

Übersetzung, Validierung und kritische Diskussion zweier Ad-hoc-konstruierter  
Skalen zum emotionalen Erleben

Dissertation

Zur Erlangung des akademischen Grades

Doktor der Medizinischen Wissenschaften (Dr. rer. medic.)  
für das Fachgebiet Psychologie

vorgelegt

der Medizinischen Fakultät

der Martin-Luther-Universität Halle-Wittenberg

von Christian Pradel  
geboren am 22.12.1994 in Halle an der Saale

Betreuer: apl. Prof. Dr. rer. nat. Stefan Watzke

Gutachter\*innen: PD Dr. Frank Pillmann, Halle (Saale)  
Prof. Dr. Peter Brieger, Haar

Datum der Verteidigung: 04.03.2024

## Referat

Testverfahren sind in der psychologischen Diagnostik und Forschung die Grundlage für Begutachtungen und empirischen Erhebungen. Sie müssen qualitative psychometrische Standards gewährleisten und es gleichzeitig erlauben, möglichst effizient zu messen. Um eine hohe Qualität der verwendeten Verfahren zu gewährleisten, müssen sie regelmäßig überprüft und mit wissenschaftlichen Standards abgeglichen werden. Die aktuelle Informationslage bezüglich der Verfahrensqualität ist allerdings unzureichend. Das gilt insbesondere für weniger bekannte oder neu entwickelte Testinstrumente.

Zur ökonomischen und validen Erfassung von Belastungsphänomenen entwickelten die australischen Autoren Rogers, Barblett, Cruickshank und Robinson 2016 die Brief Emotional Experience Scale (BEES) sowie die Brief Emotional Physical Experience Scale (BEEPS). Die beiden englischsprachigen Instrumente wurden in das Deutsche übersetzt und anhand ihrer psychometrischen Qualitätsmerkmale bewertet.

Die Verfahren wurden einer medizinstudentischen Stichprobe von  $n = 208$  Personen als Online-Fragebogen appliziert. Anhand von Cronbachs Alpha und McDonalds Omega wurde die interne Konsistenz als Schätzer der Reliabilität bestimmt. Durch die Betrachtung konvergenter und diskriminanter Validitäten zu bestehenden Belastungsmaßen sowie durch inhaltliche Überlegungen wurde die Validität der neu übersetzten Verfahren eingeordnet. Durch explorative und konfirmatorische Faktorenanalysen konnte die 2-Faktoren-Struktur der BEES belegt werden, die Faktorenstruktur der BEEPS ist hingegen ungeklärt. Es folgte der Vergleich von BEES und BEEPS mit bestehenden Testverfahren für eine Einordnung hinsichtlich ihrer Verwendbarkeit. Die vorliegende Arbeit kommt zu dem Schluss, dass weder BEES noch BEEPS gängigen psychometrischen Standards gerecht werden. Die Verfahren sind unzureichend theoretisch konzipiert, eingeordnet und zu bestehenden Konstrukten abgegrenzt. Es existiert keinerlei Interpretationsregel und daher können die Testwerte nicht interpretiert werden. Die Messintention ist nicht eindeutig beschrieben und die vorliegenden Validitätsbelege sind unzureichend. Die Verwendung in klinischen oder Forschungskontexten kann nicht empfohlen werden. Bereits existente und etablierte Maße zur Belastungsmessung sind vorzuziehen.

Pradel, Christian: Übersetzung, Validierung und kritische Diskussion zweier Ad-hoc-konstruierter Skalen zum emotionalen Erleben, Halle (Saale), Martin-Luther-Universität, Medizinische Fakultät, Dissertation mit 53 Seiten, 2023

## Abstract

Tests are widely used in diagnostics and research. They form the basis of assessments and empirical research. Tests must meet the requirement of efficient measurement while ensuring qualitative psychometric standards. The quality of the procedures used needs to be monitored. The current state of information in this regard is inadequate, especially in the case of lesser-known or newly developed instruments. The Australian authors Rogers, Barblett, Cruickshank and Robinson developed the Brief Emotional Experience Scale (BEES) and the Brief Emotional Physical Experience Scale (BEEPS) for the economic and valid assessment of stress phenomena. The two English-language instruments were translated into German and their psychometric properties were evaluated. The instruments were administered as an online questionnaire to a sample of medical students ( $n = 208$ ). Cronbach's alpha and McDonald's omega were used to determine internal consistency as an estimate of reliability. The validity of the newly translated procedures was classified by considering convergent and discriminant validities with existing stress measures, as well as by content considerations. Exploratory and confirmatory factor analyses were used to determine the factor structure of the BEES and BEEPS. The 2-factor structure of the BEES could be demonstrated, the factor structure of the BEEPS is still unclear. The translated instruments were then compared with existing tests and their usability was assessed.

The present study concludes that neither the BEES nor the BEEPS meet current psychometric standards. The tests are insufficiently theoretically designed, classified and delimited from existing constructs. The test results cannot be interpreted. The measurement intentions are not clearly described and the available validity evidence is inadequate. Use in clinical or research contexts cannot be recommended. Existing and established measures of stress are preferable.

Pradel, Christian: Übersetzung, Validierung und kritische Diskussion zweier Ad-hoc-konstruierter Skalen zum emotionalen Erleben, Halle (Saale), Martin-Luther-Universität, Medizinische Fakultät, Dissertation mit 53 Seiten, 2023

1. <u>Einleitung</u> .....	1
1.1. Relevanz des Themas .....	1
1.2. Depression .....	3
1.2.1. Bestehende Depressionsfragebögen .....	3
1.2.2. Beck Depression Inventar – II .....	3
1.2.3. Depressions-Angst-Stress-Skalen .....	4
1.2.4. Bestehende Screeningverfahren .....	4
1.3. Emotion .....	6
1.3.1. Definition Emotion .....	6
1.3.2. Emotion und Stimmung .....	6
1.3.3. Klassifizierung von Emotionen .....	7
1.3.4. Komponenten von Emotionen .....	7
1.3.5. Bedeutung der Emotionsmessung .....	8
1.3.6. Probleme der Messung .....	8
1.3.7. Bestehende Fragebögen .....	9
1.3.8. Zweifaktorielle Emotionscreenings .....	9
1.3.8.1. Positive and Negative Affect Scale .....	10
1.3.8.2. Scale of Positive and Negative Experience .....	11
1.4. Brief Emotional Experience Scale .....	12
1.4.1. Aufbau BEES .....	12
1.4.2. Faktorenstruktur BEES .....	14
1.5. Brief Emotional Experience Physical Scale .....	15
1.6. Methoden der Reliabilitätsschätzung .....	16
1.6.1. Cronbachs Alpha .....	16
1.6.2. McDonalds Omega .....	17
1.7. Standardisierte Testbeurteilungen .....	17
2. <u>Zielstellung</u> .....	18

3. <u>Methode</u> .....	19
3.1. Stichprobe .....	19
3.1.1. Rekrutierung .....	19
3.1.2. deskriptive Stichprobenbeschreibung .....	19
3.2. Untersuchungsmethode .....	20
3.2.1. Fragebogenaufbau .....	20
3.2.2. Übersetzungsmethodik BEES und BEEPS .....	21
3.3. Untersuchungsplan .....	21
3.4. Untersuchungsablauf .....	21
3.5. statistische Hypothesen .....	22
3.6. Auswertungsplan .....	23
3.7. Operationalisierung der psychometrischen Qualitätsmerkmale .....	23
3.7.1. Operationalisierung der Objektivität .....	23
3.7.2. Operationalisierung der Reliabilität .....	23
3.7.3. Operationalisierung der Validität .....	23
3.7.4. Operationalisierung der Faktorenstrukturprüfung .....	25
4. <u>Ergebnisse</u> .....	26
4.1. Deskriptive Darstellung .....	26
4.2. Objektivitätsergebnisse .....	27
4.3. Reliabilitätsergebnisse .....	27
4.4. Validitätsergebnisse .....	28
4.5. Faktorenstrukturprüfung .....	30

5. <u>Diskussion</u> .....	34
5.1. Deskriptive Daten .....	34
5.2. Objektivität .....	34
5.3. Reliabilität .....	35
5.4. Validität .....	36
5.5. Faktorenanalyse .....	39
5.6. Bewertung nach DIN .....	40
5.7. Verwendbarkeit .....	40
5.8. Limitationen .....	42
6. <u>Fazit</u> .....	42
7. <u>Literaturverzeichnis</u> .....	43
8. <u>Thesen</u> .....	53

Anhang

Eidesstattliche Versicherung

## Abkürzungsverzeichnis

BDI-II – Beck Depression Inventory II

BEES – Brief Emotional Experience Scale

BEEPS – Brief Emotional Physical Experience Scale

DASS-21 – Depression-Angst-Stress-Skalen-21

DSM-V – Diagnostisch-statisches-Manual-V

NA – Negativer Affekt

NEO-FFI – Neo-Fünf-Faktoren-Inventar

PA – Positiver Affekt

PANAS – Positive and Negative Affect Scale

SKID II – Strukturiertes klinisches Interview II

SPANE – Scale of Positive and Negative Experience

## 1. Einleitung

### 1.1. Relevanz des Themas

Befindensfragebögen sind neben Leistungstests und qualifizierten Interviews die meistverwendete Methode der psychologischen Diagnostik (Schmidt-Atzert & Amelang, 2006; Schmidt-Atzert, Krumm & Amelang, 2022). Alle drei Formen der Informationsgewinnung sind wissenschaftliche Methoden, die messtheoretisch fundiert sein müssen (Diagnostik- und Testkuratorium, 2018). Objektive, reliable und valide Messinstrumente sind die notwendige Grundlage für Eigenschaftszuweisungen und daher im klinischen als auch im forschungsorientierten Kontext von entsprechend hoher Bedeutung. Seit 1995 steigt die Nachfrage nach solchen Tests stetig (Wottawa, 2002). Die fachkorrekte Auswahl, methodenkompetente Durchführung und Interpretation eines Testverfahren entscheidet zwischen einer fachlich-fundierten, qualitätsgesicherten Diagnostik und zufallsbasierten Fehlanwendungen (Kersting & Beauducel, 2022; Kersting & Petri, 2022, Sparfeldt et al., 2022). Die Qualität von Testverfahren lässt sich anhand verschiedener Gütekriterien bewerten. Allerdings kann das aktuelle Informationsangebot diesbezüglich weder quantitativ noch qualitativ überzeugen (Kersting, 2006): Testverfahren werden zu selten in ihrer Qualität überprüft, da diese Rezensionen einerseits sehr aufwendig und andererseits als „nachrangig in ihrem Stellenwert als wissenschaftliche Leistung“ (Kersting) betrachtet werden.

Für die Erfassung von emotionalen Wohlbefinden, Belastungserscheinungen bis hin zu klinisch relevanten Depressionsphänomenen existiert eine Vielzahl an Tests und Screenings. Diese haben unterschiedliche Vor- und Nachteile und sind oft für einen spezifischen Einsatzzweck optimiert. Manche dieser Verfahren decken sehr große Bereiche ab: Strukturierte Interviews wie beispielsweise der SKID II (Wittchen, Zaudig, Fydrich, 1997) erfassen jede Kategorie potenzieller psychischer Erkrankung und sind in Durchführung, Auswertung und Interpretation entsprechend komplex und aufwendig. Andere Verfahren sind bedeutend ökonomischer, erfassen aber nur bestimmte Teile oder können in der Auswertung kein detailliert-differenziertes Bild liefern.

Testinstrumente mit einem eher ökonomischen Ansatz spielen ihre Stärke bei Langzeituntersuchungen oder Screenings großer Populationen aus, wie sie beispielsweise bei Depressionen durchgeführt werden: Da die gesamtgesellschaftlichen sowie individuellen Belastungen durch Depressionen kontinuierlich zunehmen (DAK Psychreport, 2021; Grobe & Bressel, 2021; Jacobi, Klose & Wittchen, 2004, Murray & Lopez, 1996; Üstün, Ayuso-Mateos,



Chatterji, Mathers & Murray, 2000) und frühzeitiges Detektieren depressiver Symptome die Wahrscheinlichkeit einer Diagnose und erfolgreichen Behandlung (Center et al., 2003) erhöht, ist das Screening stark betroffener Populationen sinnvoll. Eine der übermäßig stark betroffenen Gruppen sind Medizinstudierenden: Internationale Studien belegen erhöhte Prävalenzraten (Jurkat et. Al., 2011; Moir, Yelder, Sanson & Chen, 2018, Rotenstein et. al., 2016). Puthran, Zhang, Tam und Ho (2016) postulieren in ihrer Metaanalyse, dass 28% aller Medizinstudierenden weltweit von Depressionen betroffen sind. Bereits seit 2017 untersucht daher die medizinische Fakultät der Martin-Luther-Universität die Depressions- und Belastungserscheinungen bei Studierenden der Medizin (Pukas et al., 2022), Zahnmedizin (Rehnisch et al, 2021), Psychologie (Kindt et al., 2022) und Rechtswissenschaften (Rabkow et al., 2020).

---

Die australische Forschergruppe um Rogers, Barblett, Cruickshank und Robinson entwickelte seit 2016 zwei Verfahren zum Screening von Populationen, die besonders von Belastungen betroffen sind: die *Brief Emotional Experience Scale* (BEES) und die *Brief Emotional Experience Physical Scale* (BEEPS). Diese Instrumente sollten auch bei dem oben beschriebenen Forschungsvorhaben der Martin-Luther-Universität eingesetzt und dafür übersetzt und evaluiert werden.

Um die theoretische Konzeption, die psychometrische Qualität und letztendlich die Verwendbarkeit dieser neuen Tests zu beurteilen, sollen im Folgenden die dafür notwendigen Konzepte erläutert werden.

Es werden dafür zunächst Depressionen und Emotionen als integraler Symptombestandteil sowie ihre Messung besprochen. Darauf folgen ein Überblick zur Testtheorie und die Beschreibung der neuen Testverfahren BEES und BEEPS. Abschließend sollen im theoretischen Teil die Gütekriterien für psychometrische Forschung sowie deren Überprüfung in empirischen Erhebungen gezeigt werden. Anhand dieser Kriterien werden BEES und BEEPS schließlich bewertet und eingeordnet.

## 1.2. Depression

Depressionen können jeden Menschen betreffen und sie haben global massive Auswirkungen auf die Gesundheitssysteme (DAK Psychreport, 2021; Grobe & Bressel, 2021; Jacobi, Klose & Wittchen, 2004, Murray & Lopez, 1996; Üstün, Ayuso-Mateos, Chatterji, Mathers & Murray, 2000). Depressive Störungen bilden ein eigenes Kapitel im Diagnostischen und Statistischen Manual psychischer Störungen (DSM-5) (American Psychiatric Association, 2013) und sind gekennzeichnet durch das gemeinsame Merkmal einer traurigen oder reizbaren Stimmung oder des Gefühls der Leere. Dieses gemeinsame Merkmal macht deutlich, dass Emotionen den essenziellen Symptomteil bestimmen. Daraus folgt die hohe Bedeutung einer möglichst exakten Emotionsmessung für die Beschreibung und Diagnostik von depressiven Störungen. Auf die weiteren Komponenten von depressiven Störungen soll an dieser Stelle nicht weiter eingegangen werden.

### 1.2.1. Bestehende Depressionsfragebögen

Es existiert eine Vielzahl an Fragebögen zur Erfassung von depressiven Symptomatiken. Beispielhaft werden das *Beck Depressions-Inventar-II* (BDI-II) von Beck, Steer und Brown (1996) und die *Depression Anxiety Stress Scales* (DASS) von Lovibond und Lovibond (1993) vorgestellt.

### 1.2.2. Beck Depressions-Inventar-II

Das Beck Depressions-Inventar-II (BDI-II) ist das am häufigsten eingesetzte Selbstbeurteilungsinstrument zur Beurteilung der Depressionsschwere (Hautzinger, Keller & Kühner, 2006). Seine psychometrischen Qualitätsmerkmale gelten als vollständig gegeben, seine Reliabilität und Validität wurde mannigfaltig belegt (Glischinski, von Brachel & Hirschfeld, 2019; Herzberg, Goldschmidt & Heinrichs, 2008; Huang & Chen, 2015; Kühner et al., 2022; Wang & Gorenstein, 2013). Der BDI-II besteht aus 21 Items, die auf einer 4-stufigen Antwortskala bewertet werden. Aus den Scores der einzelnen Items lässt sich ein Gesamtscore bilden, der Aussagen über den Schweregrad der depressiven Symptomatik ermöglicht. Die Faktorenstruktur des BDI-II ist zweifaktoriell. Die Faktoren „kognitiv“ und „somatisch-affektiv“

sind hochkorreliert und decken die unterschiedlichen Symptommuster der Depression ab (Keller, Hautzinger & Kühner, 2008). Die deutschsprachige Adaption stammt von Hautzinger, Keller und Kühner (2006). Herzberg, Goldschmidt und Heinrichs (2008) empfehlen die Verwendung des BDI-II als klinisches Testverfahren.

### 1.2.3. Depression Anxiety Stress Scales

Ziel der 1993 von Lovibond und Lovibond konstruierten Depression Anxiety Stress Scale (DASS) war die Integration von Depression, Angst sowie dem relativ neu hinzugezogenen Konstrukt Stress in einen einzigen Fragebogen bei gleichzeitiger Orthogonalität der einzelnen Faktoren zueinander (Lovibond & Lovibond, 1995). In ihrer ursprünglichen Form umfasst die DASS 42 satzlange Items. Diese erfassen negative emotionale Symptome und werden auf einer 4-stufigen-Häufigkeitsskala bewertet. Der DASS ist dreifaktoriell aufgebaut mit den Faktoren Depression, Angst und Stress. Vollständige Orthogonalität der Faktoren konnte nicht belegt werden. Des Weiteren nehmen die Autoren einen Faktor 2. Ordnung an, der die generelle psychologische Belastung darstellt (Henry & Crawford, 2005). Lovibond (1998) betont hierbei, dass die Konstrukte der Stressskala und das Konstrukt des Negativen Affekts (NA) sich unterscheiden. Die DASS ist ein weltweit eingesetztes und etabliertes Verfahren zur Erfassung generalisierter Belastung (Scholten et al., 2017; Zanon et al., 2020).

Die DASS-21 ist die auf 21 Items gekürzte Fassung der DASS. Die DASS-21 ist ein vielfältig eingesetztes, reliables und validiertes Verfahren (Henry & Crawford, 2005; Oei, Sawang, Goh & Mukhtar, 2013; Osman et al., 2012) zur Erfassung von Depression, Angst und Stress und findet weltweit Verwendung.

### 1.2.4. Bestehende Screeningverfahren

Abhängig von Untersuchungsgegenstand und Anwendungsfall müssen Tests unterschiedliche Anforderungen erfüllen: Screeningverfahren beispielsweise müssen besonders effizient sein, also mit minimaler Anzahl an Items die notwendigen Informationen erfassen. Wie wirtschaftlich ein psychometrischer Test ist, hängt von den erforderlichen Investitionen des Testleiters (Testvorbereitung, -durchführung und -auswertung) und der zeitlichen

Beanspruchung der Testpersonen ab. Eine geringere Belastung der Versuchspersonen durch eine Komprimierung und Reduktion der Items kann im Sinne einer konstanteren Testmotivation zur Minimierung der Drop-out-Rate innerhalb einer Stichprobe beitragen (Denker, Schütte, Kersting, Weppert & Stegt; 2023; Hoerger, 2010; Speer, King & Grossenbacher, 2016). Gleichzeitig erleichtert die Verwendung eines Testinstruments mit maximal ökonomisiertem Informationsgewinn und niedrigschwelligem Zeitaufwand Verlaufsbeobachtungen mit mehrmaligen Messwiederholungen und deren Dokumentation. Schließlich reduzieren sie auch die sprachlichen Anforderungen an die Getesteten aufgrund des geringen notwendigen Vokabulars zur Testdurchführung.

Ein Beispiel: Im Bereich der Verlaufsmessung von Symptomen depressiver Störungen wurde gezeigt, dass eine engmaschige – zum Beispiel tägliche – Symptommessung positive Auswirkungen auf die Behandlung von Patienten haben kann (Clark et al., 2018; Titov et al., 2016). Screeningverfahren stellen eine Ergänzung zu Forschungsmethoden dar: Online- oder appgestützte Erhebungen ermöglichen täglich mehrfache Befragungen der Probanden bei gleichzeitig geringer zeitlicher Belastung. Sie erlauben eine hohe Anzahl an Messzeitpunkten zur genauen Verlaufsanalyse, etwa zur Bewertung des Schweregrads eines Symptoms.

In der Forschung haben sich einige solcher Instrumente breitflächig etabliert, wie z.B. der *nine-item Patient Health Questionnaire* (Spitzer, R., Kroenke, K. & Williams, J., 1999) oder die *6-item Kessler Psychological Distress Scale* (Kessler et al., 2002) und wurden in verschiedene Sprachen übersetzt (Costantini et al., 2020; Staples et al., 2019). Sie sind änderungssensitiv, weisen optimale Sensitivitäts- und Spezifitätswerte auf und erreichen somit einen hohen Nutzen. Allerdings messen sie ausschließlich Depressionssymptome und lassen somit keine Aussage über positiven Affekt zu. Um die Messung positiven Affekts verständlich und greifbar zu machen, wird im Folgenden das Konzept der Emotionen, deren Messung und die damit verbundene Probleme eingeführt.

### 1.3. Emotion

#### 1.3.1. Definition Emotion

Kleinginna und Kleinginna stellten bereits 1981 eine große Bandbreite an Definitionen des Emotionsbegriffs fest. Aus 92 bestehenden Definitionen extrahierten die Autoren die folgende:

*„Emotion is a complex set of interactions among subjective and objective factors, mediated by neural hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal-directed, and adaptive.“*

*„Emotion ist eine komplexe Reihe von Interaktionen zwischen subjektiven und objektiven Faktoren, vermittelt durch neurale Hormonsysteme, die (a) affektive Erfahrungen wie Gefühle der Erregung, Lust/Unlust hervorrufen können; (b) kognitive Prozesse wie emotional relevante Wahrnehmungseffekte, Bewertungen, Einschätzungsprozesse erzeugen; (c) weit verbreitete physiologische Anpassungen an die Erregungsbedingungen aktivieren; und (d) zu einem Verhalten führen, das oft, aber nicht immer, expressiv, zielgerichtet und adaptiv ist.“*

Demnach haben Emotionen sowohl subjektiv erfahrbare als auch objektiv erfassbare Komponenten, die zielgerichtetes Verhalten begleiten bzw. fördern, das dem Organismus eine Anpassung an seine Lebensbedingungen ermöglichen (Brandstätter, Schüler, Puca & Lozo, 2013).

#### 1.3.2. Emotionen und Stimmung

Morris und Schnur unterscheiden 1989 in ihrer Monografie Emotionen und Stimmungen. Emotionen sind den Autoren zufolge Reaktion auf ein konkretes Objekt oder Ereignis, daher zeitlich begrenzt, jedoch von höherer Intensität als Stimmungen. Stimmungen hingegen sind ein allgemeiner Hintergrund, vor dem sich konkrete Denkprozesse und somit auch Emotionen

abspielen. Sie sind zeitlich ausgedehnter, jedoch von geringerer Intensität (Brandstätter et al., 2013). Stimmungen sind zeitlich überdauernd und beeinflussen global kognitive Vorgänge (Frijda, N. 1993; Morris & Schnurr, 1989). Die Unterscheidung zwischen Emotion und Stimmung ist wichtig für die Messung: Messungen, die sich auf einen langen Zeitraum beziehen, erfassen die grundlegenden Stimmungen. Beziehen sie sich auf einen kürzeren Zeitraum oder wird die Messung gar mehrfach an einem Tag wiederholt, werden Emotionen erfasst.

### 1.3.3. Klassifizierung

Die moderne Emotionsforschung geht von einer Menge an Basisemotionen aus, die sich zu komplexen Emotionen zusammensetzt. Basisemotionen werden definiert als Emotionen mit universellen, das heißt kulturübergreifendem, mimischem Ausdruck. Sie werden kulturunabhängig gezeigt und verstanden (Brandstätter et al., 2013).

Welche Emotionen als Basisemotionen definiert werden ist bis heute Gegenstand der Emotionsforschung. Beispielsweise definiert Ekman (1982) Freude, Trauer, Überraschung, Ekel, Furcht und Wut als Basisemotionen, Schmidt-Atzert (1996) ergänzen diese um „Scham“ als Ergebnis ihrer Metaanalyse. Diese Unterschiede sollen verdeutlichen, wie wichtig die autorenabhängige Auswahl eines Testverfahrens ist: Je nach Autorenavorstellung werden die Emotionen unterschiedlich kategorisiert und dementsprechend unterschiedlich gemessen. Daher benötigen die Testverfahren eine eindeutige Zuordnung zu einer autorenspezifischen Emotionstheorie und deren konkreten Komponenten.

### 1.3.4. Komponenten von Emotionen

Aus den vorangegangenen Definitionen und Klassifizierungen von Emotionen lassen sich die einzelnen Bestandteile von Emotionen ableiten:

#### a) Subjektive Komponente

Das Erleben der konkreten Emotion ist per definitionem subjektiv.

#### b) Physiologische Komponente

Emotionen gehen einher mit physiologischen Veränderungen im Organismus, zum Beispiel der Veränderung der Herzfrequenz.

### c) Verhaltenskomponente

Emotionen sind mit konkretem Verhalten assoziiert, beginnend mit der Bewegung einzelner Gesichtsmuskeln bis zu komplexen Verhaltensreaktionen. Die Kenntnis der einzelnen Komponenten ist wichtig für deren Messung und zieht verschiedene Probleme mit sich, die im Folgenden dargestellt werden.

#### 1.3.5. Bedeutung der Emotionsmessung

Emotionen betreffen jeden Bereich menschlichen Erlebens und haben großen Einfluss auf verschiedene Komponenten der menschlichen Kognition: Emotionen lenken Aufmerksamkeit (Wentura, Rothermundt & Bak, 2000), beeinflussen das Gedächtnis (McGaugh, 1992), Urteile und Entscheidungen (Clark & Williamson, 1989) sowie die Fähigkeit zum Problemlösen (Isen, 2000).

Ebenso betreffen Emotionen die psychologische Pathologie: Während sie bei affektiven Störungen wie Depressionen oder bipolaren Störungen und Angststörungen den Kern der Pathologie ausmachen, können sie beispielweise bei schizophrenen oder Persönlichkeitsstörungen betroffen sein.

Aufgrund der vielfältigen Auswirkungen von Emotionen ist ihre genaue Messung essenziell für die psychologische Diagnostik.

#### 1.3.6. Probleme der Messung

Während physiologische und behaviorale Komponenten leicht durch physikalische Messungen oder Verhaltensbeobachtungen erfasst werden können, kann die subjektive Komponente Emotion nur durch direkte Befragung der Person erfasst werden. Gleichzeitig sind die Möglichkeiten zur Emotionsinduktion im Laborsetting aufgrund ethischer Bedenken vor allem im Bereich negativer Emotionen stark limitiert (Brandstätter et al., 2013).

Damit entsteht eine Vielzahl an Schwierigkeiten für die Emotionsmessung:

- Emotionen können nur von Personen erfasst werden, die sich verbal oder schriftlich äußern können.
- Verbale Äußerungen über Emotionen unterliegen immer einem sozialen Kontext und sind daher verfälschungsanfällig.

- Trotz einer kulturübergreifenden Definition von Basisemotionen sind die verbalen Beschreibungen immer kulturabhängig.
- Da subjektive Berichte über in der Vergangenheit erlebte Emotionen Gedächtnisverzerrungen und Rekonstruktionsfehlern unterliegen, sind diese weniger valide als zeitlich direkt an das erlebte anknüpfende Befragungen.
- Da Emotionen als unabhängige Variablen in Forschungssettings nicht systematisch variiert werden können (Induktion negativer Emotion ist verbunden mit ethischen Problemen), sind die gefundenen Zusammenhänge korrelativer Art und schwer in eindeutige Kausalzusammenhänge überführbar.

### 1.3.7. Bestehende Fragebogenverfahren

Aufgrund der Bedeutung der exakten Emotionsmessung existiert bereits eine Fülle an etablierten und viel verwendeten Testverfahren. Diese unterscheiden sich stark in ihrer theoretischen Konzeption. Wie beschrieben, fehlt es an einer einheitlichen Definition von „Emotion“. Entsprechend unterscheiden sich auch die Tests stark, etwa bei der angenommenen Faktorenstruktur des Konstrukts „Emotion“ oder der Konstruktions- und Auswahlverfahren der Items. Beispiele sind die Eigenschaftswörterliste von Janke und Debus (1978) oder die *Differential Emotions Scale* von Izard, Dougherty, Bloxom und Kotsch (1974) sowie deren deutsche Adaptierung (Brandenburg & Backhaus, 2015).

### 1.3.8. Zweifaktorielle Emotionsscreenings

Exemplarisch näher erläutert werden sollen die *Positive and Negative Affect Scale* (PANAS) von Watson, Clark & Tellegen (1988) sowie die *Scale of Positive and Negative Experience* (SPANE) von Diener et al. (2009). Gemeinsam ist ihnen die zweifaktorielle Gliederung: Sowohl PANAS als auch SPANE bestehen aus dem Faktor „positiver Affekt“ sowie dem Faktor „negativer Affekt“. Positiver Affekt (PA) und Negativer Affekt (NA) sind konzeptioniert als dispositionelle Eigenschaften. Hohe NA-Werte gehen einher mit hohem subjektiv erlebtem Stress und prinzipiellen Unwohlsein (Henry & Crawford, 2005). Sie stehen außerdem in enger Verbindung mit Symptomen und Diagnosen von Störungen sowohl aus dem depressiven als



auch dem ängstlichen Spektrum (Brown, Chorpita, Korotitsch, & Barlow, 1997; Clark & Watson, 1991; Watson & Clark, 1984).

#### 1.3.8.1. Positive and Negative Affect Scale

Angelehnt an theoretische Vorarbeiten von Watson und Tellegen (1985) wird Emotion auf der *Positive and Negative Affect Scale* (PANAS) als Zwei-Faktoren-Modell konzeptioniert mit den Faktoren „Positiver Affekt“ und „Negativer Affekt“. Watson, Clark und Tellegen (1988) benutzten zunächst faktorenanalytische Verfahren zur Identifikation ihrer Items. Diese faktorenanalytischen Verfahren beschrieben sie als sinnvollen Weg gegenüber der „ad hoc Konstruktion von Tests ohne unterstützende Reliabilitäts- und Validitätsdaten“. Aus insgesamt 87 Items extrahierten die Autoren die 10 Items pro Faktor, die die stärksten Faktorenladungen aufwiesen. So gelangten sie zu einem informationsreichen und trotzdem ebenfalls ökonomischen Verfahren. Der Nachweis der erhöhten Ökonomie bei gleichem Informationsgehalt führten gelang in mehreren Studien (Clark & Watson, 1986; Watson, 1988).

Der Test beginnt mit der Aufforderung an die Versuchspersonen, die weiter unten aufgeführten Items für einen spezifischen Zeitraum einzuschätzen. Dieser Zeitraum wird vom Versuchsleiter gewählt und kann von einer situativ-aktuellen bis hin zu einer global-generalen Zeitspanne reichen. Die insgesamt 20 Items werden auf einer 5-stufigen Likert-skala mit den möglichen Antworten „1 *very slightly or not at all*“, „2 *a little*“, „3 *moderately*“, „4 *quite a bit*“ und „5 *extremely*“ bewertet.

Die Reliabilität der PANAS wird anhand von Cronbachs Alpha und einer Retestreliabilität mit einem Zeitabstand von 8 Wochen betrachtet und weist laut den Autoren in beiden Bereichen sehr gute Werte auf. Um eine breite Anwendung der PANAS zu ermöglichen, erfolgten verschiedene Erhebungen an je einer studentischen, einer werktätigen und einer klinischen Stichprobe. Diese legten nahe, dass nur Unterschiede im positiven Affekt einer klinischen Stichprobe gegenüber anderen Stichproben zu erwarten seien.

Im Bereich der Validitätsuntersuchung wurde anhand der empirischen Daten die theoretisch vorhergesagte Zweifaktorenstruktur nachgewiesen. Die externe Validität wurde anhand korrelativer Betrachtungen zu verschiedenen Maßen belegt. Zusammenfassend präsentierten Watson, Clark und Tellegen 1988 ein reliables, präzises und normiertes Screening zum

Erfassen von negativem und positivem Affekt mit insgesamt 20 Items bestehend aus je einem Wort. Die PANAS ist somit ein valides und reliables Messinstrument, dessen hohe qualitative Testgüte in vielen Untersuchungen belegt wurde, das als ökonomisch gilt und daher weltweit vielfältig im Einsatz ist (Crawford & Henry, 2004; Heubeck & Boulter, 2021; Heubeck & Wilkinson, 2019; Kanter, Garrido, Moretti & Medrano, 2021).

#### 1.3.8.2 Scale of Positive and Negative Experience

Diener et al. stellten 2009 mit der *Scale of Positive and Negative Experience* (SPANE) ein 12 Ein-Wort-Items umfassendes Screeningverfahren zur Erfassung positiven und negativen Affekts vor. Die SPANE ist wie die PANAS ein etabliertes und validiertes Maß, das bereits in verschiedene Sprachen übersetzt wurde (Espejo, Checa, Perales-Puchalt & Lison, 2020; Jovanovic et al., 2022; Li, Bai & Wang, 2013; Martin-Carbonell, Checa, Daza, Paternina & Espejo, 2021; Rahm, Heise & Schuldt, 2017; Senol-Durak & Durak, 2019; Yaaqeb et al., 2022). Die SPANE verfolgt das Ziel, bei hoher Ökonomie die volle Bandbreite möglicher Gefühle zu erfassen, die das Wohlbefinden bestimmen. Da die SPANE mehr nach einem prinzipiellen anstelle mehrerer, konkreter Einzelgefühle fragt, ist sie laut den Autoren besser global einsetzbar. Diener et al. (2009) begründen dies damit, dass die meisten in der westlichen Hemisphäre konstruierten Emotionstests kulturspezifische Gefühle ignorieren oder ihnen keine ausreichende Bedeutung zuweisen. Die SPANE folgt in seiner Struktur der PANAS, kritisiert aber an dieser, dass einzelne Items keine Gefühle erfassen, zum Beispiel das Item „*active*“. Die SPANE beginnt mit der Aufforderung: „*Please think about what you have been doing and experiencing during the past 4 weeks. Then report how much you experienced each of the following feelings, using the scale below. For each item, select a number from 1 to 5, and indicate that number on your response sheet.*“ Das festgelegte Zeitintervall ist laut den Autoren mit vier Wochen kurz genug, um die aktuelle Situation zu erfassen und gleichzeitig lang genug, um daraus auch Rückschlüsse über die generelle Stimmung zu ziehen. Die insgesamt 12 Items werden anschließend auf einer 5-stufigen Likert-Skala von 1 „*very rarely or never*“ bis 5 „*very often or always*“ beantwortet.

Abschließend können aus den Zahlenwerten der Items ein Positive Score (SPANE-P) und ein Negative Score (SPANE-N) berechnet werden. Aus der Differenz von SPANE-P – SPANE-N wird ein Balanced Score (SPANE-B) als genereller Indikator errechnet.

#### 1.4. Brief Emotional Experience Scale (BEES)

Rogers, Barblett und Robinson entwickelten 2016 die *Brief Emotional Experience Scale* (BEES). Die Autoren wollten im Rahmen eines groß angelegten Forschungsprojektes die Belastung von Schülern, Lehrer und Eltern im Zusammenhang mit dem australischen Schulvergleichstest „NAPLAN“ messen. Bestehende Maße zur Erfassung von Prüfungsangst waren den Autoren zufolge nicht ausreichend kurz und hätten keinen Vergleich zwischen den drei Stichproben erlaubt. Rogers, Cruickshank und Nosaka (2020) begründen die Notwendigkeit der Konstruktion der BEES damit, dass es kein kurzes Selbstbeurteilungsinstrument zur Erfassung von sowohl positiven Affekt (PA) als auch negativen Affekt (NA) gebe.

Die BEES wurde daraufhin mehrfach von den Autoren verwendet (Rogers & Cruickshank, 2020; Skead & Rogers, 2016; Skead & Rogers, Doraisamy, 2018; Skead, Rogers & Johnson; 2018; Turner et al., 2021) und über die verschiedenen Studien an ca. 3000 Versuchspersonen angewendet. Die Kürze der BEES sowie die inhaltliche Überschneidung zu den universitätsinternen Forschungsprojekten der Martin-Luther-Universität Halle-Wittenberg zu Belastungserscheinungen Studierender machten die BEES interessant für eine Verwendung im deutschen Sprachraum. Die bereits mehrfache Verwendung der BEES und die Veröffentlichung in Fachzeitschriften mit dem Peer-Review-Prinzip legte nahe, dass die BEES ein belastbares und verwendbares Messinstrument sei.

##### 1.4.1. Aufbau BEES

Das Screeningverfahren BEES besteht aus sechs Items, dargestellt durch einzelne Adjektive. Die BEES enthält eine positive Affektskala mit den Items [deutsche Übersetzung nach Pradel] „Happy [Glücklich]“, „Calm [Gelassen]“ und „Confident [Selbstsicher]“, sowie eine negative Affektskala mit den Items „Worried [Besorgt]“, „Sad [Traurig]“ und „Afraid [Ängstlich]“. Alle Items werden auf einer vierstufigen Likert-Skala mit den Antwortmöglichkeiten „[1] not at all [gar nicht]“, „[2] a little bit [ein wenig]“, „[3] quite a bit [ziemlich]“ und „[4] a lot [sehr]“ für einen vom Versuchsleiter beliebig wählbaren Zeitraum bewertet. Anschließend können drei Scores berechnet werden: BEES (pos.) ist die Summe der [x]-Werte/3 der positiven Affektskala, BEES (neg.) ist das Gegenstück der [x]-Werte/3 der negativen Affektskala. Abschließend kann aus BEES (pos.) – BEES (neg.) der BEES (bal.) gebildet werden.

In der Studie von Rogers, Barblett und Robinson (2016) werden bei der Erstapplikation der BEES keinerlei konkrete Aussagen zur Itemauswahl getroffen. Die Autoren berichten von keinem faktorenanalytischen oder theoretischen Hintergrund, der die Itemauswahl erklärt. Die BEES ist so aufgebaut, dass die sechs Adjektive in ihrer Qualität gegenteilige Pole abbilden, die sich paarweise aus einem Adjektiv der positiven und der negativen Affektskala zusammensetzen. Es ergeben sich die drei Bereiche Depression, Stress und Ängstlichkeit, wobei jeder Bereich einer Subskala der DASS (Lovibond & Lovibond, 1995) zugeordnet wird. Konkret werden der Subskala Depression das Paar „Happy [Glücklich]“ und „Sad [Traurig]“, der Subskala Ängstlichkeit „Confident [Selbstsicher]“ und „Afraid [Ängstlich]“ sowie der Subskala Stress „Calm [Gelassen]“ und „Worried [Besorgt]“ zugeteilt.

Die Reliabilitätsbetrachtung erfolgt ausschließlich über Cronbachs Alpha und wird den Richtlinien von Acock (2013) folgend mit  $\alpha = 0.84$  als gut eingestuft. In weiteren Erhebungen (Skead & Rogers, 2016; Skead, Rogers und Doraisamy, 2018) wird die Reliabilität der BEES erneut durch Betrachtungen von Interitemkorrelationen und Cronbachs Alpha als gut eingeschätzt.

Der größte Vorteil der BEES liegt in ihrer Ökonomie: Mit 6 Worten, die auf einer Skala bewertet werden müssen, ist die Testdurchführungszeit extrem kurz und die Belastung der Versuchspersonen gering. Dies ermöglicht Verlaufsbetrachtungen mit mehreren Messzeitpunkten am gleichen Tag, wobei aufgrund der niedrigen Belastung der Versuchspersonen anzunehmen ist, dass die drop-out-Raten gering ausfallen (Hoerger, 2010; Speer, King & Grossenbacher, 2016).

Trotz der Verwendung von Cronbachs Alpha machen die Autoren keine Aussagen zur Faktorenstruktur des BEES, was eine notwendige Voraussetzung für seine Verwendung darstellt (Cho, 2016). Da Cronbachs Alpha als Schätzer interner Konsistenz genutzt wird, und die Autoren über keine Verletzung der Voraussetzungen berichten, liegt die Schlussfolgerung nahe, dass der BEES 2016 als unidimensional eingeschätzt wird. Welcher dieser Faktor ist und was er letztendlich darstellt, wird von den Autoren allerdings nicht beschrieben. Es irritiert zudem, dass die Items einerseits drei verschiedenen Faktoren der DASS zugeordnet werden, andererseits aber keine weiteren Aussagen über die Zusammenhänge der Faktoren untereinander getroffen werden. Die Validitätsuntersuchung besteht aus der Betrachtung der konvergenten Validität anhand der Korrelation zur DASS von Lovibond und Lovibond (1995). Auffällig ist, dass von Lovibond (1998) explizit gesagt wurde, dass die DASS keinen NA misst.

2020 formulieren Rogers, Cruickshank & Nosaka hingegen explizit den Anspruch, mit der BEES PA und NA zu messen. Wenn nun die BEES NA messen soll, der DASS dies aber explizit nicht tut, bleibt offen, was eine Korrelation zwischen BEES und DASS als Maß der Validierung aussagt. Beide Autoren konstruieren ihre Konstrukte explizit anders, aber trotzdem korrelieren sie hoch.

Der Validierungsprozess wird von Rogers, Barblett und Robinson (2016) selbst als „limitiert“ bezeichnet. Ergänzt wird diese „limitierte“ Validierung durch Skead und Rogers (2016): Die BEES wird hier in einer weiteren Erhebung verwendet und erneut wird an hohen Cronbachs Alpha- und Interkorrelationswerten der Items untereinander eine valide Messung bescheinigt. Dasselbe Vorgehen findet sich bei Skead, Rogers und Doraisamy (2018). Vier Jahre nach der Erstveröffentlichung und -verwendung der BEES untersuchen Rogers, Cruickshank & Nosaka (2020) ihre Reliabilität, Validität und Faktorenstruktur. Sie beschreiben 2020, dass die BEES konstruiert wurde in Anlehnung an die DASS sowie die SPANE (Diener et al., 2010). Im 2016 veröffentlichten Artikel über die Erstverwendung wird dieses Vorgehen nicht erklärt. Die DASS wird in dieser Veröffentlichung lediglich als externes Kriterium zur Validierung genutzt, die SPANE findet 2016 keine Erwähnung.

#### 1.4.2 Faktorenstruktur BEES

In der Veröffentlichung zur Erstverwendung der BEES aus dem Jahr 2016 gibt es seitens der Autoren keine Aussagen zur Faktorenstruktur. Rogers, Cruickshank und Nosaka schreiben 2020, dass die BEES in Anlehnung an die SPANE konstruiert wurde, und folgern aus der Struktur der SPANE selbst sowie der parallelen Konstruktion der BEES eine zweifaktorielle Struktur der BEES. Sie berechnen eine ein-, eine zwei- und eine drei-faktorielle Lösung und vergleichen diese miteinander. Das Ergebnis der konfirmatorischen Faktorenanalyse weist den besten Modellfit für die zwei-faktorielle Lösung auf. Die Autoren postulieren daher eine 2-Faktoren-Struktur der BEES mit einem Faktor für positive und einem Faktor für negative Emotion.

Mit dieser postulierten Faktorenstruktur und den vorhergegangenen, explizit erläuterten Voraussetzung für die Verwendung von Cronbachs Alpha, ist die Qualität der bescheinigten Reliabilität anhand Cronbachs Alpha höchst fraglich.

## 1.5. Brief Emotional Experience Physical Scale (BEEPS)

Skead und Rogers entwickelten 2016 die *Brief Emotional Experience Physical Scale* (BEEPS) zur Erfassung des „physical distress“. Dafür sollen die Probanden die Frage „Wie haben Sie sich im vorangegangenen Semester physisch gefühlt?“ beantworten, in dem sie sechs Items des BEEPS auf einer vier-stufigen Skala bewerteten. Die Items [deutsche Übersetzung nach Pradel] lauten dabei „Healthy [Gesund]“, „Lethargic [Lethargisch]“, „Strong [Stark]“, „Unfit [Unfit]“, „Energetic [Energiegeladen]“ und „Tense“ – wobei „Tense“ 2018 von Skead, Rogers und Doraisamy als Item verworfen und durch „Weak [Schwach]“ ersetzt wurde. Die für die Probanden möglichen Antworten sind „[1] not at all [gar nicht]“, „[2] a little bit [ein wenig]“, „[3] quite a bit [ziemlich]“ sowie „[4] a lot [sehr]“.

Die Items „Strong [Stark]“, „Healthy [Gesund]“ und „Energetic [Energiegeladen]“ sollen die positive Affektskala abdecken, die Items „Lethargic [lethargisch]“, „Weak [Schwach]“ und „Unfit [Unfit]“ die negative Affektskala.

Nach dieser sehr kurzen theoretischen Einführung und knappen Informationen zur BEEPS verwenden die Autoren den Test direkt für eine Erhebung. Es fehlen Aussagen über das gemessene Konstrukt oder eine Definition davon sowie über die Konstruktions- und Auswahlprinzipien der Items. Die Faktorenstruktur wird nicht thematisiert. Nach der Erstapplikation der BEEPS an eine Stichprobe erfolgt die „Validierung“ der BEEPS ausschließlich durch die Betrachtung von Cronbachs Alpha. Mit  $\alpha = 0.82$  wird der BEEPS von Skead und Rogers (2016) als gut eingestuft. Da das Item „Tense“ die interne Konsistenz senkte, wurde es von der Auswertung gestrichen. Diese Form der Validierung wird von den Autoren nicht als Limitation angesehen und im entsprechenden Abschnitt des Primärartikels nicht kritisch erwähnt. Der 2016 gefundene korrelative Zusammenhang zwischen dem BEES und BEEPS von  $r = 0.54$  wird von den Autoren als „moderat“ eingestuft. Die Autoren ziehen aus dieser Korrelation keine weiteren Schlüsse. 2018 erfolgt durch Skead, Rogers und Doraisamy eine Validierung der BEEPS ausschließlich durch die Betrachtung von Cronbachs Alpha. In der Verwendung der BEEPS durch Rogers und Cruickshank (2020) fällt auf, dass keine weiteren Aussagen zur Testgüte erfolgen.

Über alle drei Studien, die die BEEPS verwenden, erfolgt keinerlei Betrachtung externer Validität. Der Test wird zu keiner bestehenden Theorie anderer Autoren oder bestehenden Verfahren eingeordnet oder abgegrenzt.

## 1.6. Methoden der Reliabilitätsschätzung

Es existiert eine Vielzahl an Methoden zur Reliabilitätsschätzung. Die gängigsten Verfahren sind Split-Half-Reliabilität, Parallel-Test-Reliabilität, Test-Retest-Reliabilität, Cronbachs Alpha und McDonalds Omega. Auf die beiden letztgenannten Kennwerte als statischer Ausdruck interner Konsistenz einer Skala soll im Folgenden genauer eingegangen werden. Neben den genannten häufig verwendeten Koeffizienten existieren viele weitere (Cho, 2016).

### 1.6.1. Cronbachs Alpha

Cronbachs Alpha ist das in der Psychologie meistverwendete Maß zur Betrachtung der internen Konsistenz einer Skala als Schätzer ihrer Reliabilität (Elousa & Iliescu, 2012; McNeish, 2017). Sie ist definiert als die durchschnittliche Korrelation der einzelnen Items eines Gesamttests zueinander. Alpha-Werte von  $>0.6$  werden als akzeptabel,  $>0.7$  als gut,  $>0.8$  als sehr gut,  $>0.9$  als Hinweis auf möglicherweise redundante Items verstanden. Dieser Einteilung folgt Streiner (2003).

Für die Verwendung von Cronbachs Alpha sind seine Bedingungen einzuhalten (Yang & Green, 2011):

1. Cronbachs Alpha geht von einem tau-äquivalenten Messmodell aus,
2. die Items der Skala sind auf einer kontinuierlichen Skala und normalverteilt,
3. die Fehler der einzelnen Items kovariieren nicht und
4. die Skala ist unidimensional.

Daraus folgend ist die Verwendung von Cronbachs Alpha für multidimensionale Fragestellungen nicht gestattet. Trotz des hohen Stellenwerts und der vielfältigen Verwendung von Cronbachs Alpha finden sich in der psychologischen Fachliteratur viele Fehlverwendungen von Cronbachs Alpha (Cho & Kim, 2015; Dune, Baguley & Brunnsden, 2013; McNeish, 2017; Yang & Green, 2013). Die vorangeführten Autoren stellen sehr deutlich die Diskrepanz zwischen dem theoretischen Grundwissen bezüglich Cronbachs Alpha und seiner letztendlichen forschungspraktischen Verwendung dar. Zwar wird Cronbachs Alpha sehr häufig verwendet, oft ist er jedoch entweder nicht der für den konkreten Fall beste Koeffizient oder seine Bedingungen werden verletzt (Cho, 2016). Cho (2016) zeigt überblicksartig die häufigsten Misskonzeptionen von Cronbachs Alpha.

Graham (2006) bezeichnet das „Ausmaß der Ignoranz gegenüber fundamentalen Messgrundlagen als endemisch“. Die Verletzung der Voraussetzungen von Cronbachs Alpha führen zu systematischen Fehlern der Reliabilitätsschätzung (McNeils, 2017). Die Nicht-Beachtung der Unidimensionalitätsbedingung beispielsweise führt zu einer systematischen Unterschätzung der Reliabilität, bei der Verletzung der Voraussetzung des tau-äquivalenten Messmodels ist Cronbachs Alpha nicht zu verwenden (Cho, 2016; McNeils, 2017).

#### 1.6.2. McDonalds Omega

Omega (McDonald, 1970; McDonald, 1999) ist als Reliabilitätskoeffizient für kongenerische Skalen konstruiert. Daher entfällt bei ihm die Voraussetzung einer tau-äquivalenten Messung. Das bedeutet, dass Omega unterschiedliche Faktorenladungen zulässt während Alpha von gleichen Faktorenladungen als Modellbedingung ausgeht. Es gibt verschiedene Subkonzeptionen von Omega mit verschiedenen Einsatzbereichen. So kann „Omega hierarchisch“ für Skalen mit mehreren Dimensionen genutzt werden (McNeish, 2017).

Eine vollständige Übersicht über die mathematischen Konzeptionen, ihre Vor- und Nachteile sowie ihren spezifischen Einsatzbereiche findet sich bei Kelley und Pornprasertmanit (2016), McNeish (2017) oder Cho (2016).

Zusammenfassend lässt sich sagen, dass auch Omega-Koeffizienten aktuell einen hohen Stellenwert im Bereich der Skalvalidierung besitzen (Garcia-Garzon & Garrido, 2020; Viladrich, Angulo-Brunet & Doval, 2017; Zinbarg & Alden, 2015), da sie aufgrund realistischerer Voraussetzungen und geringerer Modellrestriktionen gegenüber Cronbachs Alpha aussagekräftigere Reliabilitätskoeffizienten darstellen (Dunn, Baguley & Brunson, 2013).

#### 1.7. Standardisierte Testbeurteilungen

Das Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation deutscher Psychologen dient der Qualitätsbeurteilung, -sicherung und -optimierung von psychologischen, theoretisch fundierten Verfahren (Diagnostik- und Testkuratorium, 2018). Das System bietet einen standardisierten Ablauf, der eine objektive Beurteilung und eine Vergleichbarkeit der zu bewertenden Tests ermöglicht. Kersting (2018) verfasste hierfür eine



DIN SCREEN Checkliste für Instrumente zur Erfassung von menschlichem Leben und Verhalten, explizit nicht jedoch für eignungsdiagnostische Verfahren. Anhand dieser Checkliste ist es möglich, standardisiert Tests in ihrer Qualität zu bewerten. Dieses Testbeurteilungssystem und die Checkliste von Kersting werden in dieser Arbeit zur finalen Qualitätsbewertung der Testverfahren BEES und BEEPS verwendet.

## 2. Zielstellung

Die von Rogers, Barblett, Cruickshank und Robinson neu entwickelten Messinstrumente BEES und BEEPS sind den Autoren zufolge valide und ökonomische Verfahren zur Erfassung sowohl von Positivem und Negativem Affekt als auch von psychischer und physischer Belastung. Aufgrund inhaltlicher Überschneidungen zu bestehenden Forschungsprojekten in der Martin-Luther-Universität Halle-Wittenberg zu Belastungserscheinungen bei Studierenden verschiedener Fachrichtungen sollten die Screeningverfahren im deutschsprachigen Raum nutzbar gemacht werden.

Zunächst sollten BEES und BEEPS nach fachlichem Standard übersetzt werden. Anschließend wurden die neu übersetzten Verfahren erstmalig appliziert. Die Erhebung wurde ergänzt durch bereits etablierte Maße wie dem BDI-II oder die DASS-21. BEES und BEEPS sollten hinsichtlich Objektivität, Reliabilität und Validität anhand der empirisch gefundenen Daten sowie den Richtlinien des Diagnostik- und Testkuratoriums der Föderation deutscher Psychologen überprüft werden. Abschließend wurde die Verwendbarkeit bewertet und die neu übersetzten Testverfahren wurden zu bestehenden Testverfahren eingeordnet.

### 3. Methode

#### 3.1. Stichprobe

##### 3.1.1. Rekrutierung

Im Voraus wurde die benötigte Stichprobengröße mit G-Power 3.1.9.7 (Faul, Erdfelder, Buchner & Lang, 2009) berechnet. Diese ergab bei festgesetzten  $\alpha = 0.05$  und  $1-\beta = 0.95$  und einer gewünschten Effektstärke von  $f = 0.25$  eine optimale Gesamtstichprobengröße von  $N = 400$ . Die Stichprobe bestand aus Studierenden der Humanmedizin des vierten und zehnten Semesters mit einer Grundgesamtheit von 446 Personen. Die Studierenden wurden im Rahmen ihrer Online-Lehrveranstaltungen gebeten, einen Online-Fragebogen auf der universitätsinternen Internetplattform „studip“ der Martin-Luther-Universität Halle-Wittenberg auszufüllen. Die Erhebung war ursprünglich als physischer Fragebogen in den Kontaktlehrstunden angedacht, aufgrund der zum Erhebungszeitpunkt geltenden Corona-Eindämmungsverordnungen war dies jedoch nicht möglich.

##### 3.1.2. Deskriptive Stichprobenbeschreibung

Von der Grundgesamtheit  $n = 446$  bearbeiteten 228 den Fragebogen und gaben ihn ab. 20 Versuchspersonen wurden von der Auswertung ausgeschlossen, da ihre Angaben mehr als zehn Auslassungen enthielten. In die Auswertung flossen somit die Daten von 208 Versuchspersonen (79,3% weiblich, 20,7% männlich) in einem Altersbereich von 19-45 Jahren ( $m = 23,47$ ;  $SD = 4,16$ ) ein. Die erwünschte Stichprobengröße konnte jedoch nicht erreicht werden. Als Hauptgrund wird die geringere Rücklaufquote bei Onlineerhebungen gegenüber Präsenzerhebungen angesehen.

## 3.2. Untersuchungsmethode

### 3.2.1. Fragebogaufbau

Der Fragebogen umfasste insgesamt 13 Seiten und war in sechs Abschnitte untergliedert:

1. Soziodemografische Faktoren
2. Risiko- und Resilienzfaktoren bezüglich Belastungserscheinungen
3. Screening zum emotionalen (BEES) und körperlichen Wohlbefinden (BEEPS)
4. Becks Depression Inventory (BDI-II)
5. Neurotizismusskala des Fünf-Faktoren Model der Persönlichkeit (NEO-FFI)
6. Depression-Angst-Skala (DASS-21)

Die Bearbeitungsdauer lag im Zeitbereich von 25 bis 35 Minuten pro Person. Alle Versuchspersonen bildeten anfangs einen Versuchspersonencode, der eine Zuordnung in der Zukunft liegender Erhebungen zur der aktuellen ermöglicht. Der vollständige Fragebogen findet sich in Anhang A1.

Als soziodemografische Faktoren wurden das Alter, Körpergröße, biologisches sowie soziales Geschlecht, Gewicht, Familienstand, Anzahl leiblicher Kinder, Ort des Aufwachsens sowie eigener Bildungsstand und der Bildungsstand der Eltern erfasst. Die Erhebung fand im Rahmen einer longitudinalen Studie über klinische und subklinische Belastungen von Studierenden verschiedener Fachrichtungen an der Martin-Luther-Universität Halle-Wittenberg statt. Da in deren Fokus unter anderem die Risiko- und Resilienzfaktoren (2) für das Erleben von Belastungen stehen, wurden diese für die Auswertung in späteren Arbeiten miterhoben. Die anfänglich erfragten, soziodemografischen Daten (1), sowie die Risiko- und Resilienzfaktoren folgten vorangegangenen institutsinternen Arbeiten und wurden gleich gehalten.

Des Weiteren wurden diverse Risiko- und Resilienzfaktoren bezüglich psychischer Belastungserscheinungen erfasst. Diese sind für die vorliegende Arbeit nicht relevant und werden daher nicht weiter betrachtet. Abhandlungen zu diesen Faktoren und ihren Auswirkungen auf Belastungserscheinungen finden sich beispielweise in Rehnisch et al., (2021), Kindt et al., (2022) sowie Rabkow et al. (2021).

BEES und BEEPS (3) sowie die bereits etablierten Belastungsmaße BDI-II (4) und DASS-21 (6) sind bereits im Theorieteil eingeführt wurden und werden daher nicht erneut genauer beschrieben.

Ein weiterer Bestandteil des Fragebogens ist die Neurotizismusskala des NEO-FFI (5) nach Borkenau und Ostendorf (1994). Die Neurotizimusskala beschreibt hierbei die psychische Stabilität als Persönlichkeitseigenschaft. Hohe Werte der Neurotizimusskala korrelieren mit Anfälligkeit für höheres Stress- und Belastungserleben.

### 3.2.2. Übersetzungsmethodik BEES und BEEPS

A priori erfolgte die Übersetzung der Instrumente BEES und BEEPS aus dem Englischen ins Deutsche. Hierbei wurde die „Forward-Backward“-Übersetzungsmethode nach den Richtlinien zur Übersetzung psychologischer Instrumente (Gudmunsson et al., 2009) verwendet. Die primäre Übersetzung aus dem Englischen ins Deutsche wurde durch den Autor Christian Pradel vorgenommen. Anschließend hat die Gerichtsübersetzerin Dr. Thurid Chapman (Bundesverband der Übersetzer und Dolmetscher e.V., BDÜ) in das Englische rückübersetzt. Die Über- und Rückübersetzung wurde abschließend vom Primärautor Dr. Shane Rogers, sowie seiner deutschsprachigen PhD-Studentin Valeska Berg bestätigt. Die Übersetzungen von BEES und BEEPS finden sich in vollständiger Form im Anhang A1.

### 3.3. Untersuchungsplan

Der Ausgangspunkt der vorliegenden Arbeit war ein Forschungsprojekt zur Messung von Belastungserscheinungen von einer Kohorte Medizinstudierender der Martin-Luther-Universität Halle-Wittenberg durch verschiedene Belastungsinstrumente. Es sollen die empirisch gefundenen Zusammenhänge zwischen den erhobenen Maßen betrachtet werden, um letztendlich Aussagen über die psychometrische Qualität der Instrumente BEES und BEEPS zu treffen. Dazu soll der vorher eingeführte Fragebogen verwendet werden.

### 3.4. Untersuchungsablauf

Nach der positive Votierung des Untersuchungsdesigns durch die lokale Ethikkommission erfolgte die Datenerhebung zwischen Mai und Juli 2021 an einer Kohorte Medizinstudierender aus zwei verschiedenen Semestern der Martin-Luther-Universität Halle-Wittenberg. Die Erhebung erfolgte freiwillig und vollständig anonymisiert mittels eines

Versuchspersonencodes, welcher die Zuordnung der Daten zu späteren Messzeitpunkten ermöglicht. Durch das Format der Onlineerhebung konnten die Versuchspersonen Zeitpunkt und Ort der Bearbeitung selbst wählen. Es erfolgte keine Erfassung des Bearbeitungsortes oder Bearbeitungszeitpunktes. Die Versuchspersonen konnten während der Bearbeitung pausieren und zu einem beliebigen Zeitpunkt fortfahren.

### 3.5. Statistische Hypothesen

Folgende inhaltliche Hypothesen sollen durch interferenzstatistische Auswertung betrachtet und anschließend als statistische Hypothesen formuliert werden.

#### Hypothese A

Da die Merkmalsverteilung der psychischen Belastung in der deutschen und der australischen Stichprobe gleich sind, sollen auch die Mittelwerte der Maße beider Stichproben gleich sein.

$$m_{bees/aus} = m_{bees/dt}$$

$$m_{beeps/aus} = m_{beeps/dt}$$

$$m_{dass-21/aus} = m_{dass-21/dt}$$

#### Hypothese B

Die Sprache hat keinen Einfluss auf die interne Konsistenz, geschätzt durch Cronbachs Alpha von BEES und BEEPS.

$$\alpha_{bees/aus} = \alpha_{dt/aus}$$

$$\alpha_{beeps/aus} = \alpha_{beeps/dt}$$

#### Hypothese C

Es existiert eine hohe positive Korrelation zwischen der BEES (neg.) und dem BDI-II sowie zwischen der BEES (neg.) und der DASS-21.

$$r_{beesneg/bdi-II} > 0.7$$

$$r_{beesneg/dass21} > 0.7$$

## Hypothese D

Es existiert eine hohe negative Korrelation zwischen der BEES (pos.) und dem BDI-II sowie zwischen der BEES und der DASS-21.

$$r_{beespos/bdi-II} > -0.7$$

$$r_{beesüps/dass21} > -0.7$$

## 3.6. Auswertungsplan

Die Datenanalyse fand unter Verwendung von SPSS 26 (IBM Corp., 2019), R (R Core Team, 2013) sowie Lavaan (Rosseel, 2012) statt. Die erhobenen Daten wurden im ersten Schritt auf Vollständigkeit geprüft, Versuchspersonen mit mehr als zehn Auslassungen wurden von der Auswertung gestrichen – das betraf 20 Teilnehmer. Für alle nachfolgenden interferenzstatistischen Verfahren wurde  $\alpha = 0.05$  festgesetzt, sofern nicht explizit anders angegeben.

## 3.7. Operationalisierung der psychometrischen Qualitätsmerkmale

Im Folgenden werden die Operationalisierungen der psychometrischen Qualitätsmerkmale Objektivität, Reliabilität und Validität vorgestellt.

### 3.7.1. Operationalisierung der Objektivität

Die Objektivität von BEES und BEEPS soll hinsichtlich der Durchführungs-, Auswertungs- und Interpretationsobjektivität betrachtet werden. Diese Einteilung folgt Borkenau (2015). Die Testverfahren werden anhand der folgenden Frage bewertet:

- Existiert eine Standardisierung der Testdurchführung und der Testinstruktionen?
- Existieren präzise Auswertungsrichtlinien?
- Existieren exakte Regeln, wie Befunde zu interpretieren sind? Sind diese Befunde genormt?
- Erreichen die Instruktionen die Unabhängigkeit von Testleiter gegenüber Testergebnis und Testinterpretation?

### 3.7.2. Operationalisierung der Reliabilität

Verschiedene gängige Verfahren der Reliabilitätsschätzung können im vorliegenden Fall nicht verwendet werden. Da für die BEES und die BEEPS keine parallelen Formen existieren, kann die Paralleltest-Reliabilität nicht bestimmt werden. Aufgrund der Kürze von nur sechs Items entfällt ebenfalls die Split-half-Reliabilität. Da es nur einen Messzeitpunkt gibt, ist keine Retest-Reliabilität erfassbar. Schließlich gibt es auch keine Menge an Expertenurteilen, sodass die Inter-Rater-Reliabilität ebenfalls nicht verwendbar ist.

Daher wird die Reliabilität von BEES und BEEPS über Maße der internen Konsistenz betrachtet. Um Vergleichbarkeit zu Arbeiten mit BEES und BEEPS im englischsprachigen Original zu ermöglichen, wird Cronbachs Alpha berechnet. Aufgrund eingangs beschriebener Probleme bei der Verwendung von Cronbachs Alpha wird zusätzlich McDonalds Omega als Schätzer der internen Konsistenz verwendet. Alpha-Werte von  $>0.6$  werden als akzeptabel,  $>0.7$  als gut,  $>0.8$  als sehr gut,  $>0.9$  als Hinweis auf möglicherweise redundante Items verstanden. Diese Einteilung folgt Streiner (2003).

### 3.7.3. Operationalisierung der Validität

Die Validierung von BEES und BEEPS soll hinsichtlich der Inhalts-, Kriteriums- und Konstruktvalidität erfolgen. Konkret werden die beiden Verfahren anhand folgender Kriterien bewertet:

- Repräsentativität der Items für das zu erfassende Merkmal,
- Enge des Zusammenhangs zwischen Test- und Kriteriumswerten,
- Bewährung des Tests im Hinblick auf mit dem Merkmal verbundene theoretische Annahmen.

Da BEES und BEEPS sowohl PA und als auch NA erfassen sollen, wären psychopathologische Befunde ein mögliches Kriterium. Da solche oder Ähnliches für die Probanden aber nicht vorliegt, soll als Kriterium die konvergente und diskriminante Validität zu den mitapplizierten Verfahren genutzt werden. Konvergente Validität „gilt als nachgewiesen, wenn die Korrelationen der Messungen eines Merkmals zwischen jeweils zwei Methoden statistisch signifikant von null verschieden und hoch sind“ (Schermelleh-Engel & Schweizer, 2003). Konkret sollen BEES und BEEPS hoch korrelieren mit den Maßen des BDI-II und der DASS-21.

Niedrig bzw. hoch negativ korrelieren sollen BEES (pos.) und BEEPS (pos.) zu BDI-II und DASS-21. Laut Campbell und Fiske (1959) existiert kein absolutes Maß der Korrelationshöhe für konvergente Validitäten.

#### 3.7.4. Operationalisierung der Faktorenstrukturprüfung

Da für den BEES bereits Überprüfungen zur Faktorenstruktur von Rogers, Cruickshank und Nosaka (2020) existieren, soll die Faktorenstruktur der neu übersetzten BEES analog betrachtet werden. In drei verschiedenen konfirmatorischen Faktorenanalysen sollen eine einfaktorielle, eine zweifaktorielle und eine dreifaktorielle Lösung berechnet und miteinander verglichen werden. Dem Postulat der Primärautoren folgend, ist eine zweifaktorielle Struktur als bestes Modell anzunehmen.

Für den BEEPS existiert seitens der Urheber keinerlei Aussage zur Faktorenstruktur. Das neu übersetzte Verfahren soll daher mit exploratorischen Faktorenanalysen bezüglich seiner Struktur untersucht werden.

Die gefundenen Modellparameter werden den Richtlinien von Acock (2013) ( $RMSEA \leq 0.08$ ,  $CFI \geq 0.95$ ,  $SRMR \leq 0.08$ ) folgend bewertet.



#### 4. Ergebnisse

Im Folgenden werden die Ergebnisse der empirischen Erhebung zunächst deskriptiv dargestellt und anschließend entsprechend der Zielstellung interferenzstatistisch ausgewertet.

##### 4.1. Deskriptive Darstellung

Es erfolgt die deskriptive Betrachtung aller Daten nach den üblichen Maßen Mittelwert, Standardabweichung, Varianz, Minimum, Maximum und Schiefe. Die für die Aufgabenstellung relevanten Ergebnisse werden im Folgenden dargestellt. Tabelle 1 gibt Auskunft über Mittelwerte inkl. Standardabweichungen, Skalenbereiche, empirisch gefundene Minima und Maxima sowie Schiefe der Verteilung.

Tabelle 1

*Darstellung der Mittelwerte, Skalarange, Minima, Maxima und Schiefe der erhobenen Maße sowie deren Subskalen*

Skala	Mittelwert (SD)	Skalarange	Min - Max	Schiefe (SE)
BEES (bal.)	.425 (1.13)	-3 – 3	-2,67 – 3	-.589 (.17)
BEES (pos.)	2.47 (.61)	1 – 4	1 – 4	-0.01 (.17)
BEES (neg.)	2.04 (.64)	1 – 4	1 – 4	.843 (.17)
BEEPS (bal.)	-.71 (1.16)	-3 – 3	-3 – 2.33	.32 (.17)
BEEPS (pos.)	2.65 (.65)	1 – 4	1 – 4	-.053 (.17)
BEEPS (neg.)	1.94 (.66)	1 – 4	1 – 4	.61 (.17)
DASS 21 (total)	15,75 (11.42)	0 – 63	0 – 54	1.11 (.17)
DASS 21 (anx.)	3.84 (3.89)	0 – 21	0 – 18	1.45 (.17)
DASS 21 (str.)	7.2 (4.56)	0 – 21	0 – 19	.6 (.17)
DASS 21 (dep.)	4.72 (4.54)	0 – 21	0 – 21	1.37 (.17)
BDI-II (total)	11.31 (8.64)	0 – 63	0 – 35	1.03 (.17)
NEO-FFI-N	21.7 (10.08)	0 – 48	2 – 43	.211 (.17)

Da für die BEES keine Normierungstabelle existiert, können die empirischen, deskriptiven Daten nur mit anderen Erhebungen der BEES verglichen werden. Hierzu sollen die Werte aus Rogers, Cruickshank und Nosaka (2020) verwendet werden. Die numerischen Werte der deutschen sowie der australischen Stichprobe zeigt Tabelle 2.

Für die BEEPS verzichteten die Autoren in den australischen Anwendungen auf den Bericht von statistischen Standardmaßen wie Mittelwert und Standardabweichungen. Ein Vergleich zu australischen Stichproben ist darum nicht möglich.

Tabelle 2

*Darstellung der Mittelwerte (Standardabweichung) in deutscher und australischer Stichprobe*

Skala	Mittelwert (SD)	Mittelwert (SD)
	dt. Stichprobe	austr. Stichprobe
BEES (bal.)	.425 (1.13)	.41 (1.19)
BEES (pos.)	2.47 (.61)	2.59 (.66)
BEES (neg.)	2.04 (.64)	2.18 (.68)

#### 4.2. Objektivitätsergebnisse

Die Überprüfung der Objektivität von BEES und BEEPS erfolgt nicht durch statistisch-mathematische, sondern durch inhaltliche Überlegungen. Diese Abhandlung folgt im Diskussionsteil (Kapitel 5.2).

#### 4.3. Reliabilitätsergebnisse

Die Reliabilitätsanalyse von BEES und BEEPS erfolgte anhand Cronbachs Alpha und McDonalds Omega. Auskunft über die empirisch gefundenen  $\alpha$ -Werte von BEES und BEEPS sowie deren Subskalen in deutschsprachiger und englischsprachiger Applikation gibt Tabelle 3. Die Berechnungen für Omega ergaben bei festgesetzter Faktorenstruktur mit zwei Faktoren und einem Generalfaktor  $\omega_h = 0.66$  sowie  $\omega_t = 0.84$  für die BEES. Die Omegastatistiken für die BEEPS wurden analog berechnet und ergaben  $\omega_h = 0.8$  sowie ein  $\omega_t = 0.94$ .

Tabelle 3

*Darstellung der Cronbachs Alpha-Werte für BEES und BEEPS sowie deren Subskalen jeweilig für die deutschsprachige Adaption und das Original*

Skala	$\alpha$ dt.	$\alpha$ engl. (Skead & Rogers, 2016)	$\alpha$ engl. (Rogers, Cruickshank, Nosaka, 2020)
BEES (bal.)	.817	.83	.84
BEES (pos.)	.715	k. A.	.81
BEES (neg.)	.71	k. A.	.76
BEEPS (bal.)	.807	.78	k. A.
BEEPS (pos.)	.768	k. A.	k. A.
BEEPS (neg.)	.687	k. A.	k. A.

#### 4.4. Validitätsergebnisse

Für eine Überprüfung der konvergenten und diskriminanten Validität sollen die Pearson Interkorrelationen zwischen der BEES, der DASS-21 und dem BDI-II sowie deren Subskalen betrachtet werden. Eine Übersicht über die gefundenen Korrelationen gibt Tabelle 4.

Tabelle 4

*Darstellung der Pearson Interkorrelationen zwischen der BEES, DASS-21 und dem BDI-II sowie deren Subskalen; alle Korrelationen sind auf dem Niveau 0.01 zweiseitig signifikant*

	BEES (bal.)	BEES (pos.)	BEES (neg.)	BEES (anx.)	BEES (str.)	BEES (dep.)	DASS 21 (total)	DASS 21 (anx.)	DASS 21 (str.)	DASS 21 (dep.)	BDI II
BEES (bal.)	1										
BEES (pos.)	.902	1									
BEES (neg.)	-.913	-.648	1								
BEES (anx.)	-.848	-.751	.788	1							
BEES (str.)	-.887	-.831	.78	.638	1						
BEES (dep.)	-.819	-.721	.765	.503	.622	1					
DASS 21 (total)	-.740	-.626	.715	.601	.661	.631	1				
DASS 21 (anx.)	-.561	-.439	.575	.510	.503	.414	.835	1			
DASS 21 (str.)	-.663	-.588	.615	.480	.667	.55	.914	.665	1		
DASS 21 (dep.)	-.717	-.609	.69	.594	.622	.682	.884	.578	.726	1	
BDI II	-.753	-.667	.699	.601	.640	.688	.843	.608	.745	.852	1

Für die Überprüfung konvergenter und diskriminanter Validitäten der BEEPS soll analog dem Vorgehen bei der BEES die Pearson Interkorrelationen zur DASS-21 und zum BDI-II betrachtet werden. Eine Übersicht über die gefundenen Korrelationen gibt Tabelle 5.

Tabelle 5

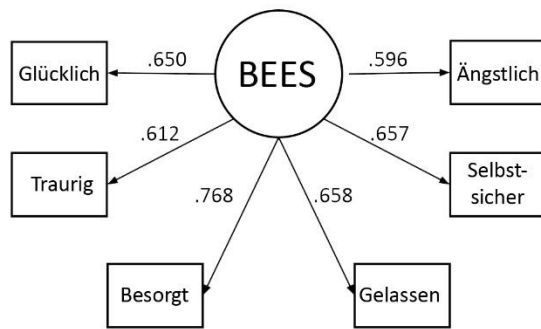
*Darstellung der Pearson Interkorrelationen der BEEPS mit dem DASS-21 und dem BDI-II sowie deren Subskalen, alle Korrelationen sind auf dem Niveau 0.01 zweiseitig signifikant*

	BEEPS (bal.)	BEEPS (pos.)	BEEPS (neg.)	DASS 21 (total)	DASS 21 (anx.)	DASS 21 (str.)	DASS 21 (dep.)	BDI II
BEEPS (bal.)	1							
BEEPS (pos.)	.579	1						
BEEPS (neg.)	-.527	-.574	1					
DASS 21 (total)	-.740	-.511	.541	1				
DASS 21 (anx.)	-.561	-.347	.385	.835	1			
DASS 21 (str.)	-.663	-.475	.489	.914	.665	1		
DASS 21 (dep.)	-.717	-.511	.542	.884	.578	.726	1	
BDI II	-.753	-.576	.583	.843	.61	.745	.852	1

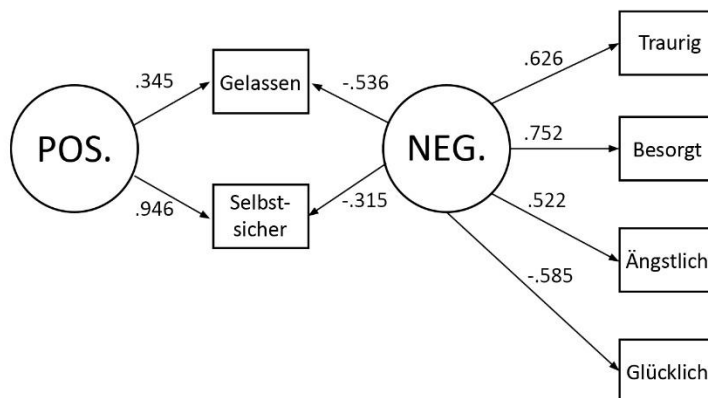
#### 4.5. Faktorenstrukturprüfung

Rogers, Cruickshank und Nosaka (2020) nutzten zur Untersuchung der Faktorenstruktur der BEES Strukturgleichungsmodelle, um konfirmatorische Faktorenanalysen durchzuführen. In ihren zwei Teilerhebungen aus dem Jahr 2020 werden ein-, zwei- und drei-faktorielle Lösungen verglichen. Diesem Vorgehen soll in der Faktorenstrukturprüfung der deutschen BEES gefolgt werden. Dazu wurden drei verschiedene konfirmatorische Faktorenanalysen berechnet, die gefundenen Fit-Werte werden in Abbildung 1 dargestellt.

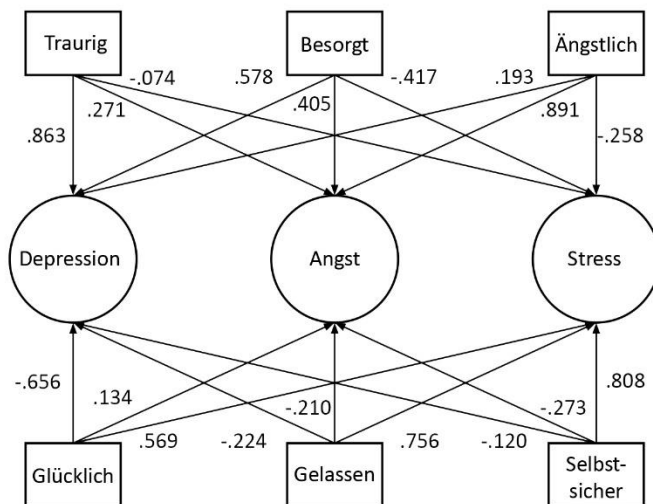
Abbildung 1. Vergleich verschiedener Faktorenstrukturen der BEES



1 Faktor Lösung  
 Chi-Quadrat (9) = 49.989 ; p = 0.0001  
 RMSEA = 0.148  
 CFI = 0.893  
 TLI = 0.822  
 SRMR = 0.058  
 AIC = 2680.727  
 BIC = 2682.756 sample-size adjusted



Chi-Quadrat (15) = 10.315; p = 0.244  
 RMSEA = 0.037  
 CFI = 0.994  
 TLI = 0.988  
 SRMR = 0.028  
 AIC = 2582.483  
 BIC = 2584.681 sample-size adjusted

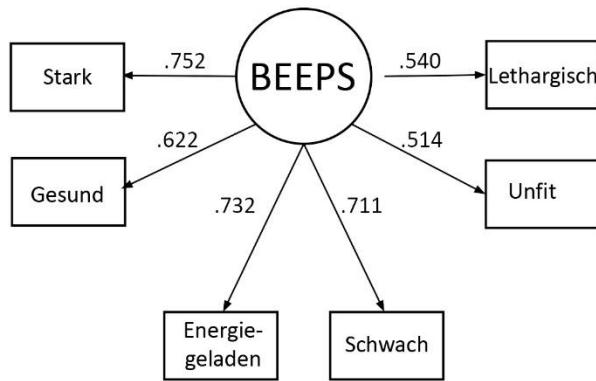


Chi-Quadrat (6) = 9.463; p = 0.149  
 RMSEA = 0.053  
 CFI = 0.990  
 TLI = 0.976  
 SRMR = 0.026  
 AIC = 2585.631  
 BIC = 2588.166 sample-size adjusted

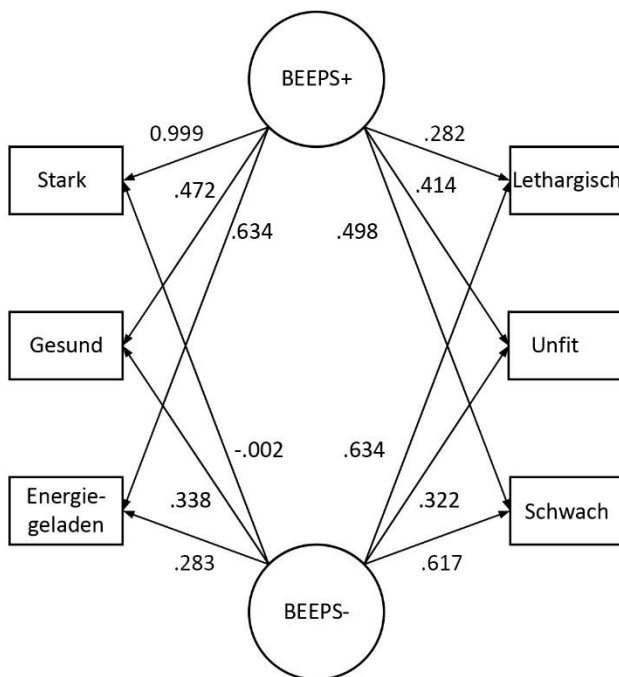
Abbildung 1. Darstellung der verschiedenen Faktorenstrukturen der BEES sowie der dazugehörigen Fit-Indizes und Faktorenladungen. Die 2-Faktoren-Lösung weist den besten Fit auf.

Für die BEEPS gab es im englischsprachigen Original keine Faktorenstrukturanalysen und keine Aussagen zu einer theoretischen Faktorenstruktur. Daher wurde eine explorative Faktorenanalyse durchgeführt, um eine potenzielle Struktur innerhalb der BEEPS zu identifizieren. Zunächst wurde die explorative Faktorenanalyse mit einer maximum likelihood extraktion und einer varimax Rotation für eine freie Faktorenanzahl durchgeführt. Diese stellte anhand des Eigenwertes  $\geq 1$  eine 1-Faktor-Lösung fest. Auf Basis rein inhaltlicher Überlegungen ist jedoch eher eine 2-Faktoren-Lösung anzunehmen, da die Items analog der BEES auf einen positiven und einen negativen Faktor ausgerichtet konstruiert wurden. Darum wurde anschließend eine entsprechende explorative Faktorenanalyse mit festgelegter Faktorenanzahl  $n = 2$  durchgeführt. Abschließend wurden beide Faktorenlösungen mit einer konfirmatorischen Faktorenanalyse überprüft. Die gefundenen Ladungsmuster sowie die dazugehörigen Fit-Indizes sind in Abbildung 2 dargestellt.

Abbildung 2. Explorative Faktorenanalyse BEEPS



1 Faktor Lösung  
 Chi-Quadrat (9) = 49.989 ; p = 0.0001  
 RMSEA = 0.148  
 CFI = 0.893  
 TLI = 0.822  
 SRMR = 0.058  
 AIC = 2680.727  
 BIC = 2682.756 sample-size adjusted



2 Faktor Lösung  
 Chi-Quadrat (9) = 23.982 ; p = 0.002  
 RMSEA = 0.098  
 CFI = 0.958  
 TLI = 0.922  
 SRMR = 0.038  
 AIC = 2656.720  
 BIC = 2658.918 sample-size adjusted

Abbildung 2. Darstellung zweier möglicher Faktorenlösungen der BEEPS sowie der dazugehörigen Fit-Indizes und Faktorenladungen



## 5. Diskussion

Im Folgenden sollen die Ergebnisse mit Blick auf die Zielstellung eingeordnet und kritisch diskutiert werden.

### 5.1. Deskriptive Daten

Die empirisch erhobenen Mittelwerte und Standardabweichungen für BEES und BEEPS sind in Tabelle 2 und Tabelle 3 dargestellt. Ein Vergleich der Werte zwischen der neu erhobenen deutschen und in der Literatur angegebenen australischen Stichprobe zeigt, dass sie numerisch nah aneinander liegen. Dies gilt sowohl für den Gesamtscore als auch für die einzelnen Subscores beider Instrumente. Die ähnlichen Testwerte in den verschiedenen Stichproben weisen auf eine ähnliche Merkmalsverteilung zwischen den Stichproben hin, die statistische Hypothese A ist daher anzunehmen.

Die Maße BDI-II sowie DASS-21 erreichen in ihren Maxima Ausprägungen, die so interpretiert werden, dass in der deutschen Stichprobe einzelne Probanden Merkmale schwerer klinischer Depressionen aufweisen. Wie eingangs beschrieben sind medizinische Populationen übermäßig von depressiven Symptomen betroffen.

### 5.2. Objektivität

Ein Testverfahren ist laut Schmidt-Atzert und Amelang (2012) dann objektiv, wenn „die Ergebnisse eines diagnostischen Verfahrens unabhängig davon zustande kommen, wer die Untersuchung, die Auswertung und die Interpretation durchführt“. Die Objektivität wird unterteilt in Durchführungs-, Auswertungs- und Interpretationsobjektivität.

Im englischsprachigen Original sowie in der deutschsprachigen Übersetzung ist die BEES vollständig verschriftlicht und standardisiert. Es existieren eindeutige Auswertungsrichtlinien für das Errechnen der Summenscores. Jedoch gibt es keine Regel, wie die empirisch gefundenen Werte zu interpretieren sind: Der empirisch gefundene Punktrohwert und die errechneten Summenscores können zu keinem inhaltlichen Kriterium in Verbindung gebracht werden, sondern höchstens interpersonell verglichen werden. Da keine Interpretationsregeln oder irgendeine andere Form der Normierung existieren, bleibt unklar, was der gefundene Punktrohwert über die Versuchsperson aussagt.

Zusammenfassend lassen sich sowohl für die englisch- und als auch die deutschsprachige BEES und BEEPS Durchführungs- und Auswertungsobjektivität feststellen, jedoch in keiner Form Interpretationsobjektivität. Ohne Interpretationsregeln vonseiten der Autoren ist jede Interpretation von Testergebnissen immer bedingt durch den Testleiter. Das Ziel der Unabhängigkeit von Testinterpretation und Testleiter ist somit verfehlt. bDurch diese Probleme in der Interpretationsobjektivität erfüllen weder BEES noch BEEPS die notwendigen Anforderungen an die Objektivität, wie sie für psychometrische Testverfahren Standard sind.

### 5.3. Reliabilität

Die Reliabilität beschreibt den Grad der Messgenauigkeit und seine Beeinflussung durch unsystematische Fehler (Schmidt-Atzert & Amelang, 2012). Die Betrachtung der Reliabilität von BEES und BEEPS erfolgte durch Cronbachs Alpha und McDonalds Omega. Cronbachs Alpha für den BEES-Gesamtscore bewegt sich mit  $\alpha = 0.817$  im sehr guten, die der Subskalen BEES (+) mit  $\alpha = 0.715$  und BEES (-) mit  $\alpha = 0.71$  im guten Bereich. Keiner der  $\alpha$ -Werte würde sich durch das Weglassen eines Items verbessern.

Für den Gesamtscore der BEEPS beträgt  $\alpha = 0.807$ , für die Subskala BEEPS (+)  $\alpha = 0.768$ , für die Subskala BEEPS (-)  $\alpha = 0.687$ . Die Subskala BEEPS (+) würde beim Verzicht auf das erste Item „Gesund“ ein  $\alpha = 0.775$  aufweisen. Aufgrund der sehr geringen numerischen Differenz sowie der Kürze des Tests ist ein Itemausschluss nicht sinnvoll und würde zudem keinen ökonomischen Vorteil bringen. Im englischsprachigen Original beträgt  $\alpha = 0.83$  (Skead & Rogers, 2016). Die  $\alpha$ -Koeffizienten der verschiedensprachigen Stichproben sind sehr ähnlich, die statistische Hypothese B kann angenommen werden. Die in der Literatur postulierte und in dieser Arbeit anschließend überprüfte Faktorenstruktur der BEES ist eine zwei-Faktorenstruktur (Rogers, Cruickshank und Nosaka, 2020). Da dies die Unidimensionalitätsbedingung von Cronbachs Alpha verletzt und anhand der Faktorenladungen nicht von einer essenziellen Tau-Äquivalenz der Skala (Graham, 2006), erkennbar an den numerisch nicht identischen Faktorenladungen der einzelnen Items (Abbildung 1), ausgegangen werden kann (McNeish, 2017), soll die Reliabilität der Skala zusätzlich mit McDonalds Omega Total und Hierachical Omega betrachtet und überprüft werden (Kelly & Pornprasertmanit, 2016). Diese haben aktuell einen hohen Stellenwert im Bereich der Skalvalidierung (Garcia-Garzon & Garrido, 2020; Viladrich, Angulo-Brunet &

Doval, 2017; Zinbarg & Alden, 2015), da sie aufgrund realistischerer Voraussetzungen und geringerer Problemtendenzen gegenüber Cronbachs Alpha aussagekräftigere Reliabilitätskoeffizienten darstellen (Dunn, Baguley & Brunson, 2013).

Die Berechnungen für Omega der BEES ergaben bei festgesetzter Faktorenstruktur mit zwei Faktoren sowie einem Generalfaktor  $\omega_h = 0.66$  sowie  $\omega_t = 0.84$ . Da sowohl  $\alpha$  als auch  $\omega_t$  eine Tendenz zur Überschätzung der Reliabilität haben (Trizano-Hermosilla, Galvez-Nieto, Alvarado, Saiz & Salvo-Garrido, 2021), gibt  $\omega_h$  die fehlerfreiste Schätzung der Reliabilität an. Der numerisch kleiner ausgeprägte Reliabilitätskoeffizient  $\omega_h$  begründet sich in der Mehrfaktorenstruktur der BEES. Der deutlich verringerte  $\omega_h = 0.66$  gibt die realistischere interne Konsistenzschätzung gegenüber Cronbachs Alpha ab.

Analog zu diesem Vorgehen wurde die Reliabilität der BEEPS unter Verwendung der Omegastatistiken untersucht. Diese ergaben ein  $\omega_h = 0.8$  sowie ein  $\omega_t = 0.94$ .

Zusammenfassend lässt sich daraus ableiten, dass Cronbachs Alpha-Werte zwar numerisch hoch ausfallen, jedoch nur bedingt verwendbar sind, da die Voraussetzungen seiner Anwendung verletzt werden. Die Omega-Statistiken beschreiben die interne Konsistenz der BEES als akzeptabel, der BEEPS als gut.

Sowohl Cronbachs Alpha als auch McDonalds Omega erreichen gute bis sehr gute Werte. Aufgrund verletzter Anwendungsbedingungen von Cronbachs Alpha sind die Werte nur bedingt aussagekräftig. Weitere Reliabilitätsanalysen sind notwendig. Beispielhaft könnte die Retest-Reliabilität durch weitere Erhebungen erfasst werden. Dazu wäre eine Änderung des Paradigmas bzw. Fragebogens notwendig. Die Maße der internen Konsistenz, vor allem McDonalds Omega, belegen die BEES und die BEEPS als reliable Messinstrumente.

#### 5.4. Validität

Validität beschreibt die Übereinstimmung von Messintention und letztendlichem Messen. Ob die Items der BEES das Merkmal vollständig repräsentieren, muss anhand inhaltlicher Überlegungen beantwortet werden. Im Aufbau der BEES wird der Anspruch formuliert, PA und NA zu erfassen. Als theoretische Konzeption für die BEES werden die drei Subskalen der DASS-21 angeführt. Da die BEES für jede der drei Subskalen jeweils zwei Items enthält und insgesamt drei Items für PA und drei Items für NA existieren, kann rein auf Basis dieser inhaltlichen Betrachtung angenommen werden, dass Konstrukt sei vollständig repräsentiert. Diese

Schlussfolgerung ist jedoch stark limitiert. Eine reelle Überprüfung der Repräsentativität der Items für das Konstrukt kann methodisch-mathematisch nur durch faktorenanalytische oder item-response-theoriebasierte Ansätze erfolgen. Folgend dem analytischen Vorgehen beispielweise der Sedimentationshypothese des NEO-FFI oder der Konstruktionsweise des PANAS, hätten aus einer großen Itemmenge zunächst jene Items mit den stärksten Faktorenladungen identifiziert werden müssen, um die informationsreichsten Items zu extrahieren. So kann sichergestellt werden, dass das Konstrukt vollständig repräsentiert wird. Bei einer Itemauswahl, die den gegebenen Informationen folgend ad hoc und ohne Vorbereitung geschah, kann keine Repräsentativität angenommen werden. Welches Konstrukt zu welchem Anteil hier repräsentiert wird, kann anhand der Daten und Aussagen der Originalautoren der BEES nicht beantwortet werden. Durch die mangelnde Einbindung der BEES in einen theoretischen Hintergrund, keinerlei Definition des zu messenden Konstrukts und keiner trennscharfen Abgrenzung zu bestehenden Konstrukten bleibt im englischsprachigen Original unklar, was die BEES letztendlich genau misst. Damit kann keine inhaltliche oder Konstruktvalidität erfasst werden. Erneut deutlich erwähnt werden muss hier, dass 2016 die BEES zur Erfassung von Prüfungsangst konzeptioniert, anschließend zur Erfassung von PA und NA validiert, aber parallel von den gleichen Autoren zur Messung psychischer Belastung genutzt wurde. Zwar haben alle dieser Konstrukte inhaltliche Überschneidungen, sind jedoch nicht identisch.

Für eine Überprüfung der konvergenten und diskriminanten Validität sollen die Pearson Interkorrelationen zwischen der BEES, der DASS-21 und dem BDI-II sowie deren Subskalen betrachtet werden. Eine Übersicht über die gefundenen Korrelationen bietet Tabelle 4. Die Validität von BEES und BEEPS wurde in der vorliegenden Arbeit ausschließlich durch konvergente Validitäten betrachtet. Diese konvergenten Validitäten zu BDI-II und DASS-21 sind numerisch hoch und wurden bereits im englischsprachigen Original als Validitätsbelege gewertet. Diese Arbeitsweise ist stark limitiert. Korrelationen zwischen Maßen bedingen zum einen keine Kausalität, zum anderen ist vollkommen unklar, was genau korreliert und ob die gefundenen Korrelationen pure Zufälle sind.

Die konvergenten Validitäten zwischen BEES, DASS-21 und BDI-II erreichen mit  $r = -0.74$  und  $r = -0.753$  die in der statistischen Hypothese geforderte numerische Höhe. Die zwei Items der BEES, die jeweilig einer Subskala der DASS-21 zugeordnet sind, erzielen auf dieser Subskala die höchsten Korrelationskoeffizienten. Es fällt jedoch auf, dass die Korrelationen zu den

jeweilig anderen beiden Subskalen numerisch zwar kleiner, jedoch ähnlich ausfallen. Der Konzeption von PA und NA folgend, korreliert die BEES (neg.) hoch mit den Maßen der DASS-21 und des BDI-II. Dies wird als Zeichen hoher konvergenter Validität gewertet. Die statistische Hypothese C ist anzunehmen. Gleichzeitig korreliert die BEES (pos.) hoch negativ mit den Werten von DASS-21 und BDI-II. Dies wird als Beleg der diskriminanten Validität der BEES gewertet. Darauffolgend ist die statistische Hypothese D anzunehmen.

Bei Überlegungen zur Inhalts- und Konstruktvalidität der BEEPS treten die gleichen Probleme wie bereits bei der BEES auf. Durch eine mangelnde Einbettung der BEEPS in einen theoretischen Hintergrund ist das gemessene Konstrukt und eine Definition davon unklar. Es kann keine Aussage darüber getroffen werden, ob das Konstrukt vollständig repräsentiert ist. Es gibt seitens der Primärautoren keinerlei Angaben, was genau die BEEPS misst. Konstrukt- und Inhaltsvalidität kann damit weder überprüft noch bestätigt werden.

Für die Überprüfung konvergenter und diskriminanter Validitäten soll analog dem Vorgehen bei der BEES die Pearson Interkorrelationen zu DASS-21 und BDI-II betrachtet werden. Eine Übersicht über die gefundenen Korrelationen bietet Tabelle 5. Als Beleg einer konvergenten Validität könnte die positiven Korrelationen zwischen BEEPS (neg.) und der DASS (total) von  $r = .541$  und zwischen BEEPS (neg.) und dem BDI-II von  $r = .583$  angeführt werden. Diese Korrelationen sind aber auch Beispiel für die unklare Abgrenzung der BEEPS: Erfasst dieser Depression? Eine mögliche Interpretation wäre, dass die BEEPS die somatisch-affektiven Anteile von depressiven Symptomatiken erfassen kann, jedoch nicht die kognitiven. Die mangelnde Informationslage ermöglicht nur Spekulationen darüber, welches Konstrukt hier letztendlich überhaupt erfasst wird. Durch unklare theoretische Abgrenzungen und ebenfalls bedingt durch die „ad hoc“-anmutende Konstruktion von BEES sowie BEEPS kann keine valide Messung erfolgen.

Das Problem bei der Validitätsanalyse von BEES und BEEPS ist fundamentaler Art: Durch die mangelnde Einbettung der Skalen in theoretische Konstrukte fehlt die trennscharfe Abgrenzung gegenüber anderen Konstrukten und es bleibt unklar, was exakt BEES und BEEPS messen sollen. Erst wenn diese Frage eindeutig beantwortet wurde, entweder durch die Originalautoren oder durch nachfolgende Anwender, können Fragen nach Inhalts- oder Konstruktvalidität sinnvoll beantwortet werden. Da die Messintention von BEES und BEEPS nicht exakt genannt und nicht theoretisch abgegrenzt wird, können Validitätsbetrachtungen kaum sinnvoll durchgeführt werden.

## 5.5. Faktorenanalyse

Die BEES wurde entsprechend der in der Literatur postulierten Faktorenstruktur einer konfirmatorischen Faktorenanalyse unterzogen. Parallel dem Vorgehen von Rogers, Cruickshank und Nosaka (2020), wurden drei verschiedene Modelle mit festgelegter Faktorenanzahl berechnet und die modell-fit-Parameter verglichen. Den Richtlinien von Acock (2013) zufolge weist die zwei-Faktoren-Lösung die besten Parameter auf. Diese Lösung entspricht auch inhaltlich der Konzeption der BEES mit den Faktoren PA und NA. Zusammenfassend konnte die Faktorenstruktur der BEES des englischsprachigen Originals in der deutschen Übersetzung repliziert und bestätigt werden. Auffällig sind jedoch zwei Dinge: Erstens lädt das Item „Glücklich“ im 2-Faktoren-Modell ausschließlich auf den Faktor „Negativer Affekt“. Die Ladung ist damit gegensätzlich der Konstruktionsintention, da das Item eindeutig dem Faktor PA zugeordnet ist. Die Untersuchung von Rogers, Cruickshank und Nosaka (2020) findet ebenfalls ausschließlich Ladungen des Items „Glücklich“ auf den Faktor PA. Zweitens weisen die Items „Gelassen“ und „Selbstsicher“ neben den Hauptladungen auf den Faktor „Positiver Affekt“ Querladungen auf den Faktor „Negativer Affekt“ auf. Da die Items auf beide Faktoren laden, legt dies nahe, dass die Faktoren PA und NA nicht unabhängig voneinander sind. Der Anspruch von SPANE und PANAS war es, die beiden Faktoren orthogonal, dementsprechend unabhängig voneinander zu konzipieren. Da der BEES nach deren Vorbild konstruiert wurde, sollten auch hier in der Faktorenanalyse eine Unabhängigkeit von PA und NA gefunden werden. Die aus den empirisch erhobenen Daten gefundene Faktorenstruktur zeigt jedoch auf, dass die Faktoren nicht unabhängig voneinander sind (Borkenau, 2015).

Im Bereich der BEEPS gibt es seitens der Originalautoren keine Aussagen zur vermuteten Faktorenstruktur. Um einen ersten Überblick über potenzielle Faktorenstrukturen innerhalb der BEEPS zu bekommen, wurde eine exploratorische Faktorenanalyse berechnet. Diese sollte dazu dienen, die Anzahl der konsekutiven konfirmatorischen Faktorenanalysen zu begrenzen. Anschließend wurden konfirmatorische Faktorenanalysen mit festgelegter Faktorenanzahl von 1 und 2 berechnet und verglichen. Keines der beiden berechneten Modelle weist dabei die nach Acock geforderten Fit-Indizes auf. Daher müssen beide Modelle verworfen werden, die finale Faktorenstruktur der BEEPS ist damit weiterhin unklar.

## 5.6. Bewertung nach DIN

Die DIN Screen Checklisten von Kersting (2017) stellen standardisierte Anforderungen an die Bewertung von Testverfahren zur Erfassung menschlichen Erlebens und Verhaltens. Anhand dieser sollen BEES und BEEPS abschließend betrachtet werden. Die vollständige Checkliste findet sich im Anhang A2. Dort finden sich auch die Einzelantworten auf die jeweiligen Items der Checkliste.

Zusammenfassend wird festgestellt, dass der Einteilung des Diagnostik- und Testkuratoriums der Föderation deutscher Psychologen (2018) folgend, das Prädikat „nicht prüffähig“ für BEES und BEEPS ausgestellt werden muss, da die festgelegten Anforderungen bezüglich Information und Dokumentation nicht erfüllt sind.

Den Verfahren fehlen wesentliche Informationen und es erfolgt keine trennscharfe Einordnung in einen psychologischen Theorierahmen. Die BEES und BEEPS genügen nicht den Anforderungen des Diagnostik- und Testkuratoriums der Föderation deutscher Psychologen. Die Verfahren sind so mangelhaft mit Informationen untermauert und in keinen theoretischen Hintergrund trennscharf eingeordnet, dass sie keine schlechte Bewertung erhalten, sondern bereits die elementaren Anforderungen an irgendeine Bewertung nicht erfüllen.

## 5.7. Verwendbarkeit

Wie jeder Test haben BEES und BEEPS spezifische Vor- und Nachteile. Abhängig vom vorliegenden Einsatzzweck können diese mehr oder weniger ins Gewicht fallen. Ob die BEES bzw. BEEPS für den jeweiligen Einsatz geeignet ist, muss immer die verwendende Person entscheiden. Die vorliegende Arbeit mit der Bewertung soll Informationen für diese Einschätzung liefern.

Die größte Stärke der Skalen liegt in der Ökonomie. Mit jeweils nur sechs Items liegt die Bearbeitungsdauer im Sekundenbereich. Diese Kürze ist besonders wichtig in Erhebungen mit großen Stichproben und mehreren Messzeitpunkten. Rogers, Cruickshank und Nosaka konnten 2020 zeigen, dass durch die BEES beispielweise drei tägliche Messungen bei gleichzeitig geringer Belastung der Versuchspersonen möglich sind. Prinzipiell ist anzunehmen, dass bei geringerer Belastung der Versuchspersonen die Drop-out-Raten niedriger sind und Stichproben vollständiger werden (Denker, Schütte, Kersting, Weppert &

Stegt; 2023; Hoerger, 2010; Speer, King & Grossenbacher, 2016). Mit letzterem ist gemeint, dass bei Drop-out-Raten, bedingt durch die Belastungen der Erhebung, mit einer systematischen Verzerrung der Stichprobe zu rechnen ist, da ein bestimmter Teil der Stichprobe, zum Beispiel Personen mit niedriger Gewissenhaftigkeit, eher dazu tendiert, die Erhebung abubrechen.

Diese ökonomischen Vorteile müssen gegen bestehende Nachteile der Konzeption von BEES und BEEPS abgewogen werden. Im Vergleich zum PANAS oder SPANE sind die Verfahren bedeutend weniger eingebettet in theoretische Konstrukte, weniger verwendet und letztendlich weniger reliabel und valide. Wie unter 5.2. bis 5.4. dargestellt, können in den Bereichen der psychometrischen Qualitäten BEES und BEEPS vor allem im Vergleich zu bestehenden Testverfahren keine guten Bewertungen erzielen. Somit muss sich der Verwender der Verfahren mit der Frage auseinandersetzen, ob die zeitliche Ersparnis in ausgewogenem Verhältnis zu den damit einhergehenden Problemen steht. Deutlich sollte geworden sein, dass BEES und BEEPS für klinische Individualdiagnostik vollkommen ungeeignet sind. Für diesen Zweck wurden sie aber auch nie konstruiert.

Betrachtet man anschließend die geringe zeitliche Differenz zwischen dem Bearbeiten von BEES (6 Items) und dem SPANE (12 Items), aber gleichzeitig den großen qualitativen Unterschieden in den Konstruktionen, Validierungen und Normierungen, muss noch deutlicher werden, wie klein der verbleibende Einsatzbereich der BEES ist.

Die Verwendbarkeit der Skalen BEES und BEEPS kann jedoch erhöht werden: Beginnend mit expliziten Definitionen des Konstrukts, über neue Erhebungsmodalitäten, die beispielsweise neue Reliabilitätsbetrachtungen zulassen, bis hin zu Normierungen und externen Kriterien besteht die Möglichkeit, das theoretische und empirische Fundament der Skalen zu verbessern. Aber auch hier muss zunächst abgewogen werden, ob die notwendige Menge an Arbeit, bis BEES und BEEPS zu objektiven, reliablen und validen Instrumenten werden, in Relation steht zu dem zeitlichen Gewinn gegenüber der Verwendung des SPANE. Denn der SPANE erfüllt bereits jetzt diese Ansprüche, ist normiert, mannigfaltig verwendet und überprüft. Abschließend zeigt sich bezüglich BEES und BEEPS: Das spontane Konstruieren von Testverfahren für spezifische Einsatzzwecke ohne das Berücksichtigen der komplexen Grundlagen für Testkonstruktion, führt zu Messungen, bei denen unklar bleibt, was gemessen wird.



## 5.8. Limitationen

Im Bereich der Übersetzungsüberprüfung wäre es sinnvoll gewesen, die Probanden neben der neu übersetzten deutschen Variante von BEES und BEEPS ebenfalls die englische Originalvariante ausfüllen zu lassen. Korrelationen der Testgesamt- sowie Itemeinzelwerte zwischen englischer und deutscher Variante hätten genutzt werden können, um die Parallelität der beiden Testverfahren sprachunabhängig zu belegen.

Zur Betrachtung von Reliabilität in Form einer Retestreliabilität wäre eine weitere Erhebung der gleichen Stichprobe zu einem späteren Zeitpunkt sinnvoll gewesen. Da von den Probanden ein Versuchspersonencode gebildet wurde, wäre eine Zuweisung der Datenpunkte konkret zu den Versuchspersonen auch zu einem späteren Zeitpunkt möglich gewesen. Solche Messwiederholungen würden auch in engeren Zeitabständen die mögliche Änderungssensitivität der Skalen BEES und BEEPs erfassbar machen.

Validitätsuntersuchungen sollten ergänzt werden durch Multimethod-Multitrait-Methoden (Borkenau, 2015), um aussagestärkere konvergente und diskriminante Validitäten zu betrachten. Vor allem die diskriminanten Validitäten wurden in der vorliegenden Arbeit unzureichend untersucht.

Hier wäre insbesondere nützlich gewesen, weitere Testverfahren aus dem Bereich der PA und NA-Messung mit in das Untersuchungsdesign einfließen zu lassen. Die zusätzliche Verwendung von PANAS und SPANE im Fragebogen hätte größere Rückschlüsse über die letztendliche Verwendbarkeit von BEES und BEEPs ermöglicht. Durch eine breitere Aufstellung der Verwendung etablierter Verfahren wäre damit auch eine genauere Abgrenzung der BEES und BEEPS diesen gegenüber greifbar geworden.

## 6. Fazit

Zur ökonomischen und validen Erfassung von Belastungsphänomenen entwickelten die australischen Autoren Rogers, Barblett, Cruickshank und Robinson 2016 die Brief Emotional Experience Scale (BEES) sowie die Brief Emotional Physical Experience Scale (BEEPS). Die beiden englischsprachigen Instrumente wurden in der vorliegenden Arbeit in das Deutsche übersetzt und in dieser Form erstmalig einer deutschen Stichprobe appliziert. Anhand der empirisch gefundenen Daten und inhaltlicher Überlegungen wurden die psychometrischen Qualitätsmerkmale der beiden Skalen bewertet.

Beide Skalen werden dabei den Ansprüchen von psychometrisch-qualitativ-gesicherter Messung nicht gerecht. In ihren vorliegenden Formen sind weder die BEES noch die BEEPS in der Lage, objektiv, reliabel und valide zu messen. Gefundene Ergebnisse können nicht eindeutig interpretiert werden. Die Beziehung zwischen den Testrohwerten und dem eigentlich gemessenen Konstrukt der Belastungsphänomene ist nicht zweifelsfrei nachvollziehbar.

Für beide Skalen wird von der Verwendung abgeraten. Es existieren qualitativ höherwertige Instrumente zur Messung von Belastung, die eindeutig vorzuziehen sind. Verfahren wie die SPANE oder die PANAS sind theoretisch besser konstruiert und empirisch fundiert. Für diese Verfahren existieren Normierungen, eindeutige Interpretationsregeln und eine Vielzahl an Forschungsarbeiten und Artikeln, die ihre Validität bestätigen.

## 7. Literaturverzeichnis

- Acock, A. (2013). *Discovering structural equation modeling using Stata*. Stata Press
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334. doi: 10.1007/BF02310555
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC.
- Brandenburg, S. & Backhaus, N. (2015). *Zur Entwicklung einer deutschen Version der modified Differential Emotions Scale (mDES)*. Conference Paper. Johannes-Gutenberg-Universität Mainz.
- Brandstätter, V., Schüler, J., Puca, R. & Lozo, L. (2013). *Motivation und Emotion*. Berlin-Heidelberg: Springer.
- Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy*, *35*, 79–89.
- Borkenau, P. (2015). Vorlesung psychologische Diagnostik. Martin-Luther-Universität, Halle.
- Borkenau, P. & Ostendorf, F. (1994). *Neo-Fünf-Faktoren-Inventar nach Costa & McCrae*. Göttingen: Hogrefe.
- Center, C., Davis, M., Detre, T., Ford, D., Hansbrough, W., Hendin, H., Laszlo, J., Litts, D., Mann, J., Mansky, P., Michels, R., Proujansky, R., Reynolds, C. & Silverman, M. (2003). Confronting Depression and Suicide in Physicians. A Consensus Statement. *Journal of the Medical Association*, *18*, Vol. 289, 3161-3166.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

Cho, E. (2016). Making Reliability reliable: A systematic approach to Reliability Coefficients. *Organizational Research Methods*, 1-32. Doi: 10.1177/1094428116656239

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods*, 18(2), 207-230.

Clark, M., Canvin, L., Green, J., Layard, R., Pilling, S. & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *The Lancet*, 391. 679-686.

Clark, M. S., & Williamson, G. M. (1989). Mood and social judgements. In H. Wagner & A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 347–370). Chichester: Wiley.

Clark, L. A., & Watson, D. (1991). *Theoretical and empirical issues in differentiating depression from anxiety*. In: J. Becker & A. Kleinman (Eds.), *Psychosocial aspects of depression* (pp. 39–65). Hillsdale, NJ: Erlbaum.

Costantini, L., Pasquarella, C., Odone, A., Colucci, M., Costanza, A., Serafini, G., Aguglia, A., Murri, M., Brakoulias, V., Amore, M., Ghaemi, S. & Amerio, A. (2020). Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. *Journal of Affective Disorders*. doi: <https://doi.org/10.1016/j.ja>

Psychreport 2021, Entwicklung der psychischen Erkrankungen im Job: 2010-2020; („Daten der DAK Gesundheit 1997-2020“; Report, 2021)

Diagnostik- und Testkuratorium. (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 03. Jan. 2018. *Psychologische Rundschau*, 69, 109– 116.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-W., Oishi, S., & Biswas-Diener, R. (2009). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, 97, 143-156. <https://doi.org/10.1007/s11205-009-9493-y>

Dunn, T., Baguley, T. & Brunsten, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 1-13, doi:10.1111/bjop.12046

Ekman, P. (1982). Methods for measuring facial action. In: K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 45–90). Cambridge: Cambridge University Press.

Elosua, P. & Iliescu, D. (2012). Tests in Europe: Where we are and where we should go. *International Journal of Testing*, 12:2, 157-175. DOI:10.1080/15305058.2012.657316

Espejo, B., Checa, I., Perales-Puchalt, J. & Lisón, J. (2020). Validation and Measurement Invariance of the Scale of Positive and Negative Experience (SPANE) in a Spanish General Sample. *Int J Environ Res Public Health*. 12;17 doi: 10.3390/ijerph17228359.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160.

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J. L., Vos, T. & Whiteford, H. A. (2013). Burden of Depressive Disorder by Country, sex Age and Year: Findings from the Global Burden of Disease Study 2010. *Psychological Medicine*, 43, 471-481. Doi: 10.1017/S003/S0033291712001511

Frijda, N. (1993). The Place of Appraisal in Emotion.

Fox-Harding, C., Harris, S., Rogers, S., Vial, S., Beranek, P., Turner, M., & Cruickshank, T. (2021). A Survey to Evaluate the Association of COVID-19 Restrictions Perceived Mood and Coping in Australian Community Level Athletes. *Frontiers in Sports and Active Living*, 3. Doi: 10.3399/fspor.2021.624267

Garcia-Garzon, E. & Garrido, F. (2020). On Omega Hierarchical Estimation: A Comparison of Exploratory Bi-Factor Analysis Algorithms. *Multivariate Behavioral Research*. Doi: 10.1080/00273171.2020.1736977

Glischinski, M., Von Brachel, R. & Hirschfeld, G. (2019). How depressed is „depressed“? A systematic review and a diagnostic meta-analyses of optimal cut points for the Beck Depression-Inventory-II. *Quality of Life Research*, 28, 1111-1118.

Graham, J. (2006). Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability: What They Are and How to Use Them. *Educational and Psychological Measurement*. 66. doi: 10.1177/0013164406288165

Grobe, T. & Bessel, S. (2021). Gesundheitsreport Arbeitsunfähigkeiten. 20 Jahre Gesundheitsberichterstattung der Techniker Krankenkasse. Hamburg: Techniker Krankenkasse.

Gustavsson, A., Svensson, M., Jacobi, F., Allgulander, C., Alonso, J., Beghi, E., Dodel, R., Ekman, M., Faravelli, C., Fratiglioni, L., Gannon, B., Jones, D. H., Jenum, P., Jordanova, A., Jönsson, L., Karampampa, K., Knapp, M., Kobelt, G., Kurth, T., Lieb, R., Linde, M., Ljungcrantz, C., Maercker, A., Melin, B., Moscarelli, M., Musayev, A., Norwood, F., Preisig, M., Pugliatti, M., Rehm, J., Salvador-Carulla, L., Schlehofer, B., Simon, R., Steinhausen, H. C., Stovner, L. J., Vallat, J. M., Van den Bergh, P., van Os, J., Vos, P., Xu, W., Wittchen, H. U., Jönsson, B. & Olesen, J. (2010). Cost of disorders of the brain in Europe 2010. *European Neuropsychopharmacol*, 21(10), 718-79. doi: 10.1016/j.euroneuro.2011.08.008.

Hautzinger, M., Keller, F. & Kühner, C. (2006). *BDI-II Beck Depressions-Inventar-Revision*. München: Pearson.

Huang, C. & Chen, J. (2015). Meta-Analysis of the Structure of the Beck Depression-Inventory-II. *Assessment*, 22 (4), 459-472. doi: 10.1177/1073191114548873.

Herzberg, P. Y., Goldschmidt, S., & Heinrichs, N. (2008). TBS-TK Rezension: Beck Depressions-Inventar (BDI-II). Revision: Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen (2006). *R E P O R T Psychologie*, Fachwissenschaftlicher Teil, pp. 301–302.

Hoerger, M. (2010). Participant Dropout as a Function of Survey Length in Internet-Mediated University Studies: Implications for Study Design and Voluntary Participation in Psychological Research. *Cyberpsychology, Behavior and Social Networking*, 13. DOI: 10.1089=cyber.2009.0445

IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp

Isen, A. M. (2000). Positive affect and decision making. In: M. Lewis & J. Haviland-Jones (Eds.), *Handbook of Emotions* (pp. 417–435). New York: Guilford Press.

Izard C. E., Dougherty F. E., Bloxom B. M. and Kotsch N. E. (1974) *The Differential Emotions Scale: a Method of Measuring the Subjective Experience of Discrete Emotions*. Vanderbilt Univ. Press, Nashville, Tenn.

Jacobi, F., Klose, M. & Wittchen, H. U. (2004). Psychische Störungen in der deutschen Allgemeinbevölkerung: Inanspruchnahme von Gesundheitsleistungen und Ausfalltage. *Bundesgesundheitsblatt für Gesundheitsforschung und Gesundheitsschutz*, 47, 736-744  
- 47 -

Janke, W. & Debus, G. (1978). *Die Eigenschaftswörterliste (EWL). Handanweisung*. Göttingen: Hogrefe.

Jurkat, H. B., Bedau, S. & Leweke, F. (2010). Depressivität und Stressbewältigung bei Medizinstudierenden. *Der Nervenarzt*, *82*, 646-652.

Jovanović, V., Joshanloo, M., Martín-Carbonell, M., Caudek, C., Espejo, B., Checa, I., Krasko, J., Kyriazos, T., Piotrowski, J., Rice, S., Junça-Silva, A., Singh, K., Sumi, K., Tong, K., Yıldırım, M. & Žemojtel-Piotrowska, M. (2022). Measurement Invariance of the Scale of Positive and Negative Experience Across 13 Countries. *Assessment*, *29*(7), 1507-1521.  
doi: 10.1177/10731911211021494.

Kanter, P., Garrido, L., Moretti, L. & Medrano, L. (2021). A modern network approach to revisiting the Positive and Negative Affect Schedule (PANAS) construct validity. *Journal of Clinical Psychology*, *77*, 2370-2404. Doi: 10.1177/1073191114548873.

Keller, F., Hautzinger, M. & Kühner, C. (2008). Zur faktoriellen Struktur des deutschsprachigen BDI-II. *Zeitschrift für klinische Psychologie und Psychotherapie*, *37* (4), 245-254. Doi: 10.1026/1616-3443.37.4.245

Kelley, K. & Pornprasertmanit, S. (2016). Confidence Intervals for Population Reliability Coefficients: Evaluation of Methods, Recommendations, and Software for Composite Measures. *Psychological Methods*, *21*, 69-92. <http://dx.doi.org/10.1037/a0040086>

Kersting, M. (2018). Zur Information über und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens – Die DIN SCREEN Checkliste 1, Version 3  
In: Diagnostik- und Testkuratorium (Hrsg.) *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430* (S. 223-244).  
Berlin: Springer. (DOI 10.1007/978-3-662-53772-5)

Kersting, M. & Beauducel, A. (2022). Gewollt ist nicht gleich gekonnt. (Persönlichkeitsfragebögen in Trainings und Coachings). *Wirtschaft + Weiterbildung*, *2/2022*, 22-27.

Kindt, T., Rabkow, N., Pukas, L., Keuch, L., Sapalidis, A., Leskien, A., Röhler, J., Proyer, R. & Watzke, S. (2022). A Comparison of Depressive Symptoms in Medical and Psychology Students in Germany – Associations with Potential Risk and Resilience Factors. *Journal of Medical Psychology*, *24*. Doi: 10.3233/JMP-200023c

Kleinginna, P. & Kleinginna, A. (1981). A categorized List of Emotion Definitions, with Suggestions for a Consensual Definition. *Motivation and Emotion*, *Vol. 5, No. 4*, 345-379.  
<https://doi.org/10.1007/BF00992553>

Kühner, C., Keller, F., Schricker, I., Bedding, T., Huffziger, S., Timm, C., Rachota-Ubl, B., Hautzinger, M. & Diener, C. (2022). Diagnostische Performanz und Validität des deutschsprachigen BDI-II – Eine Sekundäranalyse aus klinischen und nichtklinischen Stichproben. *Psychiatrische Praxis (eFirst)*. DOI: 10.1055/a-1753-2298

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, 64, 241-256.

Li, F., Bai, X., & Wang, Y. (2013). The Scale of Positive and Negative Experience (SPANE): Psychometric properties and normative data in a large Chinese Sample. *PLoS ONE*, 8(4), e61137. <https://doi.org/10.1371/journal.pone.0061137>

Lovibond, P. F. (1998). Long-term stability of depression, anxiety, and stress syndromes. *Journal of Abnormal Psychology*, 107, 520–526.

Lovibond, S. H., & Lovibond, P. E. (1993). *Manual for the depression anxiety stress scales*. Wales: Psychology Foundation Monograph.

Lovibond, S. H. & Lovibond, P.E. (1995). The Structure of Negative Emotional States: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav. Res. Ther.*, 33,3, 335-343.

Marcon, G., Massaro Carneiro Monteiro, G., Ballester, P., Cassidy, R., Zimmerman, A., Brunoni, A., von Diemen, L., Hauck, S. & Passos, I. (2020). Who attempts suicide among medical students? *Acta Psychiatr Scand*. Mar;141(3):254-264. doi: 10.1111/acps.13137. Epub 2019 Dec 23. PMID: 31797353.

Martín-Carbonell, M., Checa, I., Fernández-Daza, M., Paternina, Y.& Espejo, B. (2021) Adaptation and Psychometric Properties of the Scale of Positive and Negative Experience (SPANE) in the General Colombian Population. *Int J Environ Res Public Health*,118(12). doi: 10.3390/ijerph18126449.

McNeish, D. (2017). Thanks Coefficient Alpha, We'll take it from here. *Psychological Methods*, Advance Online Publication. <http://dx.doi.org/10.1037/met0000144>

Malhi, G. S., & Mann, J. J. (2018). Depression. *Lancet*, 392, 2299-2312.

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical & Statistical Psychology*, 23, 1–21. Doi: 10.1111/j.2044-8317.1970.tb00432.x



McDonald, R. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.

McGaugh, J. L. (1992). Affect, neuromodulatory systems, and memory storage. In: S. A. Christianson (Ed.). *The handbook of emotion and memory: Research and theory* (pp. 245–268). Hillsdale, NJ: Erlbaum.

Moir, F., Yields, J., Sanson, J., & Chen, Y. (2018). Depression in medical students: current insights. *Advances in medical education and practice*, 9, 323–333.  
<https://doi.org/10.2147/AMEP.S137384>

Morris, W. & Schnurr, P. (1989). *Mood: The Frame of Mind*. New York: Springer.

Murray, C. & Lopez, Al. (1996). *The Global Burden of Disease A Comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge: The Harvard School of Public Health on Behalf of the World Health Organisation. Harvard University Press.

Pukas, L., Rabkow, N., Keuch, L., Ehring, E., Fuchs, S., Stoevesandt, D., Sapalidis, A., Pelzer, A., Rehnisch, C. & Watzke, S. (2022). Prävalenz und Risikofaktoren depressiver Symptome unter Medizinstudierenden in Deutschland – eine Querschnittsstudie. *GMS Journal for Medical Education*, 39 (1).

Puthran, R., Zhang, M., Tam, W. & Ho, R. (2016). Prevalence of depression amongst medical students: a meta-analysis. *Medical Education*, 50, 456-468. Doi: 10.1111/medu.12962

Rabkow, N., Pukas, L., Sapalidis, A., Ehring, E., Keuch, L., Rehnisch, C., Feußner, O., Klima, I. & Watzke, S. (2020). Facing the truth – A report on the mental health situation of German law students. *International Journal of Law and Psychiatry*, 71.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.  
URL <http://www.R-project.org/>.

Rehnisch, C., Feußner, O., Klima, I., Rabkow, N., Sapalidis, A., Ehring, E., Keuch, L. & Watzke, S. (2021). Depressive Symptome unter Zahnmedizinstudierenden – Prävalenz, Risikofaktoren und Resilienzfaktoren. *Deutsche Zahnärztliche Zeitschrift International*, 3, 266-275.

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology*, 67(3), 525–539.

Rogers, S., Barblett, L. & Robinson, K. (2016). Investigating the impact of NAPLAN on students, parent and teacher emotional distress in independent schools. *The Australian Education Researcher*. doi: 10.1007/s13384-016-0203-x

Rogers, S. L., Cruickshank, T., & Nosaka, K. (2020). The Brief Emotional Experience Scale (BEES) as a measure of positive and negative affect, *PsyArXiv*

Rotenstein, L. S., Ramons, M., Torre, M., Segal, B., Peluso, M., Guille, C., Sen, S. & Mata, D., (2016). Prevalence of Depression, Depressive Symptoms, and Suicidal Ideation Among Medical Students: A Systematic Review and Meta-Analysis. *JAMA*, 316 (21), 2214-2236. doi:10.1001/jama.2016.17324

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://www.jstatsoft.org/v48/i02/>.

Schermelleh-Engel, K. & Schweizer, K. (2003). Diskriminante Validität. In: K. D. Kubinger & Jäger, R. (Hrsg). *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim: Beltz.

Schmidt-Atzert, L. (1996). *Lehrbuch der Emotionspsychologie*. Stuttgart: Kohlhammer.

Schmidt-Atzert, L. & Amelang, M. (2006). *Psychologische Diagnostik*. Berlin Heidelberg New York: Springer Verlag.

Scholten, S., Velten, J., Bieda, A., Zhang, X. & Margraf, J. (2017). Testing Measurement Invariance of the Depression, Anxiety and Stress Scale (DASS-21) Across four Countries. *Psychological Assessment*, 29, 1376-1390. Doi: 10.1037/pas0000440

Senol-Durak, E., & Durak, M. (2019). Psychometric properties of the Turkish version of the Flourishing Scale and the Scale of Positive and Negative Experience. *Mental Health, Religion & Culture*, 22(10), 1021-1032. <https://doi.org/10.1080/13674676.2019.1689548>

Skead, N., Rogers, S. & Doraisamy, J. (2018). Looking beyond the mirror: Psychological distress; disordered eating, weight and shape concerns; and a maladaptive eating habit in lawyers and law students. *International Journal of Law and Psychiatry*, 61, 90-102. <https://doi.org/10.1016/j.ijlp.2018.06.002>

Skead, N., Rogers, S. & Johnson, W. (2020). The role of place, people and perception in law student well-being. *International Journal of Law and Psychiatry*, 73. <https://doi.org/10.1016/j.ijlp.2020.101631>

Sparfeldt, J., Becker, N., Greiff, S., Kersting, M., König, C., Lang, J. & Beauducel, A. (2022). Intelligenz(tests) verstehen und missverstehen. *Psychologische Rundschau*, 73 (3), 161-172.

Spitzer R., Kroenke, K. & Williams, J. B. (1999). The Patient Health Questionnaire Primary Care Study Group. Validation and Utility of a Self-report Version of PRIME-MD: The PHQ Primary Care Study. *JAMA*. 1999;282(18):1737–1744. doi:10.1001/jama.282.18.1737

Staples, L. G., Dear, B. F., Gandy, M., Fogliati, V., Fogliati, R., Karin, E., Niessen, O., & Titov, N. (2019). Psychometric properties and clinical utility of brief measures of depression, anxiety, and general distress: The PHQ-2, GAD-2, and K-6. *General Hospital Psychiatry*, 56, 13-18. <https://doi.org/10.1016/j.genhosppsy.2018.11.003>

Streiner, D. (2003). Starting at the beginning an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99-103.

Titov, N., Dear, B., Staples, L., Bennet-Levy, J., Klein, B. & Rapee, B. (2016). The first 30 months of the MindSpot Clinic: evaluation of a national e-mental health service against project objectives. *The Australian and New Zealand journal of psychiatry*, 51, 1227-1239.

Turner, M., Beranek, P., Rogers, S., Nosaka, K., Girard, O. & Cruickshank, T. (2021). Influence of the COVID-19 Pandemic on Mood and Training in Australian Community Tennis Players. *Frontiers in Sports and Active Living*, 3, 1-8. Doi: 10.3399/fspor.2021.589617#

Üstün, T.B., Ayuso-Mateos, J.L., Chatterji, S., Mathers, C. & Murray, C.J.L. (2000). Global Burden of depressive Disorders in the year 2000. *British Journal of Psychiatry*, 184, 386-392.

Viladrich, C., Angulo-Brunet, A. & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Annals of Psychology*, 33 (3), 755-782. Doi: 10.6018/analesps.33.3.268401

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96, 465–490.

Watson, D. & Clark, L. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.

Wang, Y. P. & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Braz J Psychiatry*, 35(4):416-31. doi: 10.1590/1516-4446-2012-1048

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of behavior-related social information. *Journal of Personality and Social Psychology*, 78, 1024–1037.

Wittchen, H. U. & Hoyer, J. (2006). *Klinische Psychologie und Psychotherapie*. Berlin Heidelberg New York: Springer Verlag.

Wittchen, H.U, Zaudig, M. & Fydrich, T. (1997). *SKID. Strukturiertes Klinisches Interview für DSM-IV. Achse I und II. Handanweisung*. Göttingen: Hogrefe.

World Health Organisation (2004). *The Global Burden of Disease. Update*. Geneva: World Health Organisation.

World Health Organisation WHO (2017) *Depression and Other Common Mental Disorders: Global Health Estimates*. Geneva: World Health Organization; Licence: CC BY-NC-SA 3.0 IGO.

Wottawa, H. (2002). Einige wichtige Entwicklungen der psychologischen Diagnostik im letzten Jahrzehnt. *Psychologie in Österreich*, 2-3, 3.5.

Wundt, W. (1905). *Grundriss der Psychologie*. Leipzig: Engelmann.

## 8. Thesen

1. Die BEES und BEEPS sind neu konstruierte Screening-Instrumente zur Erfassung von psychischen Belastungen.
2. Die Verfahren wurden ins Deutsche übersetzt und einer Stichprobe von n = 208 Medizinstudenten der Martin-Luther-Universität Halle Saale appliziert.
3. Die neu übersetzten Verfahren werden hinsichtlich theoretischer Konzeption, Objektivität, Reliabilität, Validität, Faktorenstruktur und Nützlichkeit unter Verwendung der DIN-Checkliste überprüft.
4. Die theoretische Konzeption beider Verfahren im englischsprachigen Original ist unzureichend, die genaue Messintention nicht trennscharf beschrieben.
4. Beide Verfahren weisen Durchführungs- und Auswertungsobjektivität auf, sind jedoch nicht interpretationsobjektiv.
5. Die anhand von Cronbachs Alpha und McDonalds Omega bestimmte interne Konsistenz als Schätzer der Reliabilität ist für beide Verfahren hoch.
6. Die anhand konvergenter und diskriminanter Validität zu BDI-II und DASS-21 betrachtete Validität für die BEES fällt hoch aus. Für die BEEPS liegen keine verlässlichen Validitätsschätzer vor.
7. Für die BEES konnte die Zwei-Faktoren-Struktur des Originals in der deutschsprachigen Übersetzung belegt werden, die Faktorenstruktur der BEEPS ist unklar.
8. Die Verwendung von BEES und BEEPS in ihrer vorliegenden Form kann nicht empfohlen werden. Die Nachteile ihrer Verwendung aufgrund schlechter Konzeption und mangelnder Validitätsbelege stehen in unzureichenden Verhältnis gegenüber ihrer hohen Ökonomie.

## Anhang

### A1 Vollständiger Fragebogen



Universitätsklinikum  
Halle (Saale)

Klinik und Poliklinik für  
Psychiatrie, Psychotherapie  
und Psychosomatik

## **Teilnehmerinformation zur Studie**

### **Belastungserscheinungen und depressive Symptome unter Studierenden der Martin-Luther-Universität Halle Saale**

Liebe Studierende,

als angehende Mediziner\*innen sind Sie während Ihres Studiums und auch im späteren Berufsalltag zahlreichen Risikofaktoren für Belastungserscheinungen bis hin zu depressiven Erkrankungen ausgesetzt (z.B. Stress, Leistungsdruck sowie verschiedene Belastungssituationen). Aktuelle Studiendaten weisen darauf hin, dass Medizinstudierende ein höheres Ausmaß an Stress sowie depressiver Symptome gegenüber der Allgemeinbevölkerung berichten. Die derzeit geltenden, zahlreichen Maßnahmen zur Eindämmung des Coronavirus SARS-CoV-2 haben einen zusätzlichen starken Einfluss auf die Lebensführung in allen Bereichen und somit auf das individuelle Wohlbefinden und Stresserleben.

#### Studienziel

Das Hauptziel der vorliegenden Studie ist die Erfassung der Prävalenz von Belastungs- und Depressionssymptomen unter den Medizinstudenten der Martin-Luther-Universität. Die Studie ist Teil einer seit 2017 stattfindenden Studienreihe. Bereits in den vergangenen Jahren wurden hierzu Fragebögen zu Belastungserscheinungen bei Studierende verschiedener Studienabschnitte der Medizin appliziert. Die Studienreihe strebt Nachuntersuchungen im jährlichen Turnus an.

#### Studienumfang

Der vorliegende Fragebogen enthält Fragen Ihrer persönlichen Lebenssituation, zu Ihrem Medizinstudium und vor allem zu aktuell auftretenden Belastungserscheinungen. Des Weiteren wird ein Teil Ihrer Persönlichkeitsstruktur, konkret Ihre emotionale Belastbarkeit erfasst.

Die Belastungserscheinungen werden hierbei über verschiedene Instrumente erfasst, daher können Ihnen manche Fragen redundant erscheinen.

#### Datenschutz

Die vorliegende Erhebung ist vollständig anonymisiert. Weder Versuchsleiter\*innen noch anderen Teilnehmer\*innen ist es möglich, konkrete Daten mit einer konkreten Versuchsperson in Verbindung zu bringen. Um eine anonyme Teilnahme und Rückgabe zu gewährleisten, stehen ausgewiesene Rückgabekästen im Lehrsekretariat zur Verfügung. Die Rückgabe der ausgefüllten Bögen erfolgt im verschlossenen Umschlag.

Die Daten aller befragten Personen werden zentral an der Klinik und Poliklinik für Psychiatrie, Psychotherapie und Psychosomatik des Universitätsklinikums Halle gesammelt. Die Speicherung und Auswertung der Informationen erfolgt so, dass statt Ihres Namens ein **anonymer Studienteilnehmercode** verwendet wird.

Die Vorschriften über die ärztliche Schweigepflicht und den Datenschutz werden somit im Rahmen der Studie eingehalten.

#### Seitenzahl

### Freiwilligkeit

Die **Teilnahme** an dieser Untersuchung erfolgt **freiwillig**. Sie haben zu jedem Zeitpunkt das Recht, Ihre Einwilligung für die Durchführung der Untersuchung insgesamt zu widerrufen. Ihre Entscheidung, ob Sie an der Studie teilnehmen oder nicht hat keinerlei Auswirkungen auf Ihr Studium oder sonstige Kontakte zu unseren Einrichtungen.

Es ist jedoch darauf hinzuweisen, dass bei einem nachträglichen Widerruf der Teilnahme eine Löschung bereits erfasster Daten aufgrund der anonymisierten Studienteilnahme nicht mehr möglich ist.

### Persönliche Bitte

Sollten Sie bemerken, dass **Sie unter einer übermäßigen Belastung und/oder depressiven Stimmung leiden**, möchten wir Sie bitten, **Ihren Hausarzt / Ihre Hausärztin** zu kontaktieren oder sich mit verfügbaren Hilfsangeboten in Verbindung zu setzen. Sie können hierfür folgende Wege nutzen:

**Info-Telefon der Deutschen Depressionshilfe: 0800 / 33 44 533** – hier werden Ihnen Kontaktadressen und Hilfsangebote Ihrer Region genannt.

Die **Ambulanz unserer Klinik für Psychiatrie, Psychotherapie und Psychosomatik** erreichen Sie telefonisch unter **0345 / 557 3640**, persönliche Ansprechpartner sind rund um die Uhr da.

**Natürlich können Sie sich auch per E-Mail an den Studienleiter Prof. Dr. Watzke wenden ([stefan.watzke@uk-halle.de](mailto:stefan.watzke@uk-halle.de)).**

Wenn Sie Fragen zu der vorliegenden Erhebung haben, kommen Sie bitte auf uns zu.

Mit bestem Dank für Ihr Interesse.

Die Studienleitung

# Belastungserscheinungen und depressive Symptome unter Studierenden der Martin-Luther-Universität Halle Saale

Ihre Gesundheit als angehende Ärztinnen und Ärzte ist uns wichtig.

Wir wissen, dass Sie großen Belastungen im Rahmen Ihres Medizinstudiums ausgesetzt sind und sich daraus Risiken u.a. für depressive Erkrankungen ergeben können.

Um Sie zukünftig besser unterstützen zu können, bitte wir Sie, die folgenden Fragebögen auszufüllen und in die bereitstehenden Boxen einzuwerfen.

Natürlich ist Ihre Teilnahme an der Studie freiwillig und vollständig anonymisiert.

Um die Daten Ihres Fragebogens besser mit potentiellen Folgebefragungen verknüpfen zu können, ohne dass jemand herausfinden kann, wer die Fragen beantwortet hat, bitten wir Sie um die Erstellung eines persönlichen Codes:

- Erster und letzter Buchstabe des Vornamens Ihrer Mutter      \_\_\_ \_\_\_
- Tag des Geburtsdatums Ihrer Mutter (zweistellig)      \_\_\_ \_\_\_
- Erster und zweiter Buchstabe des Geburtsortes Ihrer Mutter      \_\_\_ \_\_\_

Beispiel: der Code einer Person, deren Mutter **M**argrit heißt und die am **09.07.1951** in **H**amburg geboren wurde, würde z.B. lauten: **MT09HA**)

**Wenn Sie nicht an unserer Studie teilnehmen möchten, können Sie uns Gründe dafür mitteilen:**

---

---

---

---



# Fragebogen zur Selbstbeschreibung FRAGEN

# SOZIODEMOGRAPHISCHE

# BELASTUNGEN & RESILIENZFAKTOREN

PNUM: (wird durch das Studienteam vergeben)

Datum:

In welchem Fachsemester studieren Sie derzeit?

Bitte beantworten Sie jeden der folgenden Punkte durch Ankreuzen der zutreffenden Alternative bzw. durch Eintragen des zutreffenden Wertes

### 1. Biologisches Geschlecht

weiblich   
männlich

### 3. Alter in Jahren

### 2. Soziales Geschlecht

weiblich   
männlich   
beides   
weder noch  
anders

### 4. Größe in cm

### 5. Gewicht in kg

### 6. Zivilstand

ledig   
verheiratet   
getrennt lebend oder geschieden   
verwitwet

### 7. Partnerschaft unabhängig vom Zivilstand

kein Partner oder Beziehung kürzer  
als 3 Monate / ohne Partner lebend   
verheiratet und zusammenlebend   
nicht verheiratet, aber zusammenlebend   
Partnerschaft ohne Zusammenleben  
(> 3 Monate)

### 8. Dauer der Partnerschaft in Monaten

(0 bei fehlender Partnerschaft)

### 9. Anzahl leiblicher Kinder

### 10 In welchem Bundesland sind Sie aufgewachsen?

Baden-Württemberg   
Bayern   
Berlin   
Brandenburg   
Bremen   
Hamburg   
Hessen   
Mecklenburg-Vorpommern

Niedersachsen   
Nordrhein-Westfalen   
Rheinland-Pfalz   
Saarland   
Sachsen   
Sachsen-Anhalt   
Schleswig-Holstein   
Thüringen

### 11 Welchen höchsten schulischen Abschluss hat Ihre Mutter?

Ohne Schulabschluss   
Hauptschule   
Realschule (10. Klasse)   
Abitur

### 12. Welchen höchsten Berufsabschluss hat Ihre Mutter?

ungelernt   
Berufsausbildung   
Fach-/Hochschulstudium

**13 Welcher Art Tätigkeit geht Ihre Mutter aktuell nach?**

Ohne berufliche Tätigkeit	<input type="checkbox"/>
angestellt	<input type="checkbox"/>
selbständig	<input type="checkbox"/>

**14. Welchen höchsten schulischen Abschluss hat Ihr (sozialer) Vater?**

Ohne Schulabschluss	<input type="checkbox"/>
Hauptschule	<input type="checkbox"/>
Realschule (10. Klasse)	<input type="checkbox"/>
Abitur	<input type="checkbox"/>

**15 Welchen höchsten Berufsabschluss hat Ihr (sozialer) Vater?**

ungelernt	<input type="checkbox"/>
Berufsausbildung	<input type="checkbox"/>
Fach-/Hochschulstudium	<input type="checkbox"/>

**16. Welcher Art Tätigkeit geht Ihr (sozialer) Vater aktuell nach?**

Ohne berufliche Tätigkeit	<input type="checkbox"/>
angestellt	<input type="checkbox"/>
selbständig	<input type="checkbox"/>

**17 Wie viele Geschwister haben Sie?**

**18. In welcher Geschwisterposition stehen Sie?**

(Geben Sie eine 1 an, wenn Sie das älteste (oder einzige) Kind Ihrer Eltern sind; eine 2, wenn Sie das zweitälteste sind etc.)

**19 Haben Sie ein Elternteil durch Trennung oder Tod verloren?**

nein	<input type="checkbox"/>
Trennung der Eltern	<input type="checkbox"/>
Tod der Mutter	<input type="checkbox"/>
Tod des Vaters	<input type="checkbox"/>

**20. Bei wem sind Sie aufgewachsen?**

bei beiden leiblichen Eltern	<input type="checkbox"/>
bei der Mutter aufgewachsen	<input type="checkbox"/>
beim Vater aufgewachsen	<input type="checkbox"/>
mit Mutter/Vater und neuer Partnerin / Partner	<input type="checkbox"/>
bei Pflegeeltern aufgewachsen	<input type="checkbox"/>

**21 Welche Bedeutung spielte Religion in Ihrer Erziehung?**

keine Bedeutung	<input type="checkbox"/>
marginale Bedeutung	<input type="checkbox"/>
mittelmäßige Bedeutung	<input type="checkbox"/>
extrem wichtig	<input type="checkbox"/>

**22. Welche Bedeutung spielt Religion in Ihrem heutigen Leben?**

keine Bedeutung	<input type="checkbox"/>
marginale Bedeutung	<input type="checkbox"/>
mittelmäßige Bedeutung	<input type="checkbox"/>
extrem wichtig	<input type="checkbox"/>

**23 War / ist eines Ihrer Familienmitglieder aufgrund einer psychischen Erkrankung in Behandlung?**

	Mutter	Vater	Geschwister	Großeltern
weiß ich nicht	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
nein	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Demenz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Abhängigkeit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Psychose	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bipolare Störung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Depression	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Angststörung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Essstörung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Persönlichkeitsstörung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**24. Waren oder sind Sie selbst wegen einer psychischen Erkrankung in Behandlung?**

nein	<input type="checkbox"/>
Störungen im Kindesalter	<input type="checkbox"/>
Abhängigkeit	<input type="checkbox"/>
Psychose	<input type="checkbox"/>
Bipolare Störung	<input type="checkbox"/>
Depression	<input type="checkbox"/>
Angststörung	<input type="checkbox"/>
Essstörung	<input type="checkbox"/>
Persönlichkeitsstörung	<input type="checkbox"/>

---

## Fragen zu Ihrem Studium

---

### 25 Berufsausbildung oder Studium

#### . vor dem Medizin-Studium

keine	<input type="checkbox"/>
begonnen, aber nicht abgeschlossen	<input type="checkbox"/>
abgeschlossen	<input type="checkbox"/>

### 26. Art der Ausbildung

keine	<input type="checkbox"/>
angelernt oder Teilfcharbeiter	<input type="checkbox"/>
Lehre in Handwerk, Produktion, kaufmännischem Bereich	<input type="checkbox"/>
Fachschul Ausbildung, Beamtenausbildung	<input type="checkbox"/>
Hochschul Ausbildung, Fachhochschulstudium	<input type="checkbox"/>

---

<p><b>27 Tätigkeit vor Aufnahme des Medizin - Studiums</b></p> <p>schulische Ausbildung <input type="checkbox"/></p> <p>berufliche Ausbildung <input type="checkbox"/></p> <p>anderes Hochschulstudium <input type="checkbox"/></p> <p>berufliche Tätigkeit <input type="checkbox"/></p> <p>FSJ oder Auslandsjahr <input type="checkbox"/></p> <p>keine Tätigkeit <input type="checkbox"/></p>	<p><b>28. Dauer der letzten Tätigkeit</b> in Jahren <input type="text"/></p>
<p><b>29 Was waren für Sie die Hauptgründe, sich für das Medizin -Studium zu entscheiden?</b> (Bitte wählen Sie die wichtigsten zwei)</p> <p>Ich fand das Fachgebiet spannend <input type="checkbox"/></p> <p>Ich wollte Menschen helfen <input type="checkbox"/></p> <p>Ich suchte nach einem Beruf, der mir hohen Respekt und Anerkennung verschafft <input type="checkbox"/></p> <p>Ich suchte nach einem Beruf mit guten Verdienstmöglichkeiten und Zukunftssicherheit <input type="checkbox"/></p> <p>Meine Eltern / Familie haben mir das Studium nahegelegt <input type="checkbox"/></p> <p>Ich wusste nicht, was ich sonst hätte studieren sollen <input type="checkbox"/></p>	<p><b>30. Gab es für Sie einen denkbare alternativen Berufs- oder Studienwunsch?</b> ja <input type="checkbox"/> nein <input type="checkbox"/></p> <p><b>31. Wie viele Jahre haben Sie auf den Beginn des Studiums warten müssen bzw. haben Sie anderweitig überbrückt?</b> <input type="text"/></p> <p><b>32. An wievielter Stelle war Halle unter Ihren Studienwunschornten?</b> <input type="text"/></p>
<p><b>33 Macht Ihnen Ihr Studium Spaß?</b></p> <p>ja, fast immer <input type="checkbox"/></p> <p>meist <input type="checkbox"/></p> <p>selten <input type="checkbox"/></p> <p>nein, fast nie <input type="checkbox"/></p>	<p><b>34. Sind Sie sicher, Ihr Studium erfolgreich abzuschließen?</b></p> <p>ja, fast immer <input type="checkbox"/></p> <p>meist <input type="checkbox"/></p> <p>selten <input type="checkbox"/></p> <p>nein, fast nie <input type="checkbox"/></p>
<p><b>35 Würden Sie sich aus heutiger Sicht wieder für das Medizin -Studium entscheiden?</b></p> <p>ich bin unglücklich über meine Entscheidung <input type="checkbox"/></p> <p>ich würde mich heute anders entscheiden die Perspektive des späteren Berufes lässt mich die Hürden überwinden <input type="checkbox"/></p> <p>ich würde mich auf jeden Fall wieder so entscheiden <input type="checkbox"/></p>	<p><b>36. Gibt es Dinge, die Sie es Ihnen aktuell schwer machen, mit Ihrer Studien-Entscheidung glücklich zu sein?</b> (Mehrfachnennung möglich)</p> <p>unsichere Zukunftsperspektive <input type="checkbox"/></p> <p>Konkurrenz unter den Studierenden <input type="checkbox"/></p> <p>wenig Flexibilität in der Studiengestaltung <input type="checkbox"/></p> <p>Zeitnot <input type="checkbox"/></p> <p>einzelne Lehrveranstaltungen sind sehr belastend <input type="checkbox"/></p> <p>finanzielle Belastung durch das Studium <input type="checkbox"/></p> <p>Überforderung <input type="checkbox"/></p> <p>Einsamkeit <input type="checkbox"/></p> <p>Leistungsdruck <input type="checkbox"/></p>
<p><b>37 Wie wohnen Sie aktuell?</b></p> <p>allein / eigene Wohnung <input type="checkbox"/></p> <p>Wohnung mit Partnerin / Partner <input type="checkbox"/></p> <p>in einer WG <input type="checkbox"/></p> <p>Zimmer im Studentenwohnheim <input type="checkbox"/></p> <p>bei den Eltern <input type="checkbox"/></p>	<p><b>38. Wenn Sie für Ihr Studium umgezogen sind, wie nehmen Sie den Wohnortwechsel wahr?</b></p> <p>Positiv <input type="checkbox"/></p> <p>Neutral <input type="checkbox"/></p> <p>Negativ <input type="checkbox"/></p> <p>Mein Wohnort hat sich nicht verändert <input type="checkbox"/></p>

<p><b>39. Wie finanzieren Sie Ihr Studium?</b> (Mehrfachnennung möglich)</p> <p>eigene berufliche Tätigkeit / Nebenjob <input type="checkbox"/></p> <p>finanzielle Unterstützung durch Familie <input type="checkbox"/></p> <p style="padding-left: 100px;">BAföG <input type="checkbox"/></p> <p style="padding-left: 100px;">Stipendium <input type="checkbox"/></p> <p style="padding-left: 100px;">Sozialleistungen des Staates <input type="checkbox"/></p>	<p><b>40. Stehen Ihnen ausreichende finanzielle Mittel zur Verfügung?</b></p> <p>mehr als genügend <input type="checkbox"/></p> <p>ausreichend <input type="checkbox"/></p> <p>manchmal zu wenig <input type="checkbox"/></p> <p>oft zu wenig <input type="checkbox"/></p> <p>ich stehe meist unter großem finanziellen Druck <input type="checkbox"/></p>
<p><b>41. Haben Sie seit Beginn des Studiums mehr oder weniger Kontakt zu Freunden?</b></p> <p>mehr als zuvor <input type="checkbox"/></p> <p>unverändert viel <input type="checkbox"/></p> <p>unverändert wenig <input type="checkbox"/></p> <p>weniger als zuvor <input type="checkbox"/></p>	<p><b>42. Haben Sie seit Beginn des Studiums mehr oder weniger Kontakt zu Ihrer Familie?</b></p> <p>mehr als zuvor <input type="checkbox"/></p> <p>unverändert viel <input type="checkbox"/></p> <p>unverändert wenig <input type="checkbox"/></p> <p>weniger als zuvor <input type="checkbox"/></p>
<p><b>43. Erfahren Sie ausreichend emotionale Unterstützung?</b></p> <p>Ich bekomme <b>mehr als genügend</b> emotionale Unterstützung <input type="checkbox"/></p> <p>Ich bekomme <b>ausreichend</b> emotionale Unterstützung <input type="checkbox"/></p> <p>Ich bekomme <b>manchmal zu wenig</b> emotionale Unterstützung <input type="checkbox"/></p> <p>Ich bekomme <b>oft zu wenig</b> emotionale Unterstützung <input type="checkbox"/></p> <p>Ich bekomme <b>keine</b> emotionale Unterstützung <input type="checkbox"/></p>	<p><b>44. Wie viele Stunden pro Woche verbringen Sie ca. mit studienassoziierten Tätigkeiten (Lernen, Referatvorbereitung, Prüfungsvorbereitung) außerhalb des Universitätsalltags?</b> <input style="width: 100px; height: 20px;" type="text"/></p> <p><b>45. Wie viele Stunden pro Woche verbringen Sie ca. mit Tätigkeiten im Rahmen eines Nebenjobs? (0 wenn Sie keinem Nebenjob nachgehen)</b> <input style="width: 100px; height: 20px;" type="text"/></p>
<p><b>46. Wie viele Stunden pro Woche verbringen Sie ca. mit Ihren Freunden, Ihrer Familie oder Ihrer Partnerin / Ihrem Partner?</b> <input style="width: 100px; height: 20px;" type="text"/></p>	<p><b>47. Wie viele Stunden pro Woche verbringen Sie ca. mit Ihren Hobbies?</b> <input style="width: 100px; height: 20px;" type="text"/></p>
<p><b>48. Wie viele Stunden pro Woche treiben Sie aktiv Sport?</b> <input style="width: 100px; height: 20px;" type="text"/></p>	<p><b>49. Wie viele Stunden pro Woche musizieren Sie aktiv?</b> <input style="width: 100px; height: 20px;" type="text"/></p>
<p><b>50. Wie bewerten Sie die Ihnen für das Studium zur Verfügung stehende Zeit?</b></p> <p>Ich habe stets <b>mehr als genügend Zeit</b> für die Anforderungen des Studiums <input type="checkbox"/></p> <p>Ich habe stets <b>ausreichend Zeit</b> für die Anforderungen des Studiums <input type="checkbox"/></p> <p>Ich habe <b>manchmal zu wenig Zeit</b> für die Anforderungen des Studiums <input type="checkbox"/></p> <p>Ich habe <b>oftmals zu wenig Zeit</b> für die Anforderungen des Studiums <input type="checkbox"/></p> <p>Ich stehe <b>meist unter Zeitdruck</b>, um den Anforderungen des Studiums zu genügen <input type="checkbox"/></p>	<p><b>51. Wie bewerten Sie die Ihnen für Freunde, Familie und Partnerschaft zur Verfügung stehende Zeit?</b></p> <p>Ich habe stets <b>mehr als genügend Zeit</b> für Freunde, Familie und Partnerschaft <input type="checkbox"/></p> <p>Ich habe stets <b>ausreichend Zeit</b> für Freunde, Familie und Partnerschaft <input type="checkbox"/></p> <p>Ich habe <b>manchmal zu wenig Zeit</b> für Freunde, Familie und Partnerschaft <input type="checkbox"/></p> <p>Ich habe <b>oftmals zu wenig Zeit</b> für Freunde, Familie und Partnerschaft <input type="checkbox"/></p> <p>Ich stehe <b>meist unter Zeitdruck</b>, um Freundschaften, familiäre Beziehungen und Partnerschaft zu pflegen <input type="checkbox"/></p>

**52. Wie bewerten Sie die Ihnen für Ihre Hobbies zur Verfügung stehende Zeit?**

- Ich habe stets **mehr als genügend Zeit** für meine Hobbies
- Ich habe stets **ausreichend Zeit** für die Anforderungen des Studiums
- Ich habe **manchmal zu wenig Zeit** für meine Hobbies
- Ich habe **oftmals zu wenig Zeit** für meine Hobbies
- Ich stehe **meist unter Zeitdruck**, um meinen Hobbies nachgehen zu können

**53. Wie zufrieden sind Sie insgesamt mit ...**

- |  | sehr zufrieden           |                          | sehr unzufrieden         |                          |
|--|--------------------------|--------------------------|--------------------------|--------------------------|
| ... Ihrem Studium                          | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| ... Ihrer Partnerschaft                    | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| ... Ihren familiären Beziehungen           | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| ... Ihren Freundschaften                   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| ... Ihrer Möglichkeit, Hobbies nachzugehen | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

**54. Sind Sie mit Ihrer Ernährung zufrieden?**

- ja
- nein

**55. Sind Sie mit Ihrem Gewicht zufrieden?**

- ja
- nein

**56. Nehmen Sie regelmäßige Mahlzeiten zu sich?**

- ja
- meistens
- manchmal
- nein

**57. Achten Sie auf eine gesunde Ernährung? (Halten Sie z.B. eine bestimmte Diät ein?)**

- ja
- meistens
- manchmal
- nein

**58. Nutzen Sie spezifische Techniken zur Entspannung? (z.B. Yoga, PMR, Autogenes Training, Meditation o.ä.)**

- ja
- nein

**59. An wie vielen Tagen pro Woche trinken Sie mindestens ein Alkoholgetränk?**

**60. Wie oft trinken Sie mehr als 6 (Frauen) oder 8 (Männer) Drinks pro Gelegenheit? (1 Drink = 250 ml Bier oder 1 Glas Wein oder 2 cl Schnaps)**

- nie
- selten
- monatlich
- wöchentlich
- (fast) täglich

**61. Wie oft konnten Sie sich im vergangenen Jahr nach Alkoholkonsum nicht mehr (vollständig) an einen Abend erinnern („Filmriss“)?**

- nie
- selten
- monatlich
- wöchentlich
- (fast) täglich

**62. Wie oft konnten Sie im vergangenen Jahr am Morgen nach Alkoholkonsum nicht das tun, was Sie normalerweise getan hätten?**

- nie
- selten
- monatlich
- wöchentlich
- (fast) täglich

**63. Wie oft hat Sie im vergangenen Jahr jemand auf Ihren Alkoholkonsum angesprochen z. B. mit der Empfehlung den Konsum zu reduzieren?**

- nie
- einmal
- mehrmals

**64. Wie viele Zigaretten rauchen Sie durchschnittlich pro Tag?**

---

**65 Nutzen Sie Substanzen oder Medikamente zur Verbesserung von Konzentration und Leistung?** (Mehrfachnennung möglich)

- Koffeinhaltige Getränke
- Medikamente
- aktivierende Drogen
- keine

**66. Nutzen Sie Substanzen oder Medikamente zur Beruhigung oder um schlafen zu können?**

- (Mehrfachnennung möglich)
- Nahrungsmittel (z.B. Schokolade)
  - Medikamente
  - sedierende Drogen
  - keine

---

**67 Welche der folgenden Symptome haben Sie während und aufgrund Ihres Studiums erlebt?** (Mehrfachnennung möglich)

- Prokrastination
- Erschöpfung
- Überforderung
- Ängste
- Psychosomatische Beschwerden (z.B. Magen-Darm-Probleme, Verspannung, Neurodermitis)
- Lern- und Leistungsstörungen
- Identitäts- und Selbstwertprobleme
- Gereiztheit
- depressive Verstimmung

**68. Haben Sie mit jemandem über diese Symptome gesprochen?** (Mehrfachnennung möglich)

- nein, ich habe mich niemandem anvertraut
- ja, ich habe mich jemandem anvertraut, und zwar...
  - ...meiner Familie
  - ...meinen Freunden
  - ...meiner Partnerin / meinem Partner
  - ...meinen Kommilitonen
  - ...meinen Dozenten
  - ...der Studienberatung
  - ...einer/m Psychotherapeutin / -therapeuten
  - ... einer Ärztin / einem Arzt

### Screening zum emotionalen und körperlichen Wohlbefinden

Das nachfolgende Screening enthält insgesamt 12 Aussagen zu Ihrem Befinden, die ersten 6 Fragen beziehen sich dabei auf ihr **emotionales**, die letzten 6 Fragen auf ihr **körperliches** Wohlbefinden. Bitte kreuzen Sie die Zahlen 0, 1, 2 oder 3 an, die angeben soll, wie sehr das entsprechende Adjektiv auf die Sie in den letzten Wochen zutrifft. Es gibt hierbei keine richtigen oder falschen Antworten. Versuchen Sie, sich spontan für eine Zahl zu entscheiden.

Bitte geben Sie an, wie Sie sich während der letzten Wochen **emotional** gefühlt haben.

	1 Gar nicht	2 ein wenig	3 ziemlich	4 Sehr
1. Glücklich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Besorgt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Gelassen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Traurig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Selbstsicher	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Ängstlich	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Bitte geben Sie an, wie Sie sich während der letzten Wochen **körperlich** gefühlt haben.

	1 Gar nicht	2 ein wenig	3 ziemlich	4 Sehr
1. Gesund	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Lethargisch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Stark	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Unfit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Energiegeladen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Schwach	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



## Beck-Depressions-Inventar II

Der folgende Fragebogen besteht aus 21 Gruppen von Aussagen. Lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und suchen Sie sich dann in jeder Gruppe **eine Aussage**, die am besten beschreibt, wie Sie sich **in den letzten zwei Wochen, einschließlich heute, gefühlt haben**. Kreuzen Sie die Zahl neben der Aussage an, die Sie herausgesucht haben. Wenn in einer Gruppe mehrere Aussagen gleichermaßen auf Sie zutreffen, kreuzen Sie die Aussage mit der höheren Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppen 16 (Veränderung der Schlafgewohnheiten) oder Gruppe 18 (Veränderung des Appetits).

### 1. Traurigkeit

- 0 Ich bin nicht traurig.
- 1 Ich bin oft traurig.
- 2 Ich bin ständig traurig.
- 3 Ich bin so traurig oder unglücklich, dass ich es nicht aushalten kann.

### 2. Pessimismus

- 0 Ich bin nicht mutlos, was meine Zukunft angeht.
- 1 Ich bin mutloser als früher, was meine Zukunft angeht.
- 2 Ich glaube nicht, dass sich meine Lage verbessert.
- 3 Ich habe das Gefühl, dass es keine Hoffnung gibt für meine Zukunft und es nur noch schlimmer wird.

### 3. Frühere Misserfolge

- 0 Ich fühle mich nicht als Versager.
- 1 Ich habe öfter versagt als ich sollte.
- 2 Wenn ich zurückblicke, sehe ich eine Menge Misserfolge.
- 3 Ich fühle mich persönlich als totaler Versager.

### 4. Verlust von Freude

- 0 Ich habe so viel Freude wie immer an den Dingen, die mir Spaß machen.
- 1 Ich habe nicht mehr so viel Spaß an den Dingen wie früher.
- 2 Ich habe sehr wenig Freude an den Dingen, die mir früher Spaß gemacht haben.
- 3 Ich habe keine Freude an den Dingen, die mir früher Spaß gemacht haben.

### 5. Schuldgefühle

- 0 Ich habe keine besonderen Schuldgefühle.
- 1 Ich habe bei vielen Dingen, die ich getan habe oder hätte tun sollen, Schuldgefühle.
- 2 Ich habe die meiste Zeit Schuldgefühle.
- 3 Ich habe ständig Schuldgefühle.

### 6. Gefühle, bestraft zu werden

- 0 Ich habe nicht das Gefühl, für etwas bestraft zu werden.
- 1 Ich habe das Gefühl, dass ich vielleicht für etwas bestraft werde.
- 2 Ich glaube, dass ich für etwas bestraft werde.
- 3 Ich habe das Gefühl, für etwas bestraft zu werden.

### 7. Abneigung gegen sich selbst

- 0 Meine Gefühle mir gegenüber sind die gleichen geblieben.
- 1 Ich habe das Vertrauen in mich verloren.
- 2 Ich bin von mir selbst enttäuscht.
- 3 Ich mag mich nicht.

### 8. Selbstvorwürfe

- 0 Ich bin mir gegenüber nicht kritischer als sonst und mache mir nicht mehr Vorwürfe als sonst.
- 1 Ich bin mir selbst kritischer als früher.
- 2 Ich mache mir Vorwürfe für alle meine Fehler.
- 3 Ich gebe mir die Schuld für alles Schlimme, was passiert.

### 9. Selbstmordgedanken oder -wünsche

- 0 Ich denke nie daran, mich umzubringen.
- 1 Ich habe Selbstmordgedanken, aber ich würde sie nicht ausführen.
- 2 Ich möchte mich umbringen.
- 3 Ich würde mich umbringen, wenn ich die Möglichkeit hätte.

### 10. Weinen

- 0 Ich weine nicht mehr als früher.
- 1 Ich weine mehr als früher.
- 2 Ich weine wegen jeder Kleinigkeit.
- 3 Mir ist nach Weinen zumute, aber ich kann nicht.

---

### 11. Unruhe

- 0 Ich bin nicht unruhiger und erregter als sonst.
- 1 Ich bin unruhiger und erregter als sonst.
- 2 Ich bin so unruhig oder erregt, dass es schwer ist, mich nicht zu bewegen.
- 3 Ich bin so unruhig oder erregt, dass ich ständig in Bewegung bleiben oder etwas tun muss.

---

### 12. Interessenlosigkeit

- 0 Ich habe das Interesse an anderen Menschen oder an Tätigkeiten nicht verloren.
- 1 Ich bin weniger an anderen Menschen oder Dingen interessiert als vorher.
- 2 Ich habe mein Interesse an anderen Menschen oder Dingen zum größten Teil verloren.
- 3 Es ist schwer, für irgendetwas Interesse aufzubringen.

---

### 13. Entschlussunfähigkeit

- 0 Ich treffe Entscheidungen etwa so leicht wie immer.
- 1 Es fällt mir schwerer als sonst, Entscheidungen zu treffen.
- 2 Ich habe viel größere Schwierigkeiten, Entscheidungen zu treffen als früher.
- 3 Ich habe Mühe, überhaupt Entscheidungen zu treffen.

---

### 14. Wertlosigkeit

- 0 Ich fühle mich nicht wertlos.
- 1 Ich halte mich nicht für so wertvoll und nützlich wie früher.
- 2 Ich habe das Gefühl, weniger wert zu sein als andere Menschen.
- 3 Ich habe das Gefühl, völlig wertlos zu sein.

---

### 15. Verlust an Energie

- 0 Ich habe so viel Energie wie immer.
- 1 Ich habe weniger Energie als früher.
- 2 Ich habe nicht genügend Energie, sehr viel zu tun.
- 3 Ich habe nicht genügend Energie, irgendetwas zu tun.

---

### 16. Veränderung der Schlafgewohnheiten

- 0 Meine Schlafgewohnheiten haben sich nicht geändert.
- 1a Ich schlafe etwas mehr als sonst.
- 1b Ich schlafe etwas weniger als sonst.
- 2a Ich schlafe viel mehr als sonst.
- 2b Ich schlafe viel weniger als sonst.
- 3a Ich schlafe die meiste Zeit des Tages.
- 3b Ich wache 1-2 Stunden zu früh auf und kann dann nicht mehr einschlafen.

---

### 17. Reizbarkeit

- 0 Ich bin nicht reizbarer als sonst.
- 1 Ich bin reizbarer als sonst.
- 2 Ich bin viel reizbarer als sonst.
- 3 Ich bin ständig reizbar.

---

### 18. Veränderung des Appetits

- 0 Mein Appetit hat sich nicht verändert.
- 1a Mein Appetit ist etwas kleiner als sonst.
- 1b Mein Appetit ist etwas größer als sonst.
- 2a Mein Appetit ist viel kleiner als sonst.
- 2b Mein Appetit ist viel größer als sonst.
- 3a Ich habe überhaupt keinen Appetit.
- 3b Ich habe ständig großen Hunger.

---

### 19. Konzentrationsschwierigkeiten

- 0 Ich kann mich so gut konzentrieren wie immer.
- 1 Ich kann mich nicht so gut konzentrieren wie sonst.
- 2 Es fällt mir schwer, mich sehr lange auf etwas zu konzentrieren.
- 3 Ich kann mich auf gar nichts konzentrieren.

---

### 20. Müdigkeit

- 0 Ich bin nicht müder als sonst.
- 1 Ich werde schneller müde als sonst.
- 2 Ich bin für viele Dinge, die ich früher getan habe zu müde.
- 3 Ich bin für die meisten Dinge, die ich früher getan habe zu müde.

---

### 21. Verlust des Interesses an Sex

- 0 Ich habe in der letzten Zeit keine Veränderungen meines Interesses am Sex bemerkt.
  - 1 Ich habe weniger Interesses am Sex als früher.
  - 2 Ich habe jetzt viel weniger Interesse am Sex.
  - 3 Ich habe das Interesse am Sex völlig verloren.
-

## NEO-FFI - Neurotizismus

---

Der nachfolgende Fragebogen umfasst 12 Aussagen, welche sich zur Beschreibung Ihrer eigenen Person eignen können. Lesen Sie bitte jede dieser Aussagen aufmerksam durch und überlegen Sie, ob diese Aussage auf Sie persönlich zutrifft oder nicht. Zur Bewertung jeder der 12 Aussagen steht Ihnen eine fünffach abgestufte Skala zur Verfügung. Kreuzen Sie bitte an:

**Starke Ablehnung**, wenn Sie der Aussage auf keinen Fall zustimmen oder Sie für völlig unzutreffend halten.

**Ablehnung**, wenn Sie der Aussage eher nicht zustimmen oder sie für unzutreffend halten.

**Neutral**, wenn die Aussage weder richtig noch falsch, also weder zutreffend noch unzutreffend ist.

**Zustimmung**, wenn Sie der Aussage eher zustimmen oder Sie für zutreffend halten.

**Starke Zustimmung**, wenn Sie der Aussage nachdrücklich zustimmen oder sie für völlig zutreffend halten.

Bitte lesen Sie jede Aussage genau durch und kreuzen Sie als Antwort die Kategorie an, die Ihre Sichtweise am besten ausdrückt. Es gibt bei diesem Fragebogen keine „richtigen“ oder „falschen“ Antworten. Falls Sie Ihre Meinung nach dem Ankreuzen einmal ändern sollten, streichen Sie Ihre erste Antwort bitte deutlich durch. **Lassen Sie keine Aussage aus.** Auch wenn Ihnen einmal die Entscheidung schwerfallen sollte, kreuzen Sie trotzdem immer eine Antwort an, und zwar die, welche noch am ehesten auf Sie zutrifft.

	Starke Zustimmung	Zustimmung	Neutral	Ablehnung	Starke Ablehnung
1. Ich bin nicht leicht beunruhigt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Ich fühle mich anderen oft unterlegen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Wenn ich unter starkem Stress stehe, fühle ich mich manchmal, als ob ich zusammenbräche.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Ich fühle mich selten einsam oder traurig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Ich fühle mich oft angespannt und nervös.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Manchmal fühle ich mich völlig wertlos.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Ich empfinde selten Furcht oder Angst.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Ich ärgere mich oft darüber, wie andere Leute mich behandeln.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Zu häufig bin ich entmutigt und will aufgeben, wenn etwas schiefgeht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Ich bin selten traurig oder deprimiert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Ich fühle mich oft hilflos und wünsche mir eine Person, die meine Probleme löst.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Manchmal war mir etwas so peinlich, dass ich mich am liebsten versteckt hätte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Der Nachfolgende Fragebogen enthält 21 Fragen zu Ihrem Befinden. Bitte lesen Sie jede Aussage und kreuzen Sie die Zahl 0, 1, 2, oder 3 an, die angeben soll, wie sehr die Aussage während der letzten Wochen auf Sie zutraf. Es gibt keine richtigen oder falschen Antworten. Versuchen Sie, sich spontan für eine Antwort zu entscheiden.

- 0 Traf gar nicht auf mich zu
- 1 Traf bis zu einem gewissen Grad auf mich zu oder manchmal
- 2 Traf in beträchtlichem Maß auf mich zu oder ziemlich oft
- 3 Traf sehr stark auf mich zu oder die meiste Zeit

1 Gar Nicht  
 2 Manchmal  
 3 Ziemlich oft  
 4 die meiste Zeit

- |  |   |
|--|---|
| 1. Ich fand es schwer, mich zu beruhigen.  | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 2. Ich spürte, dass mein Mund trocken war.   | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 3. Ich konnte überhaupt keine positiven Gefühle mehr erleben   | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 4. Ich hatte Atemprobleme (z. B. übermäßig schnelles Atmen, Atemlosigkeit ohne Körperliche Anstrengung     | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 5. Es fiel mir schwer, mich dazu aufzuraffen, Dinge zu erledigen.  | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 6. Ich tendierte dazu, auf Situationen überzureagieren.  | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 7. Ich zitterte (z. B. an den Händen).   | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 8. Ich fand alles anstrengend.   | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 9. Ich mache mir Sorgen über Situationen, in denen ich in Panik geraten und mich lächerlich machen könnte. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 10. Ich hatte das Gefühl, dass ich mich auf nichts mehr freuen konnte.                                     | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 11. Ich bemerkte, dass ich mich schnell aufregte.  | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 12. Ich fand es schwierig, mich zu entspannen.   | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 13. Ich fühlte mich niedergeschlagen und traurig.  | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 14. Ich reagierte ungehalten auf alles, was mich davon abhielt, meine momentane Tätigkeit fortzuführen.    | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 15. Ich fühlte mich einer Panik nahe.  | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 16. Ich war nicht in der Lage, mich für irgendetwas zu begeistern.   | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 17. Ich fühlte mich als Person nicht viel Wert.  | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 18. Ich fand mich ziemlich empfindlich.  | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |

19. Ich habe meinen Herzschlag gespürt, ohne dass ich mich körperlich angestrengt habe (z. B. Gefühl von Herzrasen oder Herzstolpern).

20. Ich fühlte mich grundlos ängstlich.

21. Ich empfand das Leben als sinnlos.

A2 Din Screen Checkliste nach Kersting (2018)

Checkliste nach Kersting genutzt für BEES

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
A1	In den Handhabungshinweisen ist die Zielsetzung des Verfahrens verständlich beschrieben	S		X			
A2	In den Handhabungshinweisen sind die Anwendungsbereiche verständlich benannt.	S		X			
V1	Verzweigungsfrage: Sind missbräuliche Anwendungen des Verfahrens zur Eignungsbeurteilung naheliegend? Falls nein, weiter zu V2			X			
A3	Die Handhabungshinweise enthalten spezifische warnende Hinweise vor missbräuchlicher Anwendung zur Eignungsbeurteilung	S			X		
V2	<i>Verzweigungsfrage: Erfordert die Handhabung des Verfahrens besondere Qualifikationen? Falls „nein“, weiter zu A5</i>		X				
A4	Die für die Handhabung des Verfahrens erforderlichen besonderen Qualifikationen sind genannt	S		X			
	Die Handhabungshinweise liefern Informationen, aus denen die Anwenderin/der Anwender den hinsichtlich der folgenden Aspekte entstehenden Aufwand abschätzen kann:						
A5	• Materialien	S	X				
A6	• Personal	S	X				
A7	• Räumlichkeiten	S	X				
	Die Handhabungshinweise liefern Informationen, aus denen die Anwenderin/						

	der Anwender den hinsichtlich der folgenden Aspekte entstehenden zeitlichen Aufwand abschätzen kann:						
A8	• für die Kandidatin/den Kandidaten	S	X				
A9	• für die Anwenderin/den Anwender bei der Routinevorbereitung	S	X				
A10	• für die Anwenderin/den Anwender bei der Durchführung	S	X				
A11	• für die Anwenderin/den Anwender bei der Auswertung	S	X				
V3	<i>Verzweigungsfrage 3: Gibt es eine Interaktion mit der Kandidatin/dem Kandidaten? Falls „nein, weiter zu A13</i>			X			
A12	Die Handhabungshinweise beinhalten verständliche Instruktionen für die Kandidatin/den Kandidaten, die dazu beitragen, die Wahrscheinlichkeit von Nachfragen zu verhindern	S			X		
	Die Handhabungshinweise sind so gestaltet, dass verschiedene Personen mit den erforderlichen Qualifikationen in der Lage sind, die Verfahren allein aufgrund dieser Handhabungshinweise auf die gleiche Art und Weise:						
A13	• durchzuführen	M	X				
A14	• auszuwerten	M	X				
A15	• und deren Ergebnisse zu interpretieren	M		X			
<p>-zu Anforderung A2: Es sollte z.B. angegeben sein, bei welcher Personengruppe (z.B. Bildungsstand) das Verfahren eingesetzt werden kann.</p> <p>-zu Anforderung A12: Beispiele für häufige, aber (durch entsprechende Instruktionen zu Beginn des Verfahrens) vermeidbare Nachfragen: Darf man sich Notizen machen? Wird die zur Verfahrensbearbeitung zur Verfügung stehende Zeit bekannt gegeben? Darf man Teilaufgaben überspringen? Gibt es Minuspunkte bzw. Abzüge für falsche Antworten?</p>							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
B1	Die theoretischen Grundlagen des Verfahrens sind beschrieben	M		X			
B2	In den Verfahrenshinweisen ist angemessen (im Sinne von ausführlich, verständlich und nachvollziehbar) dargestellt, wie das standardisierte Verfahren konstruiert wurde <sup>3</sup>	M		X			
B3	In den Verfahrenshinweisen werden die Ergebnisse einer oder mehrerer empirischen Untersuchungen berichtet	M		X			
H1	<i>Hinweis: Falls „nein“ gewählt wird, muss bei den Anforderungen B4 bis B13 ebenfalls „nein“ angekreuzt werden</i>				X		
B4	Alle in den Verfahrenshinweisen aufgeführten relevanten empirischen Untersuchungen sind nachvollziehbar beschrieben/dokumentiert	M		X			
	Der Bericht über empirische Untersuchungen enthält:			X			
B5	• eine Angabe über das Jahr der Datenerhebung	M		X			
B6	• deskriptive Statistiken über die Merkmale der Untersuchungsteilnehmer/-innen	M		X			
B7	• Angaben, mit welchem Ziel der Test von Teilnehmerinnen/Teilnehmern bearbeitet wurde	M		X			
B8	• Angaben, ob die Datenerhebung unter Aufsicht oder unter nicht kontrollierten Bedingungen (z.B. über das Internet von „zu Hause“ aus) stattgefunden hat	M		X			
B9	• Angaben, ob und wie die Teilnahme (z.B. ergebnisorientiert) belohnt (z.B. vergütet) wurde	M		X			
	Der Bericht über empirische Untersuchungen enthält:						



B10	• Informationen über den Stichprobenplan	S		X			
B11	• Informationen zu den Teilnehmerquoten	S		X			
B12	Die Dokumentation der empirischen Arbeit folgt den üblichen Kriterien für wissenschaftliche Publikationen	S		X			
B13	Die Anzahl der in den empirischen Studien untersuchten Personen ist für die jeweilige Fragestellung <sup>7</sup> angemessen	M		X			
V4	<i>Verzweigungsfrage V4: Ist mit einer Verfälschung des Verfahrens zu rechnen? Falls „nein“ weiter zu V5</i>			X			
B14	Es ist aufgeführt, ob und wie einer Verfälschung durch die Art der Verfahrensvorgabe und -durchführung – sowie ggf. auch bei der Auswertung – entgegengewirkt werden kann	S			X		
V5	<i>Verzweigungsfrage V5: Erfolgt die Auswertung manuell? Falls „nein“ -&gt; Bitte weiter bei Verzweigungsfrage V6</i>		X				
B15	In den Verfahrenshinweisen sind Regeln aufgestellt, wie bei der Auswertung mit nicht bearbeiteten Fragen bzw. (Teil-) Aufgaben umgegangen wird	M		X			
V6	<i>Verzweigungsfrage V6: Handelt es sich um ein Verfahren, welches einen Vergleich mit Normwerten anbietet? Falls „nein“ -&gt; Bitte weiter bei Anforderung B22</i>			X			
B16	Die Bezugsgruppe, an der die Normdaten gewonnen wurden, entspricht hinsichtlich zentraler Merkmale der Personengruppe, für die das Verfahren laut Verfahrenshinweisen eingesetzt wird/werden soll, oder es wird nachgewiesen, dass die vorhandenen Normdaten für die Zielgruppe verwendet werden können	M			X		
<p>-zu Anforderung B2: Das heißt, es wird z.B. erläutert, wie und warum die Fragen eines Fragebogens oder die Aufgaben eines Tests ausgewählt oder konstruiert wurden.</p> <p>-zu Anforderung B6: Erläuterung: z.B. Angaben zu Alter, Geschlecht, Bildung, Status (z.B. Schülerinnen/Schüler, Studentinnen/Studenten, Azubis, Berufstätige usw.)</p> <p>-zu Anforderung B7: z.B. ohne für die Teilnehmer/-innen relevantes Ziel, zum Zwecke der persönlichen Orientierung oder im Zusammenhang mit Personalentscheidungen.</p>							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
B1 7	Die Angemessenheit der Normwerte wurde in den letzten acht Jahren überprüft <sup>9</sup>	S			X		
V7	<i>Verzweigungsfrage V7: Wurde die Angemessenheit der Normwerte in den letzten acht Jahren überprüft? Falls „ja“-&gt; Bitte weiter bei Verzweigungsfrage V8</i>				X		
B1 8	In den Verfahrenshinweisen wird begründet, warum und unter welchen Umständen das Verfahren für einen Anwendungsfall ausgewählt werden kann, obwohl die Angemessenheit der Normwerte nicht in den letzten acht Jahren überprüft wurde <sup>10</sup>	M			X		
V8	<i>Verzweigungsfrage V8: Zielt das Verfahren auf die Erfassung eines Eignungsmerkmals ab, dessen Ausprägung in der Referenzgruppe möglicherweise relativ kurzfristigen Veränderungen unterliegt?<sup>11</sup> Falls „nein“-&gt; Bitte weiter bei Verzweigungsfrage V9</i>				X		
B1 9	Die Angemessenheit der Normwerte wurde bereits vor Ablauf der Acht-Jahres-Frist empirisch gezeigt	M			X		
V9	<i>Verzweigungsfrage V9: Besteht die Möglichkeit, die Werte einer Person anhand unterschiedlicher Normgruppen (auch gruppenspezifische Normen genannt) zu bewerten?<sup>12</sup> Falls „nein“ -&gt; Bitte weiter bei Aussage B22</i>				X		
B2 0	Zur Sicherung der Interpretationsobjektivität sind eindeutige Hinweise gegeben, wie die Entscheidung, welche Normgruppe in welchem Fall heranzuziehen ist, zu treffen ist	S			X		

B2 1	Die Effekte der Anwendung dieser gruppenspezifischen Normen <sup>13</sup> sind nachvollziehbar erläutert	S			X		
<p>-zu Anforderung B17: Erläuterung: Es geht nur um eine Überprüfung der Angemessenheit der Normwerte. Ob eine Neunormierung durchgeführt werden muss, ergibt sich in Abhängigkeit von den Ergebnissen der Überprüfung. In der DIN 33430 wird nicht gefordert, dass spätestens alle acht Jahre neu normiert werden muss.</p> <p>-zu Anforderung B18: Die Anforderung lautet im Text der DIN: „Wurde die Angemessenheit der Normwerte in den letzten acht Jahren nicht überprüft, muss begründet werden, warum das Verfahren dennoch ausgewählt wird.“ Diese Begründung obliegt den Eignungsdiagnostikern/Eignungsdiagnostikerinnen. Die hier in der Checkliste gewählte Formulierung richtet sich hingegen an die Verantwortlichen für die Verfahrenshinweise.</p> <p>-zu Verzweigungsfrage V8: z.B. EDV-Kenntnisse</p> <p>-zu Verzweigungsfrage V9: unterschiedliche Normgruppen wären beispielsweise bildungsspezifische und bildungsunspezifische Normen</p> <p>-zu Anforderung B21: z.B. Bildungsnorm</p>							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
B2 2	In den Verfahrenshinweisen werden Angaben zur Zuverlässigkeit des Verfahrens gemacht, die aus empirischen Studien abgeleitet wurden	M	X				
H2	<i>Hinweis H2: Falls keine Angaben zur Zuverlässigkeit des Verfahrens gemacht werden, ist die Anforderung B22 mit „nein“ zu beantworten</i>				X		
V1 0	<i>Verzweigungsfrage V10: Wurden Angaben zur Zuverlässigkeit des Verfahrens gemacht? Falls „nein“ -&gt; Bitte weiter bei Anforderung B27</i>		X				
B2 3	Die Angemessenheit der für die Zuverlässigkeitsbestimmung genutzten Methode(n) wird erläutert. Die Erläuterung berücksichtigt die Art der untersuchten Eignungsmerkmale und der angestrebten Entscheidung ebenso wie die jeweiligen Anwendungs- und Untersuchungsbedingungen	S		X			

V1 1	<i>Verzweigungsfrage V11: Sollen mit dem Verfahren Merkmale erfasst werden, für die eine zumindest relative Zeit- und Situationsstabilität angenommen wird? Falls „nein“ -&gt; Bitte weiter bei Anforderung B27</i>			X			
B2 4	Die Zuverlässigkeit wurde über die Retest-Methode bestimmt oder die Retest-Reliabilität wurde durch einen geeigneten Untersuchungsplan geschätzt	S			X		
B2 5	Der aktuellste Nachweis der Geltung der Zuverlässigkeitskennwerte ist jünger als 8 Jahre	S			X		
V1 2	<i>Verzweigungsfrage V12: Wurde die Geltung der Zuverlässigkeitskennwerte in den letzten 8 Jahren überprüft? Falls „nein“ weiter zu B27</i>						
B2 6	In der Verfahrenshinweisen wird begründet, warum und unter welchen Umständen das Verfahren für einen Anwendungsfall gewählt werden kann, obwohl die Geltung der Zuverlässigkeitskennwerte in den letzten acht Jahren nicht überprüft wurde.	M			X		
<p>-zu Anforderung B23: Erläuterung: Die Bestimmung der internen Konsistenz ist beispielsweise keine angemessene Art der Zuverlässigkeitsbestimmung für Verfahren mit heterogenen Inhalten; die Bestimmung der Retest-Reliabilität ist keine angemessene Art der Zuverlässigkeitsbestimmung für Verfahren zur Messung rasch veränderlicher Merkmale (z.B. Stimmungen).</p> <p>-zu Anforderung B26: Die Anforderung lautet im Text der DIN: „Wurden die Zuverlässigkeitskennwerte in den letzten 8 Jahren nicht überprüft, muss begründet werden, warum das Verfahren dennoch ausgewählt wird.“ Diese Begründung obliegt den Eignungsdiagnostikern/Eignungsdiagnostikerinnen. Die hier in der Checkliste gewählte Formulierung richtet sich hingegen an die Verantwortlichen für die Verfahrenshinweise.</p>							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
B2 7	In den Verfahrenshinweisen werden Angaben zur Gültigkeit des Verfahrens gemacht, die aus empirischen Studien abgeleitet wurden	M	X				
V1 3	<i>Verzweigungsfrage V13: Wurden Angaben zur Gültigkeit des Verfahrens gemacht?</i>		X				

	-> Falls „nein“ endet die Checkliste mit dem folgenden Hinweis H3						
H3	Hinweis H3: Falls keine empirisch fundierten Angaben zur Gültigkeit des Verfahrens gemacht werden, sind die Anforderung B27 bis B34 und B40 bis B41 mit „nein“ zu beantworten				X		
B28	Aus den Verfahrenshinweisen wird deutlich, welche empirischen Nachweise der Inhalts- und/oder Kriteriums- und/oder Konstruktgültigkeit eine Anwendung des Verfahrens bzw. der Verfahrensklasse für den laut Verfahrenshinweisen intendierten Anwendungszweck rechtfertigen			X			
	In den Verfahrenshinweisen wird angegeben, welche Gültigkeitswerte:				X		
B29	• in Bezug zu welchem Kriterium erzielt wurden	M		X			
B30	• für welche Referenzgruppen erzielt wurden	M		X			
B31	• in welcher Untersuchung erzielt wurden	M	X				
B32	• für welches Verfahrensergebnis erzielt wurden	M		X			
B33	• zu welchem Zeitpunkt erzielt wurden	M	X				
B34	Der aktuellste Nachweis über die Gültigkeit des Verfahrens ist jünger als acht Jahre	S	X				
-zu Anforderung B29: Bewährungskriterium, z.B. Vorgesetztenbeurteilung zum Berufserfolg -zu Anforderung B32: Bezieht sich der Gültigkeitswert beispielsweise auf das Gesamtergebnis oder auf ein Teilergebnis (etwa auf eine einzelne Skala oder einzelne Items)? Bezieht sich der Gültigkeitswert auf einen Rohwert oder auf einen standardisierten Wert?							
		Soll oder Muss	ja	nein	Nicht zu bewerten	Anm.	Quelle
V14	Verzweigungsfrage V14: Wurden zur Bestimmung der Gültigkeit Methoden der			X			

	<i>statistischen Adjustierung/ Optimierung eingesetzt18? Falls „nein“ -&gt; Bitte weiter bei Verzweigungsfrage V15</i>						
	Bei der Dokumentation der Analysen zur Gültigkeit:				X		
B3 5	• sind sowohl die ursprünglich erhaltenen als auch die korrigierten Kennwerte aufgeführt	M			X		
B3 6	• werden alle in Zusammenhang mit der Adjustierung verwendeten Statistiken genannt	M			X		
B3 7	• werden neben den statistisch optimierten Schätzungen auch die einfachen Schätzungen angegeben	M			X		
B3 8	• wurden die optimierten Schätzungen auf eine andere Personengruppe aus dem Geltungsbereich des Verfahrens angewendet und in ihrer Gültigkeit bestätigt (Kreuzvalidierung)	S			X		
B3 9	• sind die statistischen Optimierungen in handlungsleitende Beurteilungsregeln umgesetzt	S			X		
V1 5	<i>Verzweigungsfrage V15: Wird der Gültigkeitsanspruch damit begründet, dass Gültigkeitshinweise aus anderen Untersuchungen in Anspruch genommen werden? Falls „nein“-&gt; Bitte weiter bei Anforderung B42</i>			X			
	Es wird nachvollziehbar aufgeführt:				X		
B4 0	• welche Befunde generalisiert werden können (Darstellung der entsprechenden Studien, Literaturübersichten und Metaanalysen)	S			X		
B4 1	• weshalb (und in welchem Ausmaß) sich die Gültigkeitshinweise übertragen lassen, die sich aus anderen Studien ergeben	S			X		
<p>-zu Verzweigungsfrage V14: (z.B. Minderungskorrektur, Varianzeinschränkungskorrektur, multiple Regression)</p> <p>-zu Anforderung B37: statistisch optimierte Schätzungen sind z.B. multiple Regressionen; einfache Schätzungen sind z.B. einfache Korrelationen</p> <p>-zu Anforderung B39: Wenn beispielsweise gezeigt wird, dass die multiple Vorhersagbarkeit eines Kriteriums unter Einbezug mehrerer Prädiktoren (z.B. mehrere Skalen eines Tests) deutlich höher ist als die einfache Korrelation zwischen einzelnen Prädiktoren und diesem Kriterium, so sollte der Anwenderin/dem Anwender erläutert werden, wie sie/er die verschiedenen Prädiktoren so kombinieren/gewichten kann, dass der Vorteil praktisch nutzbar wird.</p>							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
B4 2	Aufgrund von inhaltlichen Überlegungen wird dargelegt, wie sich das fragliche Konstrukt zu ähnlichen Konstrukten verhält (konvergente Gültigkeit)	S		X			
B4 3	Aufgrund von empirischen Ergebnissen wird dargelegt, wie sich das fragliche Konstrukt zu ähnlichen Konstrukten verhält (konvergente Gültigkeit)	S	X				
B4 4	Aufgrund von inhaltlichen Überlegungen wird dargelegt, wie sich das fragliche Konstrukt zu unähnlichen Konstrukten verhält (diskriminante Gültigkeit)	S		X			
B4 5	Aufgrund von empirischen Untersuchungen wird dargelegt, wie sich das fragliche Konstrukt zu unähnlichen Konstrukten verhält (diskriminante Gültigkeit)	S		X			
B4 6	Bei der Analyse der Kriteriumsgültigkeit des Verfahrens wird beschrieben, warum das in der Analyse jeweils verwendete Kriterium angemessen ist und valide erfasst wurde	M		X			
B4 7	Sowohl die Objektivität als auch die Zuverlässigkeit jedes verwendeten Kriterienmaßes wird nach Möglichkeit dargestellt	S		X			
B4 8	Die Angemessenheit der für die Analyse der Kriteriumsgültigkeit herangezogenen Untersuchungsgruppe wird erläutert	M		X			
-zu Anforderung B48: Beispielsweise sollten die demografischen Merkmale der Untersuchungsgruppe (z.B. Bildungsstand, Alter, Berufserfahrung usw.) vor dem Hintergrund der als Zielgruppe des Verfahrens genannten Gruppe diskutiert werden.							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
V1 6	<i>Wird für das Verfahren Inhaltsgültigkeit in Anspruch genommen? Falls „nein“ endet die Checkliste an dieser Stelle</i>		X				
B4 9	Der im Verfahren abgebildete Inhaltsbereich wird nachvollziehbar beschrieben	S		X			
B5 0	Die Kriterien zur Beschreibung des dem Verfahren zugrunde liegenden, hypothetischen Itemuniversums sind angegeben	S		X			
B5 1	Die Regeln, nach denen das Verfahren als systematisch zusammengestellte Itemstichprobe aus dem Itemuniversum abgeleitet wurde, werden dargestellt	S		X			
V1 7	<i>Wurde die Frage, ob das Verfahren den definierten Inhaltsbereich repräsentiert, durch Expertinnen/Experten beurteilt? Falls „nein“ endet die Checkliste an dieser Stelle</i>			X			
	Es wird:						
B5 2	• Sowohl der fachbezogene Ausbildungsstand als auch die Erfahrung als auch die Qualifikation der beteiligten Expertinnen/Experten beschrieben	S			X		
B5 3	• erläutert, wie die Expertinnen/Experten zu ihrer Einschätzung gekommen sind	S			X		
B5 4	• angegeben, inwieweit die Expertinnen-/Expertenbeurteilungen übereinstimmen	S			X		



Checkliste nach Kersting (2018) genutzt für BEEPS

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
A1	In den Handhabungshinweisen ist die Zielsetzung des Verfahrens verständlich beschrieben	S		X			
A2	In den Handhabungshinweisen sind die Anwendungsbereiche verständlich benannt.	S		X			
V1	Verzweigungsfrage: Sind missbräuliche Anwendungen des Verfahrens zur Eignungsbeurteilung naheliegend? Falls nein, weiter zu V2			X			
A3	Die Handhabungshinweise enthalten spezifische warnende Hinweise vor missbräuchlicher Anwendung zur Eignungsbeurteilung	S			X		
V2	<i>Verzweigungsfrage: Erfordert die Handhabung des Verfahrens besondere Qualifikationen? Falls „nein“, weiter zu A5</i>		X				
A4	Die für die Handhabung des Verfahrens erforderlichen besonderen Qualifikationen sind genannt	S		X			
	Die Handhabungshinweise liefern Informationen, aus denen die Anwenderin/der Anwender den hinsichtlich der folgenden Aspekte entstehenden Aufwand abschätzen kann:						
A5	• Materialien	S	X				
A6	• Personal	S	X				
A7	• Räumlichkeiten	S	X				
	Die Handhabungshinweise liefern Informationen, aus denen die Anwenderin/						

	der Anwender den hinsichtlich der folgenden Aspekte entstehenden zeitlichen Aufwand abschätzen kann:						
A8	• für die Kandidatin/den Kandidaten	S	X				
A9	• für die Anwenderin/den Anwender bei der Routinevorbereitung	S	X				
A10	• für die Anwenderin/den Anwender bei der Durchführung	S	X				
A11	• für die Anwenderin/den Anwender bei der Auswertung	S	X				
V3	<i>Verzweigungsfrage 3: Gibt es eine Interaktion mit der Kandidatin/dem Kandidaten? Falls „nein, weiter zu A13</i>			X			
A12	Die Handhabungshinweise beinhalten verständliche Instruktionen für die Kandidatin/den Kandidaten, die dazu beitragen, die Wahrscheinlichkeit von Nachfragen zu verhindern	S			X		
	Die Handhabungshinweise sind so gestaltet, dass verschiedene Personen mit den erforderlichen Qualifikationen in der Lage sind, die Verfahren allein aufgrund dieser Handhabungshinweise auf die gleiche Art und Weise:						
A13	• durchzuführen	M	X				
A14	• auszuwerten	M	X				
A15	• und deren Ergebnisse zu interpretieren	M		X			
<p>-zu Anforderung A2: Es sollte z.B. angegeben sein, bei welcher Personengruppe (z.B. Bildungsstand) das Verfahren eingesetzt werden kann.</p> <p>-zu Anforderung A12: Beispiele für häufige, aber (durch entsprechende Instruktionen zu Beginn des Verfahrens) vermeidbare Nachfragen: Darf man sich Notizen machen? Wird die zur Verfahrensbearbeitung zur Verfügung stehende Zeit bekannt gegeben? Darf man Teilaufgaben überspringen? Gibt es Minuspunkte bzw. Abzüge für falsche Antworten?</p>							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
B1	Die theoretischen Grundlagen des Verfahrens sind beschrieben	M		X			
B2	In den Verfahrenshinweisen ist angemessen (im Sinne von ausführlich, verständlich und nachvollziehbar) dargestellt, wie das standardisierte Verfahren konstruiert wurde <sup>3</sup>	M		X			
B3	In den Verfahrenshinweisen werden die Ergebnisse einer oder mehrerer empirischen Untersuchungen berichtet	M		X			
H1	<i>Hinweis: Falls „nein“ gewählt wird, muss bei den Anforderungen B4 bis B13 ebenfalls „nein“ angekreuzt werden</i>				X		
B4	Alle in den Verfahrenshinweisen aufgeführten relevanten empirischen Untersuchungen sind nachvollziehbar beschrieben/dokumentiert	M		X			
	Der Bericht über empirische Untersuchungen enthält:			X			
B5	• eine Angabe über das Jahr der Datenerhebung	M		X			
B6	• deskriptive Statistiken über die Merkmale der Untersuchungsteilnehmer/-innen	M		X			
B7	• Angaben, mit welchem Ziel der Test von Teilnehmerinnen/Teilnehmern bearbeitet wurde	M		X			
B8	• Angaben, ob die Datenerhebung unter Aufsicht oder unter nicht kontrollierten Bedingungen (z.B. über das Internet von „zu Hause“ aus) stattgefunden hat	M		X			
B9	• Angaben, ob und wie die Teilnahme (z.B. ergebnisorientiert) belohnt (z.B. vergütet) wurde	M		X			
	Der Bericht über empirische Untersuchungen enthält:						

B10	• Informationen über den Stichprobenplan	S		X			
B11	• Informationen zu den Teilnehmerquoten	S		X			
B12	Die Dokumentation der empirischen Arbeit folgt den üblichen Kriterien für wissenschaftliche Publikationen	S		X			
B13	Die Anzahl der in den empirischen Studien untersuchten Personen ist für die jeweilige Fragestellung <sup>7</sup> angemessen	M		X			
V4	<i>Verzweigungsfrage V4: Ist mit einer Verfälschung des Verfahrens zu rechnen? Falls „nein“ weiter zu V5</i>			X			
B14	Es ist aufgeführt, ob und wie einer Verfälschung durch die Art der Verfahrensvorgabe und -durchführung – sowie ggf. auch bei der Auswertung – entgegengewirkt werden kann	S			X		
V5	<i>Verzweigungsfrage V5: Erfolgt die Auswertung manuell? Falls „nein“ -&gt; Bitte weiter bei Verzweigungsfrage V6</i>		X				
B15	In den Verfahrenshinweisen sind Regeln aufgestellt, wie bei der Auswertung mit nicht bearbeiteten Fragen bzw. (Teil-) Aufgaben umgegangen wird	M		X			
V6	<i>Verzweigungsfrage V6: Handelt es sich um ein Verfahren, welches einen Vergleich mit Normwerten anbietet? Falls „nein“ -&gt; Bitte weiter bei Anforderung B22</i>			X			
B16	Die Bezugsgruppe, an der die Normdaten gewonnen wurden, entspricht hinsichtlich zentraler Merkmale der Personengruppe, für die das Verfahren laut Verfahrenshinweisen eingesetzt wird/werden soll, oder es wird nachgewiesen, dass die vorhandenen Normdaten für die Zielgruppe verwendet werden können	M			X		
<p>-zu Anforderung B2: Das heißt, es wird z.B. erläutert, wie und warum die Fragen eines Fragebogens oder die Aufgaben eines Tests ausgewählt oder konstruiert wurden.</p> <p>-zu Anforderung B6: Erläuterung: z.B. Angaben zu Alter, Geschlecht, Bildung, Status (z.B. Schülerinnen/Schüler, Studentinnen/Studenten, Azubis, Berufstätige usw.)</p> <p>-zu Anforderung B7: z.B. ohne für die Teilnehmer/-innen relevantes Ziel, zum Zwecke der persönlichen Orientierung oder im Zusammenhang mit Personalentscheidungen.</p>							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
B1 7	Die Angemessenheit der Normwerte wurde in den letzten acht Jahren überprüft <sup>9</sup>	S			X		
V7	<i>Verzweigungsfrage V7: Wurde die Angemessenheit der Normwerte in den letzten acht Jahren überprüft? Falls „ja“-&gt; Bitte weiter bei Verzweigungsfrage V8</i>				X		
B1 8	In den Verfahrenshinweisen wird begründet, warum und unter welchen Umständen das Verfahren für einen Anwendungsfall ausgewählt werden kann, obwohl die Angemessenheit der Normwerte nicht in den letzten acht Jahren überprüft wurde <sup>10</sup>	M			X		
V8	<i>Verzweigungsfrage V8: Zielt das Verfahren auf die Erfassung eines Eignungsmerkmals ab, dessen Ausprägung in der Referenzgruppe möglicherweise relativ kurzfristigen Veränderungen unterliegt?<sup>11</sup> Falls „nein“-&gt; Bitte weiter bei Verzweigungsfrage V9</i>				X		
B1 9	Die Angemessenheit der Normwerte wurde bereits vor Ablauf der Acht-Jahres-Frist empirisch gezeigt	M			X		
V9	<i>Verzweigungsfrage V9: Besteht die Möglichkeit, die Werte einer Person anhand unterschiedlicher Normgruppen (auch gruppenspezifische Normen genannt) zu bewerten?<sup>12</sup> Falls „nein“ -&gt; Bitte weiter bei Aussage B22</i>				X		
B2 0	Zur Sicherung der Interpretationsobjektivität sind eindeutige Hinweise gegeben, wie die Entscheidung, welche Normgruppe in welchem Fall heranzuziehen ist, zu treffen ist	S			X		

B2 1	Die Effekte der Anwendung dieser gruppenspezifischen Normen <sup>13</sup> sind nachvollziehbar erläutert	S			X		
<p>-zu Anforderung B17: Erläuterung: Es geht nur um eine Überprüfung der Angemessenheit der Normwerte. Ob eine Neunormierung durchgeführt werden muss, ergibt sich in Abhängigkeit von den Ergebnissen der Überprüfung. In der DIN 33430 wird nicht gefordert, dass spätestens alle acht Jahre neu normiert werden muss.</p> <p>-zu Anforderung B18: Die Anforderung lautet im Text der DIN: „Wurde die Angemessenheit der Normwerte in den letzten acht Jahren nicht überprüft, muss begründet werden, warum das Verfahren dennoch ausgewählt wird.“ Diese Begründung obliegt den Eignungsdiagnostikern/Eignungsdiagnostikerinnen. Die hier in der Checkliste gewählte Formulierung richtet sich hingegen an die Verantwortlichen für die Verfahrenshinweise.</p> <p>-zu Verzweigungsfrage V8: z.B. EDV-Kenntnisse</p> <p>-zu Verzweigungsfrage V9: unterschiedliche Normgruppen wären beispielsweise bildungsspezifische und bildungsunspezifische Normen</p> <p>-zu Anforderung B21: z.B. Bildungsnorm</p>							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
B2 2	In den Verfahrenshinweisen werden Angaben zur Zuverlässigkeit des Verfahrens gemacht, die aus empirischen Studien abgeleitet wurden	M	X				
H2	<i>Hinweis H2: Falls keine Angaben zur Zuverlässigkeit des Verfahrens gemacht werden, ist die Anforderung B22 mit „nein“ zu beantworten</i>				X		
V1 0	<i>Verzweigungsfrage V10: Wurden Angaben zur Zuverlässigkeit des Verfahrens gemacht? Falls „nein“ -&gt; Bitte weiter bei Anforderung B27</i>		X				
B2 3	Die Angemessenheit der für die Zuverlässigkeitsbestimmung genutzten Methode(n) wird erläutert. Die Erläuterung berücksichtigt die Art der untersuchten Eignungsmerkmale und der angestrebten Entscheidung ebenso wie die jeweiligen Anwendungs- und Untersuchungsbedingungen	S		X			

V1 1	<i>Verzweigungsfrage V11: Sollen mit dem Verfahren Merkmale erfasst werden, für die eine zumindest relative Zeit- und Situationsstabilität angenommen wird? Falls „nein“ -&gt; Bitte weiter bei Anforderung B27</i>			X			
B2 4	Die Zuverlässigkeit wurde über die Retest-Methode bestimmt oder die Retest-Reliabilität wurde durch einen geeigneten Untersuchungsplan geschätzt	S			X		
B2 5	Der aktuellste Nachweis der Geltung der Zuverlässigkeitskennwerte ist jünger als 8 Jahre	S			X		
V1 2	<i>Verzweigungsfrage V12: Wurde die Geltung der Zuverlässigkeitskennwerte in den letzten 8 Jahren überprüft? Falls „nein“ weiter zu B27</i>						
B2 6	In der Verfahrenshinweisen wird begründet, warum und unter welchen Umständen das Verfahren für einen Anwendungsfall gewählt werden kann, obwohl die Geltung der Zuverlässigkeitskennwerte in den letzten acht Jahren nicht überprüft wurde.	M			X		
<p>-zu Anforderung B23: Erläuterung: Die Bestimmung der internen Konsistenz ist beispielsweise keine angemessene Art der Zuverlässigkeitsbestimmung für Verfahren mit heterogenen Inhalten; die Bestimmung der Retest-Reliabilität ist keine angemessene Art der Zuverlässigkeitsbestimmung für Verfahren zur Messung rasch veränderlicher Merkmale (z.B. Stimmungen).</p> <p>-zu Anforderung B26: Die Anforderung lautet im Text der DIN: „Wurden die Zuverlässigkeitskennwerte in den letzten 8 Jahren nicht überprüft, muss begründet werden, warum das Verfahren dennoch ausgewählt wird.“ Diese Begründung obliegt den Eignungsdiagnostikern/Eignungsdiagnostikerinnen. Die hier in der Checkliste gewählte Formulierung richtet sich hingegen an die Verantwortlichen für die Verfahrenshinweise.</p>							

		Soll oder Muss	ja	nein	Nicht zu bewerten	Anm.	Quelle
B2 7	In den Verfahrenshinweisen werden Angaben zur Gültigkeit des Verfahrens gemacht, die aus empirischen Studien abgeleitet wurden	M	X				
V1 3	<i>Verzweigungsfrage V13: Wurden Angaben zur Gültigkeit des Verfahrens gemacht?</i>		X				

	-> Falls „nein“ endet die Checkliste mit dem folgenden Hinweis H3						
H3	Hinweis H3: Falls keine empirisch fundierten Angaben zur Gültigkeit des Verfahrens gemacht werden, sind die Anforderung B27 bis B34 und B40 bis B41 mit „nein“ zu beantworten				X		
B28	Aus den Verfahrenshinweisen wird deutlich, welche empirischen Nachweise der Inhalts- und/oder Kriteriums- und/oder Konstruktgültigkeit eine Anwendung des Verfahrens bzw. der Verfahrensklasse für den laut Verfahrenshinweisen intendierten Anwendungszweck rechtfertigen			X			
	In den Verfahrenshinweisen wird angegeben, welche Gültigkeitswerte:				X		
B29	• in Bezug zu welchem Kriterium erzielt wurden	M		X			
B30	• für welche Referenzgruppen erzielt wurden	M		X			
B31	• in welcher Untersuchung erzielt wurden	M	X				
B32	• für welches Verfahrensergebnis erzielt wurden	M		X			
B33	• zu welchem Zeitpunkt erzielt wurden	M	X				
B34	Der aktuellste Nachweis über die Gültigkeit des Verfahrens ist jünger als acht Jahre	S	X				
-zu Anforderung B29: Bewährungskriterium, z.B. Vorgesetztenbeurteilung zum Berufserfolg -zu Anforderung B32: Bezieht sich der Gültigkeitswert beispielsweise auf das Gesamtergebnis oder auf ein Teilergebnis (etwa auf eine einzelne Skala oder einzelne Items)? Bezieht sich der Gültigkeitswert auf einen Rohwert oder auf einen standardisierten Wert?							
		Soll oder Muss	ja	nein	Nicht zu bewerten	Anm.	Quelle
V14	Verzweigungsfrage V14: Wurden zur Bestimmung der Gültigkeit Methoden der			X			



	<i>statistischen Adjustierung/ Optimierung eingesetzt18? Falls „nein“ -&gt; Bitte weiter bei Verzweigungsfrage V15</i>						
	Bei der Dokumentation der Analysen zur Gültigkeit:				X		
B3 5	• sind sowohl die ursprünglich erhaltenen als auch die korrigierten Kennwerte aufgeführt	M			X		
B3 6	• werden alle in Zusammenhang mit der Adjustierung verwendeten Statistiken genannt	M			X		
B3 7	• werden neben den statistisch optimierten Schätzungen auch die einfachen Schätzungen angegeben	M			X		
B3 8	• wurden die optimierten Schätzungen auf eine andere Personengruppe aus dem Geltungsbereich des Verfahrens angewendet und in ihrer Gültigkeit bestätigt (Kreuzvalidierung)	S			X		
B3 9	• sind die statistischen Optimierungen in handlungsleitende Beurteilungsregeln umgesetzt	S			X		
V1 5	<i>Verzweigungsfrage V15: Wird der Gültigkeitsanspruch damit begründet, dass Gültigkeitshinweise aus anderen Untersuchungen in Anspruch genommen werden? Falls „nein“-&gt; Bitte weiter bei Anforderung B42</i>			X			
	Es wird nachvollziehbar aufgeführt:				X		
B4 0	• welche Befunde generalisiert werden können (Darstellung der entsprechenden Studien, Literaturübersichten und Metaanalysen)	S			X		
B4 1	• weshalb (und in welchem Ausmaß) sich die Gültigkeitshinweise übertragen lassen, die sich aus anderen Studien ergeben	S			X		
<p>-zu Verzweigungsfrage V14: (z.B. Minderungskorrektur, Varianzeinschränkungskorrektur, multiple Regression)</p> <p>-zu Anforderung B37: statistisch optimierte Schätzungen sind z.B. multiple Regressionen; einfache Schätzungen sind z.B. einfache Korrelationen</p> <p>-zu Anforderung B39: Wenn beispielsweise gezeigt wird, dass die multiple Vorhersagbarkeit eines Kriteriums unter Einbezug mehrerer Prädiktoren (z.B. mehrere Skalen eines Tests) deutlich höher ist als die einfache Korrelation zwischen einzelnen Prädiktoren und diesem Kriterium, so sollte der Anwenderin/dem Anwender erläutert werden, wie sie/er die verschiedenen Prädiktoren so kombinieren/gewichten kann, dass der Vorteil praktisch nutzbar wird.</p>							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
B4 2	Aufgrund von inhaltlichen Überlegungen wird dargelegt, wie sich das fragliche Konstrukt zu ähnlichen Konstrukten verhält (konvergente Gültigkeit)	S		X			
B4 3	Aufgrund von empirischen Ergebnissen wird dargelegt, wie sich das fragliche Konstrukt zu ähnlichen Konstrukten verhält (konvergente Gültigkeit)	S		X			
B4 4	Aufgrund von inhaltlichen Überlegungen wird dargelegt, wie sich das fragliche Konstrukt zu unähnlichen Konstrukten verhält (diskriminante Gültigkeit)	S		X			
B4 5	Aufgrund von empirischen Untersuchungen wird dargelegt, wie sich das fragliche Konstrukt zu unähnlichen Konstrukten verhält (diskriminante Gültigkeit)	S		X			
B4 6	Bei der Analyse der Kriteriumsgültigkeit des Verfahrens wird beschrieben, warum das in der Analyse jeweils verwendete Kriterium angemessen ist und valide erfasst wurde	M		X			
B4 7	Sowohl die Objektivität als auch die Zuverlässigkeit jedes verwendeten Kriterienmaßes wird nach Möglichkeit dargestellt	S		X			
B4 8	Die Angemessenheit der für die Analyse der Kriteriumsgültigkeit herangezogenen Untersuchungsgruppe wird erläutert	M		X			
-zu Anforderung B48: Beispielsweise sollten die demografischen Merkmale der Untersuchungsgruppe (z.B. Bildungsstand, Alter, Berufserfahrung usw.) vor dem Hintergrund der als Zielgruppe des Verfahrens genannten Gruppe diskutiert werden.							

		Soll oder Mus s	ja	nein	Nicht zu bewerte n	Anm .	Qu ell e
V1 6	<i>Wird für das Verfahren Inhaltsgültigkeit in Anspruch genommen? Falls „nein“ endet die Checkliste an dieser Stelle</i>		X				
B4 9	Der im Verfahren abgebildete Inhaltsbereich wird nachvollziehbar beschrieben	S		X			
B5 0	Die Kriterien zur Beschreibung des dem Verfahren zugrunde liegenden, hypothetischen Itemuniversums sind angegeben	S		X			
B5 1	Die Regeln, nach denen das Verfahren als systematisch zusammengestellte Itemstichprobe aus dem Itemuniversum abgeleitet wurde, werden dargestellt	S		X			
V1 7	<i>Wurde die Frage, ob das Verfahren den definierten Inhaltsbereich repräsentiert, durch Expertinnen/Experten beurteilt? Falls „nein“ endet die Checkliste an dieser Stelle</i>			X			
	Es wird:						
B5 2	• Sowohl der fachbezogene Ausbildungsstand als auch die Erfahrung als auch die Qualifikation der beteiligten Expertinnen/Experten beschrieben	S			X		
B5 3	• erläutert, wie die Expertinnen/Experten zu ihrer Einschätzung gekommen sind	S			X		
B5 4	• angegeben, inwieweit die Expertinnen-/Expertenbeurteilungen übereinstimmen	S			X		

## Eidesstattliche Versicherung

### **Erklärungen**

(1) Ich erkläre, dass ich mich an keiner anderen Hochschule einem Promotionsverfahren unterzogen bzw. eine Promotion begonnen habe.

(2) Ich erkläre, die Angaben wahrheitsgemäß gemacht und die wissenschaftliche Arbeit an keiner anderen wissenschaftlichen Einrichtung zur Erlangung eines akademischen Grades eingereicht zu haben.

(3) Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst habe. Alle Regeln der guten wissenschaftlichen Praxis wurden eingehalten; es wurden keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht.

Ort, Datum