.93)	.86	.94
30. Geduldverlust bei Misslingen	EB	1.08 (1.16)	.70	.94

Anmerkung. *r*_{IT} = ordinale, korrigierte Item-Skala-Korrelation; α = ordinale Cronbachs Alpha; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden. EB = Emotionale Beschwerden. *N* = 392.

Anhang D: Faktorenladungen mit postulierter Faktorenanzahl

Tabelle D1

Faktorladungen der Items der BOSS I mit 4 Faktoren

Nr.	Item	Skala	Faktor			
			1	2	3	4
1	Sorge, Job nicht zu bewältigen	B	.53	.20	.49	.08
2	vermehrtes Kontrollieren, misstrauischer	B	.39	.19	.51	.11
3	Arbeiten bis spät in die Nacht	B	.13	.22	.54	.23
4	keine nachhaltige Freude über Erfolge	B	.60	.20	.32	.15
5	Fassung verlieren, wenn nicht nach Plan	B	.48	.12	.40	.12
6	Dauerstress durch lange Arbeitswege	B	.37	.13	.40	.12
7	Verlust der Freude an der Arbeit	B	.80	.00	.11	.21
8	wenig treffsicher in berufl. Entscheidungen	B	.71	.10	.15	.26
9	Alkohol/Medikamente zum Stressabbau	B	.37	.34	-.03	.13
10	gefährdetes positives Image im Job	B	.64	.20	.26	.22
11	sich selbst fremd geworden sein	EP	.69	.26	.30	.20
12	keine Lebensfreude mehr empfinden können	EP	.70	.29	.18	.22
13	herabgesetzte Frustrationstoleranz	EP	.65	.22	.40	.15
14	zunehmende Sorgen um Gesundheit	EP	.38	.22	.42	.38
15	häufige banale Infekte	EP	.27	.11	.38	.34
16	kleine Anforderungen bereits belastend	EP	.63	.15	.44	.31
17	schlechter Schlaf	EP	.40	.18	.42	.32
18	Verlust der Routine, alles schwerer	EP	.68	.17	.45	.32
19	zunehmende Konzentrationsschwierigkeiten	EP	.53	.20	.36	.36
20	Schwierigkeiten beim Entspannen	EP	.42	.28	.45	.40
21	Unstimmigkeiten in Familie/Partnerschaft	FA	.24	.69	.25	.10
22	zu wenig Zeit für Gemeinsamkeiten	FA	.04	.70	.40	.19
23	drohende Entfremdung	FA	.18	.85	.19	.16
24	Sexualität/Erotik kommt zu kurz	FA	.11	.81	.07	.19
25	weniger Anteil an Familie/Partnerschaft	FA	.29	.66	.18	.37
26	besorgte Rückmeldungen von Freunden	FR	.18	.40	.30	.62
27	häufiger Rückzug aus sozialen Kontakten	FR	.40	.32	.22	.61
28	Belastungen sind einziges Gesprächsthema	FR	.46	.28	.25	.56
29	weniger Anteil an Freundschaften	FR	.46	.37	.20	.58
30	Schwierigkeiten, sich abzugrenzen	FR	.25	.25	.49	.42

Anmerkung. Nr. = Itemnummerierung der BOSS I; B = Beruf; EP = Eigene Person; FA = Familie; FR = Freunde. Hauptachsenanalyse mit Varimax-Rotation, basierend auf polychorischen Korrelationen. Die Faktorenanzahl wurde auf die Anzahl postulierter Skalen (4) festgelegt. $N = 387$. Bedeutsame Ladungen ($\geq |.40|$) sind fettgedruckt.

Tabelle D2*Faktorladungen der Items der BOSS II mit 3 Faktoren*

Nr.	Item	Skala	Faktor		
			1	2	3
1	heftige Kopfschmerzen	KOEB	.22	.28	.59
2	häufig wiederkehrende Infekte	KOEB	.25	.24	.68
3	ständige Muskelverspannungen	KOEB	.29	.48	.60
4	beklemmende Enge über der Brust	KOEB	.34	.74	.20
5	Atembeschwerden	KOEB	.23	.83	.23
6	Schmerzen	KOEB	.21	.57	.39
7	Herzbeschwerden	KOEB	.31	.81	.11
8	Magen-Darmbeschwerden	KOEB	.21	.55	.32
9	Schlafstörungen	KOEB	.38	.59	.14
10	hohe oder schwankende Blutdruckwerte	KOEB	.19	.56	.14
11	Konzentrationschwierigkeiten	KOB	.52	.47	.20
12	zu negative Bewertungen	KOB	.69	.28	.31
13	Verlust der Entschlussfreudigkeit	KOB	.76	.30	.17
14	vermehrte Fehlerneigung	KOB	.70	.34	.20
15	nur noch Reagieren statt Agieren	KOB	.77	.30	.20
16	Verlust der Kreativität	KOB	.76	.34	.10
17	wiederkehrender, quälender Gedanke	KOB	.69	.32	.24
18	verstärkte Selbstkontrolle	KOB	.57	.19	.32
19	Verlust der Souveränität	KOB	.81	.38	.00
20	verminderte Durchsetzungsfähigkeit	KOB	.81	.27	.15
21	Zukunftsängste	EB	.71	.22	.18
22	eingeschränkte Selbstwahrnehmung	EB	.79	.31	.14
23	sozialer Rückzug	EB	.72	.23	.25
24	gedrückte Stimmung	EB	.76	.38	.24
25	Antriebslosigkeit	EB	.84	.25	.12
26	Gereiztheit	EB	.65	.35	.36
27	erhöhtes Misstrauen	EB	.63	.29	.34
28	Scham	EB	.79	.06	.24
29	erhöhte Ängstlichkeit	EB	.82	.23	.21
30	Geduldverlust bei Misslingen	EB	.59	.23	.36

Anmerkung. Nr. = Itemnummerierung der BOSS II; KOEB = Körperliche Beschwerden; KOB = Kognitive Beschwerden; EB = Emotionale Beschwerden. Hauptachsenanalyse mit Varimax-Rotation, basierend auf polychorischen Korrelationen. Die Faktorenanzahl wurde auf die Anzahl postulierter Skalen (3) festgelegt. $N = 392$. Bedeutsame Ladungen ($\geq |.40|$) sind fettgedruckt.