

Beschwerdenvalidierung bei psychosomatischen Patienten

**Inauguraldissertation**  
zur  
Erlangung der Würde  
einer Doktorin der Philosophie  
vorgelegt der  
Fakultät für Psychologie  
der Universität Basel

von

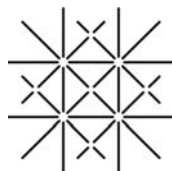
Carole Keppler, M.Sc.

aus Basel-Stadt

Basel, Mai 2016

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel

[edoc.unibas.ch](http://edoc.unibas.ch)



UNI  
BASEL

Genehmigt von der Fakultät für Psychologie

auf Antrag von

Prof. Dr. rer. nat. Dipl.-Psych. Rolf-Dieter Stieglitz

Prof. Dr. med. Ralph Mager

Basel, den 12. Dezember 2016

---

Prof. Dr. phil. Roselind Lieb



## ERKLÄRUNG ÜBER DIE SELBSTÄNDIGKEIT

Die zur Promotion eingereichten Zeitschriftenbeiträge wurden in Zusammenarbeit mit den jeweiligen Koautoren angefertigt. Es handelt sich um zwei Originalarbeiten und einen Übersichtsartikel, die zur Veröffentlichung an die Zeitschriften eingereicht, bzw. veröffentlicht wurden. Die Arbeiten wurden weder von den beteiligten Autoren noch von anderen Personen an anderer Stelle veröffentlicht. Es wurden nur die angegebenen Hilfsmittel verwendet und die Zitate gekennzeichnet.

Die folgenden Publikationen sind in die vorliegende Dissertation eingegangen und in der unten aufgeführten Reihenfolge im Anhang A-1 bis A-3 angefügt:

### **Publikation 1**

Keppler, C., Plohmann, A., Pflüger, M., Rabovsky K., Delahaye, M., Langewitz, W. & Mager, R. (eingereicht). Beschwerdenvalidierung in der versicherungsmedizinischen Begutachtung. *Fortschritte der Neurologie Psychiatrie*.

### **Publikation 2**

Keppler, C., Rosburg, T., Lemoine, P., Pflüger, M., Gyr, N. & Mager, R. (akzeptiert). Functional somatic syndromes: skin temperatures and activity measurements under ambulatory conditions. *Applied Psychophysiology and Biofeedback*.

### **Publikation 3**

Delahaye, M., Stieglitz, R., Graf, M., Keppler, C., Maes, J. & Pflüger, M. (2015). Deutsche Übersetzung und Validierung des Stress Appraisal Measure (SAM). *Fortschritte der Neurologie Psychiatrie*, 83, 276-283.

Basel, Mai 2016

Carole Keppler

## INHALTSVERZEICHNIS

<b>ERKLÄRUNG ÜBER DIE SELBSTÄNDIGKEIT</b> .....	<b>1</b>
<b>ZUSAMMENFASSUNG</b> .....	<b>3</b>
<b>1 EINLEITUNG</b> .....	<b>4</b>
<b>2 BESCHWERDENVALIDIERUNG IN DER VERSICHERUNGSMEDIZINISCHEN BEGUTACHTUNG</b> .....	<b>8</b>
2.1 EINLEITUNG .....	8
2.2 THEORETISCHER HINTERGRUND .....	8
2.3 METHODEN .....	9
2.4 ERGEBNISSE UND DISKUSSION.....	9
<b>3 FUNCTIONAL SOMATIC SYNDROMES: SKIN TEMPERATURES AND ACTIVITY MEASUREMENTS UNDER AMBULATORY CONDITIONS</b> .....	<b>17</b>
3.1 EINLEITUNG .....	17
3.2 THEORETISCHER HINTERGRUND .....	17
3.3 METHODEN .....	18
3.4 ERGEBNISSE UND DISKUSSION.....	19
<b>4 DEUTSCHE ÜBERSETZUNG UND VALIDIERUNG DES STRESS APPRAISAL MEASURE (SAM)</b> .....	<b>22</b>
4.1 EINLEITUNG .....	22
4.2 THEORETISCHER HINTERGRUND .....	22
4.3 METHODEN .....	23
4.4 ERGEBNISSE UND DISKUSSION.....	24
<b>5 ALLGEMEINES FAZIT</b> .....	<b>27</b>
<b>6 REFERENZEN</b> .....	<b>28</b>
<b>ANHÄNGE</b> .....	<b>39</b>

## ZUSAMMENFASSUNG

Aufgrund hoher Prävalenzraten nicht-authentischer Beschwerden in der versicherungsmedizinischen Begutachtung und damit verbunden das Risiko nicht zielkonform eingesetzter, limitierter geldwerter Leistungen, befasst sich diese kumulative Dissertation mit der Beschwerdenvalidierung, insbesondere bei psychosomatischen Patienten.

In der **ersten Publikation** werden deshalb die best-evaluierten Methoden und Instrumente zur Beschwerdenvalidierung in der Begutachtung vorgestellt. Dabei wird versucht, deren Validität insbesondere in Bezug auf die Anwendung im Begutachtungskontext nach der aktuellen Datenlage darzustellen. Wir kommen zu dem Schluss, dass eine lege artis durchgeführte strukturierte Beschwerdenvalidierung die Qualität klinischer Urteile des Gutachters erheblich verbessern kann. Die qualifizierte Integration dieser zusätzlich erhobenen Befunde in eine psychologische oder ärztliche Expertise ist jedoch anspruchsvoll und bedarf einer weiteren wissenschaftlichen Begleitung.

Die **zweite Publikation** validierte physiologische Beschwerden in einer empirischen Studie mit Patienten mit funktionellen somatischen Syndromen. Bei dieser Patientengruppe ist die distale Hauttemperatur während Basislinie-Messungen (ohne Stressinduktion) normalerweise tiefer im Vergleich zu Kontrollprobanden, was durch die dauernde Präsenz von schmerzbedingtem Stress bedingt sein könnte. Das Ziel der Studie war es zu untersuchen, ob diese Patienten veränderte Hauttemperaturen auch unter Alltagsbedingungen aufweisen. Es konnte eine Dissoziation zwischen objektiven Hauttemperaturmessungen und subjektiven Einschätzungen des Körperwärmekomforts beobachtet werden, was die Theorie eines Alexisomie-Anteils (reduzierte Körperwahrnehmung) bei funktionellen somatischen Syndromen unterstützt. Ausserdem zeigen die Resultate, dass subjektive Klagen über Müdigkeit und Anspannung nicht unbedingt in physischem Vermeidungsverhalten resultiert.

Der beschwerdenbedingte Stress bei psychosomatischen Patienten kann im Rahmen von Stressbewältigungstrainings angegangen werden. Stressfragebögen können dabei die Trainingsfortschritte evaluieren. Deshalb wurde in der **dritten Publikation** der Stressfragebogen Stress Appraisal Measure (SAM), ein kurzer Fragebogen zur Bewertung eines aktuellen, stressauslösenden Ereignisses, ins Deutsche übersetzt und validiert. Unsere Studie konnte die Ergebnisse der Originalstudie im Wesentlichen replizieren, weshalb der SAM nun auch in deutscher Sprache eingesetzt werden könnte.

## 1 EINLEITUNG

„Psychosomatik bedeutet, dass Körper und Seele zwei untrennbar miteinander verbundene Aspekte des Menschen sind, die nur aus methodischen Gründen oder zum besseren Verständnis unterschieden werden. (...) Ein einheitliches Modell für die Wechselwirkungen zwischen Körper, psychischen Prozessen und Umwelt existiert nicht. Meist werden Teilaspekte beschrieben, die von unterschiedlichen Theorien aufgenommen werden.“ (Schweickhardt, Fritzsche, & Wirsching, 2005, S.5 und 7).

Psychosomatische Erkrankungen werden im Diagnosesystem ICD-10 als somatoforme Störungen bezeichnet und werden wie folgt definiert:

- Wiederholte Darbietung körperlicher Symptome
- Hartnäckige Forderung nach medizinischen Untersuchungen trotz wiederholter negativer Ergebnisse und Versicherung der Ärzte, dass die Symptome nicht körperlich begründbar sind
- Bei Vorhandensein von somatischen Störungen, wird die Art und das Ausmass der Symptome, das Leiden und die innerliche Beteiligung des Patienten dadurch nicht erklärt

Wie ersichtlich, ist das Vorhandensein eines medizinischen Krankheitsfaktors an sich kein Ausschlusskriterium für eine somatoforme Störung. Die somatoformen Störungen werden in verschiedene Diagnosen unterteilt: Somatisierungsstörung, undifferenzierte Somatisierungsstörung, hypochondrische Störung, somatoforme autonome Funktionsstörung, anhaltende somatoforme Schmerzstörung, sonstige somatoforme Störungen und die nicht näher bezeichnete somatoforme Störung. Die Bezeichnung „somatoform“ hat sich in vielen Bereichen der somatischen Medizin nicht durchgesetzt, es wird dort von funktionellen Störungen gesprochen. Diese beziehen sich auf die betroffenen Körperfunktionen und werden in der jeweiligen ICD-10-Sektion klassifiziert. Die funktionellen Störungen sind vor allem für meine zweite Publikation von Relevanz. Die Bekanntesten sind das Reizdarmsyndrom, das Fibromyalgiesyndrom, der chronische Rückenschmerz, die nichtulzeröse Dyspepsie, der Spannungskopfschmerz, der psychogene Schwindel, das prämenstruelle Syndrom, der nichtkardiale Brustschmerz sowie das chronische Erschöpfungssyndrom (Gaab, 2014; Henningsen, Zipfel, & Herzog, 2007).

Folgende Merkmale zeichnen eine funktionelle Störung aus:

- Sie resultiert aus Störungen von Körperfunktionen, ohne Nachweis einer organpathologischen Veränderung
- Seelische Faktoren lösen diese Symptomatik aus und tragen zur Aufrechterhaltung bei

Aus diesen Beschreibungen wird ersichtlich, dass eine klare Abgrenzung fehlt und eine deutliche Überlappung besteht zwischen den verschiedenen diagnostischen Herangehensweisen, was sich diagnostisch als schwierig erweist (Aaron & Buchwald, 2001; Escobar, Interian, Diaz-Martinez, & Gara, 2006).

Die somatoformen Störungen werden im neuen DSM-5 (APA, 2013) als somatische Belastungsstörung und verwandte Störungen (Somatic Symptom and Related Disorders) aufgeführt. Dabei wurde das Konzept der somatoformen Störung abgeschafft. Die Somatisierungsstörung, die undifferenzierte somatoforme Störung und die Schmerzstörung wurden zur somatischen Belastungsstörung zusammengefasst. Als zentrales Merkmal der somatischen Belastungsstörung gilt das Vorhandensein von körperlichen Symptomen mit einhergehend klinisch bedeutsamer subjektiver Belastung durch die Beschwerden. Dabei findet keine Unterscheidung mehr zwischen somatoformen (medizinisch unerklärt) und körperlichen Symptomen (organmedizinisch begründet) statt. Neue Einschlusskriterien sind hingegen psychologische Faktoren. Als verwandte Störung gilt nach wie vor die Konversionsstörung, die Hypochondrie wurde jedoch durch die Krankheitsangststörung (Illness Anxiety Disorder) ersetzt. Die vorge-täuschte Störung und die psychologischen Faktoren, die eine medizinische Krankheit beeinflussen (z.B. Asthma bronchiale, rheumatoide Arthritis, Neurodermitis u.a.), welche sich vorher in anderen Kapiteln befanden, wurden nun ebenfalls unter dieses neue Kapitel gestellt. Kritiker bemängeln, dass durch die fehlenden differentialdiagnostischen Kriterien keine gezielte Differentialdiagnostik mehr möglich ist oder eine störungsspezifische Therapieplanung erschwert wird. Zudem stellt die neue Diagnose eine Art Sammeltopf für unterschiedliche Störungsvarianten dar (Hiller & Rief, 2014).

De Waal, Arnold, Eekhof und van Hemert (2004) berichten eine Prävalenz somatoformer Störungen (nach DSM-IV) unter Hausarztpatienten von 16 %. Bei Fink, Sorensen, Engberg, Holm und Munk-Jorgensen (1999) erfüllen 22 % der Hausarztpatienten die Kriterien einer somatoformen Störung nach ICD-10 (ohne Berücksichtigung der Diagnose „Somatoforme Störung, nicht näher bezeichnet“), und 58 % nach DSM-IV. Smith et al. (2005) berichten hingegen von einer deutlich niedrigeren Prävalenz von 4,4 % und Creed und Barsky (2004) von einer Prävalenz zwischen 0,03 % und 0,84 % für Somatisierungsstörungen. Diese Unterschiede zwischen den Studien, aber auch innerhalb ein und derselben Studie, deuten auf einen weiteren Schwachpunkt der aktuellen Diagnosekriterien hin.

Zahlreiche Studien sowie systematische Reviews (Creed & Barsky, 2004; Kroenke et al., 1994; Leiknes, Finset, Moum, & Sandanger, 2007; Mergl et al., 2007) konnten eine hohe Komorbidität zwischen somatoformen Beschwerden und depressiven Störungen und Angst-

störungen nachweisen, welche zwischen 30 und 70% liegt. Die Komorbidität mit muskuloskelettalen Störungen bei schwer ausgeprägten somatoformen Störungen liegt bei etwa 43% (Leiknes et al., 2007). Auch Harris, Orav, Bates und Barsky (2009) berichten von einer Rate von 39 % bei Patienten, die neben einer somatoformen Störung auch unter einer körperlichen Erkrankung leiden. Jedoch unterschied sich diese nicht signifikant von der bei Patienten ohne somatoforme Störungen (30 %).

Patienten mit somatoformen Störungen scheinen ihren Hausarzt häufiger aufzusuchen als Patienten ohne psychische Störung (Barsky, Orav, & Bates, 2005; Hanel et al., 2009) und generell mehr nicht-psychiatrische Gesundheitsdienste in Anspruch zu nehmen (Fink et al., 1999). Auch die fachärztliche Versorgung und Notfall- und stationäre Aufenthalte werden häufiger aufgesucht (Barsky et al., 2005; Barsky, Orav, & Bates, 2006). Dies könnte durch Interaktionsschwierigkeiten mit dem Hausarzt (Verunsicherung, Weiterverweisung) und einer Suche nach Rückversicherung begründet sein. Patienten mit somatoformen Beschwerden erhalten meist umfangreiche und aufwendige diagnostische Untersuchungen, auch nachdem die wichtigsten und möglicherweise gravierendsten Ursachen ausgeschlossen wurden. Die Inanspruchnahme von Diensten des Gesundheitssystems erhöht die Kosten, die auch dann konstant bleiben, wenn für Komorbidität kontrolliert wird (Barsky et al., 2005). Zudem stellt sie eine der behavioralen Merkmale somatoformer Beschwerden dar. Körperliche und somatoforme Symptome können sich jedoch auch in Form von Vermeidungsverhalten oder Rückversicherungssuche manifestieren. Ärztliche Rückversicherung bei Auftreten neuer unbekannter Körpersymptome ist nicht nur ein Bedürfnis des Patienten, sondern auch eine Strategie der Ärzte. Diese Strategie kann jedoch an Wirksamkeit verlieren, wenn Beschwerden chronifizieren und langfristig sogar zur Aufrechterhaltung der Beschwerden beitragen (Rief, Heitmüller, Reisberg, & Ruddel, 2006).

Beschwerdenvalidierung bezeichnet den Prozess, der zur Beurteilung der Beschwerdenuvalidität führt, und somit die Überprüfung der Authentizität von Beschwerdendarstellung und Symptompräsentation (Merten, 2011). Invalide Beschwerdendarstellung bzw. Antwortverzerrung bezeichnet ein Verhalten, das durch unzutreffende Antworten, fälschliche Angaben oder durch ein Testverhalten gekennzeichnet ist, welches nicht den tatsächlichen Möglichkeiten einer Person entspricht.

Die generelle Prävalenz von nicht-authentischen psychischen Störungen schwankt zwischen 10 und 30%, je nach Untersuchungskontext und Methodik (Rogers, 2008). Larrabee (2003) schätzte anhand einer Analyse von 11 empirischen Studien, dass in etwa 40% der Begutach-



tungsfälle relevante Antwortverzerrungen vorkommen. Gemäss einer Erhebung unter amerikanischen Neuropsychologen auf der Basis von 30000 Begutachtungen wurden bei 32.7% aller Fälle von Invaliditätsansprüchen „wahrscheinliche nicht-authentische kognitive Beschwerden“ (nach Slick, Sherman, & Iverson, 1999) vermutet, im Bereich Personenschaden bei 30.4%. Im Rahmen von forensischen Begutachtungen belief sich dieser Wert auf immerhin noch 22.7%, bei neuropsychologischen Abklärungen im Rahmen medizinischer oder psychiatrischer Untersuchungen ausserhalb eines juristischen Kontextes auf 8.1%. Von den Exploranden, die sich im Rechtsstreit befanden bzw. Schadensersatzansprüche geltend machten, standen nicht nur 41.2% der Patienten nach leichten Schädelhirntrauma unter Täuschungsverdacht, sondern auch ein erheblicher Anteil von Gutachtenspatienten, welche psychische Störungen wie Fibromyalgie oder ein chronisches Erschöpfungssyndrom (38.6%), Schmerzen oder somatoforme Störungen (33.5%), depressive Störungen (16.1%), Angststörungen (13.6%) oder dissoziative Störungen (10.5%) beklagten. Danach scheint die Auftretenshäufigkeit negativer Antwortverzerrung wesentlich vom Untersuchungskontext und den geltend gemachten Störungen bzw. den gestellten Diagnosen abhängig zu sein (Mittenberg, Patton, Canyock, & Condit, 2002). In Deutschland wurden ähnlich hohe Prävalenzraten für Gutachtenpopulationen gefunden. Merten, Friedel und Stevens (2006) und Stevens, Friedel, Mehren und Merten (2008) konnten jeweils suboptimale Leistungsanstrengungen in Performanzvalidierungsverfahren<sup>1</sup> bei über 45% aller untersuchten Probanden finden, die wegen psychiatrischer und neurologischer Störungen begutachtet wurden. Nach empirischen Schätzungen von Greve, Ord, Bianchini und Curtis (2009) liegt die Prävalenz von Täuschungsverhalten bei Patienten mit chronischen Schmerzen im medico-legalen Kontext je nach Klassifikationssystem zwischen 25% und 36%. 21% dieser nicht-authentischen Probanden wiesen objektive somatische Befunde auf. Auch bei anderen somatischen Krankheitsbildern finden sich Hinweise auf Aggravation oder Simulation, so z.B. bei 58% geltend gemachter Hörstörungen (Strebel & Brusis, 2010).

Hohe Prävalenzraten von nicht-authentischen Beschwerden und eine häufige Inanspruchnahme von umfangreichen und aufwendigen Diensten des Gesundheitssystems, welche die Kosten steigen lassen, legitimieren die Beschwerdengültigkeit, insbesondere im versicherungsmedizinischen Bereich.

---

<sup>1</sup> In der internationalen Literatur etabliert sich zunehmend eine sprachliche Unterscheidung zwischen Performanzvalidierungstests, die die Anstrengungsbereitschaft von Probanden in Leistungstests überprüfen und Beschwerdengültigkeitstests, mit denen die Beschwerdenschilderung überprüft werden soll.

## **2 BESCHWERDENVALIDIERUNG IN DER VERSICHERUNGS-MEDIZINISCHEN BEGUTACHTUNG**

### **2.1 Einleitung**

In der ersten Publikation werden standardisierte Beschwerdenvalidierungsverfahren vorgestellt und methodisch kurz evaluiert in ihrer Möglichkeit, angegebene Beschwerden im psychologisch-psychiatrischen Bereich gutachterlich nachzuvollziehen, um verbleibende Funktionen besser einschätzen zu können.

### **2.2 Theoretischer Hintergrund**

In Europa zeigen psychische und psychosomatische Störungen ähnlich wie in anderen Industrienationen eine stark ansteigende Bedeutung für die Verursachung von Absenzen am Arbeitsplatz und generieren mittlerweile einen führenden Anteil der entstehenden Versicherungszahlungen im Erwerbsalter (Huber, 2000; Schneider et al., 2010; Stadtland, Gündel, Schütt, & Nedopil, 2003). Zu welchem Anteil beruhen Arbeitsunfähigkeit und Ausgleichszahlungen wirklich auf tatsächlich bestehenden, also nicht vorgetäuschten Gesundheitsstörungen? Trotz der fortschreitenden Erkenntnisse in Medizin, Psychologie und Biologie ist der Zugang zu grossen Anteilen der Beschwerden weiterhin ein sprachlicher, der wesentlich bis ausschliesslich auf die subjektiven Angaben des Patienten zurückgeht. Das Risiko einer Fehleinschätzung der Authentizität von Symptomen und deren Zusammenhang zu Funktionseinschränkungen ist also in gewissem Sinne medizinimmanent und betrifft insbesondere psychiatrische, aber auch somatische Expertisen. Störungsbilder, bei welchen ein erheblicher Ermessensspielraum für die Einschätzung der Erwerbsunfähigkeit bestehen kann, sind Depressionen, posttraumatische Belastungsstörungen, Schmerzerkrankungen, Zustände nach HWS-Distorsionen, Tinnitus, fragliche Anfallsereignisse, nicht prüfbare Störungen des Seh-/Hörvermögens sowie Sensibilitätsstörungen. Ein Gutachter muss die Validität der geltend gemachten Beschwerden sorgfältig überprüfen und darf sich in der Beurteilung der Glaubwürdigkeit von Beschwerdenangaben nicht allein auf seine klinische Erfahrung stützen (Aamondt & Custer, 2006; Bartlett, Littlewort, Frank, & Lee, 2014; Ekman & O'Sullivan, 1991; Trueblood & Binder, 1997). Auf diesem Hintergrund ergibt sich die Notwendigkeit, Standards für die Beurteilung der Authentizität von Beschwerden zu entwickeln und diese im Gutachtensprozess zu verankern. In der ersten Publikation werden diese Standards vorgestellt.

### 2.3 Methoden

In den Suchmaschinen Pubmed, PsycINFO und PSYINDEXplus sowie direkt in bestimmten Zeitschriften wie Versicherungsmedizin wurde Literatur anhand der Schlüsselwörter „Beschwerdenvalidierung“, „Performanzvalidierung“, „Begutachtung“, „Versicherungsmedizin“ und im Englischen anhand „symptom validity testing“, „performance validity“, „independent medical evaluation“, „insurance medicine“ gesucht. Es stellte dabei keine abschliessende Suche dar. Bei den best-evaluierten Methoden und Instrumenten wurde deren Validität in Bezug auf die Anwendung im Begutachtungskontext untersucht. Der Fokus lag dabei auch auf deutschsprachigen Verfahren.

### 2.4 Ergebnisse und Diskussion

Die Tabelle 1 am Ende des Kapitels fasst die best-evaluierten Methoden und Instrumente zur Beschwerdenvalidierung in der Begutachtung zusammen. Daraus ist ersichtlich, dass es drei Kategorien von Methoden gibt: 1. Verfahren zur Validierung der Anstrengungsbereitschaft, 2. Verfahren zur Validierung der Beschwerdenschilderung und 3. Nachweis bewusster Täuschungsabsicht.

Bei den Verfahren zur Validierung der Anstrengungsbereitschaft wird zwischen Performanzvalidierungsverfahren und Beschwerdenvalidierungsparameter in neuropsychologischen Tests unterschieden. Die Performanzvalidierungsverfahren weisen die ausführlichste Validierung und die höchste Klassifikationsgenauigkeit sowie eine grosse empirische Datenbasis auf, was eine Reihe von Vergleichsdaten zu verschiedenen Explorandengruppen beinhaltet. Jedoch ist die Sensitivität zur Aufdeckung von negativer Antwortverzerrung in anderen Funktionsbereichen als Gedächtnisdefizite gering und Interkorrelationen sind vermehrt möglich. Beispiele für Performanzvalidierungsverfahren sind der Word Memory Test, der Test of Malingered Memory oder der Amsterdamer Kurzzeitgedächtnistest, wobei die Alternativwahlverfahren als die Methode der Wahl gelten. Die Vorteile bei Beschwerdenvalidierungsparameter in neuropsychologischen Tests liegen darin, dass diese meist direkt aus den Verfahren gewonnen werden können und somit eine direkte Information über die Gültigkeit der gezeigten Leistung darstellen und sie gegenüber Coaching weniger anfällig sind. Hingegen weisen sie eine niedrigere Robustheit gegenüber realen kognitiven Defiziten auf, es besteht die Gefahr einer falsch-positiven Klassifikation und der Einsatz erfordert eine detaillierte Kenntnis der Originalliteratur. Die Reliable Digit Span und das Finger Tapping stellen Beschwerdenvalidierungsparameter dar.

Bei der zweiten Kategorie, den Verfahren zur Validierung der Beschwerdenschilderung, gibt es auf der einen Seite die Fremdbeurteilungsverfahren für psychische Störungen und auf der anderen Seite die Selbstbeurteilungsverfahren und andere Screeningtests für psychische Störungen. Bei den Fremdbeurteilungsverfahren stehen als gut validierte Instrumente das Structured Interview of Reported Symptoms und der Miller Forensic Assessment of Symptoms Test zur Verfügung, wobei aktuell beide Verfahren noch nicht in einer deutschen Version vorliegen, jedoch das Erstere in Vorbereitung ist. Das Structured Interview of Reported Symptoms weist eine hervorragende Interrater-Reliabilität und eine sehr gute Retestreliabilität auf, wodurch das Interview als Referenzstandard gilt. Jedoch erfordert das Interview eine Expertise des Gutachters und Zeit für die Durchführung. Das Screeningverfahren Miller Forensic Assessment of Symptoms Test weist in forensischen Populationen eine sehr gute Reliabilität und Validität auf. Zudem ist die Durchführungs- und Auswertungszeit kurz. Bei den Selbstbeurteilungsverfahren können der Strukturierte Fragebogen simulierter Symptome als deutsche Version des Structured Inventory of Malingered Symptomatology, das Minnesota Multiphasic Personality Inventory, das Personality Assessment Inventory mit der deutschen Version des Verhaltens- und Erlebensinventar sowie der Screeningtest Morel Emotional Numbing Test for Posttraumatic Stress Disorder als validierte Verfahren genannt werden. Der Strukturierte Fragebogen simulierter Symptome weist eine hohe Retestreliabilität, eine akzeptable interne Konsistenz sowie eine gute vorhersagende und übereinstimmende Validität auf. In Abhängigkeit des gewählten Grenzwertes erreichen die Spezifität und die Sensitivität einen Wert über 0.85. Der Fragebogen eignet sich jedoch nur als Screeningverfahren. Das Minnesota Multiphasic Personality Inventory kann Antwortverzerrungen in psychischen, körperlichen und kognitiven Störungen aufdecken und weist hohe Spezifitäten im forensischen Kontext auf. Zudem gibt es eine deutsche Version. Das Selbstbeurteilungsinstrument ist aber sehr umfangreich und die Mehrheit der Validitätsskalen sind nicht ein Bestandteil der deutschen Version, was eine manuelle Auswertung anhand der Literatur erfordert. Dies ist aber letztendlich auf die Validierung bezogen wieder ein Vorteil, da störungsspezifische Trennwerte vorliegen. Das Personality Assessment Inventory weist eine hohe Reliabilität und gute Validität in verschiedenen Settings auf, ebenso hohe Effektstärken der Validitätsskalen. Zudem ist es weit weniger umfangreich wie das Minnesota Multiphasic Personality Inventory. Das Verhaltens- und Erlebensinventar entspricht weitgehend den amerikanischen Normen des Personality Assessment Inventory. Es bestehen akzeptable Schätzungen der Sensitivität und Spezifität der Skalen für negative Antwortverzerrung. Jedoch steht für die deutschsprachige Patientengruppe eine Adaption der ursprünglichen Trennwerte zur Diskussion. Der Morel Emotional Num-

bing Test for Posttraumatic Stress Disorder ist ein forced-choice Verfahren, welches als Screeningverfahren fungiert. Er weist eine hohe Spezifität und eine gute Sensitivität auf. Zudem existiert das Verfahren in mehreren Sprachen. Zusammengefasst stehen uns an hiesigen Gutachtenspopulationen validierten Verfahren derzeit nur der Strukturierte Fragebogen simulierter Symptome und der Morel Emotional Numbing Test for Posttraumatic Stress Disorder als Screeningverfahren sowie das Minnesota Multiphasic Personality Inventory und mit noch dürftiger Datenlage das Verhaltens- und Erlebensinventar zur Verfügung.

Die letzte Kategorie beinhaltet den Nachweis bewusster Täuschungsabsicht, welcher anhand von Kriterienkatalogen erbracht werden kann. Dabei stellen diese Kriterienkataloge einen Prozess dar, welcher eine sorgfältige Analyse aller verfügbaren Informationen erfordert und so das klinische Urteil mit Ergebnissen wissenschaftlich validierter Verfahren kombiniert. Es bestehen Kriterienkataloge für drei Störungsbilder, welche ein aufwendiges Verfahren gemein haben: aggravierte/simulierte neurokognitive Störungen, aggravierte/simulierte schmerzassoziierte Behinderungen und nicht-authentische posttraumatische Belastungsstörungen. Ursprünglich waren alle drei Kriterienkataloge gleich aufgebaut. Es mussten A) Evidenz eines externen Anreizes, B) Evidenzen aus Testbefunden, C) Evidenzen aus Inkonsistenzen/Diskrepanzen (aus Angaben des Exploranden, Validitätsskalen von Selbstbeurteilungsinstrumenten etc.) und D) Ausschlusskriterien (psychiatrische, neurologische oder Entwicklungsfaktoren) vorhanden sein. Bei den schmerzassoziierten Behinderungen kommt noch ein weiteres Kriterium dazu, welches Evidenzen aus somatischen Untersuchungsmethoden beinhaltet. Mittlerweile gibt es im Bereich der neurokognitiven Störungen Modifikationen.

Die zentrale Aufgabe in der gutachterlichen Arbeit besteht darin, eine subjektiv vorgebrachte Beschwerdensymptomatik in Abgleich mit medizinischen Befunden in ihrer Relevanz zur Beurteilung von Funktionseinbußen zu gewichten. Hierdurch erklärt sich neben der üblichen klinischen Konsistenzprüfung der Bedarf nach einer reliablen, gut validierten Methodik, um mehr „Objektivität“ zu erreichen in einem Feld von oft divergierenden Interessen und Haltungen.

Nach dem aktuellen Stand der Literatur stellen die Überprüfung der Anstrengungsbereitschaft (Performanzvalidierung) und die Validierung der Beschwerdenschilderung in diesem Sinne zwei Bausteine der versicherungsmedizinischen Begutachtung dar, die einen wichtigen Mehrwert generieren und die insbesondere in ihrer Kombination aussagekräftig sind (Bianchini, Etherton, Greve, Heinly, & Meyers, 2008; Larrabee, 2007a). Der zentrale Vorteil

dieser ergänzenden Verfahren besteht in der inhaltlichen und zufallskritischen Absicherung von Informationen.

Es wird allgemein empfohlen, sich auf mehrere Indikatoren zur Beschwerdenvalidierung abzustützen, da eine singuläre Anwendung zu Fehleinschätzungen aufgrund von falsch-positiv Raten führen kann. Dabei sollten die eingesetzten Verfahren inhaltlich möglichst unabhängig voneinander sein und idealerweise aus unterschiedlichen Funktionsbereichen stammen, damit sie wirklich einen Zugewinn an Information darstellen. Mit einer solchen Kombination von Validitätsparametern mit jeweils hoher Spezifität kann die diagnostische Sicherheit beträchtlich erhöht und die Rate falsch-positiver Klassifikationen gegen Null gesenkt werden (Larrabee, 2003). Ein weiteres Argument für die Kombination mehrerer Validitätsmasse besteht darin, dass Probanden bei der Vortäuschung psychopathologischer, kognitiver und körperlicher Beeinträchtigungen recht unterschiedliche Strategien verfolgen (Greiffenstein, Baker, & Gola, 1994; Greve et al., 2009; Nelson, Sweet, Berry, Bryant, & Granacher, 2007). Zudem lassen sich mit diesem Ansatz auch die Einflüsse von Coaching und erlernter Test-Prozeduren bei wiederholter Begutachtung in Grenzen halten (Blaskewitz & Merten, 2007).. Bei der fachlich korrekten Überführung der erhobenen Validierungsdaten in das Gesamtgutachten ist ein hohes Mass an Expertise und Sorgfalt notwendig und es sollten folgende Punkte bedacht werden:

1. Es gibt vielfältige, potentielle Ursachen für negative Antwortverzerrung und für die Annahme eines bewussten Täuschungsverhaltens (Aggravation/Simulation) bedarf es weiterer zu erfüllender Kriterien (Kriterienkataloge).
2. Auch Exploranden mit genuinen psychopathologischen Auffälligkeiten, neuropsychologischen Defiziten, chronischen Schmerzen und/oder organischen Befunden können aggravieren (Merten, Thies, Schneider, & Stevens, 2009).
3. Eine unzureichende Anstrengungsbereitschaft und eine invalide Selbstbeschreibung in einem Selbstbeurteilungsinventar können störungsimmanent sein. Exploranden mit mittelschweren oder schweren kognitiven Störungen können trotz bestmöglicher Kooperation falsch-positive Resultate in Performanzvalidierungsverfahren erzielen.
4. Parallel zu einem nachweisbaren externalen Anreiz kann durchaus eine davon unabhängige Motivation zur Einnahme der Krankenrolle bestehen, die weniger der bewussten Kontrolle unterliegt (z. B. bei chronischer Erschöpfung, somatoformen Störungen oder auch Schmerzverarbeitungsstörungen). Im Hinblick auf psychische Störungen allgemein liegen auf der anderen Seite aber bis dato keine empirischen Belege dafür vor, dass psychische Erkrankungen in leicht bis mittelgradiger Ausprägung bei vorhandener Anstrengungsbereitschaft krank-

heitsbezogen zu auffälligen Resultaten in Performanzvalidierungstests führen. Bislang unterliegt es häufig dem Expertenurteil des Gutachters, ob ein Explorand auf einem gedachten Kontinuum eher dem Pol der wenig bewusstseinsnahen Selbsttäuschung oder aber dem Pol der Täuschung anderer im Sinne einer bewussten Aggravation zuzuordnen ist (Merten, 2002, 2011). Dies gilt neben den häufig in der Diskussion im Vordergrund stehenden neurotischen und psychosomatischen Störungen in gleicher Weise aber auch für depressive Störungen und für Traumafolgestörungen.

In der versicherungsmedizinischen Begutachtung empfiehlt es sich, möglichst verschiedene standardisierte Verfahren zur Beschwerdenvalidierung heranzuziehen. Grundlage der versicherungsmedizinischen Beurteilung ist unverändert die Integration aller vorliegenden Informationen und der empirisch geleiteten Konsistenzprüfung, insbesondere auch aus der Exploration selbst heraus. Beschwerdenvalidierungsverfahren als Korrelat zu diesen Erkenntnissen ergeben jedoch aufgrund ihrer Aussagen zur Gültigkeit einer Testleistung und Zuverlässigkeit der Beschwerdenschilderung bei bekannten falsch-positiv Raten unzweifelhaft eine qualitative Verbesserung der Befundlage und erhöhen somit die Validität der getroffenen Aussagen. Dieser Sachlage wird auch in der Entwicklung jüngerer gutachterlicher Leitlinien Rechnung getragen (Schneider et al., 2016). Die Integration dieser zusätzlich erhobenen Befunde erfolgt aktuell jedoch noch wenig strukturiert und uneinheitlich. Hier besteht erheblicher Entwicklungsbedarf.

**Tab. 1** Methoden zur Beschwerdenvalidierung

	Grundprinzipien/ Detektionsstrategien	Verfahren	Kurze Übersicht	Verfahrens- beispiele	Vorteile	Nachteile	Literatur
Verfahren zur Validierung der Anstrengungsbereitschaft	<ul style="list-style-type: none"> <li>Alternativwahlverfahren</li> <li>Testdeckeneffekt oder Prinzip der verdeckten Leichtigkeit</li> <li>Leistungskurve</li> <li>Fehlergrösse</li> <li>Identifikation untypischer oder inkonsistenter Leistungsprofile</li> </ul>	Performanzvalidierungstests (PVT)	<ul style="list-style-type: none"> <li>Höchste Klassifikationsgenauigkeit</li> <li>Fokus auf vorgetäuschte Gedächtnisdefizite &gt; hohe Robustheit gegenüber realen kognitiven Defiziten</li> <li>Grosse empirische Datenbasis</li> <li>Ausführliche Validierung</li> <li>Geringe Sensitivität zur Aufdeckung von negativer Antwortverzerrung in anderen Funktionsbereichen</li> <li>Erhöhte Interkorrelationen möglich</li> <li>Soziodemographische Faktoren (Alter, Geschlecht, Bildungsstand) spielen kaum eine Rolle</li> </ul>	Word Memory Test (WMT)	<ul style="list-style-type: none"> <li>Alle Hartman-Kriterien erfüllt</li> <li>Hohe Sensitivität</li> </ul>	<ul style="list-style-type: none"> <li>Bei einzelnen Patientengruppen reduzierte Spezifität</li> <li>Relativ hoher Zeitaufwand &gt; 2 kürzere Verfahren entwickelt (MSVT, NV-MSVT)</li> </ul>	(Hartman, 2002), (Rogers, Harrell, & Liff, 1993), (Gorissen, Sanz, & Schmand, 2005; P. Green, 2003, revised 2005; P. Green, 2005, 2008), (Merten, 2014)
		Test of Malingered Memory (TOMM)	<ul style="list-style-type: none"> <li>Hohe Augenscheinvalidität</li> <li>Sprachfreies Material</li> </ul>	<ul style="list-style-type: none"> <li>Geringe Sensitivität</li> <li>Anfälligkeit gegenüber Coaching</li> </ul>	(Rogers et al., 1993; Tombaugh, 1996), (Merten, 2014)		
		Beschwerdenvalidierungsparameter in neuropsychologischen Tests	<ul style="list-style-type: none"> <li>Parameter können direkt aus Verfahren gewonnen werden</li> <li>Ggf. via zusätzlicher Untertests</li> <li>Geringere Robustheit gegenüber realen kognitiven Defiziten</li> <li>Überprüfung einer Vielzahl von Funktionsbereichen möglich</li> </ul>	Amsterdamer Kurzzeitgedächtnistest (AKGT)	<ul style="list-style-type: none"> <li>Erschwerte Durchschaubarkeit</li> </ul>	<ul style="list-style-type: none"> <li>Nicht geeignet für Patienten mit schon klinisch feststellbaren Störungen</li> </ul>	(Rogers et al., 1993), (Schmand & Lindeboom, 2005), (Merten, 2014)
			<ul style="list-style-type: none"> <li>Parameter können direkt aus Verfahren gewonnen werden</li> <li>Ggf. via zusätzlicher Untertests</li> <li>Geringere Robustheit gegenüber realen kognitiven Defiziten</li> <li>Überprüfung einer Vielzahl von Funktionsbereichen möglich</li> </ul>	Reliable Digit Span Finger Tapping	<ul style="list-style-type: none"> <li>Verbesserte Sensitivität ohne wesentliche Erhöhung der Untersuchungszeit</li> <li>Validierung auch im gutachterlichen Kontext</li> <li>Erhöhung der Spezifität</li> <li>Direkte Informationen über Gültigkeit der gezeigten Leistung</li> <li>Geringere Anfälligkeit gegenüber Coaching</li> </ul>	<ul style="list-style-type: none"> <li>Gefahr falsch-positiver Zuordnung bei Kombination inhaltlich überlappender Parameter</li> <li>Klassifikationsgüte variiert in Abhängigkeit vom Störungsbild</li> <li>Erfordert detaillierte Kenntnis der Primärliteratur</li> </ul>	(Greve et al., 2009), (Rogers et al., 1993), (Boone, 2007, 2009, 2013; Carone & Bush, 2013; Larrabee, 2007a, 2007b, 2008; Merten, 2014; Meyers & Meyers, 1995; Morgan & Sweet, 2009; Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012)
	<ul style="list-style-type: none"> <li>Seltene Beschwerden</li> <li>Quasi-seltene Beschwerden</li> <li>Unwahrscheinliche Beschwerden</li> <li>Beschwerdenkombinationen, welche bei echten psychiatrischen Störungen äusserst selten vorkommen</li> <li>Unübliche psychopathologische Muster</li> <li>Hohe, undifferenzierte Ausmasse an Beschwerden</li> </ul>	Fremdbeurteilungsverfahren für psychische Störungen		Structured Interview of Reported Symptoms (SIRS-2)	<ul style="list-style-type: none"> <li>Hervorragende Interrater-Reliabilität</li> <li>Sehr gute Retestreliabilität bei forensischen Patienten</li> <li>Belege für Konstruktvalidität mit dem MMPI-2</li> <li>Deutsche Version in Vorbereitung</li> </ul>	<ul style="list-style-type: none"> <li>Lange Durchführungszeit</li> <li>Expertise der Interviewer erforderlich</li> </ul>	(Cima et al., 2003; Lanquillon, Schmidt, & Stieglitz, 2011; McCusker, Moran, Serfass, & Peterson, 2003; Rogers, Bagby, & Dickens, 1992; Rogers, Sewell, & Gillard, 2010; Rogers, Vitacco, & Kurus, 2010; Schmidt, Lanquillon, & Ullmann, 2011; Vitacco, Rogers, Gabel, & Munizza, 2007)
				Miller Forensic Assessment of Symptoms Test (M-Fast)	<ul style="list-style-type: none"> <li>Sehr gute Reliabilität und Validität in forensischen Populationen</li> <li>Belege für Konstruktvalidität mit dem SIRS</li> </ul>	<ul style="list-style-type: none"> <li>Reines Screeningverfahren</li> </ul>	(Guy & Miller, 2004; Miller, 2001, 2005; G.P. Smith, 2008)



Verfahren zur Validierung der Beschwerdenschilderung	<ul style="list-style-type: none"> <li>▪ Aussergewöhnlich hohe Ausprägungen einzelner Symptome</li> <li>▪ Eindeutige Beschwerden</li> </ul>				<ul style="list-style-type: none"> <li>▪ Reduktion Anteil falsch-negativer Klassifikationen gegenüber klinischem Urteil</li> <li>▪ Kurze Durchführungs- und Auswertzeit</li> </ul>		
	<ul style="list-style-type: none"> <li>▪ Diskrepanzen zwischen beobachteten und berichteten Beschwerden</li> <li>▪ Fehlerhafte Stereotypen</li> </ul>	Selbstbeurteilungsverfahren und andere Screeningtests für psychische Störungen		Strukturierter Fragebogen simulierter Symptome (SFSS > deutsche Version des Structured Inventory of Malingered Symptomatology (SIMS))	<ul style="list-style-type: none"> <li>▪ Hohe Retestreliabilität</li> <li>▪ Akzeptable interne Konsistenz</li> <li>▪ Gute vorhersagende und übereinstimmende Validität</li> <li>▪ Sensitivität, Spezifität, Positive Predictive Power, Negative Predictive Power über 0.85 in Abhängigkeit des Trennwertes</li> </ul>	<ul style="list-style-type: none"> <li>▪ Reines Screeningverfahren</li> <li>▪ Sensitivität abhängig vom Untersuchungskontext bzw. angewandtem Trennwert</li> </ul>	(Cima et al., 2003; G. P. Smith & Burger, 1997; van Impelen, Merckelbach, Jelicic, & Merten, 2014; Wisdom, Callahan, & Shaw, 2010)
				Minnesota Multiphasic Personality Inventory (MMPI-2)	<ul style="list-style-type: none"> <li>▪ Aufdeckung von Antwortverzerrung im Bereich psychischer, körperlicher und kognitiver Störungen</li> <li>▪ Sehr hohe Spezifitäten im forensischen Kontext</li> <li>▪ Wahrscheinlichkeitsaussage über Unter- oder Übertreibung möglich</li> <li>▪ Aussagen über psychische Gesundheit und Persönlichkeitsmerkmale möglich</li> </ul>	<ul style="list-style-type: none"> <li>▪ Lange Durchführungszeit</li> <li>▪ Expertise erforderlich</li> <li>▪ Mehrheit der Validitätsskalen nicht Bestandteil der deutschen Version &gt; manuelle Auswertung anhand der Literatur erforderlich</li> </ul>	(Baer & Miller, 2002; Bagby & Marshall, 2004; Butcher et al., 2001; Engel, 2000; Gervais, Ben-Porath, Wygant, & Green, 2007, 2008; Greene, 2000, 2011; Lewis, Simcox, & Berry, 2002; Nelson, Hoelzle, Sweet, Arbis, & Demakis, 2010; Rogers & Bender, 2003; Thies, 2012)
				Personality Assessment Inventory (PAI)  Deutsche Version: Verhaltens- und Erlebensinventar (VEI)	<ul style="list-style-type: none"> <li>▪ Deutlich kürzer als MMPI-2</li> <li>▪ Inhaltliche Ausrichtung an ICD-10 und DSM-IV</li> <li>▪ Hohe Reliabilität, gute Validität in verschiedenen Settings</li> <li>▪ Hohe Effektstärken der Validitätsskalen der Originalversion</li> </ul>	<ul style="list-style-type: none"> <li>▪ Vergleichsweise geringe Sensitivität der deutschen Version</li> </ul>	(Engel & Groves, 2013; Hawes & Boccaccini, 2009; Morey, 2007; Vossler-Thies, Stevens, Engel, & Licha, 2013)
				Morel Emotional Numbing Test for Posttraumatic Stress Disorder (MENT)	<ul style="list-style-type: none"> <li>▪ Forced-choice Verfahren</li> <li>▪ Detektionsstrategie fehlerhafte Stereotypen</li> <li>▪ Hohe Spezifität, gute Sensitivität</li> <li>▪ In vielen Sprachen vorhanden</li> </ul>	<ul style="list-style-type: none"> <li>▪ Screeningverfahren</li> </ul>	(Geraerts et al., 2009; Merten et al., 2009; Morel, 1998, 2012, 2013; Rubenzer, 2009; Wiedmaier, 2011)

<p><b>Nachweis bewusster Täuschungsabsicht</b></p>	<ul style="list-style-type: none"> <li>Prozess, der sorgfältige Analyse aller verfügbaren Informationen erfordert</li> <li>Basierend auf objektiven Kriterien</li> <li>Einschluss von Indikatoren, welche über bekannte Klassifikationsgenauigkeit verfügen</li> <li>Kombination klinisches Urteil mit Ergebnissen wissenschaftlich validierter Verfahren</li> </ul>	<p>Kriterienkataloge</p>	<ul style="list-style-type: none"> <li>A: Vorhandensein eines substantiellen, externen Anreizes</li> <li>B: Hinweise aus neuropsychologischer Testdiagnostik</li> <li>C: Angaben des Exploranden und Validitätsskalen von Selbstbeurteilungsinstrumenten</li> <li>D: Ausschluss von psychiatrischen, neurologischen oder Entwicklungsfaktoren</li> <li>Zusätzlich für „sicher“ vorliegende vorgetäuschte, kognitive Störung: Zahl richtiger Antworten in Zwangswahlverfahren unter Zufallsniveau</li> </ul>	<p>Kriterienkatalog für die Diagnose aggravierter/simulierter neurokognitiver Störungen (Malingered Neurocognitive Disorder, MND)</p>	<ul style="list-style-type: none"> <li>Breit akzeptiertes Konzept</li> <li>primäre und sekundäre Aggravation/Simulation, dafür Wegfall Kriterium D/ Eindeutige Aggravation/Simulation kann nun auch den Nachweis einer hohen a-posteriori-Wahrscheinlichkeit aus Selbstbeurteilungsinstrumenten bestehen oder aus „compelling inconsistencies“</li> </ul>	<ul style="list-style-type: none"> <li>Hoher Zeitaufwand</li> </ul>	<p>(Slick &amp; Sherman, 2013; Slick et al., 1999)</p>
			<ul style="list-style-type: none"> <li>Zusätzliches B: Hinweise aus somatischen Untersuchungsmethoden</li> </ul>	<p>Kriterienkatalog für die Aufdeckung aggravierter/simulierter schmerzassoziierter Behinderungen (Malingered Pain-Related Disabilities, MPRD)</p>		<ul style="list-style-type: none"> <li>Hoher Zeitaufwand</li> </ul>	<p>(Bianchini et al., 2005; Larrabee, Greiffenstein, Greve, &amp; Bianchini, 2007)</p>
			<ul style="list-style-type: none"> <li>A: Evidenz eines externen Anreizes</li> <li>B: Evidenzen aus Testbefunden</li> <li>C: Evidenzen aus Inkonsistenzen/ Diskrepanzen</li> <li>D: Ausschlusskriterien</li> </ul>	<p>Kriterienkatalog für die Diagnose einer unglaubwürdigen, scheinbaren oder aggravierten/simulierten Posttraumatischen Belastungsstörung (Non-credible, Feigned, or Malingered Posttraumatic Stress Disorder Related Disability/Dysfunction, F-PTSDR-D)</p>	<ul style="list-style-type: none"> <li>Eindeutigere Definitionen und Abgrenzungen bestimmter Termini und möglicher Inkonsistenzen</li> <li>Regeln zur Erhebung und Gewichtung von Testdaten</li> <li>Separate Graduierung der Wahrscheinlichkeit von Antwortverzerrung für Kriterien B und C</li> <li>Nutzbar für Diagnose von aggravierten/simulierten kognitiven und/oder schmerzassozierten Störungen</li> <li>Vereinfachte Gesamtbeurteilung</li> <li>Klare Regeln mit Bezug zur klinischen Urteilsbildung</li> </ul>	<ul style="list-style-type: none"> <li>Hoher Zeitaufwand</li> </ul>	<p>(Demakis, Gervais, &amp; Rohling, 2008; Young, 2014)</p>

### **3 FUNCTIONAL SOMATIC SYNDROMES: SKIN TEMPERATURES AND ACTIVITY MEASUREMENTS UNDER AMBULATORY CONDITIONS**

#### **3.1 Einleitung**

Einen empirischen Ansatz um physiologische Beschwerden zu validieren, weist die zweite Publikation auf. Es handelt sich um eine Studie mit Patienten mit funktionellen Störungen. Es wurde dabei versucht physiologische Veränderungen (Hauttemperatur) zu objektivieren und diese mit subjektiven Angaben der Patienten zu vergleichen. Das erste Ziel war, herauszufinden, ob distale Hauttemperatur bei Patienten mit funktionellen Störungen tiefer ist in alltäglichen Lebenssituationen und zwar zu einem vergleichbaren Ausmass wie unter Laborbedingungen. Ein zweites Ziel war der Vergleich von objektiven Messungen der Hauttemperatur mit subjektiven Einschätzungen des Wärmekomforts von verschiedenen Hautregionen. Zusätzlich wurden Temperaturmessungen mit Aktivitätsmessungen kombiniert, da Aktimetrie eine nützliche Methode zu sein scheint, um Aktivität und Schlafmuster in einer natürlichen Umgebung zu messen (Korszun et al., 2002).

#### **3.2 Theoretischer Hintergrund**

Einige Autoren konnten zeigen, dass Patienten mit einem Fibromyalgiesyndrom unter Stressexposition eine herabgesetzte Sympathikusreaktion aufwiesen (Okifuji & Turk, 2002; Petzke & Clauw, 2000). Eine andere Studie zeigte eine höhere Sympathikus-Aktivität und eine niedrigere Parasympathikus-Aktivität oder beides bei Patienten mit Fibromyalgiesyndrom, chronischem Erschöpfungssyndrom, Reizdarmsyndrom und interstitielle Cystitis (Martinez-Martinez, Mora, Vargas, Fuentes-Iniestra, & Martinez-Lavin, 2014). Friederich und Kollegen (2005) halten fest, dass Personen mit einem hyporeaktiven sympathischen Nervensystem Defizite im schmerzhemmenden System aufweisen, kombiniert mit erhöhter Schmerzwahrnehmung. Forschungsergebnisse im Bereich von somatoformen (funktionellen) Störungen sind ziemlich inkonsistent und die Experimentalbedingungen relativ heterogen.

Es konnte gezeigt werden, dass distale Hauttemperatur unter psychologischem Stress sinkt (Mittelman & Wolff, 1939). Chronische Schmerzpatienten ohne Pathophysiologie oder Psychopathologie hatten in einer anderen Studie kältere Hände im Vergleich zu gesunden Personen (Wickramasekera, Kolm, Pope, & Turner, 1998). Cinciripini (1982) zeigte, dass das Niveau der Oberflächenelektromyographie hoch war und die distale Hauttemperatur tief bei chronischen Schmerzpatienten im entspannten Zustand. Patienten mit Fibromyalgiesyndrom

hatten signifikante Asymmetrien der distalen Hauttemperatur im Sinne von erniedrigter Hauttemperatur auf der dominanten Körperseite (Mitani et al., 2006).

Diese Studien messen distale Hauttemperatur in Laborbedingungen. Patienten erleben jedoch chronischen Stress während des normalen Lebensalltages, teilweise durch die Präsenz der Schmerzen selber (Bennett, Tennant, Piesse, Badcock, & Kellow, 1998; Martinez-Lavin, 2012). Aus diesem Grunde sollte distale Hauttemperatur konstant tiefer sein während den Wachzeiten. Kanbara und Kollegen (2004) fanden Dissoziationen zwischen subjektiven und objektiven Antworten bei Patienten mit funktionellen Störungen: Patienten zeigten hypofunktionale psychophysiologische Reaktionen (Oberflächenelektromyographie und Hautleitwiderstand) in einem Stresstest, aber hatten subjektiv verstärkte Anspannungsgefühle. Die Autoren schlossen aus dieser Diskrepanz, dass Patienten mit funktionellen Störungen eine reduzierte Körperwahrnehmung aufweisen. Hyporeaktive objektive Anspannung und hohe subjektive Anspannung könnten Charakteristiken von funktionellen Störungen darstellen (Okifuji & Turk, 2002).

### **3.3 Methoden**

Vierzehn Patienten mit einer funktionellen Störung und vierzehn gesunde Kontrollprobanden gaben ihr schriftliches Einverständnis zur Teilnahme an der Studie. Die Patienten (6 Männer, 8 Frauen) mussten vordefinierte Einschlusskriterien erfüllen. Acht Patienten hatten eine Diagnose eines Fibromyalgiesyndroms oder chronische Schmerzen, vier eine funktionelle Störung des kardiovaskulären Systems und zwei des Magen-Darm-Urogenital Systems. Zwei Patienten nahmen Antidepressiva, einer in Kombination mit Schmerzmittel. Vier Patienten nahmen regelmässig oder gelegentlich Schmerzmittel. Vier Patienten nahmen keine Medikamente. Die vierzehn Probanden (3 Männer und 11 Frauen) waren Teil einer grösseren Stichprobe von 31 Probanden, welcher auf die Patientengruppe bezüglich Jahreszeit der Testung (Frühling, Sommer, Herbst, Winter) und Alter abgestimmt wurden. Die gesunden Kontrollprobanden hatte keine somatische oder psychiatrische Diagnose. Die Teilnehmer mussten fünf Fragebogen über somatische und psychische Symptome ausfüllen und wurden von einem Psychiater interviewt.

Hauttemperaturen wurden während sechs aufeinanderfolgenden Tagen einmal alle 2.5 Minuten gemessen mit kabellosen, wasserfesten Temperatursonden: ein Paar von Sensoren am Fussknöchel (= distale Hautregion) und eines über der infraklavikulären Region auf jeder Körperseite (= proximale Hautregion). Die Teilnehmer trugen zudem wasserfeste Aktimeter, welche Bewegungsbeschleunigungen einmal jede Minute als einen Index für Körperbewe-

gungen aufzeichneten. Alle Teilnehmer wurden angehalten ihre normalen Tagesaktivitäten sowie die normalen Schlaf-Wach-Zeitpläne während der Studie einzuhalten. Diese Schlaf-Wach-Episoden wurden mittels eines Schlaf-Wach-Tagebuches festgehalten. Darin mussten neben den Schlafparametern auch subjektive Zustände (Müdigkeit, Anspannung, Konzentration, Hungergefühl, Stimmung, Wärmeempfinden von drei Körperregionen) eingetragen werden.

Die Temperatur- und Aktimeterdaten wurden nach der Bereinigung und Mittelung mittels einer zweifaktoriellen Varianzanalyse mit Messwiederholung mit Gruppe (Patienten vs. Kontrollen) als Inter-Subjekt-Faktor und Zeit (48 x 30-Min. Blöcke) als Messwiederholungsfaktor analysiert. Auch die subjektiven Einschätzungen des Schlaf-Wach-Tagebuches wurden anhand einer zweifaktoriellen ANOVA mit Messwiederholung mit Gruppe (Patienten vs. Probanden) als Inter-Subjekt-Faktor und Zeit (4 Zeitspannen) als Messwiederholungsfaktor analysiert. Für post-hoc Vergleiche wurden *t*-tests für unabhängige Stichproben angewendet.

### **3.4 Ergebnisse und Diskussion**

Zusammengefasst gab es drei Hauptergebnisse: Erstens, die distale Hauttemperatur der Patienten war am Nachmittag erhöht im Vergleich zu den Kontrollen. Zweitens, die subjektiven Einschätzungen des Wärmekomforts unterschieden sich nicht zwischen den Gruppen, folglich stimmten die objektiven Temperaturdaten nicht mit den subjektiven Erfahrungen der Patienten überein. Drittens, die Aktimeterdaten zeigten vergleichbare Levels von Tagesaktivität in den beiden Gruppen, trotz berichteter erhöhter Müdigkeit und Körperanspannung bei den Patienten.

In Bezug auf die Hauttemperaturen unterschieden sich die Gruppen signifikant bei der distalen Hauttemperatur am Nachmittag (13:30-16:00). Die Patientengruppe zeigte erhöhte distale Hauttemperatur, im Vergleich zu den Kontrollen, wohingegen die Studie von Wickramasekera, Kolm, Pope und Turner (1998) ein umgekehrtes Muster fand. Dies könnte aufgrund von Einschränkungen in dieser Studie (nicht identische Experimentalbedingungen, Vermischung durch unterschiedliche Temperaturnaheorte) und möglichem Stress bei den Basislinien-Messungen im Labor zustande gekommen sein. Im Gegensatz dazu fanden unsere Aufnahmen in einem ambulatorischen Rahmen und über eine viel längere Periode als vier Minuten statt. Deshalb hatten situative Stressfaktoren voraussichtlich wenig Einfluss auf die aufgenommenen distalen Hauttemperaturen auch weil wir die Daten über mehrere Aufnahmetage gemittelt haben.

Die beobachtete erhöhte distale Hauttemperatur bei den Patienten könnte als mögliche Evidenz für die Hypothese eines hypofunktionalen Stresssystems in funktionellen somatischen Syndromen betrachtet werden (Okifuji & Turk, 2002). Die Patienten reagierten wahrscheinlich aufgrund eines hypofunktionalen Stresssystems nicht in der üblichen Weise auf situativen Stress und zeigten keine Vasokonstriktion um die Körperkerntemperatur zu schützen (Kistler, Mariauzouls, & von Berlepsch, 1998). Im Gegenteil, verminderte sympathische Aktivität könnte zu Vasodilatation geführt haben und als Konsequenz zu erhöhter distaler Hauttemperatur. Diese Interpretation wird durch den signifikant kleineren DPG in der Patientengruppe während dem gleichen Zeitintervall unterstützt.

Unser zweites Hauptergebnis betrifft die subjektiven Einschätzungen von Wärmegefühl, welche sich nicht unterschieden zwischen den zwei Gruppen. Die objektiv gemessenen erhöhten distalen Hauttemperaturen in der Patientengruppe spiegelten sich folglich nicht in den subjektiven Einschätzungen wider. Bezüglich der proximalen Hauttemperatur wiesen die Patienten und Kontrollen zudem fast identische Temperaturverläufe auf, jedoch schätzten die Patienten ihren Körper morgens und nachmittags signifikant wärmer ein als die Kontrollen. Solche Dissoziationen zwischen objektiven Hauttemperaturmessungen und subjektiven Einschätzungen des Körperwärmekomforts könnten als Evidenz für Alexisomie gelten (Ikemi & Ikemi, 1986). Dieses Konzept ist eine Erweiterung des Alexithymie Konzeptes und beschreibt die beeinträchtigte Wahrnehmung und Ausdruck von Körperempfindungen. Alexithymie bezieht sich auf Wahrnehmungs- und Ausdrucksschwierigkeiten von emotionalen Gefühlen. In Übereinstimmung damit erzielte unsere Patientengruppe höhere Werte auf der allgemeinen Alexithymieskala als auch auf zwei Subskalen im Vergleich zur Kontrollgruppe. Zusammengefasst könnte dies darauf hinweisen, dass Patienten mit funktionellen somatischen Syndromen allgemeine Schwierigkeiten in der Wahrnehmung und dem Ausdruck von eigenen Körperempfindungen und Emotionen haben.

Bezüglich den Schlafparametern und dem subjektiven Wohlbefinden hatte die Patientengruppe subjektiv schlechtere Schlafqualität und fühlte sich müder und angespannter während des Tages und am Abend. Diese Ergebnisse sind kongruent mit der Literatur über somatische funktionelle Syndrome (Eich et al., 2012; Wessely et al., 1999). Zusammen mit chronischen Schmerzen könnte physisches Vermeidungsverhalten daraus resultieren. Jedoch zeigten unsere Aktimeterdaten vergleichbare Aktivitätslevel während des Tages zwischen der Patienten- und Kontrollgruppe. Eine andere Studie konnte zeigen, dass physisches Vermeidungsverhalten beim Fibromyalgiesyndrom nur mit zusätzlichen schweren depressiven Symptomen gefunden wurde (Korszun et al., 2002). Dies könnte darauf hinweisen, dass physisches Vermei-

dungsverhalten hauptsächlich aus der Kombination von depressiven Symptomen und funktionellen somatischen Syndromen entsteht.

Zu den Einschränkungen der Studie gehören u.a. die distale Messung der Fusstemperatur, keine Messung der Umgebungstemperatur und der getragenen Kleidung, keine Kontrolle der Medikamenteneinnahme sowie die kleine Gruppengröße.

Insgesamt lassen die Studienresultate auf eine Dissoziation zwischen objektiven Hauttemperaturmessungen und subjektiven Einschätzungen von Körperwärmegefühl schließen, welche die Theorie eines möglichen Alexsomie-Anteils bei funktionellen somatischen Syndromen unterstützt. Zudem zeigen die Resultate, dass subjektive Klagen über Müdigkeit und Anspannung nicht zwangsläufig in physischem Vermeidungsverhalten resultiert.

## **4 DEUTSCHE ÜBERSETZUNG UND VALIDIERUNG DES STRESS APPRAISAL MEASURE (SAM)**

### **4.1 Einleitung**

In der dritten Publikation wurde eine Übersetzung und eine deutschsprachige Validierung des anerkannten englischen Stressfragebogens „Stress Appraisal Measure“ (SAM) von Peacock und Wong (1990) durchgeführt. Der SAM ist ein vergleichsweise kurzer Fragebogen (28 Items) zur Bewertung eines aktuellen, stressauslösenden Ereignisses. Die Studie wurde im Rahmen des EU-Projektes INTERSTRESS durchgeführt. Das Akronym INTERSTRESS bedeutet: Interreality in the Management and Treatment of Stress-Related Disorders. Eine Teilaufgabe innerhalb von INTERSTRESS für unsere Forschungsgruppe lag in der Übersetzung und deutschsprachigen Validierung des englischen „Stress Appraisal Measure“ (SAM) von Peacock und Wong (1990). Ich war konkret für die Übersetzung des Fragebogens, die Auswahl der Validierungsinstrumente und Zusammensetzung der Fragebögen und die Erhebung zuständig. Zudem war ich auch an der Überarbeitung der Publikation beteiligt.

### **4.2 Theoretischer Hintergrund**

Psychosomatische Patienten erleben, wie in der zweiten Publikation ausgeführt, vielfach Stress im Alltag. Zur Erfassung und Verarbeitung von Stresserleben werden häufig Fragebögen eingesetzt. In der dritten Publikation soll der Fragebogen Stress Appraisal Measure (SAM) von Peacock und Wong (1990) in deutscher Sprache validiert werden. Der wesentliche Grund für die Übersetzung und Validierung des SAM war das Fehlen eines theoriegeleiteten, kurzen, deutschsprachigen Stressfragebogens, der auf ein aktuelles Lebensereignis fokussiert und bei dem die Einschätzung der Kontrollmöglichkeiten klar unterteilt ist. Peacock und Wong (1990) sprachen in diesem Zusammenhang von antizipatorischem Stress. Beim SAM überzeugte die klare, systematische und theoriegeleitete Operationalisierung der Items nach der kognitivtransaktionalen Stressmodell-Theorie von Lazarus und Folkman (1984). Auch eines der bekanntesten Interventions- und Präventionsprogramme bezüglich Stress, nämlich das Stressimpfungstraining von Meichenbaum (2003), baut auf der Theorie von Lazarus und Folkman (1984) auf, und wird auch als Einsatzgebiet für den SAM gesehen. So könnten im Rahmen des Stressmanagementtrainings etwa die Lernfortschritte der Teilnehmer in den einzelnen Phasen (Informations-, Übungs- und Anwendungsphase) mittels SAM bewertet werden. Dieses Stressimpfungstraining von Meichenbaum eignet sich auch für psychosomatische Erkrankungen (Stenzel, Fehlinger, & Radkovsky, 2015). Lazarus und Folkman (1984) teilten



die Reaktionen eines Menschen auf Ereignisse, die sein inneres Gleichgewicht stören, in primäre und sekundäre kognitive Bewertungen (primary and secondary appraisal) ein. Um nun eine Situation als stressvoll einzustufen, werden zwei kognitive Bewertungsprozesse durchlaufen: 1. Das primary appraisal schätzt die Situation als gefährlich ein und 2. das secondary appraisal schätzt die Situation als nicht kontrollierbar und bewältigbar anhand der eigenen Fähigkeiten ein. Peacock und Wong (1990) versuchten in ihrem Test die beiden kognitiven Prozesse (primary und secondary appraisal) zu operationalisieren. Die Skalen für das primary appraisal lauteten Bedrohung (threat), Herausforderung (challenge) und Bedeutsamkeit (centrality) und die Skalen für das secondary appraisal waren eigene Kontrolle (controllable by self), Kontrolle durch andere (controllable by others) und Unkontrollierbarkeit (uncontrollable). Außerdem wurde eine übergeordnete Skala generelles Stressempfinden (stressfulness) gebildet. Die Autoren führten insgesamt drei Studien durch. Fokus der dritten Publikation war die zweite Studie, wobei der SAM eingesetzt und stressauslösende Imaginationsszenarien benutzt wurden, um akuten Stress zu erzeugen.

### **4.3 Methoden**

Bei der Verteilung von Fragebögen an Jura- und Psychologiestudierende der Universität Basel konnten N = 85 ausgewertet werden, wobei per Zufall 50 Fragebögen (37 Frauen/13 Männer) auf die Variante HIV-Geschichte und 35 (21 Frauen/14 Männer) auf die Variante JOB-Geschichte als stressauslösende Bedingung.

Die 28 Items des ursprünglichen SAM wurden von einem professionellen Übersetzer vom Englischen ins Deutsche übersetzt. Anschließend wurde von einem weiteren Übersetzungsbüro eine Rückübersetzung vom Deutschen ins Englische durchgeführt, um zu überprüfen, ob eine Abweichung zwischen der Originalversion von Peacock und Wong (1990) und der Rückübersetzung vom Deutschen ins Englische vorlag. Die beiden englischen Versionen wichen nur sehr geringfügig voneinander ab, was als Nachweis dafür gelten kann, dass die deutsche der englischen Version semantisch entsprach.

Zur Stressinduktion wurden zwei Imaginationsgeschichten mit unterschiedlicher Stressintensität und Einflussmöglichkeiten konstruiert. Bei einer der Imaginationsgeschichten ging es um die Absage eines nicht zum Lebensunterhalt notwendigen Zusatzjobs (JOB) und bei der anderen um die Mitteilung, sich möglicherweise mit HIV infiziert zu haben (HIV).

Zur konvergenten und diskriminanten Validierung wurden der Coping Inventory for Stressful Situations (CISS; deutsche Version nach Kälin (1995)) und das State-Trait-Angstinventar (STAI (Laux, Glanzmann, Schaffner, & Spielberger, 1981)) eingesetzt. Aus dem STAI wurde

in unserer Studie lediglich der Global Score der Zustandsskala (State-Angst) verwendet, welcher das aktuelle Angstniveau als Mittelwert aus 20 Items angibt.

Spezifische Emotionen (Angst, Freude, Entspannung) wurden mittels numerischen Ratingskalen gemessen. Die Variablen wurden in folgender Reihenfolge im Fragebogen vorgegeben: Demografische Daten (Alter, Geschlecht, Zivilstand, Ausbildung), STAI, spezifische Emotionsskalen, CISS, jeweiliges Stressinduktionsszenarium (HIV oder JOB), STAI, spezifische Emotionsskalen, SAM, CISS.

Aus Gründen der Vergleichbarkeit wurden dieselben statistischen Verfahren wie bei Peacock und Wong (1990) angewendet. Um die stressauslösende Wirkung der Imaginationsgeschichten zu überprüfen, wurde der Global Score des STAI verwendet und eine Devianzanalyse auf der Basis eines gemischten linearen Modells gerechnet. Zur Ermittlung der Faktorenstruktur wurde danach eine Hauptkomponentenanalyse mit Varimax-Rotation gerechnet. Die internen Konsistenzen der sieben SAM-Skalen wurden mittels Cronbachs  $\alpha$  bestimmt. In einem weiteren Analyseschritt sollte in einer schrittweisen Regression das generelle Stressempfinden (Kriterium) vorhergesagt werden. Um die Sensitivität des SAM zu überprüfen, wurden mittels *t*-Tests die Mittelwerte der SAM-Skalen für die beiden Szenarien verglichen. Abschließend wurde die konvergente Validität des SAM anhand der Korrelationen mit STAI, CISS und den Emotionsskalen mittels Pearson-Produkt-Moment-Korrelation ermittelt.

#### **4.4 Ergebnisse und Diskussion**

Der Mittelwert im Global Score des STAI stieg nach dem Lesen der stressauslösenden Szenarien signifikant an, somit konnten die Imaginationsgeschichten als Stressinduktor für die Validierung des SAM genutzt werden. Außerdem wurde eine Jobabsage als weniger angstauslösend erlebt als die Möglichkeit, sich mit HIV infiziert zu haben.

In der anschließend durchgeführten exploratorischen Hauptkomponentenanalyse (PCA) mit Varimax-Rotation resultierte eine fünffaktorielle Lösung. Der erste Faktor umfasste alle Items der beiden Skalen Bedrohung und Bedeutsamkeit sowie zwei weitere Items (völlige Hoffnungslosigkeit, Aufregung). Dabei hatte der erste Faktor den höchsten Eigenwert und verdeutlichte, in welchem hohem Ausmaß die beiden Aspekte Bedrohung und individuelle Bedeutsamkeit bei Stresserleben miteinander verknüpft sind. Der zweite, dritte und vierte Faktor spiegelte genau wie bei Peacock und Wong (1990) die theoriegeleitete Annahme von Lazarus und Folkman (1984) wider, dass das secondary appraisal (sekundäre Bewertung) aus drei verschiedenen Dimensionen (eigene Kontrolle, Kontrolle durch andere, Unkontrollierbarkeit) besteht. Der fünfte Faktor Herausforderung bestand lediglich aus zwei Items (positive Aus-

wirkung, macht stärker) und war nur mit einem Item deckungsgleich zu der Originalstudie. In unserer Analyse wiesen somit einzelne Items eine höhere semantische Nähe zu anderen Faktoren auf als bei Peacock und Wong (1990). Es bleibt aber festzuhalten, dass die von Peacock und Wong (1990) identifizierten Faktoren jedoch im Wesentlichen repliziert werden konnten. Es sei darauf hingewiesen, dass die Faktorenstruktur des SAM sich sowohl in den Studien von Peacock und Wong (1990) als auch in anderen Studien als wenig stabil erwies (Roesch & Rowley, 2005; Rowley, Roesch, Jurica, & Vaughn, 2005). So zeigten sich neben der fünffaktoriellen Lösung auch drei-, vier- oder sogar sechsfaktorielle Strukturen. Dies muss allerdings nicht als Beleg gegen das Modell von Lazarus und Folkman (1984) gewertet werden, sondern könnte auch darauf zurückzuführen sein, dass zur Erfassung dieser konzeptuellen Differenzierungen zu wenige Items formuliert wurden oder Items, die den Inhalt der Konzepte nicht trennscharf genug treffen. Trotzdem können die Ergebnisse zum Anlass genommen werden, den theoretischen Bezug zum Transaktionalen Stressmodell von Lazarus und Folkman (1984) zu hinterfragen. Darin wurde ursprünglich angenommen, dass Bedeutsamkeit orthogonal zu Herausforderung und Bedrohung angeordnet ist. Es wurde postuliert, dass eine Bedrohung ebenso wie eine Herausforderung nur dann Beachtung findet, wenn sie als bedeutsam (zentral) für das eigene Selbstkonzept (im Sinne von „Gefahr“ bzw. „selbstwertsteigernd“) wahrgenommen wird. Verschiedene Studienergebnisse (inklusive der Originalarbeit und unserer eigenen Studie) legen nun aber nahe, dass zumindest Bedrohung und Bedeutsamkeit auf einem Faktor laden und somit nicht unabhängig sind. So sind Bedrohung und Bedeutsamkeit enger miteinander verbunden als Herausforderung und Bedeutsamkeit, bei denen die Orthogonalität bestehen bleibt. Dies würde eine Abkehr vom ursprünglichen Modell bedeuten. Eine Interpretation könnte etwa lauten, dass gefährvolle Ereignisse wesentlicher und zentraler sind als herausfordernde Ereignisse. Es gibt aber auch Studien bei denen die Orthogonalität aufrechterhalten bleibt ((Durak, 2007; Levy, Nicholls, & Polman, 2012). Die zum Teil gegensätzliche Ergebnislage führt dazu, dass eine letztliche Entscheidung über die Aufrechterhaltung bzw. Verwerfung der Theorie von Lazarus und Folkman (1984) offen bleiben und in weiteren Studien geklärt werden muss.

Die von den Autoren der Originalversion gefundenen internen Konsistenzen, die die Reliabilität der sieben Skalen beschrieb, konnten mit ausreichender Höhe ( $\alpha$ -Wert von 0,90 bis 0,73) in der übersetzten Version bestätigt werden. Lediglich die Skala Herausforderung schnitt mit einem  $\alpha$ -Wert von 0,57 schlechter ab. Bis auf den nicht befriedigenden Wert sprachen die Ergebnisse für eine weitgehende Vergleichbarkeit zu den Ergebnissen von Peacock und Wong (1990).

Die schrittweise Regression des generellen Stressempfindens führte zu einem vereinfachten Modell mit drei signifikanten Prädiktoren (Bedrohung, Bedeutsamkeit, Wechselwirkung Herausforderung und Stressinduktionsgeschichte), welche 67% der Varianz des generellen Stressempfindens aufklären konnten, was als hoch einzustufen ist (Hattie, 2009). Die Ergebnisse der Regressionsanalyse zeigten weiter, dass Kontrollattributionen im Sinne der sekundären Bewertung (secondary appraisal) wie bei Peacock und Wong (1990) keine signifikanten Prädiktoren waren. Es ist somit inhaltlich festzuhalten, dass Stresserleben zwar aus verschiedenen Facetten bestand, dass aber die Kontrollattributionen gemeinsam mit den primären Bewertungsfaktoren keinen eigenen Beitrag bei der Regression zum Stressempfinden lieferten, was in einer ähnlich angelegten Studie auch so gefunden wurde (Anshel, Robertson, & Caputi, 1997).

Die durchgeführten Mittelwertsvergleiche der SAM-Skalen zu den beiden Stressinduktionsszenarien belegten die Validität des SAM. Alle Mittelwertsvergleiche waren bis auf Kontrolle durch andere signifikant. So wurde die Möglichkeit, sich mit HIV infiziert zu haben, als insgesamt stressauslösender, bedeutsamer, bedrohlicher und herausfordernder erlebt als ein verpasster Nebenjob.

Die Analyse der konvergenten und diskriminanten Validität ergab folgendes Resultat: Der SAM korrelierte signifikant positiv und in den meisten Fällen in ausreichender Höhe mit inhaltlich nahen Konstrukten (STAI, CISS, Emotionsskala Angst) und korrelierte negativ (wenn auch nicht immer in der geforderten Höhe) mit Massen zur Bestimmung der Divergenz (Emotionsskala Freude bzw. Entspannung). Die höchste Korrelation ( $r = 0,59$ ) bestand zwischen Bedrohung und Angst. Die SAM-Skala Kontrolle durch andere korrelierte mit keiner Validierungsskala signifikant.

Insgesamt kann festgehalten werden, dass unsere Studie die Ergebnisse der Originalstudie im Wesentlichen replizieren konnte, weshalb der SAM nun auch in deutscher Sprache eingesetzt werden könnte.

## **5 ALLGEMEINES FAZIT**

Die Beschwerdenvalidierung bei psychosomatischen Patienten ist aus unterschiedlichen Perspektiven ein brisantes Thema - es ist politisch, wirtschaftlich, oft subjektiv und individuell. Die ersten beiden Publikationen liefern dazu wichtige Erkenntnisse, woraus ersichtlich wird, dass die Prävalenz nicht-authentischer Beschwerden hoch ist, aber gute Instrumente zur Beschwerdenvalidierung bei psychosomatischen Patienten vorliegen, es jedoch noch an Forschung und Leitlinien zur qualitativen Integration dieser Instrumente mangelt. Die zweite Publikation konnte zeigen, dass bei psychosomatischen Patienten störungsimmanente Anteile vorliegen, indem die subjektiven Wahrnehmungen nicht den objektiven Messungen entsprach. Diese sollten, neben vielen anderen möglichen Ursachen, von den Gutachtern immer in Betracht gezogen werden. Es wurde auch ersichtlich, dass sich psychosomatische Patienten müde und angespannt fühlen und der Stress ein präsent Thema bei psychosomatischen Patienten darstellt, was mehrfach untersucht wurde. Dazu liefert die Übersetzung des Stressfragebogens SAM (Stress Appraisal Measure) einen wichtigen Beitrag, indem er in Stressbewältigungstrainings eingesetzt werden kann und somit zur Evaluation beiträgt.

## 6 REFERENZEN

- Aamondt, M. G., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner, 15*(1), 6-11.
- Aaron, L. A., & Buchwald, D. (2001). A review of the evidence for overlap among unexplained clinical conditions. *Ann Intern Med, 134*(9 Pt 2), 868-881.
- Anshel, M. H., Robertson, M., & Caputi, P. (1997). Sources of acute stress and their appraisals and reappraisals among Australian police as a function of previous experience. *Journal of Occupational and Organizational Psychology, 70*, 337-356.
- APA. (2013). *Diagnostic and statistical manual of mental disorders, 5. Aufl.* . Arlington VA: American Psychiatric Association.
- Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: a meta-analytic review. *Psychol Assess, 14*(1), 16-26.
- Bagby, R. M., & Marshall, M. B. (2004). Assessing underreporting response bias on the MMPI-2. *Assessment, 11*(2), 115-126.
- Barsky, A. J., Orav, E. J., & Bates, D. W. (2005). Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity. *Arch Gen Psychiatry, 62*(8), 903-910.
- Barsky, A. J., Orav, E. J., & Bates, D. W. (2006). Distinctive patterns of medical care utilization in patients who somatize. *Med Care, 44*(9), 803-811.
- Bartlett, M. S., Littlewort, G. C., Frank, M. G., & Lee, K. (2014). Automatic decoding of facial movements reveals deceptive pain expressions. *Curr Biol, 24*(7), 738-743.
- Bennett, E. J., Tennant, C. C., Piesse, C., Badcock, C. A., & Kellow, J. E. (1998). Level of chronic life stress predicts clinical outcome in irritable bowel syndrome. *Gut, 43*(2), 256-261.
- Bianchini, K. J., Etherton, J. L., Greve, K. W., Heinly, M. T., & Meyers, J. E. (2008). Classification accuracy of MMPI-2 validity scales in the detection of pain-related malingering: a known-groups study. *Assessment, 15*(4), 435-449.
- Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *Spine J, 5*(4), 404-417.
- Blaskewitz, N., & Merten, T. (2007). Diagnostik der Beschwerdenvalidität--Diagnostik bei Simulationsverdacht: ein Update 2002 bis 2005. *Fortschr Neurol Psychiatr, 75*(3), 140-154.

- Boone, K. B. (2007). *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective*. New York: Guilford Press.
- Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *Clin Neuropsychol*, 23(4), 729-741.
- Boone, K. B. (2013). *Clinical Practice of Forensic Neuropsychology. An Evidence-based Approach*. New York: Guilford Press.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . Standards for Reporting of Diagnostic, A. (2003). Toward complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative. *Am J Clin Pathol*, 119(1), 18-22.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2). Manual for administration, scoring and interpretation. (rev.ed.)*. Minneapolis, MN: University of Minnesota Press.
- Carone, D., & Bush, S. S. (2013). *Mild Traumatic Brain Injury. Symptom Validity Assessment and Malingering*. New York: Springer.
- Cima, M., Hollnack, S., Kremer, K., Knauer, E., Schellbach-Matties, R., Klein, B., & Merckelbach, H. (2003). "Strukturierter Fragebogen Simulierter Symptome". Die deutsche Version des "Structured Inventory of Malingered Symptomatology: SIMS". *Nervenarzt*, 74(11), 977-986.
- Cinciripini, P. M. (1982). Relationship between frontalis muscle tension and digital skin temperature during EMG biofeedback. *Percept Mot Skills*, 54(3), 895-898.
- Creed, F., & Barsky, A. (2004). A systematic review of the epidemiology of somatisation disorder and hypochondriasis. *J Psychosom Res*, 56(4), 391-408.
- de Waal, M. W., Arnold, I. A., Eekhof, J. A., & van Hemert, A. M. (2004). Somatoform disorders in general practice: prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *Br J Psychiatry*, 184, 470-476.
- Demakis, G. J., Gervais, R. O., & Rohling, M. L. (2008). The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of post-traumatic stress disorder. *The Clinical Neuropsychologist*, 22, 879-895.
- Durak, M. (2007). *The relationship between cognitive appraisal of stress, coping strategies and psychological distress among correctional officers: personal and environmental factors*. Middle East Technical University.

- Eich, W., Hauser, W., Arnold, B., Jackel, W., Offenbacher, M., Petzke, F., . . .  
Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen, F. (2012). [Fibromyalgia syndrome. Definition, classification, clinical diagnosis and prognosis]. *Schmerz*, 26(3), 247-258.
- Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *Am Psychol*, 46(9), 913-920.
- Engel, R. R. (2000). *MMPI-2. Minnesota Multiphasic Personality Inventory-2 Manual*. Bern: Hans Huber.
- Engel, R. R., & Groves, J. A. (2013). *Verhaltens- und Erlebensinventar. Deutschsprachige Adaption des Personality Assessment Inventory (PAI) von L.C. Morey*. Bern: Hans Huber.
- Escobar, J. I., Interian, A., Diaz-Martinez, A., & Gara, M. (2006). Idiopathic physical symptoms: a common manifestation of psychiatric disorders in primary care. *CNS Spectr*, 11(3), 201-210.
- Fink, P., Sorensen, L., Engberg, M., Holm, M., & Munk-Jorgensen, P. (1999). Somatization in primary care. Prevalence, health care utilization, and general practitioner recognition. *Psychosomatics*, 40(4), 330-338.
- Friederich, H. C., Schellberg, D., Mueller, K., Bieber, C., Zipfel, S., & Eich, W. (2005). [Stress and autonomic dysregulation in patients with fibromyalgia syndrome]. *Schmerz*, 19(3), 185-188, 190-182, 194.
- Gaab, J. (2014). Go west or rest - die Behandlung chronischer Erschöpfungssyndrome im Wandel der Zeiten. *Verhaltenstherapie*, 24, 108-113.
- Geraerts, E., Kozaric-Kovacic, D., Merckelbach, H., Peraica, T., Jelicic, M., & Candel, I. (2009). Detecting deception of war-related posttraumatic stress disorder. *Journal of Forensic Psychiatry and Psychology*, 20, 278-285.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a response bias scale (RBS) for the MMPI-2. *Assessment*, 14(2), 196-208.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2008). Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. *Clin Neuropsychol*, 22(6), 1061-1079.
- Gorissen, M., Sanz, J. C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophr Res*, 78(2-3), 199-208.
- Green, P. (2003, revised 2005). *Medical Symptom Validity Test for Windows. User's Manual and Program*. Edmonton: Green's Publishing.



- Green, P. (2005). *Green's Word Memory Test. User's Manual*. Edmonton: Green's Publishing.
- Green, P. (2008). *Non-Verbal Medical Symptom Validity Test for Windows. User's Manual*. . Edmonton: Green's Publishing.
- Greene, R. L. (2000). *The MMPI-2: an interpretive manual. 2nd ed.* Boston: Allyn & Bacon.
- Greene, R. L. (2011). *The MMPI-2/MMPI-2-RF: an interpretive manual. 3rd ed.* Boston: Allyn & Bacon.
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6(3), 218-224.
- Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychological evaluation in a medico-legal context. *Arch Phys Med Rehabil*, 90(7), 1117-1126.
- Guy, L. S., & Miller, H. A. (2004). screening for malingered psychopathology in a correctional setting - Utility of the Miller-Forensic Assessment of Symptoms Test (M-FAST). *Criminal Justice and Behavior*, 31(6), 695-716.
- Hanel, G., Henningsen, P., Herzog, W., Sauer, N., Schaefer, R., Szecsenyi, J., & Lowe, B. (2009). Depression, anxiety, and somatoform disorders: vague or distinct categories in primary care? Results from a large cross-sectional study. *J Psychosom Res*, 67(3), 189-197.
- Harris, A. M., Orav, E. J., Bates, D. W., & Barsky, A. J. (2009). Somatization increases disability independent of comorbidity. *J Gen Intern Med*, 24(2), 155-161.
- Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing: neuropsychological malingering and the Word Memory Test. *Arch Clin Neuropsychol*, 17(7), 709-714.
- Hattie, J. (2009). *Visible learning a synthesis of over 800 metaanalyses relating to achievement*. London: Routledge.
- Hawes, S. W., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: a meta-analytic review. *Psychol Assess*, 21(1), 112-124.
- Henningsen, P., Zipfel, S., & Herzog, W. (2007). Management of functional somatic syndromes. *Lancet*, 369(9565), 946-955.
- Henry, M. (2009). Beschwerdenvalidierungstests in der zivil- und sozialrechtlichen Begutachtung: Verfahrensüberblick. In T. Merten & H. Dettenborn (Eds.), *Diagnostik der Beschwerdenvalidität* (pp. 118-161). Berlin: Deutscher Psychologen Verlag GmbH.

- Hiller, W., & Rief, W. (2014). Die Abschaffung der somatoformen Störungen durch DSM-5 - ein akademischer Schilfbürgerstreich. *Psychotherapeut*, 59, 448-455.
- Huber, M. (2000). Aspekte der Berufsunfähigkeit bei psychosomatischen Erkrankungen. *Versicherungsmedizin*, 52(2), 66-75.
- Ikemi, Y., & Ikemi, A. (1986). An oriental point of view in psychosomatic medicine. *Psychother Psychosom*, 45(3), 118-126.
- Kälin, W. (1995). *Deutsche 24-Item Kurzform des "Coping Inventory for Stressful Situations" (CISS) von N.S. Endler & J.D.A. Parker basierend auf der Übersetzung von N. Semmer, F. Tschan & V. Schade (unveröffentlichter Fragebogen)*. Bern: Universität, Institut für Psychologie.
- Kaluza, G. (2004). *Stressbewältigung: Trainingsmanual zur psychologischen Gesundheitsförderung*. Heidelberg: Springer.
- Kanbara, K., Mitani, Y., Fukunaga, M., Ishino, S., Takebayashi, N., & Nakai, Y. (2004). Paradoxical results of psychophysiological stress profile in functional somatic syndrome: correlation between subjective tension score and objective stress response. *Appl Psychophysiol Biofeedback*, 29(4), 255-268.
- Kistler, A., Mariauzouls, C., & von Berlepsch, K. (1998). Fingertip temperature as an indicator for sympathetic responses. *Int J Psychophysiol*, 29(1), 35-41.
- Korszun, A., Young, E. A., Engleberg, N. C., Brucksch, C. B., Greden, J. F., & Crofford, L. A. (2002). Use of actigraphy for monitoring sleep and activity levels in patients with fibromyalgia and depression. *J Psychosom Res*, 52(6), 439-443.
- Kroenke, K., Spitzer, R. L., Williams, J. B., Linzer, M., Hahn, S. R., deGruy, F. V., 3rd, & Brody, D. (1994). Physical symptoms in primary care. Predictors of psychiatric disorders and functional impairment. *Arch Fam Med*, 3(9), 774-779.
- Lanquillon, S., Schmidt, T., & Stieglitz, R. D. (2011). *Strukturiertes Interview Berichteter Beschwerden SIBB*. Bern: Huber.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *Clin Neuropsychol*, 17(3), 410-425.
- Larrabee, G. J. (2007a). *Assessment of malingered neuropsychological deficits*. Oxford: Oxford University Press.
- Larrabee, G. J. (2007b). Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *Clinical Neuropsychologist*, 17, 54-68.
- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: relationship to likelihood ratios. *Clin Neuropsychol*, 22(4), 666-679.

- Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334-371). New York: Oxford.
- Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *Das State-Trait-Angstinventar. Theoretische Grundlagen und Handanweisung*. Weinheim: Beltz.
- Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York: Springer.
- Leiknes, K. A., Finset, A., Moum, T., & Sandanger, I. (2007). Current somatoform disorders in Norway: prevalence, risk factors and comorbidity with anxiety, depression and musculoskeletal disorders. *Soc Psychiatry Psychiatr Epidemiol*, 42(9), 698-710.
- Levy, A., Nicholls, A., & Polman, R. (2012). Cognitive appraisals in sport: the direct and moderating role of mental toughness. *International Journal of Applied Psychology*, 2, 71-76.
- Lewis, J. L., Simcox, A. M., & Berry, D. T. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the structured inventory of malingered symptomatology. *Psychol Assess*, 14(2), 170-176.
- Martinez-Lavin, M. (2012). Fibromyalgia: When Distress Becomes (Un)sympathetic Pain. *Pain Res Treat*, 2012, 981565.
- Martinez-Martinez, L. A., Mora, T., Vtra, M., & Martinez-Lavin, M. (2014). Sympathetic nervous system dysfunction in fibromyalgia, chronic fatigue syndrome, irritable bowel syndrome, and interstitial cystitis: a review of case-control studies. *J Clin Rheumatol*, 20(3), 146-150.
- McCusker, P. J., Moran, M. J., Serfass, L., & Peterson, K. H. (2003). Comparability of the MMPI-2 F(p) and F scales and the SIRS in clinical use with suspected malingerers. *Int J Offender Ther Comp Criminol*, 47(5), 585-596.
- Meichenbaum, D. (2003). *Intervention bei Stress: Anwendung und Wirkung des Stressimpfungstrainings*. Bern: Huber.
- Mergl, R., Seidscheck, I., Allgaier, A. K., Moller, H. J., Hegerl, U., & Henkel, V. (2007). Depressive, anxiety, and somatoform disorders in primary care: prevalence and recognition. *Depress Anxiety*, 24(3), 185-195.
- Merten, T. (2002). Fragen der neuropsychologischen Diagnostik bei Simulationsverdacht. *Fortschr Neurol Psychiatr*, 70(3), 126-138.
- Merten, T. (2011). Beschwerdenvalidierung bei der Begutachtung kognitiver und psychischer Störungen. *Fortschr Neurol Psychiatr*, 79(2), 102-116.

- Merten, T. (2014). *Beschwerdenvalidierung. Fortschritte der Neuropsychologie, Bd. 14*. Göttingen: Hogrefe.
- Merten, T., Friedel, E., & Stevens, A. (2006). Eingeschränkte Kooperativität in der neurologisch-psychiatrischen Begutachtung: Schätzungen zur Auftretenshäufigkeit an einer Begutachtungspopulation. *Versicherungsmedizin*, 58(1), 19-21.
- Merten, T., Thies, E., Schneider, K., & Stevens, A. (2009). Symptom Validity Testing in Claimants with Alleged Posttraumatic Stress Disorder: Comparing the Morel Emotional Numbing Test, the Structured Inventory of Malingered Symptomatology, and the Word Memory Test. *Psychological Injury and Law*, 2, 284-293.
- Meyers, J., & Meyers, K. (1995). *Rey Complex Figure and Recognition Trial: Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Miller, H. A. (2001). *Miller-Forensic Assessment of Symptoms Test professional manual*. Odessa, FL: Psychological Assessment Resources.
- Miller, H. A. (2005). The Miller-Forensic Assessment of Symptoms Test (M-FAST): Test generalizability and utility across race, literacy, and clinical opinion. *Criminal Justice and Behavior*, 32(6), 591-611.
- Mitani, Y., Fukunaga, M., Kanbara, K., Takebayashi, N., Ishino, S., & Nakai, Y. (2006). Evaluation of psychophysiological asymmetry in patients with fibromyalgia syndrome. *Appl Psychophysiol Biofeedback*, 31(3), 217-225.
- Mittelman, B., & Wolff, H. (1939). Affective states and skin temperature: Experimental study of subjects with „cold hands“ and Raynaud's syndrome. *Psychosom Med*, 1, 271-292.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *J Clin Exp Neuropsychol*, 24(8), 1094-1102.
- Morel, K. R. (1998). Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. *J Pers Assess*, 70(2), 299-314.
- Morel, K. R. (2012). *Manual for the Morel Emotional Numbing Test for Posttraumatic Stress Disorder: Psychometric Properties. (3rd ed. rev.)*. Las Vegas: NV: Author.
- Morel, K. R. (2013). Cutoff Scores for the Morel Emotional Numbing Test for PTSD: Considerations for Use in VA Mental Health Examinations. *Psychological Injury and Law*, 6, 138-143.
- Morey, L. C. (2007). *The personality assessment inventory professional manual. 2nd ed.* Odessa, FL: Psychological Assessment Resources.
- Morgan, J. E., & Sweet, J. J. (2009). *Neuropsychology of malingering casebook*. New York: Psychology Press.

- Nelson, N. W., Hoelzle, J. B., Sweet, J. J., Arbisi, P. A., & Demakis, G. J. (2010). Updated meta-analysis of the MMPI-2 symptom validity scale (FBS): verified utility in forensic practice. *Clin Neuropsychol*, 24(4), 701-724.
- Nelson, N. W., Sweet, J. J., Berry, D. T., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: exploratory factor analytic evidence of distinct cognitive and psychological constructs. *J Int Neuropsychol Soc*, 13(3), 440-449.
- Okifuji, A., & Turk, D. C. (2002). Stress and psychophysiological dysregulation in patients with fibromyalgia syndrome. *Appl Psychophysiol Biofeedback*, 27(2), 129-141.
- Peacock, E. J., & Wong, P. T. P. (1990). The Stress Appraisal Measure (SAM) - a multidimensional approach to cognitive appraisal. *Stress Medicine*, 6(3), 227-236.
- Petzke, F., & Clauw, D. J. (2000). Sympathetic nervous system function in fibromyalgia. *Curr Rheumatol Rep*, 2(2), 116-123.
- Rief, W., Heitmüller, A. M., Reisberg, K., & Rüdell, H. (2006). Why reassurance fails in patients with unexplained symptoms--an experimental investigation of remembered probabilities. *PLoS Med*, 3(8), e269.
- Roesch, S. C., & Rowley, A. A. (2005). Evaluating and developing a multidimensional, dispositional measure of appraisal. *J Pers Assess*, 85(2), 188-196.
- Rogers, R. (2008). Detection strategies in malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception*. 3rd ed. (pp. 14-38). New York: Guilford.
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS) and professional manual*. Odessa, FL: Psychological Assessment Resources.
- Rogers, R., & Bender, S. D. (2003). Evaluation of malingering and deception. In I. B. Weiner & A. M. Goldstein (Eds.), *Handbook of psychology: forensic psychology* (pp. 109-131). Hoboken, NJ: Wiley & Sons.
- Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning Neuropsychological Impairment - a Critical-Review of Methodological and Clinical Considerations. *Clinical Psychology Review*, 13(3), 255-274.
- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms, 2nd ed. (SIRS-2) and professional manual*. Odessa, FL: Psychological Assessment Resources.

- Rogers, R., Vitacco, M. J., & Kurus, S. J. (2010). Assessment of malingering with repeat forensic evaluations: patient variability and possible misclassification on the SIRS and other feigning measures. *J Am Acad Psychiatry Law*, 38(1), 109-114.
- Rowley, A. A., Roesch, S. C., Jurica, B. J., & Vaughn, A. A. (2005). Developing and validating a stress appraisal measure for minority adolescents. *J Adolesc*, 28(4), 547-557.
- Rubenzler, S. (2009). Posttraumatic Stress Disorder: Assessing Response Style and Malingering. *Psychological Injury and Law*, 2, 114-142.
- Schmand, B., & Lindeboom, J. (2005). *Amsterdam Short Term Memory Test. Amsterdamer Kurzzeitgedächtnistest. Manual. Handanweisung*. Leiden: PITS.
- Schmidt, T., Lanquillon, S., & Ullmann, U. (2011). Kontroverse zu Beschwerdenvalidierungsverfahren bei der Begutachtung psychischer Störungen. *Forensische Psychiatrie, Psychologie, Kriminologie*, 5(3), 177-183.
- Schneider, W., Becker, D., Dohrenbusch, R., Freyberger, H. J., Gündel, H., & Henningsen, P. (2010). Berufliche Leistungsfähigkeit. Begutachtung bei psychischen und psychosomatischen Erkrankungen. *Psychotherapeut*, 55(5), 373-379.
- Schneider, W., Dohrenbusch, R., Henningsen, P., Freyberger, H. J., Irle, H., Köllner, V., & Widder, B. (2016). *Begutachtung bei psychischen und psychosomatischen Erkrankungen: Autorisierte Leitlinien und Kommentare. 2., überarbeitete und erweiterte Auflage*. Bern: Hogrefe.
- Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable Digit Span: a systematic review and cross-validation study. *Assessment*, 19(1), 21-30.
- Schweickhardt, A., Fritzsche, K., & Wirsching, M. (2005). *Psychosomatische Medizin und Psychotherapie*. Heidelberg: Springer.
- Slick, D. J., & Sherman, E. M. (2013). Differential Diagnosis of Malingering, in Mild Traumatic Brain Injury. In D. Carone & S. S. Bush (Eds.), *Symptom Validity Assessment and Malingering* (pp. 57-72). New York: Springer Publishing Company.
- Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *Clin Neuropsychol*, 13(4), 545-561.
- Smith, G. P. (2008). Brief screening measures for the detection of feigned psychopathology. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception. 3rd ed.* (pp. 323-339). New York: Guilford.

- Smith, G. P., & Burger, G. K. (1997). Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). *J Am Acad Psychiatry Law*, 25(2), 183-189.
- Smith, R. C., Gardiner, J. C., Lyles, J. S., Sirbu, C., Dwamena, F. C., Hodges, A., . . . Goddeeris, J. (2005). Exploration of DSM-IV criteria in primary care patients with medically unexplained symptoms. *Psychosom Med*, 67(1), 123-129.
- Stadtland, C., Gündel, H., Schütt, S., & Nedopil, N. (2003). Kriterien zur Beurteilung der quantitativen Leistungseinschränkung bei der Begutachtung funktioneller körperlicher Störungen. Eine Literaturübersicht. *Versicherungsmedizin*, 55(3), 111-117.
- Stenzel, N. M., Fehlinger, T., & Radkovsky, A. (2015). Fertigkeitentrainings in der Verhaltenstherapie. *Verhaltenstherapie*, 25, 54-66.
- Stevens, A., Friedel, E., Mehren, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: prevalence and effects in a German sample of claimants. *Psychiatry Res*, 157(1-3), 191-200.
- Strebel, M., & Brusis, T. (2010). Zur Problematik der Simulation und Aggravation in der HNO-ärztlichen Begutachtung. *HNO*, 58, 126-131.
- Thies, E. (2012). *Der deutsche MMPI-2: Effektivität der Validitätsskalen in der Aufdeckung von Antwortverzerrung*. Marburg: Tectum.
- Tombaugh, T. N. (1996). *Test of Memory Malingered (TOMM)*. North Tonawanda: Multi-Health Systems.
- Trueblood, W., & Binder, L. M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. *Arch Clin Neuropsychol*, 12(1), 13-27.
- van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A Systematic Review and Meta-Analysis. *Clin Neuropsychol*, 1-30.
- Vitacco, M. J., Rogers, R., Gabel, J., & Munizza, J. (2007). An evaluation of malingering screens with competency to stand trial patients: a known-groups comparison. *Law Hum Behav*, 31(3), 249-260.
- Vossler-Thies, E., Stevens, A., Engel, R. R., & Licha, C. (2013). Erfassung negativer Antwortverzerrungen mit der deutschen Fassung des "Personality Assessment Inventory", dem "Verhaltens- und Erlebensinventar". *Diagnostica*, 59(2), 73-85.
- Wessely, S., Nimnuan, C., & Sharpe, M. (1999). Functional somatic syndromes: one or many? *Lancet*, 354(9182), 936-939.

- Wickramasekera, I. E., Kolm, P., Pope, A., & Turner, M. (1998). Observation of a paradoxical temperature increase during cognitive stress in some chronic pain patients. *Appl Psychophysiol Biofeedback*, 23(4), 233-241.
- Wiedmaier, P. (2011). *Schätzung von Normalbefunden für drei Beschwerdenvalidierungstests: Der Morel Emotional Numbing Test (MENT), der Strukturierte Fragebogen Simulierter Symptome und die List of Improbable Problems (LIMP)*. Saarbrücken: VDM Verlag Dr. Müller.
- Wisdom, N. M., Callahan, J. L., & Shaw, T. G. (2010). Diagnostic utility of the structured inventory of malingered symptomatology to detect malingering in a forensic sample. *Arch Clin Neuropsychol*, 25(2), 118-125.
- Young, G. (2014). *Malingering, Feigning, and Response Bias in Psychiatric/Psychological Injury. Implications for Practice and Court*. (Vol. 56). Dordrecht: Springer SBM.



## **ANHÄNGE**

- A-1: Beschwerdenvalidierung in der versicherungsmedizinischen Begutachtung.
- A-2: Functional somatic syndromes: skin temperatures and activity measurements under ambulatory conditions
- A-3: Deutsche Übersetzung und Validierung des Stress Appraisal Measure (SAM)

**ANHANG A-1**

**Publikation 1:**

Keppler, C., Plohmann, A., Pflüger, M., Rabovsky K., Delahaye, M., Langewitz, W. & Mager, R. (eingereicht). Beschwerdenvalidierung in der versicherungsmedizinischen Begutachtung. *Fortschritte der Neurologie Psychiatrie*.

**BESCHWERDENVALIDIERUNG IN DER VERSICHERUNGSMEDIZINISCHEN  
BEGUTACHTUNG**

**SYMPTOM VALIDATION IN INDEPENDENT MEDICAL EVALUATIONS**

Carole Keppler<sup>1</sup>, Andrea Plohm<sup>2</sup>, Marlon Pflüger<sup>1</sup>, Kristin Rabovsky<sup>1</sup>, Wolf Langewitz<sup>3</sup>, Ralph Mager<sup>1</sup>

<sup>1</sup> Forensisch-Psychiatrische Klinik, Universitäre Psychiatrische Kliniken Basel, Wilhelm Klein-Str. 27, CH-4012 Basel, Schweiz

<sup>2</sup> Praxis für Neuropsychologie und Psychotherapie, Freie Strasse 59, CH-4001 Basel, Schweiz

<sup>3</sup> Psychomatik, Bereich Innere Medizin, Universitätsspital Basel, Hebelstr. 2, CH-4031 Basel, Schweiz

**Korrespondierender Autor:**

Prof. Dr. med. Ralph Mager

Universitäre Psychiatrische Kliniken Basel

Versicherungsmedizin

Forensisch-Psychiatrische Klinik

Wilhelm Klein-Str. 27

CH-4012 Basel

Schweiz

Tel.: +41 61 325 52 94

Fax: +41 61 325 54 92

E-mail: [ralph.mager@upkbs.ch](mailto:ralph.mager@upkbs.ch)

## Zusammenfassung

Hohe Prävalenzraten von nicht-authentischen Beschwerden in der versicherungsmedizinischen Begutachtung weisen auf das Risiko hin, dass in einer Gesellschaft zur Verfügung gestellte und limitierte geldwerte Leistungen nicht zielkonform eingesetzt werden. Es liegt dabei wesentlich in der Verantwortung der Gutachter, in geeigneter Form eine Beschwerdengültigkeit zu erbringen, um dem Rechtsanwender eine zutreffende Entscheidungsgrundlage zu bieten und medizinischen Qualitätskriterien gerecht zu werden. Traditionell geschieht dies im Wesentlichen in einer nicht standardisierten Konsistenzprüfung auf dem Hintergrund der klinischen Untersuchungsergebnisse in Abgleich mit der Aktenlage. Die nicht gesicherte Validität dieses Vorgehens mit hoher Abhängigkeit von individueller Expertise, Datenlage und Haltung hat internationale Fachgesellschaften in jüngerer Zeit dazu veranlasst, Leitlinien mit ihren Empfehlungen zu öffnen für eine standardisiertere Erfassung der Beschwerdengültigkeit. Die best-evaluierten Methoden und Instrumente hierzu werden in der vorliegenden Arbeit vorgestellt. Dabei wird versucht, deren Validität insbesondere in Bezug auf die Anwendung im Begutachtungskontext nach der aktuellen Datenlage darzustellen. Die Autoren kommen zu dem Schluss, dass eine lege artis durchgeführte strukturierte Beschwerdengültigkeit die Qualität klinischer Urteile erheblich verbessern kann. Die qualifizierte Integration dieser zusätzlich erhobenen Befunde in eine psychologische oder ärztliche Expertise ist jedoch anspruchsvoll und bedarf einer weiteren wissenschaftlichen Begleitung, einer mechanischen Beurteilung nach Punktwerten gilt es vorzubeugen.

Schlüsselwörter: Beschwerdengültigkeit, Performanzvalidierung, Begutachtung, Versicherungsmedizin

## Abstract

On the background of high prevalence rates of non-authentic complaints in the field of insurance medicine symptom validity testing has become a growing issue to prevent non targeted monetary compensations. This paper outlines the best evaluated methods and instruments. Based on the data currently available it is attempted to discuss their validity in particular in the context of medico-legal assessment. The authors conclude that symptom validity assessment allows inferences about the degree of certainty of clinical judgements on the authenticity of reported symptoms. Thus, the application of the suggested instruments enhances significantly the quality of medical and psychological expertises. However, the integration of the additional results into the overall assessment is challenging and needs to be further clarified.

Keywords: symptom validity testing, performance validity, independent medical evaluation, insurance medicine

## Einleitung

In Europa zeigen psychische und psychosomatische Störungen ähnlich wie in anderen Industrienationen eine stark ansteigende Bedeutung für die Verursachung von Absenzen am Arbeitsplatz und generieren mittlerweile einen führenden Anteil der entstehenden Versicherungszahlungen im Erwerbsalter. Betroffen sind dabei zunehmend auch jüngere Menschen [1-3]. Die im OECD-Bericht [4] festgehaltene gesamteuropäische Situation bestätigt diese neben dem individuellen Leid auch wirtschaftlich problematische Entwicklung. Ausgehend von entsprechenden Untersuchungen in den USA kommt im deutschsprachigen Raum neben den prioritären medizinischen und betrieblichen Aspekten im Gefolge dieser Entwicklung auch die Frage auf, zu welchem Anteil Arbeitsunfähigkeit und Ausgleichszahlungen wirklich auf tatsächlich bestehenden, also nicht vorgetäuschten Gesundheitsstörungen beruhen. Eine gut belegbare kritische Sicht zeigt dabei, dass trotz der fortschreitenden Erkenntnisse in Medizin, Psychologie und Biologie der Zugang zu grossen Anteilen der Beschwerden weiterhin ein sprachlicher ist, der wesentlich bis ausschliesslich auf die subjektiven Angaben des Patienten zurückgeht. Das Risiko einer Fehleinschätzung der Authentizität von Symptomen und deren Zusammenhang zu Funktionseinschränkungen ist also in gewissem Sinne medizinimmanent und betrifft insbesondere psychiatrische, aber auch somatische Expertisen. Der Umgang mit nicht authentischen Beschwerden wird dennoch weder in der medizinischen noch in der psychologischen Ausbildung systematisch adressiert.

Einschätzungen mittels Gutachterbefragungen des Bundesamtes für Sozialversicherungen in der Schweiz gehen von etwa 8-10% der IV-Renten aus, bei welchen der Verdacht besteht, dass sie nicht zielkonform zugesprochen wurden [5, 6]. Bei diesen nicht-zielkonformen Leistungen müsse bei einem kleineren Teil davon ausgegangen werden, dass sie aufgrund falscher Angaben der Exploranden (Aggravation/Simulation) hervorgegangen sind. Diese seien nachvollziehbar hauptsächlich bei Störungsbildern zu erwarten, bei welchen ein erheblicher Ermessensspielraum für die Einschätzung der Arbeitsunfähigkeit bestehen kann, wie Depressionen, posttraumatischen Belastungsstörungen und Schmerzkrankungen einschliesslich Zuständen nach HWS-Distorsionen. Diese Liste ist jedoch um alle Störungsbilder zu erweitern, bei denen der Gutachter bei der Diagnosestellung und der Einschätzung der daraus resultierenden Beeinträchtigungen der Arbeitsfähigkeit im Wesentlichen auf die subjektiven Schilderungen des Betroffenen angewiesen ist. Hierzu zählen beispielsweise auch Tinnitus, fragliche Anfallsergebnisse, nicht prüfbare Störungen des Seh-/Hörvermögens und Sensibilitätsstörungen. Ein Gutachter darf, um seinem Auftrag zu entsprechen, die Authentizität von geschilderten Beschwerden nicht einfach voraussetzen, sondern muss die Validität der geltend gemachten Beschwerden sorgfältig überprüfen. Es wurde mehrfach herausgearbeitet [7-10], dass die Beurteilung der Glaubwürdigkeit von Beschwerdeangaben allein aufgrund klinischer Erfahrung der Gutachter unter Nutzung nicht standardisierter und oft nicht transparenter individueller Analyse von Verhaltensmustern fehleranfällig ist. Festgehalten sei aber ebenfalls, dass die Authentizität von subjektiv geklagten Beschwerden auch bei Nachweis von objektiven morphologischen Abweichungen nicht automatisch gegeben ist, da beispielsweise bildgebende Ab-

weichungen von Normalbefunden und daraus resultierende Beschwerden und Funktionseinschränkungen nur sehr unzuverlässig korrelieren [11-14]. Die Praxis zeigt jedoch, dass trotz des häufig niedrigen Beweiswertes eines solchen somatischen Befundes, die Rechtsanwender diesem einen unverhältnismäßig hohen Stellenwert einräumen. Auf diesem Hintergrund ergibt sich der Bedarf, Standards für die Beurteilung der Authentizität von Beschwerden zu entwickeln und diese im Gutachtensprozess zu verankern. Insbesondere die Validierung neurokognitiver Beschwerden kann sich dabei bereits auf eine breite Studienlage und die Entwicklung spezifischer Messinstrumente stützen und hat unzweifelhaft das Potential, die Konsistenzprüfung nach klinischen Befunden und Aktenlage sinnvoll zu ergänzen.

In der vorliegenden Arbeit sollen nun standardisierte Verfahren vorgestellt und methodisch kurz evaluiert werden in ihrer Möglichkeit, angegebene Beschwerden im psychologisch-psychiatrischen Bereich gutachterlich nachzuvollziehen, um verbleibende Funktionen besser einschätzen zu können.

### **Begriffsbestimmung**

Beschwerdvalidierung bezeichnet den Prozess, der zur Beurteilung der Beschwerdvalidität führt, und somit die Überprüfung der Authentizität von Beschwerdendarstellung und Symptompräsentation [15]. Invalide Beschwerdendarstellung bzw. Antwortverzerrung (response bias) bezeichnet ein Verhalten, das durch unzutreffende Antworten, fälschliche Angaben oder durch ein Testverhalten gekennzeichnet ist, welches nicht den tatsächlichen Möglichkeiten einer Person entspricht.

Simulation, Aggravation (engl. unter malingering zusammengefasst) und im weitesten Sinne auch Verdeutlichungstendenzen fallen dabei unter die Rubrik negativer Antwortverzerrung. Unter Simulation versteht man die bewusste, auf ein externes Ziel (z.B. finanzielle Entschädigung, Straffreiheit, Dienstunfähigkeit) ausgerichtete Vortäuschung von Krankheit oder Beschwerden bei Abwesenheit von tatsächlichen Beschwerden im beklagten Bereich. In vielen Kontexten dürfte diese eher die Ausnahme sein. Viel häufiger hingegen ist die Aggravation, also die absichtlich übertreibende Darstellung vorhandener Beschwerden zur Erlangung von Vorteilen. Hiervon wird von einigen Autoren die Verdeutlichungstendenz abgegrenzt, die häufig unbewusst motiviert ist und darauf abzielt, den Gutachter vom Vorhandensein der beklagten Beschwerden zu überzeugen [16]. Auf Leistungsebene tritt negative Antwortverzerrung als unzureichende oder suboptimale Anstrengungsbereitschaft (poor effort) auf, während sie bei beklagten psychischen oder somatischen Leiden oft als übertriebene Beschwerdendarstellung (exaggeration) in Erscheinung tritt. Der Begriff Anstrengungsbereitschaft ist problematisch, da Formulierungen wie geringe oder inadäquate Anstrengungsbereitschaft implizieren, dass ein Explorand sich bei der Bearbeitung nicht genügend anstrengt. Tatsächlich aber kann die Anstrengung erheblich sein – allerdings dahin gehend, ein schlechtes Testergebnis zu erzielen. Besser wäre es daher, von der Zielrichtung gezeigter Anstrengung zu sprechen.

Positive Antwortverzerrung zeichnet sich demgegenüber durch die Leugnung, Bagatellisierung und Abwehr von Beschwerden (Dissimulation), geschönte Selbstdarstellung (impression management), aber auch Antwortstile wie die Tendenz zur sozialen Erwünschtheit aus. Diese Form der Antwortverzerrung

findet sich vor allem, wenn das Sorge- oder Besuchsrecht zur Disposition steht, das Überwiegen körperlicher gegenüber psychischen Symptomen mit Vorteilen in der Rechtsprechung verbunden ist, bei Substanzmissbrauch oder auch dort, wo psychische Störungen kulturell bedingt transformiert oder stigmatisiert werden. Dissimulation kann aber auch psychologisch begründet sein, beispielsweise im Sinne einer Verleugnung von Defiziten bei Angst vor Ausgrenzung.

Selbst bei adäquater Beschwerdendarstellung und Anstrengungsbereitschaft kann es sich formal um Simulation handeln, nämlich dann, wenn authentische Beschwerden anderen Ursprungs bewusst einem nicht ursächlichen, jedoch besser versicherten Ereignis zugeschrieben werden [17]. Auf der anderen Seite ist die, einer überzeichneten Symptomdarbietung zugrunde liegende, Motivation nicht immer bewusster Natur, sondern kann auch in einem primären Krankheitsgewinn bestehen. So zeigen bei bestimmten psychiatrischen Krankheitsbildern Exploranden ohne Täuschungsabsicht eine verzerrte Beschwerdenschilderung und/oder unzureichende Anstrengungsbereitschaft; möglich ist dies z.B. bei Patienten mit somatoformen oder dissoziativen Störungen. Eine Auseinandersetzung mit der Problematik wie auffällige Ergebnisse in der Beschwerdengültigkeit bei diesen Patientengruppen zu bewerten sind, haben erst kürzlich Merten und Merckelbach [18] vorgenommen. Die Abgrenzung zwischen Simulation bzw. Aggravation, artifizieller, somatoformer und dissoziativer Störung ist auch deshalb schwierig, weil zwischen Selbst- und Fremdtäuschung fließende Übergänge bestehen, primärer und sekundärer Krankheitsgewinn sich oft vermischen und die Bewusstheit eigener Motive unterschiedlich ausgeprägt ist.

### **Prävalenzraten negativer Antwortverzerrung**

Die generelle Prävalenz von nicht-authentischen psychischen Störungen schwankt zwischen 10 und 30%, je nach Untersuchungskontext und Methodik [19]. Larrabee [20] schätzte anhand einer Analyse von 11 empirischen Studien, dass in etwa 40% der Begutachtungsfälle relevante Antwortverzerrungen vorkommen. Gemäss einer Erhebung unter amerikanischen Neuropsychologen auf der Basis von 30000 Begutachtungen wurden bei 32.7% aller Fälle von Invaliditätsansprüchen „wahrscheinliche nicht-authentische kognitive Beschwerden“ (nach Slick, Sherman & Iverson [21]) vermutet, im Bereich Personenschaden bei 30.4%. Im Rahmen von forensischen Begutachtungen belief sich dieser Wert auf immerhin noch 22.7%, bei neuropsychologischen Abklärungen im Rahmen medizinischer oder psychiatrischer Untersuchungen ausserhalb eines juristischen Kontextes auf 8.1%. Von den Exploranden, die sich im Rechtsstreit befanden bzw. Schadensersatzansprüche geltend machten, standen nicht nur 41.2% der Patienten nach leichten Schädelhirntrauma unter Täuschungsverdacht, sondern auch ein erheblicher Anteil von Gutachtenspatienten, welche psychische Störungen wie Fibromyalgie oder ein chronisches Erschöpfungssyndrom (38.6%), Schmerzen oder somatoforme Störungen (33.5%), depressive Störungen (16.1%), Angststörungen (13.6%) oder dissoziative Störungen (10.5%) beklagten. Unter Patienten nach mittelgradigem oder schwerem Schädelhirntrauma wurde die Authentizität der beklagten Beschwerden in lediglich 8.8% der Fälle angezweifelt, bei jenen mit vaskulärer Demenz in 2.4%. Danach scheint die Auftretenshäufigkeit negativer Antwortverzerrung wesentlich vom Untersuchungskontext

und den geltend gemachten Störungen bzw. den gestellten Diagnosen abhängig zu sein [22]. In Deutschland wurden ähnlich hohe Prävalenzraten für Gutachtenpopulationen gefunden. Merten, Friedel und Stevens [23] und Stevens, Friedel, Mehren et al. [24] konnten jeweils suboptimale Leistungsanstrengungen in Performanzvalidierungsverfahren<sup>1</sup> bei über 45% aller untersuchten Probanden finden, die wegen psychiatrischer und neurologischer Störungen begutachtet wurden. Nach empirischen Schätzungen von Greve, Ord, Bianchini et al. [25] liegt die Prävalenz von Täuschungsverhalten bei Patienten mit chronischen Schmerzen im medico-legalen Kontext je nach Klassifikationssystem zwischen 25% und 36%. 21% dieser nicht-authentischen Probanden wiesen objektive somatische Befunde auf. Auch bei anderen somatischen Krankheitsbildern finden sich Hinweise auf Aggravation oder Simulation, so z.B. bei 58% geltend gemachter Hörstörungen [26].

### **Methoden zur Beschwerdvalidierung**

Ergebnisse psychologischer Testverfahren sind leicht zu verfälschen und ohne spezifische Methoden hinsichtlich ihrer Gültigkeit nicht zu beurteilen. Deshalb wurden von Neuropsychologen spezielle Performanzvalidierungsverfahren entwickelt, die inzwischen als am besten untersuchte Methodengruppe gelten. Weitere neuropsychologische Ansätze bestehen in Validitätsindikatoren, welche aus Parametern gebräuchlicher Testinstrumente abgeleitet werden können, sowie im Einsatz von Fragebögen und Selbstbeurteilungsinstrumenten, welche über spezifisch entwickelte Skalen zur Identifizierung wahrscheinlicher negativer und positiver Antwortverzerrungen verfügen. Letztere dienen vor allem der Validierung der Beschwerdenschilderung. Tabelle 1 stellt eine Übersicht über die Methoden der Beschwerdvalidierung dar.

#### **A. Verfahren zur Validierung der Anstrengungsbereitschaft**

Beschwerdvalidierungs- oder nach neuerer Terminologie Performanzvalidierungstests (PVT) sind speziell zur Erfassung von Antwortverzerrungen entwickelt worden. Im Bereich kognitiver Funktionen operieren sie vorwiegend im Bereich des Gedächtnisses, wobei sie eine hohe Robustheit gegenüber realen kognitiven Defiziten aufweisen. Aufgrund verschiedener Ansätze bei der Testkonstruktion sind die zwischenzeitlich verfügbaren Verfahren unterschiedlich gut in der Lage authentische und negativ verzerrte Testergebnisse zuverlässig zu unterscheiden. Zur Beurteilung der Güte eines PVT entwickelte Hartman [27] einen allgemein akzeptierten Kriterienkatalog. Ein guter Test sollte danach

- Anstrengungsbereitschaft messen, aber nicht für kognitive Störungen anfällig sein (Sensitivität und Spezifität)
- eine gute normative Basis aufweisen. Nur wenn die Ergebnisse von authentischen Patienten verschiedener Diagnosegruppen und mit unterschiedlichem Ausmass kognitiver Störungen bekannt sind, kann ein zuverlässiger Trennwert festgelegt werden.

---

<sup>1</sup> In der internationalen Literatur etabliert sich zunehmend eine sprachliche Unterscheidung zwischen Performanzvalidierungstests, die die Anstrengungsbereitschaft von Probanden in Leistungstests überprüfen und Beschwerdvalidierungstests, mit denen die Beschwerdenschilderung überprüft werden soll.



- beim Exploranden den Eindruck erwecken, dass es sich um eine realistische Messung der untersuchten kognitiven Funktion handelt (Augenscheinvalidität)
- auf Validierungsstudien beruhen, die gesunde Probanden, Patientengruppen und echte, sowie instruierte Simulanten einschliessen
- schwierig zu verfälschen oder zu coachen sein, um die Testsicherheit zu gewährleisten
- hinsichtlich Testanwendung, -auswertung und -interpretation möglichst einfach sein
- Symptombereiche messen, welche von Exploranden häufig übertrieben werden
- von aktuellen Forschungsergebnissen unterstützt werden.

### A.1 Detektionsstrategien zur Validierung der Anstrengungsbereitschaft

Grundlage der PVT sind Testprinzipien, die motivationale Einflüsse zufallskritisch und wahrrscheinlichkeitsbasiert abschätzen [28]. Nach Rogers, Harrell und Liff [29] lassen sich neuropsychologische Forschungs- und Diagnoseansätze zur Performanzvalidierung wie folgt klassifizieren:

#### Alternativwahlverfahren

Alternativwahlverfahren stellen die am besten untersuchte Methodengruppe mit der höchsten Klassifikationsgenauigkeit bei der Differenzierung von authentischen und nicht-authentischen Beschwerden dar. Sie sind so konzipiert, dass ein Zielreiz zusammen mit einem Distraktorreiz angeboten wird, wobei der Explorand den Zielreiz identifizieren soll. Wenn er die richtige Antwort nicht kennt, so muss geraten werden (forced-choice). Statistisch ist bei einer dichotomen Reizverteilung von Ziel- und Distraktorreiz eine 50%-ige Ratewahrscheinlichkeit als minimale Wiedererkennungsrage zu erwarten. Eine wesentlich geringere Trefferquote (signifikant below-chance) ist als Beleg für eine negative Antwortverzerrung zu interpretieren. Damit erlauben sie im Unterschied zu allen anderen Verfahren mit mathematisch begründeter und exakt zu bestimmender Sicherheit willentliche, bewusste Antwortverzerrung nachzuweisen und somit eine sichere Vortäuschung kognitiver Störungen gemäss den weiter unten beschriebenen Slick-Kriterien [21] zu identifizieren. Neben den Antwortmustern unter Zufallsniveau, welche eher selten vorkommen, wird mit Trenn- bzw. Grenzwerten (cut-offs) operiert, die Wahrscheinlichkeitsaussagen über das Vorliegen unzureichender Leistungsmotivation erlauben. Ermittelt werden solche Trennwerte bzw. als authentisch zu betrachtende Wertebereiche durch sogenannte „known-groups“ oder „criterion-groups“-Validierungen, in welchen als simulierend diagnostizierte Patienten mit ähnlichen authentischen Patienten verglichen werden, die objektiv häufig grössere kognitive Einschränkungen, schmerzassoziierte Behinderungen oder psychische Belastungen aufweisen, als die Gruppe simulierender Patienten. Eines dieser Alternativwahlverfahren stellt beispielsweise der Test of Malingered Memory (TOMM; [30]) dar.

#### Testdeckeneffekt oder Prinzip der verdeckten Leichtigkeit

Diese Tests täuschen einen Schwierigkeitsgrad vor, der von Exploranden häufig überschätzt wird. Tatsächlich sind die Aufgaben jedoch so einfach, dass sie von nahezu allen Exploranden bewältigt werden

können, auch von solchen mit authentischen kognitiven Defiziten. Aufgrund des niedrigen Schwierigkeitsgrades wird bei anstrengungsbereiten Exploranden eine Punktzahl im Bereich des Maximums (der „Decke“) erwartet. In diese Gruppe fällt z.B. der Rey Fifteen Items Test (FIT; [31]).

### Leistungskurve

Hierbei wird die Plausibilität des Antwortverhaltens innerhalb eines Tests bzw. zwischen verschiedenen Tests überprüft. Die Leistungskurve eines Exploranden soll dem realen Schwierigkeitsgrad der gestellten Aufgaben oder Anforderungen innerhalb eines Funktionsbereichs entsprechen. Danach sollten einfache Aufgaben eher bewältigt werden als deutlich schwierigere Aufgaben. Beispiele für diesen Ansatz stellen der Dot Counting Test [32, 33] oder die Gegenüberstellung von Wiedergabeleistungen zuvor gelernter Inhalte unter verschiedenen Abrufbedingungen dar. Der typische Leistungsgradient kann bei bestimmten Krankheitsbildern wie z.B. Aphasie oder Akalkulie verändert sein. Wenn keine oder nur wenige empirische Daten zu einem „typischen“ Leistungsprofil vorliegen und die Klassifikationsgüte bestimmter Antwortmuster unbekannt ist, ist ebenfalls Vorsicht geboten.

### Fehlergrösse

Dieser Ansatz betrifft das gehäufte Auftreten von Knapp-daneben-Antworten. Weicht die Antwort eines Exploranden regelmässig um eine bestimmte, geringe Konstante von der korrekten Lösung ab, so weist dies darauf hin, dass der Explorand die richtige Antwort kennt, sie aber wissentlich verfälscht. Ein solches Verhalten lässt sich beispielsweise bei Orientierungsfragen oder Rechenaufgaben beobachten.

### Identifikation untypischer oder inkonsistenter Leistungsprofile

Hierunter fällt die Überprüfung der Konsistenz und Plausibilität eines Testprofils innerhalb einer umfassenden Untersuchung oder über mehrere Testzeitpunkte hinweg. Von Interesse sind hierbei v.a. Ergebnisse von Testverfahren, die gleiche oder ähnliche Eigenschaften erfassen. Ausserhalb des Einsatzes fixer Testbatterien ist dieser Ansatz jedoch wenig formalisiert, so dass hier dieselben Einschränkungen gelten wie für die Analyse der Leistungskurve.

## A.2 Spezielle Performanzvalidierungstests

Einige der speziell entwickelten, standardisierten PVT wie der Word Memory Test (WMT; [34]) kombinieren die genannten Grundprinzipien teilweise. Kritisch ist anzumerken, dass die hohe Sensitivität des Verfahrens bei einzelnen Patientengruppen möglicherweise mit einer reduzierten Spezifität einhergeht. So erzielten in einer Untersuchung von Gorissen, Sanz und Schmand [35] 72% Schizophrener ein Ergebnis unterhalb des empirischen Trennwertes. Negativsymptome (Apathie oder Interessenverlust) korrelierten hierbei signifikant mit der Leistungsmotivation. Es wurden von Green noch zwei weitere Kurzverfahren aufgelegt - der Medical Symptom Validity Test (MSVT; [36]) und der Nonverbal Medical Symptom Validity Test (NV-MSVT; [37]) - für die, bis auf eine geringere Zahl an Vergleichsdaten, im Wesentlichen die gleichen Bewertungen gelten. Der TOMM weist als weltweit verbreitetes Verfahren eine grosse empirische Datenbasis auf, ist gut validiert, hat eine hohe Augenscheinvalidität und ist

als Computer- wie Papier-Bleistift-Version anwendbar. Als Beispiel eines nicht dem Alternativwahlprinzip folgenden Verfahrens sei der Amsterdamer Kurzzeitgedächtnistest (ASTM, dt. AKGT) von Schmand und Lindeboom [38] genannt.

### A.3 Eingebettete Beschwerdenvalidierungsparameter

Im weiteren Sinne fallen auch Indikatoren in konventionellen neuropsychologischen Tests, sogenannte eingebettete Beschwerdenvalidierungsparameter unter die Kategorie der PVT. Diese Parameter können in verschiedenen Tests direkt aus den Verfahren gewonnen werden, wie beispielsweise die Reliable Digit Span aus dem Untertest „Zahlen nachsprechen“ des Wechsler Intelligenz- oder Gedächtnistests [39]. Andere Variablen erfordern den ergänzenden Einsatz zusätzlicher Untertests, die jeweils speziell für die Beschwerdenvalidierung entwickelt wurden. Hierzu zählen z.B. zusätzliche Wiedererkennungsdurchgänge im Complex Figure Test [40]. Da die Datenlage zu den meisten dieser Indikatoren fortlaufend ergänzt wird, muss hier auf die jeweilige Originalliteratur verwiesen werden. Englischsprachige Zusammenstellungen finden sich bei Boone [41, 42], Larrabee [43], Morgan und Sweet [44] sowie Carone und Bush [45].

Da die Anstrengungsbereitschaft im Verlauf einer Begutachtung durchaus Schwankungen unterliegt und negative Antwortverzerrung sich selektiv auf einzelne Funktionsbereiche beschränken kann, erhöht sich durch den Einsatz multipler testinterner Validitätsparameter die Wahrscheinlichkeit solche Verzerrungen im Untersuchungsverlauf zu entdecken [46]. Gleichzeitig lässt sich wegen des Prinzips der kombinierten Unwahrscheinlichkeit auch die Spezifität auf deutlich über 90% erhöhen [47, 48]. Hierbei ist allerdings zu beachten, dass die als authentisch bzw. nicht-authentisch herangezogenen Parameter inhaltlich voneinander unabhängig sein müssen, da sich allein aus mathematisch-statistischen Gründen die Wahrscheinlichkeit eines auffälligen Validitätsparameter mit steigender Anzahl erhöht. Ein entscheidender Nachteil besteht darin, dass die Klassifikationsgüte eines einzelnen Indikators nicht an jene eines eigenständigen Performanzvalidierungstests heranreicht und sie damit die Gefahr einer falsch-positiven Klassifikation in sich bergen. Ihr Einsatz erfordert daher eine detaillierte Kenntnis der Originalliteratur, da die empfohlenen Grenzwerte für unterschiedliche diagnostische und ethnische Patientengruppen durchaus differieren können.

Ein umfassenderer Überblick über im deutschen Sprachraum verwendete standardisierte PVT und eingebettete Beschwerdenvalidierungsparameter findet sich bei Merten [49].

## **B. Verfahren zur Validierung der Beschwerdenschilderung**

Neben der Überprüfung der Anstrengungsbereitschaft (Performanzvalidierung) stellt auch die Validierung der Beschwerdenschilderung einen essentiellen Baustein der Begutachtung dar. Obwohl verfahrensübergreifend das Bild eines mässigen Zusammenhangs zwischen Performanzvalidierungstests und solchen zur Validierung der Beschwerdenpräsentation entsteht, scheinen beide Ansätze doch verschiedene Konstrukte zu erfassen. Aus diesem Grund empfiehlt es sich neben der Anstrengungsbereitschaft

auch die Beschwerdendarstellung einer Prüfung auf Authentizität zu unterziehen. Grundsätzlich muss in jeder Begutachtungssituation mit einer von verschiedenen Prozessen überlagerten Berichterstattung gerechnet werden. Ein Vorteil von Fragebögen besteht darin, dass einige dieser verzerrenden Einflussfaktoren in einer Begutachtungssituation erfasst und unabhängig von persönlichen Wertungen des Gutachters abgebildet werden können. Auch wenn solche Validitätsskalen keine Aussage darüber erlauben, ob und allenfalls welche Gesundheitsstörungen tatsächlich bestehen, weisen sie den Gutachter doch auf mögliche Verfälschungen der Beschwerdendarstellung hin.

### B.1 Detektionsstrategien zur Validierung der Beschwerdenschilderung

Bewährte Strategien zum Nachweis übertriebener Beschwerdenpräsentation oder vorgetäuschter psychischer Störungen wurden von Rogers [50] formuliert:

#### Seltene Beschwerden

Diese werden von authentischen Patienten äusserst selten berichtet. Beispiele für diesen Ansatz stellen u.a. die Infrequency Psychopathology Scale (Fp) und die Infrequent Somatic Complaints Scale (Fs) des Minnesota Multiphasic Personality Inventory (MMPI-2) dar, die sich für die Aufdeckung vorgetäuschter psychischer bzw. somatischer Beschwerden zu bewähren scheinen.

#### Quasi-seltene Beschwerden (quasi-rare symptoms)

Die Entwicklung solcher Skalen (z.B. MMPI-2 F-Skala und Fb-Skala) beruht auf klinischen Merkmalen, welche in der Allgemeinbevölkerung selten vorkommen.

#### Unwahrscheinliche Beschwerden

Dieser Ansatz (z.B. im Structured Inventory of Reported Symptoms (SIRS)) beruht auf der Präsentation absurder, grotesker Beschwerden. Sein Vorteil liegt darin, dass solche Beschwerden kaum eine andere Interpretation als die einer Täuschungsabsicht zulassen und eine Anwendung auch auf somatische Beschwerden möglich ist.

#### Beschwerdenkombinationen, welche bei echten psychiatrischen Störungen äusserst selten vorkommen

Exploranden mit Täuschungsabsicht geben nicht selten ungewöhnliche Kombinationen von Beschwerden an, wie z.B. Grössenideen bei gleichzeitig erhöhtem Schlafbedürfnis. Diese schwer zu durchschauende Detektionsstrategie dürfte nicht leicht zu coachen sein. Sie wird beispielsweise vom SIRS und dem Miller Forensic Assessment of Symptoms Test (M-FAST) benutzt.

#### Unübliche psychopathologische Muster

Dieser Ansatz, der sich u.a. im Personality Assessment Inventory (PAI) wiederfindet, fokussiert auf Kombinationen erhöhter Skalenwerte, welche als charakteristisch für Simulationstendenzen gelten, bei authentischen Patientengruppen jedoch äusserst ungewöhnlich sind.

#### Hohe, undifferenzierte Ausmasse an Beschwerden

Patienten mit Täuschungsabsichten können oft schon alleine durch die Vielzahl beklagter Symptome identifiziert werden. Überprüft wurde diese Detektionsstrategie im Bereich primär psychischer Beschwerden bisher nur in strukturierten klinischen Interviews wie dem SIRS.

### Aussergewöhnlich hohe Ausprägungen einzelner Symptome

Selbst schwer beeinträchtigte Patienten empfinden nur eine geringe Anzahl an Beschwerden als unerträglich oder aussergewöhnlich intensiv. Personen mit Täuschungsabsicht hingegen geben häufig für ein weites Spektrum von Symptomen extreme Ausprägungen an. Eine entsprechende Validitätsskala stellt z.B. die Lachar-Wrobel Critical Items Scale des MMPI-2 dar.

### Eindeutige Beschwerden

Dieser Ansatz, u.a. umgesetzt in der MMPI-2 Obvious-Subtle Difference Scale, beruht auf dem Gedanken, dass Personen mit Täuschungsabsicht wahrscheinlich Beschwerden beklagen, welche als klare Indikatoren einer ernsthaften psychischen Erkrankung bekannt geworden sind.

### Diskrepanzen zwischen beobachteten und berichteten Beschwerden

Täuschende Personen können oft durch die Richtung dieser Diskrepanzen identifiziert werden. In diesen Fällen fehlen die zu erwartenden, mit bestimmten beklagten Beschwerden in der Regel einhergehenden, klinischen Beobachtungen in der Untersuchungssituation (SIRS Reported vs. Observed Scale).

### Fehlerhafte Stereotypen

Diese Methode nutzt verbreitete, fehlerhafte Annahmen über die mit psychischen Störungen assoziierten klinischen Merkmale. Eine zugehörige Validitätsskala stellt die MMPI-2 Dissimulation Scale dar.

Validiert wurden die erstmals von Rogers [51] vorgestellten Kriterien sowohl in Analogstudien mit instruierten Simulanten als auch in Known-Groups Designs mit Gutachtenspopulationen und erzielten hohe bis sehr hohe Effektstärken. Für einen detaillierten Überblick sei auf Rogers [19] verwiesen.

## B.2 Fremdbeurteilungsverfahren für psychische Störungen

Das Structured Interview of Reported Symptoms (SIRS; [52], SIRS-2, revidierte Version; [53]) setzt die oben beschriebenen Validierungstechniken in der Interviewsituation um [54]. Es beinhaltet acht Primärskalen und fünf Zusatzskalen, welche psychische Störungen erheben sollen. Das Verfahren besticht durch eine hervorragende Interrater-Reliabilität [55] und eine Retestreliabilität zwischen 97.6% und 99.5% bei forensischen Patienten [56]. McCusker, Moran, Serfass et al. [57] fanden auf zwei Skalen hohe Ladungen (Aggravation von Symptomen und unglaubwürdige Beschwerden), welche hoch mit der Fp- und der F-Skala des MMPI-2 korrelierten, was als Beleg für die Konstruktvalidität des SIRS interpretiert wurde. Minuspunkte stellen die relativ lange Durchführungszeit und die nötige Expertise der Interviewer dar [58]. Eine deutsche Version ist als Strukturiertes Interview Berichteter Beschwerden (SIBB; [59]) in Vorbereitung.

Auch der Miller Forensic Assessment of Symptoms Test (M-Fast; [60]) soll simulierte psychische Störungen im Rahmen eines strukturierten Interviews aufdecken. Validiert wurde das Instrument primär in forensischen Kontexten. Das Interview lieferte in einer Studie von Smith [61] bezüglich Reliabilität und Validität sehr gute psychometrische Kennwerte. Guy und Miller [62] konnten mittels hoher Korrelationen mit dem SIRS Belege für die Konstruktvalidität des Interviews erbringen. Bei dem Instrument handelt es sich um ein reines Screening-Verfahren, auf dessen isolierten Ergebnissen eine Klassifikation

eines Exploranden nicht statthaft ist. Dem klinischen Urteil von Psychologen und Psychiatern gegenüber erwies sich das Instrument jedoch als deutlich überlegen, wobei es v.a. den Anteil falsch-negativer Klassifikationen, also den Anteil übersehener Täuschungsmanöver reduzierte [63]. Ein weiterer Pluspunkt besteht in der kurzen Durchführungs- und Auswertungszeit von zusammen etwa 15-20 Minuten.

### B.3 Selbstbeurteilungsverfahren und andere Screeningtests für psychische Störungen

Der Strukturierte Fragebogen simulierter Symptome (SFSS; [58]) als deutsche Version des Structured Inventory of Malingered Symptomatology (SIMS) von Smith und Burger [64] stellt eine leicht handhabbare Screeningmethode zur Erfassung einer Vielzahl simulierter Symptome dar. Für die englische Originalfassung (SIMS) erwies sich der Gesamtwert mit einer Sensitivität von 95.6% und einer Spezifität von 87.9% [64] als ein effizienter Indikator für nicht-authentische Beschwerdenschilderung. Die deutsche Version verfügt über eine hohe Retestreliabilität ( $r = 0.97$ ) und eine akzeptable interne Konsistenz (Cronbach's  $\alpha = 0.80$ ). Zudem weist sie eine gute vorhersagende und übereinstimmende Validität aus. Die Werte für Sensitivität, Spezifität, Positive Predictive Power (PPP) und Negative Predictive Power (NPP) liegen in Abhängigkeit vom gewählten Grenzwert über 0.85. Damit ist die diagnostische Genauigkeit recht akzeptabel, solange man den SFSS als Screeningmethode benutzt. Die falsch-positiv Rate variiert dabei mit dem Trennwert, der in Abhängigkeit von Untersuchungskontext und Fragestellung unterschiedlich hoch gesetzt werden kann [65]. Ein systematisches Review mit Meta-Analyse zum SFSS ist kürzlich erschienen [66].

Das Minnesota Multiphasic Personality Inventory (MMPI-2; [67]) bzw. dessen Nachfolger MMPI-2-RF ist ein umfassendes Persönlichkeitsverfahren mit eingebetteten, gut überprüften Validitätsskalen. Neben den klassischen Validitätsskalen (L, F und K) wurden im Laufe der Zeit weitere entwickelt, die sich die weiter oben beschriebenen, unterschiedlichen Strategien zur Aufdeckung negativer Antworttendenzen zunutze machen. Eine vergleichsweise neu entwickelte Validitätsskala umfasst Items, die eine Unterscheidung zwischen Personen mit hoher und niedriger Anstrengungsbereitschaft in Performanzvalidierungsverfahren ermöglichen (Response Bias Scale; [68, 69]). Eine Stärke des Verfahrens liegt darin, dass die entwickelten Validitätsskalen gemäss verschiedener Metaanalysen negative Antwortverzerrung im Bereich psychischer Störungen, körperlicher Störungen und kognitiver Störungen aufzudecken vermögen [70, 71]. Im forensischen Kontext fanden Lewis, Simcox und Berry [72] je nach Skala Spezifitäten zwischen 94% und 100%. Auch für die Aufdeckung positiver Antwortverzerrung existiert eine erste Metaanalyse [73], wonach die entsprechenden Skalen tatsächlich zur Unterscheidung von Dissimulation und authentischem Antwortverhalten in der Lage sind. Die jeweils effektivsten Validitätsskalen und Trennwerte variieren in Abhängigkeit von klinischen Gruppen, Untersuchungskontexten und Grundraten von Simulation, Aggravation und Dissimulation. Seit 2000 liegt eine deutsche Version des MMPI-2 vor [74], für welche inzwischen eine Schätzung der Gütekriterien der Validitätsskalen an einer realen Gutachtenspopulation erhoben wurde [75]. Unabhängig davon weisen verschiedenste

Arbeiten darauf hin, dass der MMPI-2 kulturübergreifend anwendbar ist [76-79]. Eine deutsche Übersetzung des mit nur knapp der Hälfte der Items wesentlich kürzeren MMPI-2-RF ist in Vorbereitung. Mit dem Personality Assessment Inventory (PAI; [80]) liegt seit einigen Jahren ein weiteres multidimensionales Selbstbeurteilungsinstrument zu klinischen Symptomen und Persönlichkeitsmerkmalen vor, das eine Überprüfung der Beschwerdenschilderung erlaubt. Zur Aufdeckung von vorgetäuschter Psychopathologie bzw. negativer Antwortverzerrung beinhaltet er drei Indikatoren, die auf der Aufdeckung seltener Beschwerden und unüblichen psychopathologischen Mustern beruhen. Von den acht weiteren Validitätsindikatoren des PAI schätzen drei ein, ob die Beantwortung zufällig erfolgt ist, fünf Masse evaluieren positive Antwortverzerrungen. Vorteile des Verfahrens bestehen in der gegenüber dem MMPI-2 deutlich geringen Zahl an Fragen sowie der inhaltlichen Ausrichtung an ICD-10 und DSM-IV. Positiv zu werten ist auch die hohe Reliabilität und gute Validität der Skalen in verschiedenen Settings, ebenso wie die hohen Effektstärken der Validitätsskalen [81].

Seit 2013 existiert mit dem Verhaltens- und Erlebensinventar (VEI; [82]) eine deutschsprachige Adaption des PAI, dessen Normdaten den amerikanischen Normen weitgehend zu entsprechen scheinen. Auch für das VEI wurden an kleinen Gruppen instruierter Simulanten und zu begutachtender Patienten erste akzeptable Schätzungen der Sensitivität und Spezifität der Skalen für negative Antwortverzerrung vorgelegt [83]. Angesichts ihrer Ergebnisse stellen die Autoren für deutschsprachige Patienten allerdings eine Adaption der von Morey [80] empfohlenen Trennwerte zur Diskussion.

Der Morel Emotional Numbing Test for Posttraumatic Stress Disorder (MENT; [84-86]) ist ein aus 60 Items bestehendes forced-choice Verfahren, bei welchem anhand von Gesichtsausdrücken Emotionen identifiziert werden sollen. Der Test basiert auf der suggerierten Annahme, dass Personen mit genuiner PTSD hier Schwierigkeiten aufweisen und nutzt so die oben erwähnte Detektionsstrategie falscher Stereotype. Ursprünglich an US-Veteranen validiert, umfasst das Verfahren inzwischen Daten zu unterschiedlichen Patientengruppen, auch im medico-legalen Kontext. Es verfügt über eine exzellente Spezifität zwischen 96 und 100% und eine gute Sensitivität von bis zu 85% [86, 87] und sollte nach Meinung verschiedener Autoren [88, 89] bei Begutachtungen von PTSD routinemässigen Einsatz finden. Der MENT ist u.a. in deutscher und türkischer Sprache erhältlich. Normalbefunde von 64 Probanden für den deutschen Sprachraum finden sich bei Wiedmaier [90].

### **C. Nachweis bewusster Täuschungsabsicht**

Die Diagnose von Simulation oder Aggravation beinhaltet immer eine Aussage über Bewusstheit und Motivation des Täuschungsverhaltens. Um bei nachgewiesener mangelnder Anstrengungsbereitschaft und/oder verzerrter Beschwerdenschilderung eine bewusste Täuschungsabsicht (Aggravation/Simulation) von unbewussten negativen Antwortverzerrungen z.B. im Rahmen einer Konversionsstörung abgrenzen zu können, müssen weitere Informationen mit herangezogen werden. Die Differentialdiagnose von Simulation bzw. Aggravation beinhaltet einen Prozess, der eine sorgfältige Analyse aller verfügbarer Informationen erfordert, auf objektiven Kriterien basiert, Indikatoren mit einschliesst, welche über

eine bekannte Klassifikationsgenauigkeit verfügen und das klinische Urteil mit den Ergebnissen wissenschaftlich validierter Verfahren kombiniert. Dabei ist auch den Besonderheiten der Exploranden wie Bildungsstand, kultureller Hintergrund, ethnische Zugehörigkeit und Sprachkompetenz Rechnung zu tragen.

### C.1 Kriterienkatalog für die Diagnose aggravierter/simulierter neurokognitiver Störungen

In diesem Zusammenhang hat sich ein von Slick et al. [21] bereits 1999 vorgestellter Kriterienkatalog für die Diagnose vorgetäuschter neurokognitiver Störungen (Malingered Neurocognitive Disorder, MND) in Forschung wie Einzelfalldiagnostik etabliert. Danach ist eine vorgetäuschte neurokognitive Störung als willentliche Übertreibung oder Vortäuschung kognitiver Dysfunktionen definiert, die zum Ziel hat, einen externen materiellen Gewinn zu erlangen, oder aber einer gesetzlichen Verpflichtung (z.B. Militärdienst) oder Verantwortung (z.B. Gerichtsverfahren) zu entkommen. Konzeptuell lässt sich dieser externe Anreiz als sekundärer Gewinn einordnen. Dabei fokussieren die Slick-Kriterien nicht auf psychosoziale Aspekte des Krankheitsgewinns, sondern auf bedeutsame Vorteile wie Entschädigungszahlungen oder Renten. Der Terminus „willentlich“ schliesst Bewusstheit und Absicht mit ein. Im Kontext mit der Diagnose von Aggravation/Simulation wird das Ausmass, in welchem ein Verhalten als willentlich deklariert wird, durch Ausschluss plausibler, alternativer Erklärungen bestimmt (D-Kriterium). Mit anderen Worten darf das zu beurteilende Verhalten nicht besser als unbewusstes oder unabsichtliches Handeln im Rahmen von psychiatrischen, neurologischen oder Entwicklungsstörungen erklärt werden können. Falls also ein Explorand in einem forced-choice Verfahren eine Leistung zeigt, die unter dem Zufallsniveau liegt, wird angenommen, dass er nicht in der Lage ist, eine korrekte Antwort zu unterdrücken und durch eine Falsche zu ersetzen, ohne dass ihm dies bewusst ist. Somit kann dieses Verhalten als vorsätzlich bezeichnet werden.

Slick und Sherman [91] haben kürzlich verschiedene Modifikationen und Erweiterungen des ursprünglichen Kriterienkatalogs vorgeschlagen. Zu den Neuerungen zählt zunächst die Unterscheidung in primäre und sekundäre Aggravation/Simulation. Bei letzterer sind die Kriterien für eine definitive oder wahrscheinliche Aggravation/Simulation neurokognitiver Störungen bis auf das frühere D-Kriterium erfüllt. Gibt es aber zwingende Gründe anzunehmen, dass der Explorand, sekundär zu Unreife (bei Kindern) oder bona fide Entwicklungsstörungen (z.B. geistige Behinderung), psychiatrischen oder neurologischen Erkrankungen oder Verletzungen mindestens mittelschwerer Ausprägung, zum Untersuchungszeitpunkt nicht über die kognitiven Fähigkeiten verfügte, um die moralischen, ethischen und rechtlichen Implikationen seines Verhaltens zu verstehen und/oder zudem unfähig war, sein Verhalten zu kontrollieren, so wird neu, statt des Ausschlusses der Diagnose MND, die Diagnose sekundäre Aggravation/Simulation vorgeschlagen. Bei Personen mit leichten Erkrankungen/Verletzungen wie z.B. einer MTBI kann diese Diagnose nicht gestellt werden. Der Wegfall des Ausschlusskriteriums (D) soll



der Vermeidung von Missverständnissen dienen. So könnte Kriterium D u.a. fälschlicherweise der Annahme Vorschub leisten, Patienten mit psychiatrischen, neurologischen und entwicklungsbedingten Störungen seien unfähig, Beschwerden zu übertreiben oder zu simulieren.

Änderungen wurden auch für die Definition eindeutiger bzw. definitiver und wahrscheinlicher Aggravation/Simulation neurokognitiver Störungen vorgeschlagen. Zu den eindeutigen Indikatoren für eine Aggravation/Simulation neuropsychologischer Defizite zählt neben Testleistungen unter Zufallsniveau in einem oder mehreren forced-choice Verfahren nun ebenso eine hohe a-posteriori-Wahrscheinlichkeit ( $\geq .95$ ) in einem oder mehreren gut validierten psychometrischen Indizes auch aus Selbstbeurteilungsinstrumenten (z.B. Validitätsskalen des MMPI-2), die belegen, dass die gezeigte Leistung erheblich unter dem aktuellen Fähigkeitsniveau eines Exploranden liegt. Ziel war es, den Kriterienkatalog auch für Fälle anwendbar zu machen, in welchen primär psychologische Symptome übertrieben oder vorgetäuscht werden, während für Befunde aus Leistungstests keine Aggravation/Simulation nachzuweisen ist. Der Nachweis einer hohen a-posteriori-Wahrscheinlichkeit bedingt jedoch das Vorliegen zuverlässiger Grundraten zur Auftretenswahrscheinlichkeit von Aggravation/Simulation in einer dem Exploranden vergleichbaren Population (a-priori-Wahrscheinlichkeiten) und einen gut validierten Prozess zur Bestimmung der positiven prädiktiven Vorhersagekraft. Neu sollen auch beklagte Beschwerden, welche mit dem beobachtetem Verhalten und/oder den Testleistungen eindeutig nicht vereinbar sind oder in direktem Widerspruch dazu stehen (compelling inconsistencies) als eindeutige Indikatoren für Aggravation/Simulation gelten. Beispielhaft führen die Autoren hierzu Erkenntnisse aus Beobachtungen des Gutachters (in Pausen, beim Blick aus dem Fenster usw.), Observationen, Überwachungsvideos und andere Datenquellen an. Die Kriterien für eine wahrscheinliche Aggravation/Simulation neuropsychologischer Defizite bestehen neben der Präsenz eines bedeutsamen externen Anreizes in drei oder mehr Indikatoren möglicher Aggravation/Simulation neuropsychologischer Probleme oder Defizite. Diese können sowohl aus Testdaten, die eher für unglaubwürdige Leistungen sprechen sowie markanten und unplausiblen Diskrepanzen zwischen Testleistungen und/oder Beschwerdenschilderung auf der einen Seite und beobachtbarem Verhalten, zu erwartendem Fähigkeitsniveau aufgrund der Krankengeschichte und reliablen Zusatzinformationen auf der anderen Seite bestehen.

### C.2 Kriterienkatalog für die Diagnose aggravierter/simulierter schmerzassoziierter Behinderungen

Für die Aufdeckung aggravierter/simulierter schmerzassoziierter Behinderungen (Malingered Pain-Related Disabilities, MPRD) wurden die ursprünglichen Slick-Kriterien von Bianchini, Greve und Glynn [92] modifiziert und um ein Kriterium (B) erweitert, welches Hinweise auf übertriebene Darstellung oder Vortäuschung körperlicher Behinderung aus der somatischen Untersuchung oder Behandlung (z.B. medizinische Untersuchung, EFL, Physiotherapie) liefert. Zu diesen zählen Leistungen in einem oder mehreren gut validierten Verfahren zur Erfassung physischer Fähigkeiten (z.B. Jamar Grip Test), welche mit der übertriebenen Darstellung oder Vortäuschung körperlicher Einschränkungen vereinbar sind, Diskrepanzen zwischen subjektiven Schmerzschilderungen und physiologischer Reaktivität, eindeutig

nicht-organische Befunde sowie Diskrepanzen zwischen der Art, wie sich der Patient während der formalen Untersuchung präsentiert und seinem körperlichen Funktionsvermögen, wenn er sich nicht beobachtet fühlt. Auch für die Diagnose einer definitiven Aggravation/Simulation schmerzassoziierter Behinderungen ist neben einem signifikanten externen Anreiz und dem Ausschluss von Differentialdiagnosen, ein eindeutiger Beweis des Vorsatzes notwendig. Dieser gilt nach Bianchini et al. als erbracht, wenn eine definitive negative Antwortverzerrung, also eine Leistung in einem Alternativwahlverfahren zur Erfassung kognitiver oder perzeptiver Fähigkeiten unterhalb des Zufallsniveaus ( $p < .05$ ) oder eindeutige Inkonsistenzen (compelling inconsistencies) vorliegen. Die Autoren betonen, dass Aggravation/Simulation schmerzassoziierter Beeinträchtigungen nicht nur auf der somatischen Ebene, sondern auch auf der emotionalen und kognitiven Ebene zum Ausdruck kommen kann. Hierdurch rechtfertigt sich der Einsatz von Persönlichkeitsinventaren und neuropsychologischen Testverfahren i. e. S. bei Schmerzpatienten.

Zum Nachweis einer bewussten Täuschungsabsicht hielten Larrabee, Greiffenstein und Greve [93] fest: „(...) *intent is inferred as a result of the combined improbability of events rather than relying on a single definitive indication of intent*“ (S. 338). Dieser Ansatz stellte eine wesentliche Erweiterung der ursprünglichen Slick-Kriterien dar, welcher in deren späterer Überarbeitung Berücksichtigung fand. Durch die Kombination mehrerer, inhaltlich unabhängiger PVT lässt sich nicht nur, wie oben bereits angeführt (siehe A.3), die Sensitivität erhöhen, sondern gleichzeitig auch die falsch-positiv Rate als Ganzes unter diejenige eines einzelnen PVT senken. In der Praxis wird der Grad der Unwahrscheinlichkeit in der falsch-positiv Rate bei Verwendung eines empirisch ermittelten Trennwertes, ab dem ein Testwert nicht mehr als authentisch zu betrachten ist, ausgedrückt. Ermittelt werden solche Trennwerte durch sogenannte „known-groups“ oder „criterion-groups“-Validierungen (siehe A.1).

### C.3 Kriterienkatalog für die Diagnose einer nicht-authentischen posttraumatischen Belastungsstörung

Da PTSD-Patienten häufig über kognitive Beeinträchtigungen klagen, aber auch Exploranden mit nicht-authentischer PTSD oft neuropsychologische Leistungseinbussen beklagen, ist es, wie Demakis, Gervais und Rohling [94] aufzeigten, sinnvoll, auch bei dieser Patientengruppe sowohl PVT als auch Verfahren zur Überprüfung der Authentizität der Beschwerdenschilderung einzusetzen.

Young [95] stellte 2014 entsprechende Kriterien eines Diagnosesystems für „Non-credible, Feigned, or Malingered Posttraumatic Stress Disorder Related Disability/Dysfunction (F-PTSDR-D)“ vor. Mit vier Hauptkategorien (A-Evidenz eines externen Anreizes, B-Evidenzen aus Testbefunden, C-Evidenzen aus Inkonsistenzen/Diskrepanzen und D-Ausschlusskriterien) folgte er den Modellen zur Diagnose von aggravierten/simulierten neurokognitiven Störungen (MND) bzw. schmerzassoziierter Behinderungen (MPRD), wobei er kritische Äusserungen und Verbesserungsvorschläge für beide Systeme [96-98] sowie eine Arbeit von Rubenzer [87] aufgriff und in die Entwicklung seines Diagnosesystems einfließen liess. Zu den Neuerungen zählen eindeutigere Definitionen und Abgrenzungen bestimmter Termini und

möglichen Inkonsistenzen, Regeln zur Erhebung und Gewichtung von Testdaten sowie eine jeweils separate Graduierung der Wahrscheinlichkeit von Antwortverzerrung für die Kriterien B und C. Das System ist so gestaltet, dass es sich mit gewissen inhaltlichen Anpassungen auch für die Diagnose von aggravierten/simulierten kognitiven Störungen (Feigned Neurocognitive Related Disability/Dysfunction (F-NCR-D)) und/oder schmerzassoziierten Störungen (Feigned Pain Related Disability/Dysfunction (F-PR-D)) nutzen lässt. Dadurch lassen sich bei Vorliegen von Komorbiditäten Ratings eines Systems in die beiden anderen integrieren. Eine Gesamtbetrachtung wird so wesentlich vereinfacht. Nach Meinung des Autors stellen die klaren Regeln mit Bezug zur klinischen Urteilsbildung einen bedeutenden Vorteil des Systems dar, da sie auch Berufsgruppen ohne testpsychologische Expertise zu einer reliableren und valideren Beurteilung verhelfen. Die detaillierten Operationalisierungen sollen überdies eine wissenschaftliche Überprüfung der Reliabilität, Validität und Nützlichkeit der vorgestellten Diagnosesysteme erlauben.

### **Diskussion**

Die zentrale Aufgabe in der gutachterlichen Arbeit besteht darin, eine subjektiv vorgebrachte Beschwerdesymptomatik in Abgleich mit medizinischen Befunden in ihrer Relevanz zur Beurteilung von Funktionseinbußen zu gewichten. Hierdurch erklärt sich neben der üblichen klinischen Konsistenzprüfung der Bedarf nach einer reliablen, gut validierten Methodik, um mehr „Objektivität“ zu erreichen in einem Feld von oft divergierenden Interessen und Haltungen. Es muss allgemein konstatiert werden, dass in der medizinischen Begutachtung die Schlussfolgerungen zwischen subjektiver Beschwerdenangabe, dem erhobenen Befund und der konsekutiv abgeleiteten Funktionseinschränkung bislang wenig kriteriengeleitet erfolgen und sich kaum auf Evidenz abstützen können. Dies gilt ausdrücklich für somatische wie auch für psychiatrische Fragestellungen.

Nach dem aktuellen Stand der oben referenzierten Literatur stellen die Überprüfung der Anstrengungsbereitschaft (Performanzvalidierung) und die Validierung der Beschwerdeschilderung in diesem Sinne zwei Bausteine der versicherungsmedizinischen Begutachtung dar, die einen wichtigen Mehrwert generieren und die insbesondere in ihrer Kombination aussagekräftig sind [43, 99]. Der zentrale Vorteil dieser ergänzenden Verfahren besteht in der inhaltlichen und zufallskritischen Absicherung von Informationen. Dabei gelten im Bereich der Performanzvalidierungsverfahren zur Überprüfung kognitiver Beschwerden die Alternativwahlverfahren als die Methode der Wahl. Sie basieren auf grossen Datensätzen und zeigen eine hohe Robustheit gegenüber realen kognitiven Defiziten auf. Sogenannte eingebettete Validitätsparameter, die aus ohnehin eingesetzten Testverfahren entwickelt wurden, bieten den Vorteil der Ökonomie und der spezifischen Beurteilung der Validität der in Frage stehenden Funktion. Ihre Anwendung setzt allerdings eine umfassende Sachkenntnis des Gutachters hinsichtlich der Klassifikationsgenauigkeit und der Grenzen in der Anwendbarkeit einzelner Verfahren bei bestimmten Explorationsgruppen voraus, ein unkritischer Umgang kann Fehler generieren. Die Klassifikationsgüte eines

Performanzvalidierungsverfahrens wird im Wesentlichen durch die Spezifität und die Sensitivität bestimmt. Im Allgemeinen sollten diese Gütekriterien anhand einer Stichprobe von Personen geschätzt werden, die denjenigen Exploranden ähneln, an welchen der Test in der klinischen Praxis eingesetzt werden soll [100, 101]. In Erfüllung dieser Vorgabe sind die meisten Performanzvalidierungsverfahren mittlerweile an realen Gutachtenspopulationen überprüft, so dass eine Reihe von Vergleichsdaten zu verschiedenen Explorandengruppen existieren. Demgegenüber sind die Verfahren zur Validierung der Authentizität der Beschwerdenschilderung im deutschsprachigen Raum noch weiter entwicklungsfähig. International wird bei der Validierung diesbezüglicher Instrumente häufig der weiter oben beschriebene SIRS als Referenzstandard eingesetzt. Deshalb ist es bemerkenswert, dass dieses Verfahren aktuell nicht in deutscher Sprache vorliegt. Im deutschsprachigen Raum stehen uns an hiesigen Gutachtenspopulationen validierten Verfahren derzeit nur der SFSS und der MENT als Screeningverfahren sowie der MMPI-2 und mit noch dürftiger Datenlage das VEI zur Verfügung.

Aus ethischer Sicht ist in der Umsetzung einer strukturierten Beschwerdengültigkeit eine minimale falsch-positiv Rate zu fordern, was einer submaximalen bis maximalen Spezifität entspricht. Auch wenn die ausgewiesenen falsch-positiv Raten der einzelnen Tests statistisch gesehen meist tatsächlich minimal sind, können sie bei singulärer Anwendung zu Fehleinschätzungen führen. Es wird deshalb allgemein empfohlen, sich auf mehrere Indikatoren zur Beschwerdengültigkeit abzustützen. Dabei sollten die eingesetzten Verfahren inhaltlich möglichst unabhängig voneinander sein und idealerweise aus unterschiedlichen Funktionsbereichen stammen, damit sie wirklich einen Zugewinn an Information darstellen. Mit einer solchen Kombination von Validitätsparametern mit jeweils hoher Spezifität kann die diagnostische Sicherheit beträchtlich erhöht und die Rate falsch-positiver Klassifikationen gegen Null gesenkt werden [20]. Ein weiteres Argument für die Kombination mehrerer Validitätsmasse besteht darin, dass Probanden bei der Vortäuschung psychopathologischer, kognitiver und körperlicher Beeinträchtigungen recht unterschiedliche Strategien verfolgen [25, 102, 103]. Zudem lassen sich mit diesem Ansatz auch die Einflüsse von Coaching und erlernter Test-Prozeduren bei wiederholter Begutachtung in Grenzen halten [104]. Eine hohe Klassifikationsgüte wird wie gesagt auch mit der Kombination von Performanzvalidierungsverfahren und Instrumenten zur Beschwerdengültigkeit erreicht. Beim MMPI-2 und VEI gilt ebenfalls, dass die Klassifikationsgenauigkeit mit der Zahl auffälliger, spezifischer Validitätsskalen steigt, die ebenso inhaltlich möglichst unabhängig voneinander sein sollten.

Neben der methodisch korrekten Durchführung der strukturierten Beschwerdengültigkeit, ist die fachlich korrekte Überführung der erhobenen Validierungsdaten in das Gesamtgutachten ein zentraler und kritischer Schritt, der auch immer wieder Irritationen im Umgang mit der beschriebenen Methodik hervorruft. Hinsichtlich der Interpretation auffälliger Validitätsmasse und daraus abgeleiteten Schlussfolgerungen ist ein hohes Mass an Expertise und Sorgfalt notwendig.

Von Kritikern der Testverfahren zur Authentizitätsprüfung wird häufig angeführt, dass auffällige Ergebnisse zu einem „reflexhaften“ Schluss auf Aggravation oder Simulation [104] verleiten könnten,

ohne die vielfältigen, potentiellen Ursachen für negative Antwortverzerrung ausreichend zu berücksichtigen. Ein solches Vorgehen hat in der Vergangenheit sicherlich auch dazu beigetragen, die Methodik der Beschwerdvalidierung als reines „Überführungsinstrument“ wahrzunehmen. Für die Annahme eines bewussten Täuschungsverhaltens bedarf es ohnehin weiterer zu erfüllender Kriterien, wie sie für vorgetäuschte kognitive Störungen [21] und vorgetäuschte schmerz-assoziierte Störungen [92] mit mittlerweile guter Akzeptanz angewandt werden.

Es gilt in der Tat zu berücksichtigen, dass auch Exploranden mit genuinen psychopathologischen Auffälligkeiten, neuropsychologischen Defiziten, chronischen Schmerzen und/oder organischen Befunden nicht selten aggravieren [88]. Zunächst einmal kann aus den beschriebenen Verfahren nur abgeleitet werden, dass die beklagten Beschwerden und Funktionseinbussen nicht in der angegebenen Ausprägung und/oder Qualität vorliegen. In einer Reihe von Fällen wird der Gutachter auch damit konfrontiert sein, dass eine unzureichende Anstrengungsbereitschaft und eine invalide Selbstbeschreibung in einem Selbstbeurteilungsinventar störungsimmanent sind. So können Exploranden mit mittelschweren oder schweren kognitiven Störungen je nach Art und Ausmass der Beeinträchtigung trotz bestmöglicher Kooperation auffällige und in diesem Fall falsch-positive Resultate in Performanzvalidierungsverfahren erzielen. Genannt seien hier beispielhaft schwere Frontalhirnstörungen und fortgeschrittene Demenzen. Schon Slick berücksichtigte im Rahmen seiner in 1999 formulierten Kriterien, dass auffällige Ergebnisse in Beschwerdvalidierungsverfahren durchaus auch Ausdruck einer krankheitswerten Störung sein können. Somit ergibt sich eine ganz grundlegende, bislang aus Sicht der Autoren wenig untersuchte Frage, inwieweit und unter welchen Bedingungen belegbare Beschwerdenverzerrung als krankheitszugehörig gewertet werden kann. Parallel zu einem nachweisbaren externalen Anreiz kann durchaus eine davon unabhängige Motivation zur Einnahme der Krankenrolle bestehen, die weniger der bewussten Kontrolle unterliegt. Diesbezüglich besonders in der Diskussion stehen hier aber auch im weitesten Sinne „neurotisch“ geprägte, zum Teil ohnehin schlecht validierte, versicherungsmedizinisch aber hochrelevante Diagnosen wie zum Beispiel chronische Erschöpfung, somatoforme Störungen oder auch Schmerzverarbeitungsstörungen. Diese Störungen gehen klinisch fast schon definitorisch mit einer Verdeutlichungstendenz in der Symptompräsentation einher. Im Hinblick auf psychische Störungen allgemein liegen auf der anderen Seite aber bis dato auch keine empirischen Belege dafür vor, dass psychische Erkrankungen in leicht bis mittelgradiger Ausprägung bei vorhandener Anstrengungsbereitschaft krankheitsbezogen zu auffälligen Resultaten in Performanzvalidierungstests führen. Auch das Vorliegen einer somatoformen Störung oder einer anderen neurotischen Störung sollte grundsätzlich mit einer unauffälligen Bewältigung üblicher Performanzvalidierungstests vereinbar sein. Unterschreitungen entsprechender Trennwerte stellen in diesen Fällen durchaus einen relevanten, zunächst einmal klärungsbedürftigen Befund dar. Hier braucht es zur reliablen Integration der Beschwerdvalidierungsergebnisse eine bislang nicht ausreichend vorliegende Konsensbildung, sowie die Notwendigkeit weiterer Forschungsbemühungen, um eine Einschätzung ausreichend empirisch untermauern zu können. Bislang

unterliegt es häufig dem Expertenurteil des Gutachters, ob ein Explorand auf einem gedachten Kontinuum eher dem Pol der wenig bewusstseinsnahen Selbsttäuschung oder aber dem Pol der Täuschung anderer i.S. einer bewussten Aggravation zuzuordnen ist [15, 105]. Dies gilt neben den häufig in der Diskussion im Vordergrund stehenden neurotischen und psychosomatischen Störungen in gleicher Weise aber auch für depressive Störungen und für Traumafolgestörungen. Auch angesichts der bedeutsamen Zunahme von potentiellen Trauma-Betroffenen in Deutschland (institutioneller Missbrauch, Kampfeinsätze der Bundeswehr, hohe Flüchtlingszahlen aus Kriegsgebieten) ist auch hier eine auf der beschriebenen Methodik basierende Plausibilisierung von Trauma-bezogenen Symptomkomplexen (z.B. im kognitiven Bereich) in der Lage, Aussagen abzusichern oder auch in Frage zu stellen.

Zur Frage, wie sich eine negative Antwortverzerrung in der gutachterlichen Bewertung der Funktions- bzw. Leistungsfähigkeit niederschlagen soll, gibt es bislang unterschiedliche Positionen. Dressing, Widder und Foerster [106] empfehlen, ggf. eine vorhandene „krankheitswertige neurotische Symptomatik“ zu attestieren und festzuhalten, dass „aufgrund der zusätzlich nachweisbaren Übertreibungstendenzen des Probanden das Ausmass der tatsächlichen Krankheit aber nicht abzuschätzen“ (S. 167) sei. Andere Autoren neigen dazu, zu argumentieren, dass bei negativer Antwortverzerrung bzw. fehlender Kooperativität der Nachweis einer tatsächlich vorliegenden Erkrankung nicht zu führen sei. In jedem Fall haben solche Inkonsistenzen Gewicht, können nicht einfach übergangen werden, und machen je nach gutachtlichem Gesamtbild weitergehende Abklärungen unausweichlich.

Die in der vorliegenden Arbeit vorgestellte Methodik einer Beschwerdvalidierung erbringt gesamthaft bei fachgerechter Anwendung zusätzliche Befunde, die aus Sicht der Autoren die traditionelle klinische Konsistenzprüfung in sinnvoller Weise als annähernd unabhängige Quelle ergänzen. Faktisch stellt die Methodik ein Werkzeug dar, angegebene Beschwerden und Defizite zu plausibilisieren und gegenüber Aggravation und Simulation abzugrenzen. Dieser Erkenntnis wird zunehmend Rechnung getragen durch fachärztliche und neuropsychologische Fachverbände, die Beschwerdvalidierung als wichtigen Baustein einer Begutachtung in ihre Empfehlungen aufnehmen (Konsensuskonferenz der American Academy of Clinical Neuropsychology [107], Professional Practice Board der British Psychological Society [108], Schweizerische Vereinigung der Neuropsychologinnen und Neuropsychologen [109, 110], Leitlinien für die Begutachtung bei psychischen und psychosomatischen Erkrankungen [111, 112]).

In Gesamtsicht kann konstatiert werden, dass die beschriebenen Verfahren zwar in der Lage sind, wichtige und in vielen Fällen auch unverzichtbare, unerwartete Informationen zur Beurteilung der Beschwerdvalidität zu liefern, dabei jedoch aktuell noch die Integration dieser zusätzlich erhobenen Befunde wenig strukturiert und uneinheitlich erfolgt. Mit anderen Worten besteht die Gefahr, dass in validierter Form erhobene reliable Befunde in invalider Weise interpretiert werden. Aus Sicht der Autoren besteht gerade hier ein erheblicher Entwicklungsbedarf.

### **Take home message**

In der versicherungsmedizinischen Begutachtung empfiehlt es sich aus Sicht der Autoren, möglichst verschiedene standardisierte Verfahren zur Beschwerdvalidierung heranzuziehen. Grundlage der versicherungsmedizinischen Beurteilung ist unverändert die Integration aller vorliegenden Informationen und der empirisch geleiteten Konsistenzprüfung, insbesondere auch aus der Exploration selbst heraus. Beschwerdvalidierungsverfahren als Korrelat zu diesen Erkenntnissen ergeben jedoch aufgrund ihrer Aussagen zur Gültigkeit einer Testleistung und Zuverlässigkeit der Beschwerd Schilderung bei bekannten falsch-positiv Raten unzweifelhaft eine qualitative Verbesserung der Befundlage und erhöhen somit die Validität der getroffenen Aussagen. Dieser Sachlage wird auch in der Entwicklung jüngerer gutachterlicher Leitlinien Rechnung getragen [111].

## Literatur

1. Huber M. Aspekte der Berufsunfähigkeit bei psychosomatischen Erkrankungen. *Versicherungsmedizin* 2000; 52: 66-75
2. Stadtland C, Gündel H, Schütt S et al. Kriterien zur Beurteilung der quantitativen Leistungseinschränkung bei der Begutachtung funktioneller körperlicher Störungen. Eine Literaturübersicht. *Versicherungsmedizin* 2003; 55: 111-117
3. Schneider W, Becker D, Dohrenbusch R et al. Berufliche Leistungsfähigkeit. Begutachtung bei psychischen und psychosomatischen Erkrankungen. *Psychotherapeut* 2010; 55: 373-379
4. OECD. Sick on the Job? Myths and Realities about Mental Health and Work. *Mental Health and Work*, OECD Publishing; 2012
5. Ott W, Bade S, Wapf B. Nicht zielkonforme Leistungen in der Invalidenversicherung: Bedeutung und Größenordnung. Bern: Bundesamt für Sozialversicherungen; 2007
6. Kool J, Meichtry A, Schaffert R et al. Der Einsatz von Beschwerdengültigkeitstests in der IV-Abklärung. Bern: Bundesamt für Sozialversicherungen; 2008
7. Ekman P, O'Sullivan M. Who can catch a liar? *Am Psychol* 1991; 46: 913-920
8. Trueblood W, Binder LM. Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. *Arch Clin Neuropsychol* 1997; 12: 13-27
9. Aamondt MG, Custer H. Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner* 2006; 15: 6-11
10. Bartlett MS, Littlewort GC, Frank MG et al. Automatic decoding of facial movements reveals deceptive pain expressions. *Curr Biol* 2014; 24: 738-743
11. Espeland A, Vetti N, Krakenes J. Are two readers more reliable than one? A study of upper neck ligament scoring on magnetic resonance images. *BMC Med Imaging* 2013; 13: 4
12. Khan L, Mitera G, Probyn L et al. Inter-rater reliability between musculoskeletal radiologists and orthopedic surgeons on computed tomography imaging features of spinal metastases. *Curr Oncol* 2011; 18: 282-287
13. Carrino JA, Lurie JD, Tosteson AN et al. Lumbar spine: reliability of MR imaging findings. *Radiology* 2009; 250: 161-170
14. Jarvik JG, Deyo RA. Moderate versus mediocre: the reliability of spine MR data interpretations. *Radiology* 2009; 250: 15-17
15. Merten T. Beschwerdengültigkeit bei der Begutachtung kognitiver und psychischer Störungen. *Fortschr Neurol Psychiatr* 2011; 79: 102-116
16. Dohrenbusch R. Symptom- und Beschwerdengültigkeit chronifizierter Schmerzen in sozialmedizinischer Begutachtung. Teil I: Terminologische und methodologische Zugänge. *Schmerz* 2009; 23: 231-234, 236-240
17. Iverson GL. Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Appl Neuropsychol* 2006; 13: 77-90
18. Merten T, Merckelbach H. Symptom validity testing in somatoform and dissociative disorders: a critical review. *Psychological Injury and Law* 2013; 6: 122-137
19. Rogers R. *Clinical Assessment of Malingering and Deception*. 3rd ed. New York: Guilford; 2008
20. Larrabee GJ. Detection of malingering using atypical performance patterns on standard neuropsychological tests. *Clin Neuropsychol* 2003; 17: 410-425
21. Slick DJ, Sherman EM, Iverson GL. Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *Clin Neuropsychol* 1999; 13: 545-561
22. Mittenberg W, Patton C, Canyock EM et al. Base rates of malingering and symptom exaggeration. *J Clin Exp Neuropsychol* 2002; 24: 1094-1102
23. Merten T, Friedel E, Stevens A. Eingeschränkte Kooperativität in der neurologisch-psychiatrischen Begutachtung: Schätzungen zur Auftretenshäufigkeit an einer Begutachtungspopulation. *Versicherungsmedizin* 2006; 58: 19-21
24. Stevens A, Friedel E, Mehren G et al. Malingering and uncooperativeness in psychiatric and psychological assessment: prevalence and effects in a German sample of claimants. *Psychiatry Res* 2008; 157: 191-200



25. Greve KW, Ord JS, Bianchini KJ et al. Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Arch Phys Med Rehabil* 2009; 90: 1117-1126
26. Strebel M, Brusis T. Zur Problematik der Simulation und Aggravation in der HNO-ärztlichen Begutachtung. *HNO* 2010; 58: 126-131
27. Hartman DE. The unexamined lie is a lie worth fibbing: neuropsychological malingering and the Word Memory Test. *Arch Clin Neuropsychol* 2002; 17: 709-714
28. Dohrenbusch R, Merten T. Psychologische Mess- und Testverfahren. Aussagekraft in der sozialmedizinischen Begutachtung. *Psychotherapeut* 2010; 55: 389-393
29. Rogers R, Harrell EH, Liff CD. Feigning neuropsychological impairment - a critical-review of methodological and clinical considerations. *Clinical Psychology Review* 1993; 13: 255-274
30. Tombaugh TN. *Test of Memory Malingering (TOMM)*. North Tonawanda: Multi-Health Systems; 1996
31. Rey A. *L'examen clinique en psychologie*. Paris: Presses Universitaires de Paris; 1958
32. Boone KB, Lu P, Herzberg D. *The Dot Counting Test™ (DCT™)*. Torrance (CA): Western Psychological Services; 2002
33. Boone KB, Lu P, Back C et al. Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Arch Clin Neuropsychol* 2002; 17: 625-642
34. Green P. *Green's Word Memory Test. User's Manual*. Edmonton: Green's Publishing; 2005
35. Gorissen M, Sanz JC, Schmand B: Effort and cognition in schizophrenia patients. *Schizophr Res* 2005; 78: 199-208
36. Green P. *Medical Symptom Validity Test for Windows. User's Manual and Program*. Edmonton: Green's Publishing; 2003, revised 2005
37. Green P. *Non-Verbal Medical Symptom Validity Test for Windows. User's Manual*. Edmonton: Green's Publishing; 2008
38. Schmand B, Lindeboom J. *Amsterdam Short Term Memory Test. Amsterdamer Kurzzeitgedächtnistest. Manual. Handanweisung*. Leiden: PITS; 2005
39. Schroeder RW, Twumasi-Ankrah P, Baade LE et al. Reliable Digit Span: a systematic review and cross-validation study. *Assessment* 2012; 19: 21-30
40. Meyers J, Meyers K. *Rey Complex Figure and Recognition Trial: Professional Manual*. Odessa, FL: Psychological Assessment Resources; 1995
41. Boone KB. *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective*. New York: Guilford Press; 2007
42. Boone KB. *Clinical Practice of Forensic Neuropsychology. An Evidence-based Approach*. New York: Guilford Press; 2013
43. Larrabee GJ. *Assessment of Malingered Neuropsychological Deficits*. Oxford: Oxford University Press; 2007
44. Morgan JE, Sweet JJ. *Neuropsychology of Malingering Casebook*. New York: Psychology Press; 2009
45. Carone D, Bush SS. *Mild Traumatic Brain Injury. Symptom Validity Assessment and Malingering*. New York: Springer; 2013
46. Boone KB. The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *Clin Neuropsychol* 2009; 23: 729-741
47. Larrabee GJ. Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *Clinical Neuropsychologist* 2007; 17: 54-68
48. Larrabee GJ. Aggregation across multiple indicators improves the detection of malingering: relationship to likelihood ratios. *Clin Neuropsychol* 2008; 22: 666-679
49. Merten T. *Beschwerdvalidierung. Fortschritte der Neuropsychologie, Bd. 14*. Göttingen: Hogrefe; 2014
50. Rogers R. Detection Strategies in Malingering and Defensiveness. In: Rogers R, ed. *Clinical Assessment of Malingering and Deception*. 3rd ed. New York: Guilford; 2008: 14-38
51. Rogers R. Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law* 1984; 2: 93-112
52. Rogers R, Bagby RM, Dickens SE. *Structured Interview of Reported Symptoms (SIRS) and Professional Manual*. Odessa, FL: Psychological Assessment Resources; 1992

53. Rogers R, Sewell KW, Gillard ND. Structured Interview of Reported Symptoms, 2nd ed. (SIRS-2) and Professional Manual. Odessa, FL: Psychological Assessment Resources; 2010
54. Schmidt T, Lanquillon S, Ullmann U. Kontroverse zu Beschwerdvalidierungs-verfahren bei der Begutachtung psychischer Störungen. Forensische Psychiatrie, Psychologie, Kriminologie 2011; 5: 177-183
55. Vitacco MJ, Rogers R, Gabel J et al. An evaluation of malingering screens with competency to stand trial patients: a known-groups comparison. Law Hum Behav 2007; 31: 249-260
56. Rogers R, Vitacco MJ, Kurus SJ. Assessment of malingering with repeat forensic evaluations: patient variability and possible misclassification on the SIRS and other feigning measures. J Am Acad Psychiatry Law 2010; 38: 109-114
57. McCusker PJ, Moran MJ, Serfass L et al. Comparability of the MMPI-2 F(p) and F scales and the SIRS in clinical use with suspected malingerers. Int J Offender Ther Comp Criminol 2003; 47: 585-596
58. Cima M, Hollnack S, Kremer K et al. Strukturierter Fragebogen Simulierter Symptome. Die deutsche Version des Structured Inventory of Malingered Symptomatology: SIMS. Nervenarzt 2003; 74: 977-986
59. Lanquillon S, Schmidt T, Stieglitz RD. Strukturierteres Interview Berichteter Beschwerden SIBB. Bern: Huber; 2011 [in Vorbereitung]
60. Miller HA. Miller-Forensic Assessment of Symptoms Test Professional Manual. Odessa, FL: Psychological Assessment Resources; 2001
61. Smith GP. Brief Screening Measures for the Detection of Feigned Psychopathology. In: Rogers R, ed. Clinical Assessment of Malingering and Deception. 3rd ed. New York: Guilford; 2008: 323-339
62. Guy LS, Miller HA. Screening for malingered psychopathology in a correctional setting - utility of the Miller-Forensic Assessment of Symptoms Test (M-FAST). Criminal Justice and Behavior 2004; 31: 695-716
63. Miller HA. The Miller-Forensic Assessment of Symptoms Test (M-FAST): test generalizability and utility across race, literacy, and clinical opinion. Criminal Justice and Behavior 2005; 32: 591-611
64. Smith GP, Burger GK. Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). J Am Acad Psychiatry Law 1997; 25: 183-189
65. Wisdom NM, Callahan JL, Shaw TG. Diagnostic utility of the Structured Inventory of Malingered Symptomatology to detect malingering in a forensic sample. Arch Clin Neuropsychol 2010; 25: 118-125
66. van Impelen A, Merckelbach H, Jelicic M et al. The Structured Inventory of Malingered Symptomatology (SIMS): a systematic review and meta-analysis. Clin Neuropsychol 2014: 1-30
67. Butcher JN, Graham JR, Ben-Porath YS et al. Minnesota Multiphasic Personality Inventory-2 (MMPI-2). Manual for Administration, Scoring and Interpretation. (rev.ed.). Minneapolis, MN: University of Minnesota Press; 2001
68. Gervais RO, Ben-Porath YS, Wygant DB et al. Development and validation of a Response Bias Scale (RBS) for the MMPI-2. Assessment 2007; 14: 196-208
69. Gervais RO, Ben-Porath YS, Wygant DB et al. Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. Clin Neuropsychol 2008; 22: 1061-1079
70. Rogers R, Bender SD. Evaluation of Malingering and Deception. In: Weiner IB, Goldstein AM, eds. Handbook of Psychology: Forensic Psychology. Hoboken, NJ: Wiley & Sons; 2003: 109-131
71. Nelson NW, Hoelzle JB, Sweet JJ et al. Updated meta-analysis of the MMPI-2 symptom validity scale (FBS): verified utility in forensic practice. Clin Neuropsychol 2010; 24: 701-724
72. Lewis JL, Simcox AM, Berry DT. Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the structured inventory of malingered symptomatology. Psychol Assess 2002; 14: 170-176
73. Baer RA, Miller J. Underreporting of psychopathology on the MMPI-2: a meta-analytic review. Psychol Assess 2002; 14: 16-26

74. Engel RR. MMPI-2. Minnesota Multiphasic Personality Inventory-2 Manual. Bern: Hans Huber; 2000
75. Thies E. Der deutsche MMPI-2: Effektivität der Validitätsskalen in der Aufdeckung von Antwortverzerrung. Marburg: Tectum; 2012
76. Butcher JN, Lim J, Nezami E. Objective study of abnormal personality in cross-cultural settings: The Minnesota Multiphasic Personality Inventory (MMPI-2). *Journal of Cross-Cultural Psychology* 1998; 29: 189-211
77. Dean AC, Boone KB, Kim MS et al. Examination of the impact of ethnicity on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Fake Bad Scale. *Clin Neuropsychol* 2008; 22: 1054-60
78. Hall GCN, Bansal A, Lopez IR. Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychological Assessment* 1999; 11: 186-197
79. Knaster CA, Micucci JA. The effect of client ethnicity on clinical interpretation of the MMPI-2. *Assessment* 2013; 20: 43-7
80. Morey LC. The Personality Assessment Inventory Professional Manual. 2nd ed. Odessa, FL: Psychological Assessment Resources; 2007
81. Hawes SW, Boccaccini MT. Detection of overreporting of psychopathology on the Personality Assessment Inventory: a meta-analytic review. *Psychol Assess* 2009; 21: 112-124
82. Engel RR, Groves JA. Verhaltens- und Erlebensinventar. Deutschsprachige Adaption des Personality Assessment Inventory (PAI) von L.C. Morey. Bern: Hans Huber; 2013
83. Vossler-Thies E, Stevens A, Engel RR et al. Erfassung negativer Antwortverzerrungen mit der deutschen Fassung des Personality Assessment Inventory, dem Verhaltens- und Erlebensinventar. *Diagnostica* 2013; 59: 73-85
84. Morel KR. Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. *J Pers Assess* 1998; 70: 299-314
85. Morel KR. Manual for the Morel Emotional Numbing Test for Posttraumatic Stress Disorder: Psychometric Properties. (3rd ed. rev.). Las Vegas: NV: Author; 2012
86. Morel KR. Cutoff Scores for the Morel Emotional Numbing Test for PTSD: Considerations for Use in VA Mental Health Examinations. *Psychological Injury and Law* 2013; 6: 138-143
87. Rubenzer S. Posttraumatic Stress Disorder: Assessing Response Style and Malingering. *Psychological Injury and Law* 2009; 2: 114-142
88. Merten T, Thies E, Schneider K et al. Symptom Validity Testing in Claimants with Alleged Posttraumatic Stress Disorder: Comparing the Morel Emotional Numbing Test, the Structured Inventory of Malingered Symptomatology, and the Word Memory Test. *Psychological Injury and Law* 2009; 2: 284-293
89. Geraerts E, Kozaric-Kovacic D, Merckelbach H et al. Detecting deception of war-related posttraumatic stress disorder. *Journal of Forensic Psychiatry and Psychology* 2009; 20: 278-285
90. Wiedmaier P. Schätzung von Normalbefunden für drei Beschwerdengültigkeitstests: Der Morel Emotional Numbing Test (MENT), der Strukturierte Fragebogen Simulierter Symptome und die List of Improbable Problems (LIMP). Saarbrücken: VDM Verlag Dr. Müller; 2011
91. Slick DJ, Sherman EM. Differential Diagnosis of Malingering. In: Carone D, Bush SS, eds. *Mild Traumatic Brain Injury, in Symptom Validity Assessment and Malingering*. New York: Springer Publishing Company; 2013: 57-72
92. Bianchini KJ, Greve KW, Glynn G. On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *Spine J* 2005; 5: 404-417
93. Larrabee GJ, Greiffenstein MF, Greve KW et al. Refining diagnostic criteria for malingering. In: Larrabee GJ, ed. *Assessment of malingered neuropsychological deficits*. New York: Oxford; 2007: 334-371
94. Demakis GJ, Gervais RO, Rohling ML. The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of post-traumatic stress disorder. *The Clinical Neuropsychologist* 2008; 22: 879-895
95. Young G. Malingering, Feigning, and Response Bias in Psychiatric/Psychological Injury. Implications for Practice and Court. In: *International Library of Ethics, Law, and the New Medicine*, Vol. 56. Dordrecht: Springer SBM; 2014

96. Boone KB. Clarification or confusion? A review of Rogers, Bender, and Johnson's a critical analysis of the MND criteria for feigned cognitive impairment: implications for forensic practice and research. *Psychological Injury and Law* 2011; 4: 157-162
97. Rogers R, Bender SD, Johnson SF. A critical analysis of the MND criteria for feigned cognitive impairment: Implications for forensic practice and research. *Psychological Injury and Law* 2011; 4: 147-156
98. Rogers R, Bender SD, Johnson SF. A commentary on the MND model and the Boone critique: "Saying it doesn't make it so". *Psychological Injury and Law* 2011; 4: 162-167
99. Bianchini KJ, Etherton JL, Greve KW et al. Classification accuracy of MMPI-2 validity scales in the detection of pain-related malingering: a known-groups study. *Assessment* 2008; 15: 435-449
100. Bossuyt PM, Reitsma JB, Bruns DE et al. Toward complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative. *Am J Clin Pathol* 2003; 119: 18-22
101. Henry M. Beschwerdengültigkeitstests in der zivil- und sozialrechtlichen Begutachtung: Verfahrensüberblick. In: Merten T, Dettenborn H, Hrsg. *Diagnostik der Beschwerdengültigkeit*. Berlin: Deutscher Psychologen Verlag GmbH; 2009: 118-161
102. Greiffenstein MF, Baker WJ, Gola T. Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment* 1994; 6: 218-224
103. Nelson NW, Sweet JJ, Berry DT et al. Response validity in forensic neuropsychology: exploratory factor analytic evidence of distinct cognitive and psychological constructs. *J Int Neuropsychol Soc* 2007; 13: 440-449
104. Blaskewitz N, Merten T. Diagnostik der Beschwerdengültigkeit - Diagnostik bei Simulationsverdacht: ein Update 2002 bis 2005. *Fortschr Neurol Psychiatr* 2007; 75: 140-54
105. Merten T. Fragen der neuropsychologischen Diagnostik bei Simulationsverdacht. *Fortschr Neurol Psychiatr* 2002; 70: 126-138
106. Dressing H, Widder B, Foerster K. Kritische Bestandsaufnahme zum Einsatz von Beschwerdengültigkeitstests in der psychiatrischen Begutachtung. *Versicherungsmedizin* 2010; 62: 163-167
107. Heilbronner RL, Sweet JJ, Morgan JE et al. American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *Clin Neuropsychol* 2009; 23: 1093-1129
108. McMillan TM, Anderson S, Baker G et al. Assessment of effort in clinical testing of cognitive functioning for adults. In: Professional Practice Board. Leicester: British Psychological Society; 2009: 1-27
109. Schweizerische Vereinigung der Neuropsychologinnen und Neuropsychologen. Leitlinien für neuropsychologische Gutachten. 2011
110. Frei A, Balzer CH, Gysi F et al. Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit. *Zeitschrift für Neuropsychologie* 2016; 28: in press.
111. Schneider W, Becker D, Vogel M et al. Begutachtung der Leistungsfähigkeit bei Personen mit psychischen und psychosomatischen Störungen - ein diagnostisches Modell. *Versicherungsmedizin* 2011; 63: 68-75
112. Schneider W, Dohrenbusch R, Henningsen P et al. Begutachtung bei psychischen und psychosomatischen Erkrankungen: Autorisierte Leitlinien und Kommentare. 2., überarbeitete und erweiterte Auflage. Bern: Hogrefe; 2016
113. Greene RL. *The MMPI-2: an Interpretive Manual*. 2nd ed. Boston: Allyn & Bacon; 2000
114. Greene RL. *The MMPI-2/MMPI-2-RF: an Interpretive Manual*. 3rd ed. Boston: Allyn & Bacon; 2011
115. Bagby RM, Marshall MB. Assessing underreporting response bias on the MMPI-2. *Assessment* 2004; 11: 115-126

**Tab. 1** Methoden zur Beschwerdendvalidierung

	Grundprinzipien/ Detektionsstrategien	Verfahren	Kurze Übersicht	Verfahrens- beispiele	Vorteile	Nachteile	Literatur
Verfahren zur Validierung der Anstrengungsbereitschaft	<ul style="list-style-type: none"> <li>▪ Alternativwahlverfahren</li> <li>▪ Testdeckeneffekt oder Prinzip der verdeckten Leichtigkeit</li> <li>▪ Leistungskurve</li> <li>▪ Fehlergrösse</li> <li>▪ Identifikation untypischer oder inkonsistenter Leistungsprofile</li> </ul>	Performanzvalidierungstests (PVT)	<ul style="list-style-type: none"> <li>▪ Höchste Klassifikationsgenauigkeit</li> <li>▪ Fokus auf vorgetäuschte Gedächtnisdefizite &gt; hohe Robustheit gegenüber realen kognitiven Defiziten</li> <li>▪ Grosse empirische Datenbasis</li> <li>▪ Ausführliche Validierung</li> <li>▪ Geringe Sensitivität zur Aufdeckung von negativer Antwortverzerrung in anderen Funktionsbereichen</li> <li>▪ Erhöhte Interkorrelationen möglich</li> <li>▪ Soziodemographische Faktoren (Alter, Geschlecht, Bildungsstand) spielen kaum eine Rolle</li> </ul>	Word Memory Test (WMT)	<ul style="list-style-type: none"> <li>▪ Alle Hartman-Kriterien erfüllt</li> <li>▪ Hohe Sensitivität</li> </ul>	<ul style="list-style-type: none"> <li>▪ Bei einzelnen Patientengruppen reduzierte Spezifität</li> <li>▪ Relativ hoher Zeitaufwand &gt; 2 kürzere Verfahren entwickelt (MSVT, NV-MSVT)</li> </ul>	[27], [29], [34-37], [49]
				Test of Malingered Memory (TOMM)	<ul style="list-style-type: none"> <li>▪ Hohe Augenscheinvalidität</li> <li>▪ Sprachfreies Material</li> </ul>	<ul style="list-style-type: none"> <li>▪ Geringe Sensitivität</li> <li>▪ Anfälligkeit gegenüber Coaching</li> </ul>	[29, 30], [49]
				Amsterdamer Kurzzeitgedächtnistest (AKGT)	<ul style="list-style-type: none"> <li>▪ Erschwerte Durchschaubarkeit</li> </ul>	<ul style="list-style-type: none"> <li>▪ Nicht geeignet für Patienten mit schon klinisch feststellbaren Störungen</li> </ul>	[29], [38], [49]
		Beschwerdendvalidierungsparameter in neuropsychologischen Tests	<ul style="list-style-type: none"> <li>▪ Parameter können direkt aus Verfahren gewonnen werden</li> <li>▪ Ggf. via zusätzlicher Untertests</li> <li>▪ Geringere Robustheit gegenüber realen kognitiven Defiziten</li> <li>▪ Überprüfung einer Vielzahl von Funktionsbereichen möglich</li> </ul>	Reliable Digit Span Finger Tapping	<ul style="list-style-type: none"> <li>▪ Verbesserte Sensitivität ohne wesentliche Erhöhung der Untersuchungszeit</li> <li>▪ Validierung auch im gutachterlichen Kontext</li> <li>▪ Erhöhung der Spezifität</li> <li>▪ Direkte Informationen über Gültigkeit der gezeigten Leistung</li> <li>▪ Geringere Anfälligkeit gegenüber Coaching</li> </ul>	<ul style="list-style-type: none"> <li>▪ Gefahr falsch-positiver Zuordnung bei Kombination inhaltlich überlappender Parameter</li> <li>▪ Klassifikationsgüte variiert in Abhängigkeit vom Störungsbild</li> <li>▪ Erfordert detaillierte Kenntnis der Primärliteratur</li> </ul>	[25], [29], [39-49]
	<ul style="list-style-type: none"> <li>▪ Seltene Beschwerden</li> <li>▪ Quasi-seltene Beschwerden</li> <li>▪ Unwahrscheinliche Beschwerden</li> <li>▪ Beschwerdenkombinationen, welche bei echten psychiatrischen Störungen äusserst selten vorkommen</li> <li>▪ Unübliche psychopathologische Muster</li> <li>▪ Hohe, undifferenzierte Ausmasse an Beschwerden</li> </ul>	Fremdbeurteilungsverfahren für psychische Störungen		Structured Interview of Reported Symptoms (SIRS-2)	<ul style="list-style-type: none"> <li>▪ Hervorragende Interrater-Reliabilität</li> <li>▪ Sehr gute Retestreliabilität bei forensischen Patienten</li> <li>▪ Belege für Konstruktvalidität mit dem MMPI-2</li> <li>▪ Deutsche Version in Vorbereitung</li> </ul>	<ul style="list-style-type: none"> <li>▪ Lange Durchführungszeit</li> <li>▪ Expertise der Interviewer erforderlich</li> </ul>	[52-59]
				Miller Forensic Assessment of Symptoms Test (M-Fast)	<ul style="list-style-type: none"> <li>▪ Sehr gute Reliabilität und Validität in forensischen Populationen</li> <li>▪ Belege für Konstruktvalidität mit dem SIRS</li> <li>▪ Reduktion Anteil falsch-negativer Klassifikationen gegenüber klinischem Urteil</li> <li>▪ Kurze Durchführungs- und Auswertzeit</li> </ul>	<ul style="list-style-type: none"> <li>▪ Reines Screeningverfahren</li> </ul>	[60-63]

Verfahren zur Validierung der Beschwerdenschilderung	<ul style="list-style-type: none"> <li>▪ Aussergewöhnlich hohe Ausprägungen einzelner Symptome</li> <li>▪ Eindeutige Beschwerden</li> <li>▪ Diskrepanzen zwischen beobachteten und berichteten Beschwerden</li> <li>▪ Fehlerhafte Stereotypen</li> </ul>	Selbstbeurteilungsverfahren und andere Screeningtests für psychische Störungen		<p>Strukturierter Fragebogen simulierter Symptome (SFSS &gt; deutsche Version des Structured Inventory of Malingered Symptomatology (SIMS))</p> <ul style="list-style-type: none"> <li>▪ Hohe Retestreliabilität</li> <li>▪ Akzeptable interne Konsistenz</li> <li>▪ Gute vorhersagende und übereinstimmende Validität</li> <li>▪ Sensitivität, Spezifität, Positive Predictive Power, Negative Predictive Power über 0.85 in Abhängigkeit des Trennwertes</li> </ul>	<ul style="list-style-type: none"> <li>▪ Reines Screeningverfahren</li> <li>▪ Sensitivität abhängig vom Untersuchungskontext bzw. angewandtem Trennwert</li> </ul>	[58, 64-66]
				<p>Minnesota Multiphasic Personality Inventory (MMPI-2)</p> <ul style="list-style-type: none"> <li>▪ Aufdeckung von Antwortverzerrung im Bereich psychischer, körperlicher und kognitiver Störungen</li> <li>▪ Sehr hohe Spezifitäten im forensischen Kontext</li> <li>▪ Wahrscheinlichkeitsaussage über Unter- oder Übertreibung möglich</li> <li>▪ Aussagen über psychische Gesundheit und Persönlichkeitsmerkmale möglich</li> </ul>	<ul style="list-style-type: none"> <li>▪ Lange Durchführungszeit</li> <li>▪ Expertise erforderlich</li> <li>▪ Mehrheit der Validitätsskalen nicht Bestandteil der deutschen Version &gt; manuelle Auswertung anhand der Literatur erforderlich</li> </ul>	[67-75, 113-115]
				<p>Personality Assessment Inventory (PAI)</p> <p>Deutsche Version: Verhaltens- und Erlebensinventar (VEI)</p> <ul style="list-style-type: none"> <li>▪ Deutlich kürzer als MMPI-2</li> <li>▪ Inhaltliche Ausrichtung an ICD-10 und DSM-IV</li> <li>▪ Hohe Reliabilität, gute Validität in verschiedenen Settings</li> <li>▪ Hohe Effektstärken der Validitätsskalen der Originalversion</li> </ul>	<ul style="list-style-type: none"> <li>▪ Vergleichsweise geringe Sensitivität der deutschen Version</li> </ul>	[80-83]
				<p>Morel Emotional Numbing Test for Posttraumatic Stress Disorder (MENT)</p> <ul style="list-style-type: none"> <li>▪ Forced-choice Verfahren</li> <li>▪ Detektionsstrategie fehlerhafte Stereotypen</li> <li>▪ Hohe Spezifität, gute Sensitivität</li> <li>▪ In vielen Sprachen vorhanden</li> </ul>	<ul style="list-style-type: none"> <li>▪ Screeningverfahren</li> </ul>	[84-90]

<p><b>Nachweis bewusster Täuschungs- absicht</b></p>	<ul style="list-style-type: none"> <li>▪ Prozess, der sorgfältige Analyse aller verfügbaren Informationen erfordert</li> <li>▪ Basierend auf objektiven Kriterien</li> <li>▪ Einschluss von Indikatoren, welche über bekannte Klassifikations-genauigkeit verfügen</li> <li>▪ Kombination klinisches Urteil mit Ergebnissen wissenschaftlich validierter Verfahren</li> </ul>	<p>Kriterienkataloge</p>	<ul style="list-style-type: none"> <li>▪ A: Vorhandensein eines substantiellen, externen Anreizes</li> <li>▪ B: Hinweise aus neuropsychologischer Testdiagnostik</li> <li>▪ C: Angaben des Exploranden und Validitätsskalen von Selbstbeurteilungsinstrumenten</li> <li>▪ D: Ausschluss von psychiatrischen, neurologischen oder Entwicklungsfaktoren</li> <li>▪ Zusätzlich für „sicher“ vorliegende vorgetauschte, kognitive Störung: Zahl richtiger Antworten in Zwangswahlverfahren unter Zufallsniveau</li> </ul>	<p>Kriterienkatalog für die Diagnose aggravierter/ simulierter neurokognitiver Störungen (Malingered Neurocognitive Disorder, MND)</p>	<ul style="list-style-type: none"> <li>▪ Breit akzeptiertes Konzept</li> <li>▪ Modifikationen 2013: primäre und sekundäre Aggravation/Simulation, dafür Wegfall Kriterium D/ Eindeutige Aggravation/Simulation kann nun auch den Nachweis einer hohen a-posteriori-Wahrscheinlichkeit aus Selbstbeurteilungsinstrumenten bestehen oder aus „compelling inconsistencies“</li> </ul>	<ul style="list-style-type: none"> <li>▪ Hoher Zeitaufwand</li> </ul>	<p>[21, 91]</p>
			<ul style="list-style-type: none"> <li>▪ Zusätzliches B: Hinweise aus somatischen Untersuchungsmethoden</li> </ul>	<p>Kriterienkatalog für die Aufdeckung aggravierter/ simulierter schmerzassoziierter Behinderungen (Malingered Pain-Related Disabilities, MPRD)</p>		<ul style="list-style-type: none"> <li>▪ Hoher Zeitaufwand</li> </ul>	<p>[92, 93]</p>
			<ul style="list-style-type: none"> <li>▪ A: Evidenz eines externen Anreizes</li> <li>▪ B: Evidenzen aus Testbefunden</li> <li>▪ C: Evidenzen aus Inkonsistenzen/ Diskrepanzen</li> <li>▪ D: Ausschlusskriterien</li> </ul>	<p>Kriterienkatalog für die Diagnose einer unglaubwürdigen, scheinbaren oder aggravierten/ simulierten Posttraumatischen Belastungsstörung (Non-credible, Feigned, or Malingered Posttraumatic Stress Disorder Related Disability/Dysfunction, F-PTSDR-D)</p>	<ul style="list-style-type: none"> <li>▪ Eindeutigere Definitionen und Abgrenzungen bestimmter Termini und möglicher Inkonsistenzen</li> <li>▪ Regeln zur Erhebung und Gewichtung von Testdaten</li> <li>▪ Separate Graduierung der Wahrscheinlichkeit von Antwortverzerrung für Kriterien B und C</li> <li>▪ Nutzbar für Diagnose von aggravierten/simulierten kognitiven und/oder schmerzassozierten Störungen</li> <li>▪ Vereinfachte Gesamtbetrachtung</li> <li>▪ Klare Regeln mit Bezug zur klinischen Urteilsbildung</li> </ul>	<ul style="list-style-type: none"> <li>▪ Hoher Zeitaufwand</li> </ul>	<p>[94, 95]</p>

**ANHANG A-2**

**Publikation 2**

Keppler, C., Rosburg, T., Lemoine, P., Pflüger, M., Gyr, N. & Mager, R. (akzeptiert). Functional somatic syndromes: skin temperatures and activity measurements under ambulatory conditions. *Applied Psychophysiology and Biofeedback*.



## **FUNCTIONAL SOMATIC SYNDROMES: SKIN TEMPERATURES AND ACTIVITY MEASUREMENTS UNDER AMBULATORY CONDITIONS**

Carole Keppler<sup>1</sup>, Timm Rosburg<sup>1</sup>, Patrick Lemoine<sup>1</sup>, Marlon Pflüger<sup>1</sup>, Niklaus Gyr<sup>2</sup>, Ralph Mager<sup>1</sup>

<sup>1</sup> Department of Forensic Psychiatry, University Psychiatric Clinics Basel, Wilhelm Klein-Str. 27, CH-4012 Basel, Switzerland

<sup>2</sup> Faculty of Medicine, University of Basel, Karl Jaspers-Allee 10, CH-4052 Basel, Switzerland

### **Corresponding Author:**

C. Keppler

Department of Forensic Psychiatry

University Psychiatric Clinics Basel

Wilhelm Klein-Str. 27, CH-4012 Basel

Switzerland

Tel.: +41 61 325 54 65

Fax: +41 61 325 54 92

E-mail: carole.keppler@upkbs.ch

### **Acknowledgments**

This work was supported by the Academy of Swiss Insurance Medicine (ASIM). The authors are grateful to Drs. Andreas Linde, Joachim Dietiker and Patrick Brunner for their contribution in the recruitment and interviewing of the study sample. Furthermore, the authors greatly appreciate Kurt Kräuchi's comments on the current and the previous version of the manuscript. Moreover, the authors thank the team of the Centre of Chronobiology for their technical aid.

## **Abstract**

Functional somatic syndromes are mostly associated with pain and emotional distress. As one marker for the autonomic stress response, the distal skin temperature decreases during psychological stress. In patients with functional somatic syndromes, the distal skin temperature under baseline conditions (without stress induction) is usually lower than in healthy subjects, which could be due to the sustained presence of pain-related stress in such patients. The aim of our study was to investigate whether patients with functional somatic syndromes show altered skin temperatures also under everyday life conditions.

14 patients with functional somatic syndromes and 14 matched healthy control subjects were investigated under ambulatory conditions over six consecutive days. During this time, distal and proximal skin temperatures were continuously recorded and sleep-wake cycles were monitored by actimetry and sleep-wake diaries.

Unexpectedly, the patients showed higher distal skin temperatures than control subjects only in the afternoon. The objective temperature data did not match the patients' subjective experience: ratings of thermal comfort did not vary between the two groups. Moreover, similar levels of daytime activity were recorded in the two samples, even though patients reported more tiredness and more body tension than controls.

We interpret the observed dissociation between objective skin temperature measurements and subjective ratings of the bodily thermal comfort as support for the notion of an alexisomia account (reduced bodily awareness) for functional somatic syndromes. Moreover, findings indicate that subjective complaints of tiredness and tension do not necessarily result in physical avoidance behaviour.

Keywords: Functional somatic syndromes, skin temperature, wrist activity, ambulatory measurements, stress, sympathetic nervous system

### **Introduction**

Somatic symptoms of medically unknown origin are called functional somatic syndromes, which include various diseases (e.g. irritable bowel syndrome, fibromyalgia syndrome, chronic idiopathic lower back pain, non-cardiac chest pain, chronic fatigue syndrome, tension headache, multiple chemical sensitivity, etc.). Patients with such diagnoses have many symptoms, characteristics and other clinical conditions in common, like pain, sleep disturbances, tiredness, and emotional distress (Wessely et al. 1999; Eich et al. 2012). Petzke and Clauw (2000) reported that chronic pain and stress in fibromyalgia syndrome patients have a secondary action on various autonomic nervous systems parameters (muscle sympathetic activity, microcirculation, heart rate variability, tilt table testing, endogenous pain modulation, biochemical markers), in terms of that stress exposition leads to a lowered sympathetic response. The autonomic nervous system, in particular the sympathetic nervous system, is an important component of the human stress response. However, findings on patients with functional somatic syndromes are rather inconsistent and the experimental procedures in such studies are rather heterogeneous. A recent review suggested that a sympathetic nervous system predominance, defined as statistically significant higher sympathetic activity, decreased parasympathetic activity, or both, was common in patients with fibromyalgia syndrome, chronic fatigue syndrome, irritable bowel syndrome, and interstitial cystitis (Martinez-Martinez et al. 2014). Okifuji and Turk (2002) reviewed studies on stress reactivity in patients with fibromyalgia syndrome and concluded that previous findings support the hypothesis that hypofunctional stress systems play an important role in the pathophysiology of this syndrome. In line with this theory, Friederich and colleagues (2005) stated that subjects with a hyporeactive sympathetic nervous system show deficits in pain inhibiting mechanisms combined with an increased pain perception.

Distal skin temperature has been reported to decrease by exposure to psychological stress (Mittleman and Wolff 1939). Temperatures of proximal skin regions (e.g. thigh, infraclavicular region, stomach) passively follow core body temperature with similar amplitude under controlled laboratory conditions of a constant-routine protocol (resting and alert conditions) (Kräuchi and Wirz-Justice 1994; Kräuchi et al. 2006). However, distal skin temperatures (mainly hands and feet) exhibit an inverse and higher amplitude rhythm than the body core temperature (Kräuchi et al. 2006). Importantly, under ambulatory conditions, it could be shown that proximal and distal skin temperatures were in phase and synchronized to the sleep-wake cycle, with maximal values during nighttime and minimal values during waking hours (Martinez-Nicolas et al. 2015; Kräuchi et al. 2014; Martinez-Nicolas et al. 2011; Gompfer et al. 2010). Under such real-life conditions, all body temperatures show higher diurnal amplitudes especially in the lower extremities, as compared to controlled, resting laboratory conditions (Kräuchi et al. 2014).

In ambulatory settings, certain masking properties on the thermoregulatory system can occur, for example, sleep, postural change, environmental light exposure, motor activity, food and fluid intake, and emotional events (Martinez-Nicolas et al. 2015; Kräuchi et al. 2014). Kräuchi et al. (2014) suggested that the whole palette of regularly distributed behaviours across a day are potential zeitgebers. The distal-proximal skin temperature gradient (DPG) provides a selective measure of distal skin blood flow and, hence, of efficient body heat loss via the extremities (House and Tipton 2002; Rubinstein and Sessler 1990; Severens et al. 2010). Chronic pain somatoform disorder patients without obvious pathophysiology or psychopathology were found to have colder hands than healthy subjects (Wickramasekera et al. 1998). Cinciripini (1982) showed that surface electromyographic levels were high and distal skin temperature was low in chronic pain patients at rest. Moreover, Mitani et al. (2006) reported that fibromyalgia syndrome patients had significant asymmetries of distal skin temperature in terms of decreased skin temperature at the dominant body side.

The before-mentioned studies measured distal skin temperatures at baseline and under stress in laboratory conditions. However, patients presumably experience chronic stress during their normal life routine, partially by the presence of the pain itself (Martínez-Lavín 2012; Bennett et al. 1998). Therefore, distal skin temperatures should permanently be decreased during waking hours. The first objective of this study was to examine whether the distal skin temperature of patients with functional somatic syndromes is decreased in everyday life situations to a similar degree as observed under laboratory conditions. For practical reasons, distal skin temperatures were measured at the outer ankle of the feet. To our knowledge, the present study is the first to monitor skin temperature under ambulatory conditions in patients with functional somatic syndromes. Moreover, Kanbara and colleagues (2004) found dissociations between subjective and objective responses in functional somatic syndromes patients: patients showed hypofunctional psychophysiological responses (surface electromyography and skin conductance) to a stress task, but subjectively had stronger sensations of tension. The authors concluded from this discrepancy that functional somatic syndromes patients show a reduced bodily awareness. Hyporeactive objective tension and high subjective tension might be one of the characteristics of functional somatic syndromes (Okifuji and Turk 2002). Therefore, a second aim of our study was to compare the objective skin temperature measurements with the subjective ratings of the bodily thermal comfort of different body regions. Additionally, we combined the temperature measurements with activity measurements because actimetry stated to be a useful method for studying activity and sleep pattern in a naturalistic way (Korszun et al. 2002). Measuring levels of activity during the day provides an objective indication of the level of impaired physical function.

## **Methods**

### **Participants**

Fourteen patients with functional somatic syndromes (median age: 40 years; range 22-58 years) and fourteen healthy subjects (median age: 37.5; range 25-52 years) gave written informed consent to participate in the study that was approved by the local Ethical Committee. All participants were financially compensated with CHF 400 for their participation. The patient group (6 male and 8 female) was recruited from the department of psychosomatic medicine of the University Hospital Basel and by local practicing psychiatrists and physicians. Patients needed to fulfil the defined inclusion criteria (Table 1). Eight patients had a diagnosis of fibromyalgia syndrome or chronic pain, four a functional somatic syndromes diagnosis of the cardiovascular system and two of the stomach, bowel and genitourinary system. Two patients took antidepressants, one in combination with an analgesic. Four patients each took analgesics either regularly or occasionally. Four patients did not take any medication. Healthy control subjects were recruited via an advertisement on a university webpage. The sample of fourteen subjects (3 male and 11 female) was part of a larger sample of 31 subjects and was matched to the patient group, with respect to season at the time of the measurements (spring, summer, autumn, winter) and age. Control subjects had no medical or psychiatric diagnosis.

### **Procedures**

First, all subjects completed five questionnaires about somatic (SOMS-2, Rief et al. 1997; SOMS-7T, Rief and Hiller 2003) and mental symptoms (BDI, Hautzinger et al. 1995; BSI, Franke 2000; TAS-26, Kupfer et al. 2001; STAXI, Schwenkmezger et al. 1992). Subsequently, subjects were interviewed by an experienced psychiatrist from the department of psychosomatic medicine of the University Hospital Basel, meanwhile the questionnaires were evaluated. If subjects fulfilled the inclusion criteria for taking part in the study, they received detailed instructions about study procedures. Three additional patients were not included in the study because of not fulfilling the diagnosis criteria at the time point of the interview or because of symptoms of a severe depression. Two other patients were excluded in the course of the study because of faulty temperature recordings. One subject could not be included as controls because of impaired vision and another because of a depressive episode in the last two years. Moreover, data of two controls was not considered for the study because they worked night shifts, affecting their sleep-wake cycles.

### *Skin Temperature and Actimetric Recordings*

Skin temperatures were recorded once every 2.5-min during six consecutive days with wireless, waterproof temperature sensors (DS 1922L-F5, Thermochron iButtons®; resolution 0.0625°C, accuracy 0.5°C; Maxim, Dallas, USA). The four skin temperature sensors were affixed to the skin with thin air-permeable adhesive surgical tape (Omnifix®, Hartmann, Heidenheim Germany): a pair of sensors at the ankle (on the outer calcaneus bone, above the shoes; = distal skin region) and over the infraclavicular region on each side of the body (= proximal skin region). Participants continually wore waterproof actimeters (Daqtometer; Daqtix, Oetzen, Germany) around the wrist of their non-dominant arm. The actimeters recorded movement accelerations once every minute as an index for body movement, and saved these data. When the participants removed either temperature sensors or the actimeter, for example during or after showering, they had to note this time span in the sleep-wake diary. All participants were asked to maintain their normal daily activities, including sports, and sleep wake schedules during the time of the study.

### *Sleep-Wake Diaries*

The participants kept records of their sleep and wake episodes using a sleep-wake diary (like a questionnaire). Immediately after awakening, seven items had to be completed (bedtime, lights-off, estimated sleep onset latency, number of waking up during night, number of getting up during night, wake-up time (lights-on), and get-up time in the morning). Moreover, they had to rate the sleep quality, restorative sleep, and sleepiness before lights-off on a Likert scale (1 = very poor to 8 = very good). Besides the sleep diary, the Leeds Sleep Evaluation Questionnaire had to be filled out (Parrott and Hindmarch 1980). In the evening, before lights-off, participants had to complete the wake diary, which was divided into four time spans of the day: morning (getting up-12:00), afternoon (12:00-18:00), evening (18:00-lights-off), and shortly before lights-off. Bipolar visual analogue scales (VAS) from 0 to 10 were used to rate the following subjectively states: tiredness (alert-sleepy), tension (relaxed-tense), concentration (unfocused-concentrated), feelings of hunger (satiated-hungry), and mood (good mood-bad mood), and their thermal comfort (extremely hot-extremely cold) of three different skin regions (feet, body, and hands) (Gompper et al. 2010). Participants kept records of their daily activities and made notes of physical activity time (sport, gardening etc.), times spent outdoors, times of stressful psychological situations, times of eating a larger meal and of drinking beverages with caffeine or alcohol and their amount (no limitations of meals or drinks).

### Data Analysis

The recorded temperature and activity data were screened for artefacts. Data segments with large drops of the recorded temperatures, related to the removal of the temperature sensors, were replaced by values derived from a linear interpolation procedure. Abnormal spikes (steep gradients of  $> \pm 5^\circ/5\text{min}$ ), for example caused by temporarily malfunctioning of the temperature sensor, were replaced by values derived from the average data of this time span. For actimeter data, such artefact data were replaced by the average data of the day in question. This procedure has previously been used (Kräuchi et al. 2006; Gompfer et al. 2010). Subsequently, the revised data were averaged in 30-min bins across the six recording days per subject in order to minimize fluctuations and to reduce the number of time segments, as used in previous studies on circadian temperature rhythms (Kräuchi et al. 2012; Gompfer et al. 2010). Proximal skin temperature was determined as the mean value of the two infraclavicular region sensors, whereas distal skin temperature was measured as the mean value of the two ankle sensors. Temperature and actimeter data were analysed using a two-way repeated-measures Analysis of Variance (ANOVA) with Group (patients vs. controls) as between-subject factor and Time (48 x 30-min blocks) as within-subject factor. Data of the ratings in the sleep-wake diaries were averaged across the six recording days per subject and were analysed using a two-way repeated-measures ANOVA with Group (patients vs. controls) as between-subject factor and Time (4 time spans) as within-subject factor. The data were corrected with Greenhouse-Geisser technique to account for the violations of the sphericity assumption when necessary. Unpaired *t*-tests were performed for post-hoc group comparisons. All statistical analyses were performed using SPSS 23 for Windows™ (SPSS Inc., Chicago). The alpha level was set to 0.05.

## Results

### Subjects' Characteristics

The two groups were identical in season at the time of the measurements and were matched for age ( $U = 77$ ;  $p = .334$ ). Moreover, no statistical differences were present in gender ( $p = .420$ ), language ( $p = .481$ ), employment status ( $p = .648$ ) and education ( $p = .128$ ): All but two patients spoke German as a first language, four patients had an academic education, and ten patients were employed, whereas all controls spoke German as a first language, nine had an academic education, and twelve were employed.

As shown in Table 2, patients stated more physical complaints and the intensity of the complaints was clearly higher than in the control group. Patients also experienced more mental complaints: they indicated more symptoms and more severe symptoms in a wide range of psychopathology (e.g. depression, anxiety, psychoticism, obsessive-

compulsive etc.), as compared to controls. The patients had more difficulties identifying and describing feelings, but stated comparably externally-oriented thinking as controls. On the other hand, they showed more anger expression-in and anger expression-out strategies, but they did not differ from the control group in terms of anger control strategies.

Taken together, patients showed, as expected, more physical and mental impairments, as compared to controls.

### **Subjective Measures**

*Sleep:* Table 3 summarizes the sleep parameters of the study participants, as obtained by the sleep-wake diaries. With respect to sleep characteristics, patients reported significantly more getting up during the night, less sleep quality and less restorative sleep than the control group. The two groups did not differ in time in bed before lights-off, time of lights-off or in their sleep latency. The sleep duration statistically did not differ between the two groups, neither was waking during night, time of lights-on and time in bed after lights-on.

*VAS ratings:* Main effects of Time were not of major interest for the current study purpose and are, therefore, only described anecdotally. The subjective ratings on the basis of the sleep-wake diaries showed the following: patients stated a warmer sensation of body temperature in the morning and afternoon than controls and similar sensations in the evening and before lights-off (Time x Group:  $F(2.128,55.333) = 3.411, p = .037$ ; Figure 1a). No differences were found for the experienced hands and feet temperatures (Figures 1b and 1c).

The two groups differed significantly by the ratings of tension, in the sense of that the patients generally felt more tension than controls (Group,  $F(1,26) = 8.487, p = .007$ ; Figure 1d). Moreover, the patient group felt more tired during the day and in the evening than controls (Group:  $F(1,26) = 10.438, p = .003$ ; Time x Group:  $F(1.744,45.336) = 4.577, p = .019$ , Figure 1e). The level of tiredness varied little in the patient group, whereas controls were most alert in the afternoon, got sleepier during the evening and were most tired before lights-off.

Similar to tiredness, patients' concentration was relatively constant over the day and evening, whereas controls showed some variation over the day. Controls felt concentrated most in the afternoon and showed a pronounced decrease in the evening and before lights-off (Time x Group:  $F(1.887,49.068) = 3.943, p = .028$ ; Figure 1f).

The patient group did not statistically differ from the control group in subjective ratings of feelings of hunger. All subjects estimated to be less hungry before sleep (Figure 1g).

Taken together, these findings indicate that patients had more sleep disturbances, felt more tired and tense and showed warmer sensations of body temperature during the day than controls.



### Temperature and Actimeter data

The distal and proximal skin temperature measures showed maximal values during the night and minimal values during the light phase (main effect of Time, distal:  $F(3.89, 101.17) = 44.409, p < .001$ ; proximal:  $F(4.40, 114.54) = 12.330, p < .001$ ; distal-proximal skin temperature gradient, DPG:  $F(3.47, 90.23) = 24.560, p < .001$ ; Figures 2a and 2b). As expected, the proximal skin temperature amplitude was smaller than the amplitude of the distal skin temperature. Comparison did not show a Group effect in distal skin temperature ( $F(1, 26) = 1.344, p = .257$ ). However, the diurnal course of the distal temperature showed some variation between patients and controls (Time x Group interaction:  $F(3.89, 101.17) = 2.220, p = .074$ ). Post-hoc tests revealed increased distal skin temperatures in patients at afternoon (Figure 2a, Table 4). The DPG significantly differed between the groups during this time as well (Figure 2b, Table 4). In contrast, the proximal temperatures showed neither a Group effect ( $F(1, 26) = 1.270, p = .270$ ) nor a Time x Group interaction ( $F(4.40, 114.54) = 1.199, p = .315$ ).

Further analyses concerned correlations between sleep onset latency and temperature difference (from lights-off until 120 min after lights-off) or between sleep onset latency and temperature slope (from 5 min after lights-off until 35 min after lights-off), but their results were not significant (data not shown). The actimeter data showed an inverse pattern to the recorded skin temperatures. Hence, minimal values were observed during the night and maximal values were observed at the daytime (main effect of Time:  $F(7.09, 184.44) = 38.384, p < .001$ ). However, there were neither a Group effect ( $F(1, 26) = 0.906, p = .350$ ) nor a Time x Group interaction ( $F(7.09, 184.44) = 1.438, p = .192$ ). Overall, the activity level of patients was highly similar to controls (Figure 2c).

Taken together, patients showed higher distal skin temperatures and a smaller DPG during the afternoon than controls. The two groups exhibited similar levels of daytime activity.

### Discussion

This ambulatory study over six consecutive days in patients with functional somatic syndromes and healthy subjects provided an insight into diurnal profiles of distal and proximal skin temperatures and activity measurements. We revealed three main findings: first, the distal skin temperature of patients was increased in the afternoon, as compared to controls. Second, the subjective ratings of the bodily thermal comfort did not vary between the two groups; hence, the objective temperature data did not match the patients' subjective experience. And third, the actimeter data indicated similar levels of daytime activity in the two groups, despite higher levels of reported tiredness and body tension in patients.

With respect to skin temperatures, the groups significantly differed in distal skin temperatures in the afternoon (13:30-16:00). The patient group showed increased distal skin temperature, as compared to the control subjects, whereas the study of Wickramasekera and colleagues (1998) reported a reversed pattern. In the study of Wickramasekera et al. (1998), the distal skin temperature of chronic pain somatoform disorder patients was recorded in a resting position for a 4-minute period. The temperature in patients was significantly lower than in normal subjects, as collected in another study (Blanchard et al. 1989). One crucial limitation of the study of Wickramasekera et al. (1998) is that the control group data was collected under similar but not identical experimental conditions. The lower temperatures in the patient group of Wickramasekera et al. (1998) might indeed partly be explained by differences in the experimental set-up, as compared to Blanchard et al. (1989). As shown by the recent study of Gatt et al. (2015), the temperature is gradually falling from the 2<sup>nd</sup> to the 5<sup>th</sup> digit. In the study of Wickramasekera et al. (1998), the distal temperatures were recorded from the 5<sup>th</sup> digit, whereas the skin temperature was recorded at the 2<sup>nd</sup> digit in the study of Blanchard et al. (1989). Thus, the across-studies comparison between the two studies was confounded by different recording sites. Moreover, it is not clear to what extent patients in the study of Wickramasekera et al. (1998) felt stressed in their baseline recordings (simply by the experimental setting). Heightened stress levels in the 'baseline' recordings might provide another explanation for the previously observed decreased distal skin temperatures in patients.

In contrast, our recordings took place in an ambulatory setting and over a much longer period than four minutes. Thus, situational stress factors presumably had little influence on the recorded distal skin temperatures also because we averaged the data across several recording days. Associated with ambulatory measurements are seasonal differences, which influence the distal skin temperature measurements. Martinez-Nicolas and colleagues (2015) showed significant differences between summer and winter measurements of distal skin temperature in healthy controls, in the sense of increased distal skin temperatures during summer measurements. In the current study, this circumstance was accounted for by matching the two groups with regard to the season at the time of the measurements.

The observed increased distal skin temperature in patients compared to controls might be considered as support for the hypothesis of a hypofunctional stress system in functional somatic syndromes (Okifuji and Turk 2002). Presumably due to a hypofunctional stress system, patients did not react to situational stress in a normal manner and did not show vasoconstriction to protect the body core temperature (Kistler 1998). In contrary, lowered sympathetic activity might have led to vasodilatation and, as consequence, to increased distal skin temperature. This interpretation is supported by the significantly smaller DPG in the patient group during the same time period. The DPG is constituted to be a selective measure of distal skin blood flow (House and Tipton 2002; Rubinstein and

Sessler 1990; Severens et al. 2010), which indicates an increased distal skin blood flow in our patient group. Our finding definitely argues against the assumption of a sympathetic nervous system predominance in patients with functional somatic syndromes (Martinez-Martinez et al. 2014).

Our second major finding concerns the subjective ratings of the bodily thermal comfort that did not vary between the two groups. The objectively measured increased distal skin temperatures in the patient group were, thus, not reflected in their subjective ratings. Moreover, regarding the proximal skin temperatures, patients and controls exhibited almost identical temperature patterns, but, in the morning and the afternoon, the patient group rated their body as significantly warmer than controls did. Such dissociations between objective skin temperature measurements and subjective ratings of the bodily thermal comfort might be considered as evidence for alexisomia (Ikemi and Ikemi 1986). This concept is an expansion of the alexithymia concept and describes impaired awareness and expression of bodily sensations. Moreover, there is some evidence that patients with psychosomatic diseases and somatization tend to have alexithymia (Bach and Bach 1996; Shipko 1982; Sifneos 1973). Alexithymia refers to a difficulty in the awareness and expression of emotional feelings. In line with such account, our patient group scored higher on the overall alexithymia scale as well as on two of its subscales than the control group. Combined, this might indicate that patients with functional somatic syndromes have overall difficulties in the awareness and expression of both their bodily sensations and their emotions. However, one could also speculate that the subjectively experienced body temperatures in patients were related to increased room temperatures. In the absence of ambient temperature measurements, we cannot rule out this possibility.

With respect to the sleep parameters and subjective well-being, the patient group had subjectively less sleep quality and felt more tired and tensed up during the day and evening. These results are congruent with the literature about somatic functional syndromes (Wessely et al. 1999; Eich et al. 2012). It is not clear how these findings are related: The subjective feeling of reduced sleep quality and less restorative sleep might have led to increased tiredness or the increased tension might have led to a poorer sleep quality. However, together with the subjectively feeling of more body tension and chronic pain, physical avoidance behaviour might be expected: patients are choosing to be inactive in order to avoid discomfort resulting from physical activity. This would lead to a reduction of their physical capacity, which in turn would induce even more pain and symptoms (Schäfer et al. 2005). However, our actimeter data showed similar activity levels during the day between the patient and control group. A previous study about actigraphy in patients with fibromyalgia syndrome and depression is in concordance with our finding (Korszun et al. 2002). The authors compared fibromyalgia syndrome patients without comorbidity, patients with fibromyalgia syndrome with comorbid depression, patients with recurrent major depression and healthy controls. Fibromyalgia patients without comorbidity showed high levels of daytime activity similar to the control group,

whereas patients with fibromyalgia syndrome and comorbid depression showed significantly reduced daytime activity and significantly increased daytime sleeping compared to controls. Thus, physical avoidance behaviour was only found when severe depressive symptoms were present in addition to fibromyalgia syndrome. This might indicate that physical avoidance behaviour mainly arises from the combination of depressive symptoms and functional somatic syndromes. The similar activity levels of the two groups during the day also indicate that neither decreased activity levels in patients nor associated factors (less airflow due to less activity) could have caused the higher distal temperatures.

We acknowledge some limitations in our study. First, most studies about distal skin temperatures measure finger temperatures, not feet temperatures. However, temperature recordings from the finger is not suited for recordings over several days and, therefore, we opted for feet recordings. Even though to our knowledge there is no evidence in the literature, distal fingers and feet temperatures might show some systematic differences in their stress-related responses. Second, the ambient temperature has an effect on distal skin temperature, even though it is not directly reflected in the distal skin temperature profiles (Martinez-Nicolas et al. 2015). Similarly, clothing might have some impact on the recorded skin temperatures. In ambulatory recordings, these factors are difficult to control: Ideally, temperature sensors would record both skin temperatures and ambient temperatures, but to the best of our knowledge such sensors are not available. Detailed protocolling of the worn clothes would require extremely large compliance of study participants, but, as a downside, could result in some unwanted behavioural changes. Third, some of the patients were medicated. A possible influence of medication on the current findings remains debatable, also due to the heterogeneous and variable intake of analgesics. Finally, the group size and the observed effect sizes in the current study were small. Strict correction for multiple testing would have eliminated the significance of the observed group difference in distal skin temperature and subjectively experienced thermal discomfort. Thus, further work will be needed to confirm these yet preliminary findings in larger samples, also using more rigorous surveillance schemes of medication intake and environmental factors.

Overall, the results of this study suggest dissociation between objective skin temperature measurements and subjective ratings of the bodily thermal comfort, which we interpret as support for the notion of an alexisomia account for functional somatic syndromes. Moreover, findings indicate that subjective complaints of tiredness and tension do not necessarily result in physical avoidance behaviour.

## **Compliance with Ethical Standards**

Funding: This study was funded by the Academy of Swiss Insurance Medicine (ASIM).

Conflict of Interest: Author Carole Keppler declares that she has no conflict of interest. Author Timm Rosburg declares that he has no conflict of interest. Patrick Lemoine declares that he has no conflict of interest. Marlon Pflüger declares that he has no conflict of interest. Niklaus Gyr declares that he has no conflict of interest. Ralph Mager declares that he has no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

## References

- Bach, M. & Bach, D. (1996). Alexithymia in somatoform disorder and somatic disease: A comparative study. *Psychotherapy and Psychosomatics*, 65, 150-152.
- Bennett, E.J., Tennant, C.C., Piesse C., Badcock, C.A. & Kellow J.E. (1998). Level of chronic life stress predicts clinical outcome in irritable bowel syndrome. *Gut*, 43(2), 256-261.
- Blanchard, E.B., Morrill, B., Wittrock, D.A., Scharff, L. & Jaccard, J. (1989). Hand temperature norms for headache, hypertension, and irritable bowel syndrome. *Biofeedback and Self-Regulation*, 14(4), 319-331.
- Cinciripini, P.M. (1982). Relationship between frontal muscle tension and digital skin temperature during EMG biofeedback. *Perceptual and Motor Skills*, 54 (3), 895-898.
- Eich, W., Häuser, W., Arnold, B., Jäckel, W., Offenbächer, M., Petzke, F., ... Henningsen, P. (2012). Fibromyalgia syndrome. Definition, classification, clinical diagnosis und prognosis. *Schmerz*, 26, 247-258.
- Franke, G.H. (2000) *BSI. Brief Symptom Inventory von L.R. Derogatis (Kurzform der SCL-90-R) – Deutsche Version, Manual*. Göttingen: Beltz.
- Friederich, H.C., Schellberg, D., Mueller, K., Bieber, C., Zipfel, S. & Eich, W. (2005). Stress and autonomic dysregulation in patients with fibromyalgia syndrome. *Schmerz*, 19, 184-194.
- Gatt, A., Formosa, C., Cassar, K., Camilleri, K.P., De Raffaele, C., Mizzi, A., ... Chockalingam, N. (2015). Thermographic patterns of the upper and lower limbs: Baseline data. *International Journal of Vascular Medicine*, 2015.
- Gompper, B., Bromundt, V., Orgül, S., Flammer, J., Kräuchi, K. (2010). Phase relationship between skin temperature and sleep-wake rhythms in women with vascular dysregulation and controls under real-life conditions. *Chronobiology International*, 27(9-10), 1778-1796.
- Hautzinger, M., Bailer, M., Worall, H. & Keller, F. (1995). *BDI. Beck-Depressions-Inventar, Testhandbuch 2., überarbeitete Auflage*. Bern: Hans Huber.
- House, J.R. & Tipton, M.J. (2002). Using skin temperature gradients or skin heat flux measurements to determine thresholds of vasoconstriction and vasodilatation. *European Journal of Applied Physiology*, 88, 141-145.
- Ikemi, Y. & Ikemi, A. (1986). An oriental point of view in psychosomatic medicine. *Psychotherapy and Psychosomatics*, 45, 118-126.

- Kanbara, K., Mitani, Y., Fukunaga, M., Ishino, S., Takebayashi, T. & Nakai, Y. (2004). Paradoxical results of psychophysiological stress profile in functional somatic syndrome: correlation between subjective tension score and objective stress response. *Applied Psychophysiology and Biofeedback*, 29 (4), 255-268.
- Kistler, A., Mariauzouls C. & von Berlepsch K. (1998). Fingertip temperature as an indicator for sympathetic responses. *International Journal of Psychophysiology*, 29, 35-41.
- Korszun, A., Young, E.A., Engleberg, N.C., Brucksch, C.B., Greden, J.F. & Crofford, L.A. (2002). Use of actigraphy for monitoring sleep and activity levels in patients with fibromyalgia and depression. *Journal of Psychosomatic Research*, 52, 439-443.
- Kräuchi, K. & Wirz-Justice, A. (1994). Circadian rhythm of heat production, heart rate, and skin and core temperature under unmasking conditions in men. *American Journal of Physiology*, 267, R819-R829.
- Kräuchi, K., Knoblauch, V., Wirz-Justice, A. & Cajochen, C. (2006). Challenging the sleep homeostat does not influence the thermoregulatory system in men: evidence from a nap vs. sleep-deprivation study. *American Journal of Physiology Regulatory Integrative Comparative Physiology*, 290, R1052-R1061.
- Kräuchi, K., Gompfer, B., Hauenstein, D., Flammer, J., Pflüger, M., Studerus, E., ... Orgül, S. (2012). Diurnal blood pressure variations are associated with changes in distal-proximal skin temperature gradient. *Chronobiology International*, 29 (9), 1273-1283.
- Kräuchi, K., Konieczka, K., Roescheisen-Weich, C., Gompfer, B., Hauenstein, D., Schoetzau, A., ... Flammer, J. (2014). Diurnal and menstrual cycles in body temperature are regulated differently: A 28-day ambulatory study in healthy women with thermal discomfort of cold extremities and controls. *Chronobiology International*, 31 (1), 102-113.
- Kupfer, J., Brosig, B. & Brähler, E. (2001). *TAS-26. Toronto-Alexithymie-Skala-26, Manual*. Göttingen: Hogrefe.
- Martínez-Lavín, M. (2012). Fibromyalgia when distress becomes (un)sympathetic pain. *Pain Research and Treatment*, 2012.
- Martínez- Martínez, L.A., Mora, T., Vargas, A., Fuentes-Iniestra, M. & Martínez-Lavín, M. (2014). Sympathetic Nervous System Dysfunction in Fibromyalgia, Chronic Fatigue Syndrome, Irritable Bowel Syndrome, and Interstitial Cystitis. A Review of Case-Control Studies. *Journal of Clinical Rheumatology*, 20 (3), 146-150.
- Martinez-Nicolas, A., Ortiz-Tudela, E., Madrid, J.A. & Angeles Rol, M. (2011). Crosstalk between environmental light and internal time in humans. *Chronobiology International*, 28 (7), 617-629.

- Martinez-Nicolas, A., Meyer, M., Hunkler, S., Madrid, J.A., Angeles Rol, M., Meyer, A.H., ... Kräuchi, K. (2015). Daytime variation in ambient temperature affects skin temperatures and blood pressure: Ambulatory winter/summer comparison in healthy young women. *Physiology and Behavior*, *149*, 203-211.
- Mitani, Y., Fukunaga, M., Kanbara, K., Takebayashi, N., Ishino, S. & Nakai, Y. (2006). Evaluation of psychophysiological asymmetry in patients with fibromyalgia syndrome. *Applied Psychophysiology and Biofeedback*, *31* (3), 217-225.
- Mittleman, B. & Wolff, H. (1939). Affective states and skin temperature: Experimental study of subjects with „cold hands“ and Raynaud’s syndrome. *Psychosomatic Medicine*, *1*, 271-292.
- Okifuji, A. & Turk, D.C. (2002). Stress and Psychophysiological Dysregulation in Patients with fibromyalgia syndrome. *Applied Psychophysiology and Biofeedback*, *27* (2), 129-141.
- Petzke, F. & Clauw, D.J. (2000). Sympathetic nervous system function in fibromyalgia. *Current Rheumatology Reports*, *2*, 116-123.
- Rief, W., Hiller, W. & Heuser, J. (1997). *SOMS. Das Screening für Somatoforme Störungen, Manual*. Bern: Hans Huber.
- Rief, W. & Hiller, W. (2003). A new approach to the assessment of the treatment effects of somatoform disorders. *Psychosomatics*, *44* (6), 492-498.
- Rubinstein, E. & Sessler, D. (1990). Skin-surface temperature gradients correlate with fingertip blood flow in humans. *Anesthesiology*, *73*, 541-545.
- Schäfer, R., Benedikt, G., Sauer, N., Wilke, S., Herzog, W., Szecsenyi, ... Henningsen, P. (2005). Früherkennung und Behandlung funktioneller / somatoformer Beschwerden in der Allgemeinpraxis. *Notfall und Hausarztmedizin*, *31* (12), 527-536.
- Schwenkmezger, P., Hodapp, V. & Spielberger C.D. (1992). *STAXI. Das State-Trait-Ärgerausdrucks-Inventar, Handbuch*. Bern: Hans Huber.
- Severens, N.M., van Marken Lichtenbelt, W.D., Frijns, A.J., Kingma, B.R., de Mol, B.A., van Steenhoven, A.A. (2010). Measurement of model coefficients of skin sympathetic vasoconstriction. *Physiological Measurement*, *31*, 77-93.
- Shipko, S. (1982). Alexithymia and somatization. *Psychotherapy and Psychosomatics*, *37*, 193-201.
- Sifneos, P.E. (1973). The prevalence of « alexithymic » characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, *22*, 255-262.
- Parrott, A.C. & Hindmarch, I. (1980). The Leeds Sleep Evaluation Questionnaire in psychopharmacological investigations – a review. *Psychopharmacology*, *71*, 173-179.



Wessely, S., Nimnuan, C. & Sharpe, M. (1999). Functional somatic syndromes: one or many? *Lancet*, 354, 936-939.

Wickramasekera, I.E., Kolm, P., Pope, A. & Turner, M. (1998). Observation of a paradoxical temperature increase during cognitive stress in some chronic pain patients. *Applied Psychophysiology and Biofeedback*, 23g (4), 223-241.

## Tables

**Table 1** Inclusion criteria patient group

---

Inclusion criteria	
▪ Diagnosis of functional disorder (including chronic pain)	▪ No apparent addiction disorder; no substance or drug abuse in the last two years
▪ Age between 21 and 60 years	▪ In the last twelve months, no inability to work for longer than three months
▪ Good knowledge of the German language	▪ No pension/partial pension or an apply for a pension of a disability insurance or an accident insurance
▪ No somatic diagnosis for the complained symptoms	
▪ No psychiatric disorder other than mild depression	

---

**Table 2** Somatic and mental questionnaires

Variable	Controls	Patients	Statistics
Somatization disorder (SOMS; raw scores)	0.0 (0.3)	4.0 (3.8)	$U = 13, p < .001$
Somatoform autonomic dysfunction (SOMS; raw sc.)	0.0 (0.3)	4.0 (3.0)	$U = 5.5, p < .001$
Complaint/somatization index (SOMS; raw scores)	0.0 (0.5)	12.0 (12.5)	$U = 5.5, p = .001$
Somatization symptom count (SOMS-7; raw scores)	0.0 (0.3)	10.0 (8.5)	$U = 9.5, p = .001$
Somatization Severity index (SOMS-7; raw scores)	0.0 (0.3)	2.1 (0.9)	$U = 8.5, p = .001$
Depression (BDI; raw scores)	1.5 (3.0)	13.0 (12.3)	$U = 8.5, p = .001$
Somatization (BSI; <i>T</i> values)	44.5 (9.0)	64.5 (16.0)	$U = 9.5, p = .001$
Obsessive-Compulsive (BSI; <i>T</i> values)	36.0 (11.5)	62.5 (53.0)	$U = 36.5, p = .004$
Interpersonal Sensitivity (BSI; <i>T</i> values)	38.0 (4.3)	56.0 (16.0)	$U = 14.5, p < .001$
Depression (BSI; <i>T</i> values)	40.0 (3.0)	62.5 (15.5)	$U = 3.0, p < .001$
Anxiety (BSI; <i>T</i> values)	45.0 (7.0)	61.0 (22.0)	$U = 13.0, p < .001$
Hostility (BSI; <i>T</i> values)	43.0 (10.0)	57.5 (18.5)	$U = 25.5, p = .001$
Phobic Anxiety (BSI; <i>T</i> values)	44.0 (1.0)	56.0 (16.5)	$U = 38.5, p = .004$
Paranoid Ideation (BSI; <i>T</i> values)	45.0 (10.0)	55.0 (8.0)	$U = 17.0, p < .001$
Psychoticism (BSI; <i>T</i> values)	43.0 (1.0)	51.0 (27.3)	$U = 37.0, p = .003$
Global Severity Index (BSI; <i>T</i> values)	38.5 (10.0)	62.5 (14.3)	$U = 14.0, p < .001$
Positive Symptom Total (BSI; <i>T</i> values)	38.5 (14.3)	58.0 (14.0)	$U = 0.0, p < .001$
Positive Symptom Distress Index (BSI; <i>T</i> values)	40.5 (10.5)	66.5 (13.3)	$U = 14.5, p < .001$
Difficulties identifying feelings (TAS-26; raw scores)	8.5 (5.3)	18.0 (9.5)	$U = 8.0, p < .001$
Difficulties describing feelings (TAS-26; raw scores)	10.0 (2.5)	12.5 (5.5)	$U = 24.5, p = .001$
Externally-oriented thinking (TAS-26; raw scores)	12.5 (2.5)	13.0 (7.0)	$U = 72.0, p = .226$
Alexithymia (TAS-26; raw scores)	31.0 (6.8)	47.0 (14.5)	$U = 10.5, p < .001$
Anger expression-in (STAXI; raw scores)	13.0 (6.8)	17.5 (5.5)	$U = 41.0, p = .009$
Anger expression-out (STAXI; raw scores)	9.5 (3.0)	11.0 (4.5)	$U = 55.5, p = .048$
Anger control (STAXI; raw scores)	25.0 (8.0)	24.5 (9.5)	$U = 73.5, p = .258$

Values are medians (interquartile range). *p*-values are performed by Mann-Whitney *U* test. All data were collected by questionnaires.

**Table 3** Sleep parameters of study participants

Variable	Controls	Patients	Statistics
Time in bed before lights-off (min)	5.5 (18.3)	5.0 (29.3)	$U = 97.0, p = .963$
Lights-off (h)	24.0 (1.3)	24.0 (1.4)	$U = 78.0, p = .358$
Sleep latency (min)	9.5 (14.8)	18.5 (40.3)	$U = 64.6, p = .123$
Sleep duration (h)	7.2 (1.1)	7.1 (0.5)	$U = 78.0, p = .358$
Lights-on (h)	6.5 (1.4)	7.3 (1.4)	$U = 68.5, p = .175$
Time in bed after lights-on (min)	25.5 (32.0)	38.5 (36.5)	$U = 85.5, p = .565$
Waking during night (number)	1.0 (0.5)	1.5 (2.0)	$U = 58.5, p = .087$
Getting up during night (number)	0.0 (1.0)	1.0 (1.3)	$U = 50.0, p = .015$
Sleep quality (Likert scale)	6.0 (1.3)	4.5 (2.3)	$U = 49.5, p = .022$
Restorative sleep (Likert scale)	6.0 (1.3)	4.0 (2.3)	$U = 37.5, p = .004$

Values are medians (interquartile range) over 1 week. *p*-values are performed by Mann-Whitney *U* test. All variables were collected by the use of sleep-wake diaries. Sleep quality and restorative sleep were measured by Likert scales (1 = very poor, 8 = very good).

**Table 4** Post hoc *t*-tests for skin temperatures

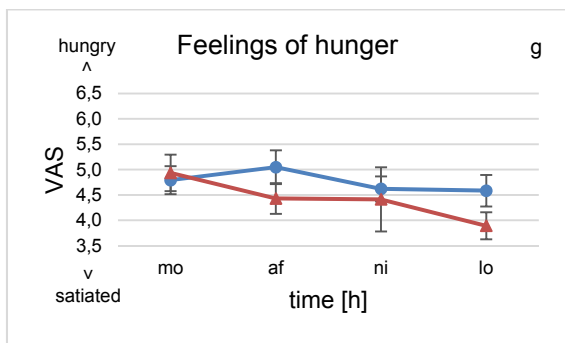
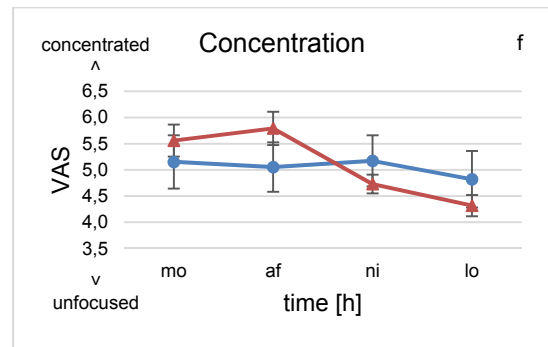
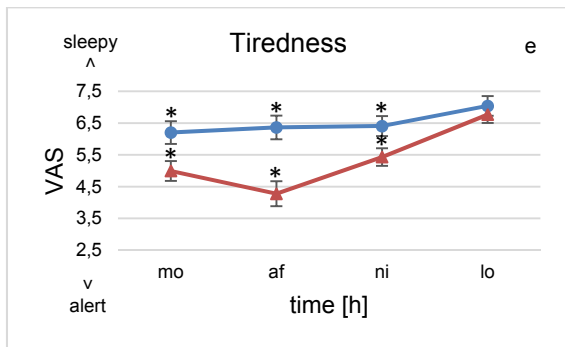
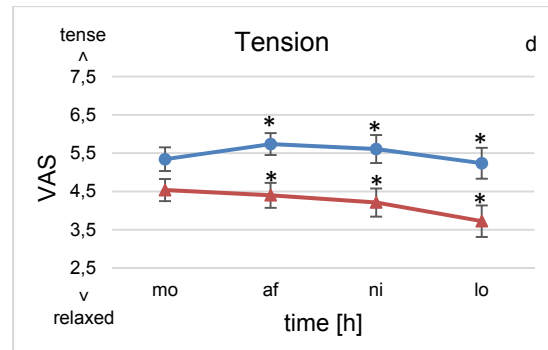
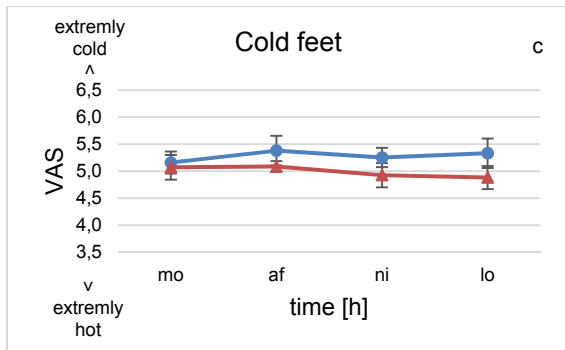
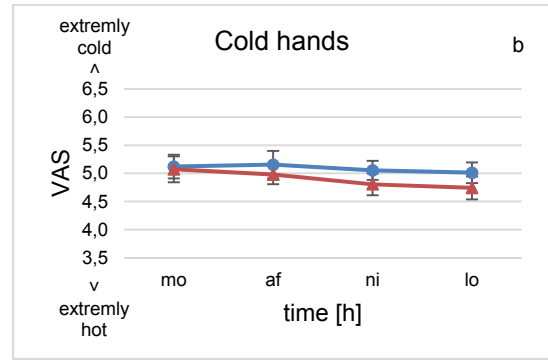
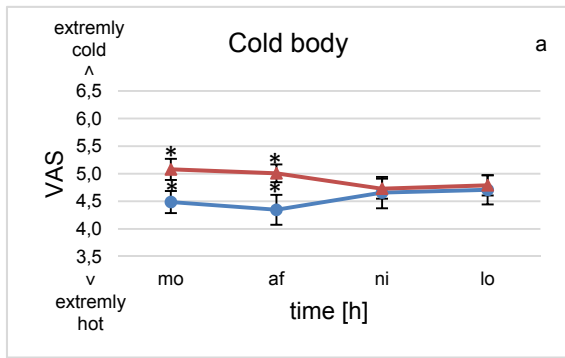
Time point	Distal	DPG
13:30	$t_{26} = 2.076$ $p = .048$	-
14:00	$t_{26} = 2.308$ $p = .029$	$t_{26} = 2.371$ $p = .025$
14:30	$t_{26} = 2.603$ $p = .015$	$t_{26} = 2.191$ $p = .038$
15:00	$t_{26} = 2.545$ $p = .017$	$t_{26} = 2.102$ $p = .045$
15:30	$t_{26} = 2.406$ $p = .024$	$t_{26} = 2.214$ $p = .036$
16:00	$t_{26} = 2.145$ $p = .041$	$t_{26} = 2.121$ $p = .044$

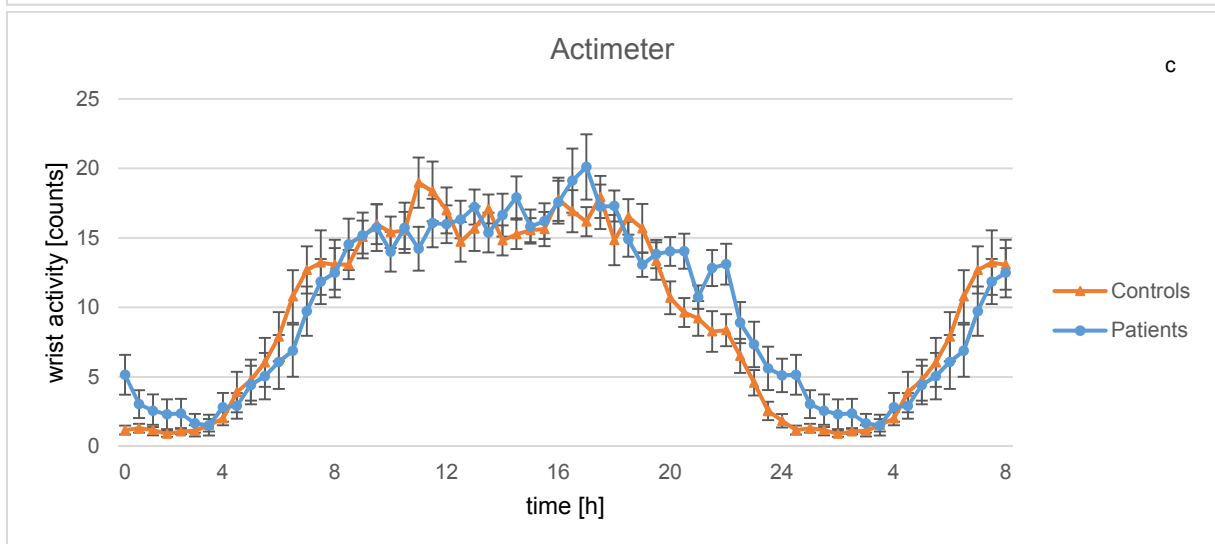
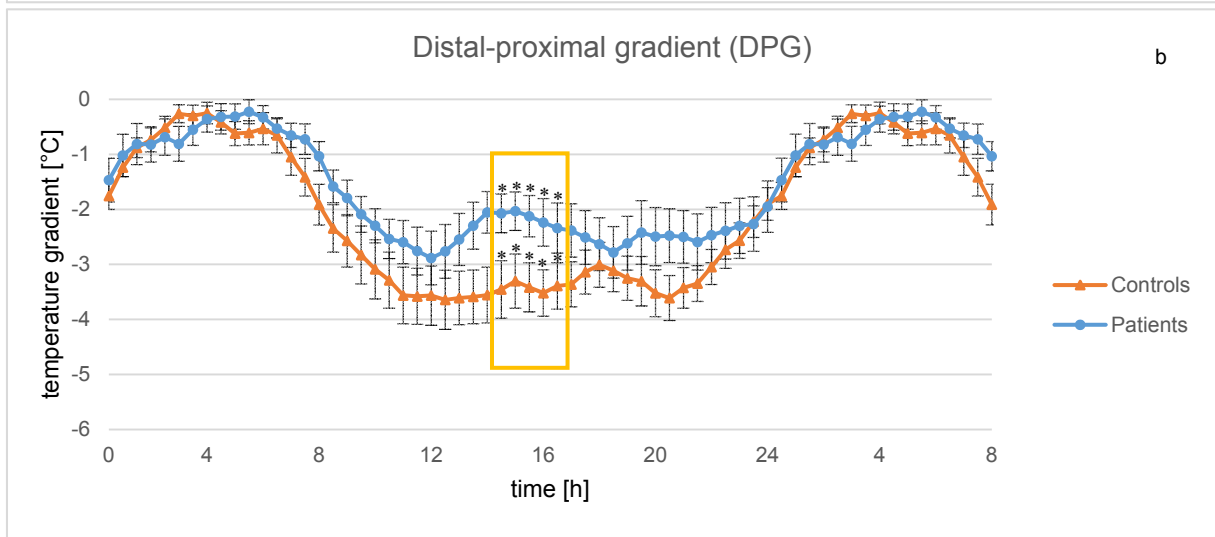
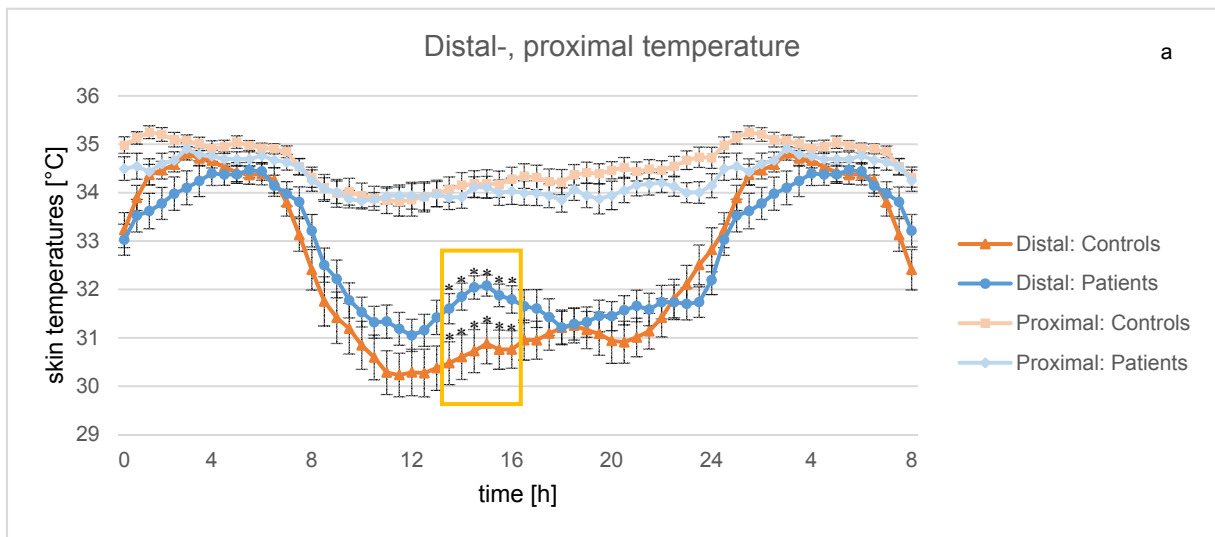
*t*-tests between groups (patients versus controls) for averaged time points for distal temperature and DPG (distal-proximal skin temperature gradient). Only the six statistically significant time points for distal and the five statistically significant time points for DPG are displayed.

## Figure-caption list

**Fig. 1a-1g** Subjective data. Subjective VAS data of thermal discomfort (cold body, cold feet, cold hands), tension, tiredness, concentration and feelings of hunger (Patients: N = 14, blue curves; Controls: N = 14, red curves). Data (mean±SEM) are displayed for time spans (mo = getting up-12:00 h; af = 12:00-18:00 h; ni = 18:00-lights-off) and at lights-off (= lo). \*  $p < .05$

**Fig. 2a-2c** Skin temperature and actimeter data (30-min bins). From top to bottom: Diurnal pattern of distal and proximal skin temperatures, DPG (distal-proximal skin temperature gradient), and actimeter data under ambulatory conditions (mean±SEM). Patients: N = 14, blue curves; Controls: N = 14, red curves. Data of the sleep episode are plotted to both sides to highlight the diurnal rhythm. \*  $p < .05$







**ANHANG A-3**

**Publikation 3**

Delahaye, M., Stieglitz, R., Graf, M., Keppler, C., Maes, J. & Pflüger, M. (2015).  
Deutsche Übersetzung und Validierung des Stress Appraisal Measure (SAM).  
*Fortschritte der Neurologie Psychiatrie*, 83, 276-283.

# Deutsche Übersetzung und Validierung des Stress Appraisal Measure (SAM)

## German Translation and Validation of the Stress Appraisal Measure (SAM)

### Autoren

M. Delahaye<sup>1</sup>, R. D. Stieglitz<sup>2</sup>, M. Graf<sup>1</sup>, C. Keppler<sup>1</sup>, J. Maes<sup>3</sup>, M. Pflueger<sup>1</sup>

### Institute

<sup>1</sup> Forensisch-Psychiatrische Klinik, Universität Basel, Universitäre Psychiatrische Kliniken, Basel

<sup>2</sup> Psychiatrie und Psychotherapie, Universitäre Psychiatrische Kliniken, Basel

<sup>3</sup> Psychologie, Universität der Bundeswehr München, Neubiberg/München

### Schlüsselwörter

- Stress
- Messung
- Fragebogen
- Deutsche Validierung

### Key words

- stress
- measurement
- questionnaire
- German-language validation

### Zusammenfassung



**Ziel der Studie:** In der vorliegenden Studie wurden eine Übersetzung und eine deutschsprachige Validierung des Stress Appraisal Measure (SAM) von Peacock und Wong an einer studentischen Stichprobe vorgenommen. Der SAM ist ein vergleichsweise kurzer Fragebogen (28 Items) zur Bewertung eines aktuellen, stressauslösenden Ereignisses. Den theoretischen Hintergrund lieferte das Stressmodell von Lazarus und Folkman.

**Methodik:** Bei 85 Studenten (Durchschnittsalter: 23; 59 Frauen, 26 Männer; Rücklaufquote: 42%) wurde anhand von kurzen Imaginationsgeschichten Stress induziert, und anschließend sollte der SAM ausgefüllt werden. Zunächst wurde die stressauslösende Wirkung der Imaginationsgeschichten geprüft. Als weiterer statistischer Analyseschritt wurden eine Faktorenanalyse sowie die Bestimmung der internen Konsistenzen der sieben SAM-Skalen durchgeführt. Außerdem wurde die konvergente Validität des SAM zum State and Trait Anxiety Inventory (STAI), zum Coping Inventory for Stressful Situations (CISS) und zu spezifischen Emotionskalen mittels Pearson-Produkt-Moment-Korrelation ermittelt.

**Ergebnisse:** Eine stressinduzierende Wirkung der Imaginationsgeschichten konnte nachgewiesen werden. Die Faktorenstruktur, die internen Konsistenzen der einzelnen Skalen sowie die konvergente Validität des SAM aus der ursprünglichen Studie konnten in der vorliegenden deutschen Version in großen Teilen repliziert werden. Einzelne Items (besonders des fünften Faktors) luden allerdings auf anderen Faktoren als bei der Studie von Peacock und Wong.

**Schlussfolgerung:** Der SAM könnte nun auch in deutscher Sprache eingesetzt werden.

### Abstract



**Objective:** In the present study, the German-language version of the Stress Appraisal Measure (SAM) by Peacock and Wong was validated in a student population. SAM is a relatively short questionnaire (28 items) that evaluates a current, stress-triggering event. The theoretical background is provided by the stress model of Lazarus and Folkman.

**Method:** 85 students (age: 23; 59 female, 26 male) were exposed to two stress scenarios in order to test whether they were suited to provoke stress. A factor analysis was performed and the internal consistency of the seven SAM scales was determined. In addition, the convergent validity of SAM with State and Trait Anxiety Inventory (STAI), Coping Inventory for Stressful Situations (CISS) and specific emotion scales was investigated via Pearson's product-moment correlation.

**Results:** The two stress scenarios were suited to evoke stress. The factor structure and the internal consistency of the individual scales, as well as the convergent validity of SAM were replicated with minor limitations in the present German version. Some items (especially from the fifth factor) were only replicated partially.

**Conclusion:** SAM can also be employed in the German language version.

### Bibliografie

**DOI** <http://dx.doi.org/10.1055/s-0034-1399727>  
 Fortschr Neurol Psychiatr 2015; 83: 276–285 © Georg Thieme Verlag KG Stuttgart · New York · ISSN 0720-4299

### Korrespondenzadresse

**Marcel Delahaye**  
 Forensik, UPK Basel  
 Wilhelm Klein-Str. 27  
 4052 Basel  
 marcel.delahaye@upkbs.ch

## Einleitung



Zur Erfassung und Verarbeitung von Stresserleben werden häufig Fragebögen eingesetzt (z. B. COPE [1], PASA [2], FERUS [3], SVF-120 [4], TICS [5]). In diesem Artikel soll der Fragebogen Stress Appraisal Measure (SAM) von Peacock und Wong [6] vorgestellt, übersetzt und in deutscher Sprache validiert werden.

Der wesentliche Grund für die Übersetzung und Validierung des SAM war das Fehlen eines theoriegeleiteten, kurzen, deutschsprachigen Stressfragebogens, der auf ein aktuelles Lebensereignis fokussiert und bei dem die Einschätzung der Kontrollmöglichkeiten klar unterteilt ist. Im Vergleich mit anderen Stresserfassungsinstrumenten überzeugte der SAM für diesen Einsatz am meisten, da viele andere Fragebögen entweder einen längeren Zeitraum, bzw. generelle Copingstile (z. B. TICS [5], SVF 120 [4]) oder eher Handlungsstrategien (z. B. CISS [8], COPE [1]) messen. Beim SAM steht die Erfassung von aktuell stattfindenden, kognitiven Verarbeitungsmechanismen bei akutem Stress im Vordergrund. Peacock und Wong [1] sprachen in diesem Zusammenhang von antizipatorischem Stress. Beim SAM überzeugte die klare, systematische und theoriegeleitete Operationalisierung der Items nach der Theorie von Lazarus und Folkman [7]. Außerdem ist die genauere Aufteilung der Kontrollattributionen (z. B. im Gegensatz zum PASA [4]) hervorzuheben.

Die Autoren sehen als Einsatzgebiet für den SAM beispielsweise das Stressimpfungstraining sensu D. Meichenbaum (Stress inoculation training) [9]. So könnten im Rahmen des Stressmanagementtrainings etwa die Lernfortschritte der Teilnehmer in den einzelnen Phasen (Informations-, Übungs- und Anwendungsphase) mittels SAM bewertet werden. Beispielsweise könnten vor und nach dem Erlernen der einzelnen Phasen Stress hervorrufende Ereignisse mittels SAM bewertet werden. Ein weiteres Einsatzgebiet lässt sich aus der Studie von Levy [10] bei Sportlern ableiten. So lautete die zentrale Aussage, dass die mentale Stärke eines Athleten enormen Einfluss darauf hat, ob ein Wettkampf als Herausforderung oder Bedrohung erlebt wird. Hier könnte der SAM etwa dazu dienen, die Änderung des mentalen Reifungsprozesses eines Athleten während der Vorbereitung auf den Wettkampf zu dokumentieren.

Als Grundlage zur Testkonstruktion zur Messung von „Stress“ bzw. „Stressverarbeitung“ werden je nach Autor eine oder mehrere Theorien herangezogen. Gängige Stresstheorien sind das salutogenetische Modell von Antonovsky [11, 12], das Selbstmanagementkonzept von Kanfer, Reinecker und Schmelzer [13], die Theorie der Selbstwirksamkeit von Bandura [14], die Depressionstheorie von Beck [15], das Modell zur sozialen Unterstützung nach Sommer und Fydrich [16], die Theorie der Ressourcenerhaltung nach Hobfoll und Buchwald [17] sowie das „Model of behavioral Self regulation“ von Carver und Scheier [18]. Das kognitiv-transaktionale Stressmodell von Lazarus und Folkman [7] ist sicherlich eines der gebräuchlichsten in der Stressforschung und wird nach wie vor erfolgreich angewendet [19, 20].

Sowohl für den SAM [6] als auch für zahlreiche andere Stresstests war diese Theorie Grundlage der Testkonstruktion (CISS [8], COPE [1], SVF 120 [4], PASA [2]). Auch eines der bekanntesten Interventions- und Präventionsprogramme bezüglich Stress, nämlich das Stressimpfungstraining von Meichenbaum [21], baut auf der Theorie von Lazarus und Folkman [7] auf.

Lazarus und Folkman [7] teilten die Reaktionen eines Menschen auf Ereignisse, die sein inneres Gleichgewicht stören, in primäre und sekundäre kognitive Bewertungen (primary and secondary appraisal) ein. Unter einer kognitiven Bewertung (appraisal)

werden generell Wahrnehmungs- und Interpretationsprozesse einer Situation verstanden. Um nun eine Situation als stressvoll einzustufen, werden nach Lazarus und Folkman zwei kognitive Bewertungsprozesse durchlaufen. Das primary appraisal meint die Einschätzung einer Situation entweder als gefährlich, angenehm oder irrelevant. Unter secondary appraisal wird die Einschätzung der eigenen Fähigkeiten verstanden, dieses Ereignis zu kontrollieren und zu bewältigen. Stress entsteht, wenn das Ergebnis des ersten Bewertungsschritts gefährlich lautet und das des zweiten Schritts, dass die zur Verfügung stehenden eigenen Ressourcen und Kontrollkompetenzen nicht ausreichen, um eine potenzielle Bedrohung zu bewältigen.

Peacock und Wong [6] versuchten in ihrem Test die beiden kognitiven Prozesse (primary und secondary appraisal) zu operationalisieren. Sie führten insgesamt drei Studien durch. In der ersten Studie, bei der 100 Psychologiestudenten ihr Stresserleben im Hinblick auf ein bevorstehendes Examen einschätzen sollten, erfolgte die Itemauswahl (28 Items) für den SAM. In dieser Studie wurden aus 37 Items diejenigen mit der höchsten Skalenkorelation für die von Peacock und Wong [6] zuvor theoretisch ausgewählten sieben Skalen gewählt. Die Skalen für das primary appraisal lauteten Bedrohung (threat), Herausforderung (challenge) und Bedeutsamkeit (centrality) und die Skalen für das secondary appraisal waren eigene Kontrolle (controllable by self), Kontrolle durch andere (controllable by others) und Unkontrollierbarkeit (uncontrollable). Außerdem wurde eine übergeordnete Skala generelles Stressempfinden (stressfulness) gebildet. Der SAM besteht also aus einer Skala zur Messung von allgemeinem Stress und aus sechs Skalen zur Erfassung spezifischer kognitiver Bewertungen. Peacock und Wong [6] identifizierten jeweils vier Items pro Skala. Jedes Item konnte auf einer fünfstufigen Likertskala (1 = stimme überhaupt nicht zu, 5 = stimme voll und ganz zu) beantwortet werden.

In der zweiten und dritten Studie von Peacock und Wong [6] wurde der eigentliche SAM eingesetzt, und es wurden jeweils Faktorenanalysen gerechnet. Das Ergebnis der zweiten Studie, bei der Stressinduktionsgeschichten eingesetzt wurden, waren fünf Faktoren, und das Ergebnis der dritten Studie, bei der die Angst vor einem bevorstehenden Examen bewertet werden sollte, war eine sechsfaktorielle Lösung. Fokus der vorliegenden Arbeit war die zweite Studie, da ebenfalls stressauslösende Imaginationsszenarien benutzt wurden, um akuten Stress zu erzeugen. Diese wurden von Peacock und Wong [6] allerdings nicht in ihrem genauen Wortlaut veröffentlicht.

## Methodik



Neben der Übersetzung sollten Reliabilität und Validität des SAM (erneut) geprüft werden. Dieses Vorgehen empfiehlt sich vor dem Einsatz eines übersetzten fremdsprachigen Fragebogens. Der SAM wurde also zunächst übersetzt und dann mit einer deutschsprachigen Stichprobe erneut validiert. Ziel war es, sowohl die Faktorenstruktur als auch die internen Konsistenzen analog zu Peacock und Wong [6] zu replizieren und eine zufriedenstellende konvergente Validität nachzuweisen.

## Design

Übersetzung: Die 28 Items des ursprünglichen SAM wurden für unsere Studie zunächst von einem professionellen Übersetzer vom Englischen ins Deutsche übersetzt (Anhang 1). Anschließend wurde von einem weiteren Übersetzungsbüro eine Rück-

übersetzung vom Deutschen ins Englische durchgeführt, um zu überprüfen, ob eine Abweichung zwischen der Originalversion von Peacock und Wong [6] und der Rückübersetzung vom Deutschen ins Englische vorlag. Die beiden englischen Versionen (das Original von Peacock und Wong und die vom zweiten Übersetzungsbüro rückübersetzte Version) wichen nur sehr geringfügig voneinander ab, was als Nachweis dafür gelten kann, dass die deutsche der englischen Version semantisch entspricht.

**Stressinduktion:** Zur Stressinduktion wurden in Anlehnung an die von den Autoren genannten Themen in ihrer zweiten Studie zwei Imaginationsgeschichten mit unterschiedlicher Stressintensität und unterschiedlichen Einflussmöglichkeiten konstruiert (Anhang 2 und 3). Bei einer der Imaginationsgeschichten ging es um die Absage eines nicht zum Lebensunterhalt notwendigen Zusatzjobs (JOB) und bei der anderen um die Mitteilung, sich möglicherweise mit HIV infiziert zu haben (HIV). Peacock und Wong [6] nahmen an, dass die Situationen im Hinblick auf ihre Kontrollierbarkeit und ihren Schweregrad von den Probanden unterschiedlich wahrgenommen werden. Es wurde also ein ein-faktorielles Prä-Post-Design angewendet. Der Faktor Stressinduktion enthielt zwei Stufen (HIV- bzw. JOB-Imaginationsgeschichte). Die Teilnehmer erhielten als Stressauslöser jeweils eine der beiden Imaginationsgeschichten.

**Validierung:** Im Gegensatz zu Peacock und Wong [6], die in ihrer Originalstudie Instrumente zur Messung von „psychological symptoms“ sowie „dysphoric mood“ einsetzten, wurden in unserer Studie deutsche Verfahren gewählt, die sich inhaltlich stärker an akute Stressmessung anlehnen und als etablierte Validierungsinstrumente in ähnlichen Studien eingesetzt wurden

[22–24]. Zur Bestimmung der konvergenten sowie diskriminanten Validität wurden der CISS (Coping Inventory for Stressful Situations [8]), der STAI (State-Trait-Angstinventar [25]) sowie fünf Einzelitems mit Augenscheinvalidität zur Messung spezifischer Emotionen eingesetzt. Die Reihenfolge der Fragebögen ergab sich wie in **Tab. 1** beschrieben.

Der CISS misst im Unterschied zum SAM zwar nicht aktuelles Erleben (states), sondern generelle, zeitlich überdauernde Dispositionen zur Stressverarbeitung (traits). Dennoch ist der CISS als Validierungsinstrument geeignet, da sich die situationsspezifischen states (wie im SAM gemessen) aus konsistenten, zeitlich überdauernden, typischen Stress-Reaktionsmustern (den traits) ableiten lassen (wie beim CISS gemessen). Die Grundlage zu dieser Annahme bildete die Latent-State-Trait-Theorie von Kelava und Schermelleh-Engel [26].

Der Einsatz des STAI als zweites Validierungsinstrument lag darin begründet, dass Angst inhaltlich einem intensiven negativen Stresserleben nahesteht, obwohl Stress emotional facettenreicher ist. In der Definition der APA (American Psychological Association) besteht akuter Stress emotional aus einer Kombination aus Wut bzw. Reizbarkeit, Depression und vor allem aus Angst [27].

### Messinstrumente

**CISS:** Die deutsche Version des CISS nach Kälin [8] besteht aus 33 Items, denen auf einer fünfstufigen Likertskala (von 1 = „sehr untypisch“ bis 5 = „sehr typisch“) zugestimmt werden kann. Zur Auswertung werden nach Kälin (1995) die Skalenmittelwerte gebildet. Der CISS besteht aus einer Skala zum emotionsorientierten Coping, bei dem die Einschätzung stressauslösender Gefühle

	Faktorladungen				
	I	II	III	IV	V
<i>SAM-Skala Bedrohung (threat)</i>					
Bedrohliche Situation (SAM 20)	0,88				
Ängstlich (SAM 5)	0,74				
Negatives Ende (SAM 11)	0,60				
Negative Auswirkungen (SAM 28)	0,82				
<i>SAM-Skala Bedeutsamkeit (centrality)</i>					
Wichtigkeit (SAM 6)	0,85				
Auswirkung (SAM 9)	0,62				
Gravierende Folgen (SAM 13)	0,85				
Langfristige Folgen (SAM 27)	0,89				
<i>SAM-Skala Eigene Kontrolle (controllable by self)</i>					
Eigene Fähigkeiten (SAM 12)		0,80			
Eigenes Können (SAM 14)		0,72			
Problemlösekompetenz (SAM 22)		0,67			
Nötige Fertigkeiten (SAM 25)		0,75			
<i>SAM-Skala Kontrolle durch andere (controllable by others)</i>					
Sich an andere wenden (SAM 4)			0,86		
Verfügbare Hilfe (SAM 15)			0,75		
Genügend Ressourcen (SAM 17)			0,51		
Unterstützendes Umfeld (SAM 23)			0,86		
<i>SAM-Skala Unkontrollierbarkeit (uncontrollable)</i>					
Völlige Hoffnungslosigkeit (SAM 1)	0,52				
Unkontrollierbares Ende (SAM 3)				0,65	
Ohne jeglichen Einfluss (SAM 18)				0,80	
Unlösbar (SAM 21)				0,66	
<i>SAM-Skala Herausforderung (challenge)</i>					
Positive Auswirkung (SAM 7)					0,85
Ehrgeiz (SAM 8)		0,69			
Macht stärker/härtet ab (SAM 10)					0,76
Aufregung (SAM 19)	0,73				

**Tab. 1** PCA: Faktorenstruktur sowie Faktorladungen des SAM (nur Ladungen  $\geq 0,50$ ) geordnet nach ursprünglichen SAM-Skalen.

im Vordergrund steht, und zum anderen aus einer Skala zum aufgabenorientierten Coping, bei dem der Fokus auf der Bewältigung der Aufgabe liegt.

STAI: Aus dem STAI sollte in unserer Studie lediglich die Zustandsskala (State-Angst) mit dem Global Score verwendet werden. Der Global Score im STAI gibt das aktuelle Angstniveau als Mittelwert aus 20 Items an.

Messung spezifischer Emotionen: Es wurden drei numerische Ratingskalen vorgegeben, auf denen die aktuelle Intensität der Emotionen Angst, Freude und Entspannung mit einem Wert zwischen 1 (überhaupt nicht) bis 7 (sehr) angegeben werden konnte. Die drei Emotionen sollten wie auch der CISS und der STAI jeweils vor und nach dem Lesen der Stressinduktionsgeschichte bearbeitet werden.

Die Variablen wurden in folgender Reihenfolge im Fragebogen vorgegeben: Demografische Daten (Alter, Geschlecht, Zivilstand, Ausbildung), STAI, Spezifische Emotionsskalen, CISS, jeweiliges Stressinduktionszenarium, STAI, Spezifische Emotionsskalen, SAM, CISS.

### Studienteilnehmer

Nach der Genehmigung zur Durchführung der Studie durch die Ethikkommission der Universität Basel (Prüfnummer EK: 10/11 vom 11. Jan. 2011) erfolgte die Rekrutierung der Teilnehmer durch Verteilung der Fragebögen in einer Vorlesung für Jura- und Psychologiestudierende. Es wurden ca. 200 Fragebögen verteilt, von denen N = 85 (42,5 % Rücklauf) zurückgesendet wurden. Der geringe Rücklauf ist eventuell darauf zurückzuführen, dass der Fragebogen vor einer Klausur verteilt wurde und die Teilnehmer den Fragebogen mit nach Hause nehmen, in ihrer Freizeit ausfüllen und zurücksenden sollten. Insgesamt haben N = 59 Frauen und N = 26 Männer die Fragebögen bearbeitet. Das Durchschnittsalter der Frauen lag bei 22 und das der Männer bei 23 Jahren. Von den zurückgesendeten Fragebögen entfielen 50 (37 Frauen/13 Männer) auf die Variante HIV-Infektionsgeschichte und 35 Fragebögen (21 Frauen/14 Männer) auf die Variante Arbeitsplatzgeschichte als stressauslösende Bedingung. Alle Fragebögen gingen in die Auswertung ein.

### Hypothesen

1. Die Imaginationsgeschichten lösen Stress aus. Eine Überprüfung erfolgte über einen Vergleich der Vorher-Nachher-Mittelwerte im STAI.
2. Die Faktorenstruktur des deutschsprachigen SAM entspricht der des Originalinstruments. Kriterium war wie bei Peacock und Wong [6] ein Eigenwert über 1.
3. Die internen Konsistenzen sind mit denen des Originalinstruments vergleichbar.
4. Die sechs kognitiven Bewertungsskalen des SAM sind geeignet, um das Ausmaß an generellem Stressempfinden (stressfulness) aufzuklären. Die Hypothese wurde mittels Regressionsanalyse überprüft. Als Referenzwert sollten die Ergebnisse von Peacock und Wong [6] dienen, bei denen ein  $R^2 = 0,60$  mittels der Prädiktoren Bedrohung (threat) und Bedeutsamkeit (centrality) aufgeklärt werden konnte.
5. Der SAM eignet sich, um Stress differenziert zu erfassen. Da die beiden Szenarien sich gravierend in ihrer Intensität und existenziellen Bedeutung unterscheiden, würden Unterschiede im SAM für dessen Validität sprechen. Eine Überprüfung erfolgte über Mittelwertvergleiche der SAM-Skalen.
6. Der SAM verfügt über eine gute konvergente und diskriminante Validität. Aus konzeptlogischen Gründen sollten sich

positive Korrelationen mit inhaltlich nahen Skalen (z. B. STAI) sowie negative Korrelationen mit konträren Zuständen (Freude, Entspannung) in der entsprechenden Richtung ergeben.

Die entsprechenden Korrelationen sollten nach der Einteilung von Brosius mindestens im mittleren Bereich liegen (d. h.  $0,4 < r < 0,6$ ) [28].

### Statistische Auswertung

Aus Gründen der Vergleichbarkeit wurden dieselben statistischen Verfahren wie bei Peacock und Wong [6] angewendet. Alle zurückgesendeten Fragebögen wurden in die Auswertung miteinbezogen. Analyseprogramm war die „R-Statistische Rechenumgebung“ [29].

Wesentlich für die Validierung des SAM war die Überprüfung, ob die Stressszenarien überhaupt Stress auslösten. Dafür wurde im ersten Schritt der Global Score des STAI verwendet. Als statistisches Verfahren wurde eine Devianzanalyse auf der Basis eines gemischten linearen Modells gerechnet.

Im dritten Auswertungsschritt sollte die Faktorenstruktur ermittelt werden. Es wurde eine Hauptkomponentenanalyse (PCA) mit Varimax-Rotation gerechnet. Vorab wurde eine konfirmatorische Faktorenanalyse gerechnet.

In einem vierten Schritt wurden die internen Konsistenzen der sieben SAM-Skalen mittels Cronbachs  $\alpha$  bestimmt.

In einem weiteren Analyseschritt sollte in einer schrittweisen Regression das generelle Stressempfinden (Kriterium) vorhergesagt werden. Ziel war es, relevante Prädiktoren aus dem primary und secondary appraisal zur Stressentstehung zu ermitteln. Es wurde neben den Prädiktoren der Originalstudie (sechs SAM-Skalen) der Faktor Stressinduktionsgeschichte mit den Ausprägungen JOB AIDS ausgewählt. Grund für die Auswahl der Prädiktoren war die Nähe zum Modell von Lazarus. Außerdem wurde die Art des Stressors (Stressinduktionsgeschichte) als wesentlich für das Stresserleben erachtet. Als Interaktionseffekte wurden lediglich Interaktionen zwischen Stressinduktionsgeschichten und den sechs SAM-Variablen zugelassen. Vor dem Rechnen der Regression sollte eine Multikollinearitätsanalyse durchgeführt werden. Eine sukzessive Elimination von Variablen mit maximalem Variance Inflation Factor (VIF) sollte so lange erfolgen, bis der VIF-Wert für alle verbliebenen Modellvariablen  $< 7$  war [30]. Anschließend erfolgte eine weitere Variablenelimination durch Kreuzvalidierung, bis das Modell mit dem kleinsten Prädiktionsfehler gefunden war. Ziel der Regressionsanalyse ist ein möglichst knappes Modell mit einem maximalen Vorhersagewert.

Um die Sensitivität des SAM zu überprüfen, wurden mittels t-Tests die Mittelwerte der SAM-Skalen für die beiden Szenarien verglichen.

Abschließend wurde die konvergente Validität des SAM anhand der Korrelationen mit STAI, CISS und den Emotionsskalen mittels Pearson-Produkt-Moment-Korrelation ermittelt.

### Ergebnisse

#### ▼ Überprüfung der stressauslösenden Wirkung der beiden Stressszenarien

Mittels Devianzanalyse ließ sich feststellen, dass der Mittelwert im Global Score des STAI nach dem Lesen des stressauslösenden Szenariums signifikant anstieg (für das Szenarium HIV von 36,1 auf 61,1 und für das Szenarium JOB von 35,1 auf 51,2). Die Probanden beurteilten sich nach der Stressinduktion verängstigter als zuvor ( $\chi^2 = 37,1$ ,  $df = 1$ ;  $p < 0,001$ ). Außerdem wurde eine Job-

**Tab. 2** Interkorrelationen der A-priori-Skalen von Peacock und Wong.

	Herausforderung	Bedeutsamkeit	Eigene Kontrolle	Kontrolle durch andere	Unkontrollierbarkeit	Generelles Stressempfinden
Bedrohung	0,29	0,85	-0,51	0,01	0,52	0,76
Herausforderung		0,35	0,26	0,39	-0,09	0,46
Bedeutsamkeit			-0,45	0,11	0,46	0,74
Eigene Kontrolle				0,43	-0,46	-0,34
Kontrolle durch andere					-0,21	0,04
Unkontrollierbarkeit						0,41

absage als weniger angstausslösend erlebt als die Möglichkeit, sich mit HIV infiziert zu haben (Interaktion: Zeitpunkt  $\times$  Art der Stressgeschichte:  $\chi^2=6,6$ ,  $df=1$ ;  $p<0,001$ ). Das unterschiedliche Ausgangsniveau der Angst war zwar numerisch unterschiedlich, aber statistisch nicht signifikant ( $\chi^2=0,3$ ,  $df=1$ ;  $p=0,608$ ).

### Faktorenstruktur sowie Faktorladungen des SAM

Das Ergebnis der vorab gerechneten konfirmatorischen Faktorenanalyse war unbefriedigend. Faktor fünf konnte nicht repliziert werden. Die Minimum Function Test Statistic der konfirmatorischen Faktorenanalyse ergab einen Chi-Square-Wert von 536,17 ( $Df=228$ ;  $p$ -value  $<0,001$ ). Aus einer exploratorischen Hauptkomponentenanalyse (PCA) mit Varimax-Rotation resultierte eine fünffaktorielle Lösung (Kriterium war Eigenwert größer 1) (Tab. 1).

Der erste Faktor (Eigenwert: 8,14) umfasste die beiden Skalen Bedrohung (threat) und Bedeutsamkeit (centrality). Der Faktor fasste also zwei der drei Skalen zum primary appraisal zusammen. Insgesamt luden 10 Items mit ausreichend hohen Faktorladungen (zwischen 0,60 und 0,89) auf dem Faktor. Auf dem ersten Faktor luden neben den acht Items der Skalen Bedrohung (threat) und Bedeutsamkeit (centrality) außerdem das Item 1 völlige Hoffnungslosigkeit (0,52) sowie Item 19 Aufregung (0,73).

Auf dem zweiten Faktor eigene Kontrolle (controllable by self) luden alle vier Ursprungsitems sowie Item 8 Ehrgeiz (0,69) (Eigenwert von 4,26).

Der dritte Faktor (Eigenwert: 1,74) lautete Kontrolle durch andere (controllable by others), und alle vier diesbezüglichen Items luden auf diesem Faktor.

Der vierte Faktor bildete inhaltlich Unkontrollierbarkeit (uncontrollable) ab (Eigenwert: 1,45). Drei Items luden mit Faktorladungen von 0,65 bis 0,80 auf diesem Faktor. Wie bereits beschrieben, fiel das erste Item völlige Hoffnungslosigkeit auf den ersten, kombinierten Faktor Bedrohung-Bedeutsamkeit (0,52).

Der fünfte Faktor (Eigenwert von 1,13) Herausforderung (challenge) bestand in unserer Analyse lediglich aus den beiden Items 7 positive Auswirkung (0,85) und Item 10 macht stärker (0,76).

### Interne Konsistenzen der Skalen (wie von Peacock und Wong [6] festgelegt)

Als nächstes sollten die internen Konsistenzen, d. h. die Reliabilität der einzelnen Skalen, berechnet werden, wie sie von Peacock und Wong [6] mit jeweils vier Items ursprünglich festgelegt wurden. In Klammern wird der von Peacock und Wong [1] ermittelte Wert angegeben. Die Skala Bedeutsamkeit (centrality) hatte mit einem  $\alpha$ -Wert von 0,90 (0,86) die höchste interne Konsistenz. Darauf folgten eigene Kontrolle (controllable by self) mit einem  $\alpha$ -Wert von 0,86 (0,85), Bedrohung (threat) mit einem  $\alpha$ -Wert von 0,84 (0,71), Kontrolle durch andere (controllable by others) mit einem  $\alpha$ -Wert von 0,82 (0,84), Unkontrollierbarkeit (un-

controllable) mit einem  $\alpha$ -Wert von 0,73 (0,63). Die Skala Herausforderung (challenge) hatte einen  $\alpha$ -Wert von 0,57 (0,73). Generelles Stressempfinden (stressfulness) hatte eine interne Konsistenz von 0,75 (0,78).

Bildete man die Skalen nach der Vorgabe von Peacock und Wong [6], so sprachen die Interkorrelationen nicht in allen Fällen für eine relative Unabhängigkeit der Skalen. So ist zum einen die Korrelation von 0,85 zwischen den Skalen Bedrohung (threat) und Bedeutsamkeit (centrality) zu nennen, deren Items, wie bereits beschrieben, auf dem gleichen Faktor luden. Ebenso zeigten aber auch andere SAM-Skalen zum Teil mittelhohe Skaleninterkorrelationen [28]. Die maximale gemeinsame Varianz liegt allerdings bei 27%, sodass davon ausgegangen werden kann, dass die übrigen SAM-Skalen eigene Ausprägungen von Stress erfassen (Tab. 2).

### Regression des generellen Stressempfindens (stressfulness) durch die SAM-Skalen

Als nächstes wurde eine schrittweise Regression des generellen Stressempfindens (stressfulness) durch die sechs Bewertungsskalen durchgeführt. Peacock und Wong [6] trafen keine Aussagen zu Interaktionseffekten. In unserer Studie wurde dennoch eine Regression mit Berücksichtigung von Interaktionseffekten durchgeführt. Um der Gefahr zu entgehen, dass zu viele erklärende Variablen im Regressionsmodell enthalten sind, wurde eine Multikollinearitätsanalyse (Kriterium VIF-Wert  $<7$ ) durchgeführt. [30]. Es erfolgte eine weitere Variablenelimination durch Kreuzvalidierung: Alle 8 Effekte (s. o. außer Intercept) wurden in eine (nach der Größe ihres T-Werts) absteigende Rangfolge gebracht. Es wurden dann 8 genestete Modelle erstellt und Kreuzvalidiert. Das Modell mit dem kleinsten Prädiktionsfehler (0,26) wurde schließlich ausgewählt. Um dieses Modell noch weiter zu vereinfachen, wurden die drei Effekte mit dem höchsten T-Wert als Grundlage für eine weitere Vorhersage genommen. Mittels Maximum-Likelihood-Analyse konnte nun gezeigt werden, dass der Erklärungswert dieses vereinfachten Modells (mit drei signifikanten Prädiktoren) nicht signifikant von dem Modell mit sechs signifikanten Parametern verschieden war. Aufgrund dieser Analysen wurde schließlich folgendes Modell spezifiziert (Tab. 3): So wurde neben den Haupteffekten Bedrohung (threat) (Regressionskoeffizient = 0,42;  $p<0,01$ ) und Bedeutsamkeit (centrality) (Regressionskoeffizient = 0,27;  $p<0,01$ ) eine signifikante Interaktion zwischen Herausforderung (challenge) und der Stressinduktionsgeschichte (Regressionskoeffizient = 0,40;  $p<0,01$ ) gefunden. Bedrohung, Bedeutsamkeit und die Wechselwirkung Herausforderung und Stressinduktionsgeschichte sagten das Kriterium generelles Stressempfinden (stressfulness) mit einem adjusted  $R^2=0,67$  vorher ( $F_{5,78}=35,01$ ;  $p<0,001$ ).

### Überprüfung der Sensitivität des SAM

In einem weiteren Schritt wurde überprüft, ob die sieben SAM-Skalen nach dem Lesen der Stressinduktionsgeschichten unterschiedlich beantwortet wurden. Fast alle Mittelwerte der sieben Skalen für die jeweiligen Szenarien unterschieden sich signifikant voneinander, und zwar in der zu erwartenden Richtung. Die Möglichkeit, sich mit HIV infiziert zu haben, wurde als signifikant bedrohlicher, herausfordernder und bedeutsamer erlebt. Die wahrgenommene eigene Kontrolle (controllable by self) war bei einer Jobsage signifikant höher als bei der Möglichkeit, sich mit HIV infiziert zu haben. Ein Gefühl der Unkontrollierbarkeit (uncontrollable) war bei dem HIV-Szenarium signifikant stärker als bei dem Jobszenarium. Das generelle Stressempfinden (stressfulness) war bei dem HIV-Szenarium signifikant höher als bei dem Jobszenarium. Lediglich die Kontrolle durch andere (controllable by others)

unterschied sich nicht signifikant nach dem Lesen der beiden Stressinduktionsgeschichten (Tab. 4).

### Konvergente und diskriminante Validität

Abschließend sollte die konvergente sowie diskriminante Validität dargestellt werden, d. h. wie sehr der SAM mit inhaltlich nahestehenden Messverfahren korrelierte bzw. sich von ihnen unterschied.

Aus Tab. 5 ist hervorzuheben, dass die SAM-Skala generelles Stressempfinden (stressfulness) mit allen Validierungsskalen außer mit aufgabenorientiertem Coping signifikant korrelierte. Ebenso korrelierten die SAM-Skalen Bedrohung (threat), Bedeutsamkeit (centrality) sowie eigene Kontrolle (controllable by self) signifikant mit ähnlichen Validierungsskalen.

Die beiden Skalen des CISS korrelierten zum Teil mit unterschiedlichen SAM-Skalen. Das emotionsorientierte Coping des CISS korrelierte mit Bedrohung (threat), Bedeutsamkeit (centrality), generelles Stressempfinden (stressfulness), Unkontrollierbarkeit (uncontrollable) und mit negativem Vorzeichen mit eigene Kontrolle (controllable by self). Das aufgabenorientierte Coping des CISS korrelierte hingegen mit Herausforderung (challenge), eigene Kontrolle (controllable by self) und mit negativem Vorzeichen mit Unkontrollierbarkeit (uncontrollable). Die höchste Korrelation ( $r=0,59$ ) bestand zwischen Bedrohung (threat) und Angst. Die SAM-Skala Kontrolle durch andere (controllable by others) korrelierte mit keiner Validierungsskala signifikant. Die Emotionsskalen Freude und Entspannung korrelierten im Gegensatz zu STAI, CISS und Angst negativ mit den SAM-Skalen.

Tab. 3 Ergebnisse der Regressionsanalyse.

	Regressionskoeffizient	Standardfehler	T-Wert
(intercept)	1,29	0,38	3,38**
Bedrohung (threat)	0,42	0,11	3,69**
Bedeutsamkeit (centrality)	0,27	0,10	2,78**
Stressinduktionsgeschichte	0,37	0,21	1,83
Herausforderung (challenge)	0,08	0,11	0,73
Interaktion: Stressinduktionsgeschichte × Herausforderung (challenge)	0,40	0,15	2,61*

\* $p < 0,05$ , \*\*  $p < 0,01$ ,  
Adjusted R-squared: 0,67

Tab. 4 Mittelwerte und Streuung der SAM-Skalen für beide Stressinduktionsgeschichten: HIV, JOB.

	Mittelwert		Streuung		Signifikanz
	HIV	JOB	HIV	JOB	Vergleich der Mittelwerte HIV und JOB
Bedrohung (threat)	3,82	2,0	0,58	0,60	$t = 13,77$ ; $df = 69,36$ ; $p < 0,001$
Herausforderung (challenge)	3,65	3,06	0,65	0,80	$t = 3,54$ ; $df = 60,95$ ; $p < 0,001$
Bedeutsamkeit (centrality)	4,30	2,53	0,63	0,79	$t = 10,86$ ; $df = 60,68$ ; $p < 0,001$
Eigene Kontrolle (controllable by self)	3,54	4,16	0,73	0,80	$t = -3,65$ ; $df = 66,10$ ; $p < 0,001$
Kontrolle durch andere (controllable by others)	4,22	3,91	0,65	0,97	$t = 1,63$ ; $df = 52,63$ ; $p < 0,11$
Unkontrollierbarkeit (uncontrollable)	2,72	1,88	0,90	0,78	$t = 4,56$ ; $df = 77,22$ ; $p < 0,001$
Generelles Stressempfinden (stressfulness)	4,06	3,01	0,60	0,75	$t = 6,78$ ; $df = 62,53$ ; $p < 0,001$

Tab. 5 Korrelationen der SAM-Skalen mit STAI, CISS sowie den spezifischen Emotionsskalen „Angst“, „Freude“ und „Entspannung“.

	STAI Global Score	CISS emotionsorientiertes Coping	CISS aufgabenorientiertes Coping	Spezifische Emotionsskalen		
				Angst	Freude	Entspannung
Bedrohung (threat)	0,45**	0,53**	-0,10	0,59**	-0,11	-0,25*
Herausforderung (challenge)	0,22*	0,04	0,39**	0,30**	-0,02	-0,12
Bedeutsamkeit (centrality)	0,37**	0,53**	-0,13	0,49**	-0,13	-0,18
Eigene Kontrolle (controllable by self)	-0,23*	-0,44**	0,42**	-0,30**	0,22*	0,11
Kontrolle durch andere (controllable by others)	0,06	0,00	0,14	0,10	0,10	-0,04
Unkontrollierbarkeit (uncontrollable)	0,13	0,34**	-0,37**	0,19	0,01	-0,09
Generelles Stressempfinden (stressfulness)	0,51**	0,51**	-0,03	0,52**	-0,26**	-0,41**

\*  $p < 0,05$ , \*\*  $p < 0,01$ .

## Diskussion



Zunächst konnte gezeigt werden, dass die eingesetzten Stressinduktionsszenarien ihre Wirkung erzielten. Mittels Global Score des STAI (als bereits existierendem, validem Messinstrument) konnte bestätigt werden, dass beide Induktionsgeschichten (Absage eines Jobs bzw. die Möglichkeit, sich mit HIV infiziert zu haben) Angst bzw. Stress auslösten. Der Global Score des STAI war nach dem Lesen der beiden Szenarien signifikant höher als zuvor. Außerdem unterschieden sich die Werte des Global Score abhängig von der Art der Stressinduktionsgeschichte. Wie erwartet, war die Möglichkeit, sich mit HIV infiziert zu haben, angstauslösender als die Absage eines Zusatzjobs.

Zusammenfassend bleibt festzuhalten, dass das Lesen der Szenarien (und die damit verbundene Imagination) Stress beim Leser auslöste, was Hypothese 1 bestätigte. Die Szenarien konnten somit als Stressinduktor für die Validierung des SAM genutzt werden, um eine wirksame und akute Stresssituation zu erzeugen. Es sei angemerkt, dass Stress in diesem Fall durch Imagination nach dem Lesen einer stressauslösenden Geschichte ausgelöst wurde. Der SAM soll später aber auch zur Messung von realem Stress eingesetzt werden.

In einem weiteren Analyseschritt wurde zunächst eine konfirmatorische Faktorenanalyse gerechnet analog zu der Itemzuordnung und Faktoranzahl wie bei Peacock und Wong [6]. Allerdings war das Ergebnis unbefriedigend. In der anschließend durchgeführten Hauptkomponentenanalyse (PCA) wurde eine fünffaktorielle Lösung gefunden. Auf einem ersten Faktor luden alle Items der Skalen Bedrohung (threat) und Bedeutsamkeit (centrality) sowie zwei weitere Items. Dieser erste Faktor hatte den höchsten Eigenwert und verdeutlichte, in welchem hohem Ausmaß die beiden Aspekte Bedrohung und individuelle Bedeutsamkeit bei Stresserleben miteinander verknüpft sind.

Unser zweiter, dritter und vierter Faktor spiegelte genau wie bei Peacock und Wong [6] die theoriegeleitete Annahme von Lazarus und Folkman [7] wider, dass das secondary appraisal (sekundäre Bewertung) aus drei verschiedenen Dimensionen besteht. Wie angestrebt, konnten theoriekonform drei Faktoren gefunden werden, die die Skalen eigene Kontrolle (controllable by self), Kontrolle durch andere (controllable by others) und Unkontrollierbarkeit (uncontrollable) abbildeten.

Der fünfte Faktor beschrieb wie bei Peacock und Wong [6] den Herausforderungscharakter einer bedrohlichen Situation. Allerdings war dieser fünfte Faktor Herausforderung (challenge) lediglich mit dem Item 7 positive Auswirkung zu unserem Faktor deckungsgleich. Dieser fünfte Faktor blieb aber inhaltlich der Faktor Herausforderung (challenge), der einen Glauben an einen positiven Einfluss und an einen Wachstumsprozess beinhaltet. Die Existenz dieses Faktors belegte zumindest eingeschränkt die theoretische Annahme von Lazarus und Folkman [7], dass sich das primary appraisal aus unterschiedlichen Inhalten zusammensetzt.

Insgesamt zeigte unsere Faktorenstruktur lediglich leichte Abweichungen zu den Ergebnissen von Peacock und Wong [6]. So fanden Peacock und Wong [6] ebenfalls einen Faktor, der die Skalen Bedrohung (threat) und Bedeutsamkeit (centrality) beinhaltete. Außerdem wiesen in unserer Analyse einzelne Items eine höhere semantische Nähe zu anderen Faktoren auf als bei Peacock und Wong [6]. Es bleibt festzuhalten, dass die von Peacock und Wong [6] identifizierten Faktoren im Wesentlichen repliziert werden konnten. Dies spricht für Hypothese 2. Es sei darauf hingewiesen, dass die Faktorenstruktur des SAM sich sowohl in den Studien von Peacock und Wong [6] als auch in anderen Studien

als wenig stabil erwies [32, 33]. So zeigte sich in einer Studie von Peacock und Wong [6] wie bereits beschrieben eine fünf- und in einer weiteren eine sechsfaktorielle Lösung. In den erwähnten anderen Studien wurde sogar nur eine vier- [32] bzw. dreifaktorielle [33] Struktur gefunden. In einer Studie von Roesch [32] waren Bedrohung (threat), Herausforderung (challenge) und Bedeutsamkeit (centrality) getrennte Faktoren. In einer Studie von Rowley [33] luden Bedrohung (threat) und Bedeutsamkeit (centrality) wie in unserer Studie auf einem Faktor. Dritter bzw. vierter Faktor war jeweils Ressourcen, auf dem alle Variablen des secondary appraisal zusammengefasst wurden [32, 33].

Übereinstimmend mit der Studie von Peacock und Wong [6] konnten auch in unserer Studie nicht alle Differenzierungen des Transaktionalen Stressmodells von Lazarus und Folkman [7] (drei getrennte Bewertungsskalen im primary appraisal) abgebildet werden. Dies muss allerdings nicht als Beleg gegen das Modell von Lazarus und Folkman [7] gewertet werden, sondern könnte auch darauf zurückzuführen sein, dass zur Erfassung dieser konzeptuellen Differenzierungen zu wenige Items formuliert wurden oder Items, die den Inhalt der Konzepte nicht trennscharf genug treffen.

Die von den Autoren der Originalversion gefundenen internen Konsistenzen, die die Reliabilität der sieben Skalen beschrieb, konnten mit ausreichender Höhe in der übersetzten Version bestätigt werden. Die internen Konsistenzen lagen bei sechs der sieben Skalen zwischen 0,73 und 0,89 und entsprachen annähernd den Ergebnissen der Ursprungsautoren. Lediglich die Skala Herausforderung (challenge) schnitt mit einem Alphaswert von 0,57 schlechter ab (Studie von Peacock und Wong [6]: 0,73), sicherlich entspricht der Wert nicht gängigen Qualitätsmerkmalen für eine gute Reliabilität [30]. Bis auf den nicht befriedigenden Wert sprachen die Ergebnisse für eine weitgehende Vergleichbarkeit zu den Ergebnissen von Peacock und Wong [6] und bestätigten Hypothese 3.

Mittels schrittweiser Regressionsanalyse konnte gezeigt werden, dass sich durch die sechs SAM-Skalen 67% der Varianz des generellen Stressempfindens (stressfulness) aufklären ließen, was als hoch einzustufen ist [34]. Bedrohung (threat), Bedeutsamkeit (centrality) und eine Wechselwirkung zwischen dem zuvor gelesenen Szenarium und Herausforderung (challenge) waren signifikante Prädiktoren. Diese Ergebnisse stehen im Einklang mit Befunden aus der Depressionsforschung, bei denen die Einordnung eines Ereignisses als bedeutsam einen wesentlichen Einfluss auf dessen Verarbeitung hat [35]. Die Ergebnisse der Regression zeigten weiter, dass Kontrollattributionen im Sinne der sekundären Bewertung (secondary appraisal) wie bei Peacock und Wong [6] keine signifikanten Prädiktoren waren. Es ist somit inhaltlich festzuhalten, dass Stresserleben zwar aus verschiedenen Facetten bestand, dass aber die Kontrollattributionen gemeinsam mit den primären Bewertungsfaktoren keinen eigenen Beitrag bei der Regression zum Stressempfinden lieferten, was in einer ähnlich angelegten Studie auch so gefunden wurde [36]. Die von Peacock und Wong [6] aufgeklärte Varianz lag bei 60%. Die Vorhersage wurde durch unsere Studie übertroffen, was Hypothese 4 bestätigte. Die signifikante Interaktion zwischen dem zuvor gelesenen Szenarium und Herausforderung (challenge) bedeutete, dass unter der Bedingung JOB die Skala Herausforderung (challenge) das generelle Stressempfinden (stressfulness) zusätzlich erklären konnte, nicht jedoch unter der Bedingung HIV. Das erlebte Stressempfinden wurde also dadurch beeinflusst, ob ein Szenarium als herausfordernd erlebt wurde oder nicht. Die signifikante Interaktion zeigte, dass Stress-



empfinden und Kontrollüberzeugungen situationsabhängig waren und dass unterschiedliche Bewertungsprozesse stattfanden.

Die durchgeführten Mittelwertvergleiche der SAM-Skalen zu den beiden Stressinduktionsszenarien belegten die Validität des SAM. Alle Mittelwertvergleiche waren bis auf Kontrolle durch andere (controllable by others) signifikant. So wurde die Möglichkeit, sich mit HIV infiziert zu haben, als insgesamt stressauslösender, bedeutsamer, bedrohlicher und herausfordernder erlebt als ein verpasster Nebenjob. Dieser Analyseschritt wurde von Peacock und Wong [6] nicht durchgeführt. Hypothese 5 wurde bestätigt.

Die Korrelationen der eingesetzten Validierungsskalen mit den SAM-Skalen zeigten mittelhohe signifikante Korrelationen [28] zwischen dem Global Score des STAI und den SAM-Skalen generelles Stressempfinden (stressfulness) und Bedrohung (threat) sowie annähernd mittelhoch zu Bedeutsamkeit (centrality), was auf eine inhaltliche Nähe der beiden Messinstrumente hindeutet und der theoretischen Annahme entsprach. Eine zwar eher schwächere [28], aber dennoch signifikante negative Korrelation des Global Score des STAI mit eigener Kontrolle (controllable by self) entsprach der Erwartung, dass Angst entsteht, wenn keine eigenen Einflussmöglichkeiten verspürt werden. Es gab keine signifikanten Korrelationen mit Kontrolle durch andere (controllable by others) und Unkontrollierbarkeit (uncontrollable), was sich inhaltlich schwieriger interpretieren ließ. Eventuell wurden die beiden Skalen unterschiedlich wahrgenommen: Kontrolle durch andere (controllable by others) könnte sowohl psychisch entlastend wirken als auch bedrohlich, wenn auf die eigene Autonomie verzichtet werden muss. Andere Studienergebnisse bestätigen, dass die Unterstützung von anderen nicht immer als entlastend erlebt wird [37, 38]. Eine andere Interpretation wäre, dass Kontrolle durch andere (controllable by others) völlig emotionsfrei erlebt wird, was die Autoren aber für eher unwahrscheinlich halten. Ebenso könnte Unkontrollierbarkeit (uncontrollable) unterschiedlich wahrgenommen werden. Es könnte einerseits entlastend sein (da niemand etwas ändern kann bzw. sich niemand anstrengen muss), andererseits auch wiederum bedrohlich, da man sich hilflos und ausgeliefert fühlt [39, 40].

Die Skala emotionsorientiertes Coping des CISS misst stressauslösende Gefühle. Sie korrelierte signifikant auf einem oberen mittleren Niveau [28] mit den SAM-Skalen Bedrohung (threat), Bedeutsamkeit (centrality) und mit generellem Stressempfinden (stressfulness). Die durch den CISS gemessenen stressbezogenen Gefühle (z. B. Anspannung, Befürchtung, Selbstvorwürfe) zeigten somit eine hohe inhaltliche Nähe zu den Primary-appraisal-Skalen, die eher das negative Ausmaß einer Stresssituation erfassen, was für eine gute konvergente Validität des SAM sprach. Entsprechend konnte die diskriminante Validität dadurch bestätigt werden, dass keine signifikante Korrelation zwischen emotionsorientiertem Coping und Herausforderung (challenge) existierte. Herausforderung (challenge) misst eher die positiven Anteile eines Stressgeschehens und passt inhaltlich nicht zu Anspannung, Befürchtung oder Selbstvorwürfen (wie beim emotionsorientierten Coping gemessen). Die Richtungen und die Höhe der Korrelationswerte des emotionsorientierten Copings mit den SAM-Skalen eigene Kontrolle (controllable by self) sowie Unkontrollierbarkeit (uncontrollable) waren ebenfalls nachvollziehbar und ausreichend hoch (annähernd mittelhoch bzw. mittelhoch).

Im Gegensatz zum emotionsorientierten Coping misst das aufgabenorientierte Coping des CISS die kognitive Einschätzung, ob eine Aufgabe bewältigt werden kann. Es ergaben sich nachvollziehbare signifikante mittelhohe Korrelationen [28] mit den SAM-Skalen Herausforderung (challenge), eigene Kontrolle (controlla-

ble by self), Unkontrollierbarkeit (uncontrollable) mit negativem Vorzeichen. Die kognitive, d. h. reflektierte Einschätzung, ob eine Situation gemeistert werden kann (aufgabenorientiertes Coping), ist von den eigenen Kontrollmöglichkeiten abhängig. Beispielsweise ist die negative, näherungsweise mittelhohe Korrelation [28] zwischen aufgabenorientiertem Coping und Unkontrollierbarkeit (uncontrollable) sinnvoll: Je mehr man sich mit der Lösung des Problems beschäftigt, umso kontrollierbarer erscheint die Situation. Die signifikante mittelhohe Korrelation [28] zwischen aufgabenorientiertem Coping und Herausforderung (challenge) sprach ebenfalls für eine gute konvergente Validität: Wer sich Pläne zur Lösung des Stressgeschehens macht, empfindet eine Situation eher als Herausforderung. Insgesamt belegten die Zusammenhänge eine inhaltliche Nähe zwischen CISS und SAM.

Die spezifische Emotionsskala Angst als weitere Validierungsskala misst das momentane Angstgefühl. Hier ergaben sich ebenfalls mittelhohe [28], signifikante und nachvollziehbare Zusammenhänge zu den SAM-Skalen Bedrohung (threat), generelles Stressempfinden (stressfulness) und Bedeutsamkeit (centrality) sowie eine eher schwächere, aber immer noch signifikante Korrelation [28] mit eigener Kontrolle (controllable by self). Interessanterweise fand sich auch ein positiver, wenn auch eher schwächerer Zusammenhang [28] zwischen Herausforderung (challenge) und erlebter Angst. Offensichtlich kann auch die Bereitschaft, ein Problem anpacken zu wollen, mit Angst einhergehen. Dies ergibt inhaltlich Sinn: Trotz positiven Entschlusses (eine Situation als Herausforderung zu betrachten) bleibt der Ausgang der eigenen Bemühungen ungewiss. Dieses Ergebnis wird durch die Untersuchung von Besser [41] bestätigt. Die Einschätzung eines Ereignisses als herausfordernd kann stress- und angstauslösend wirken. Die Skala Freude korrelierte mit eigener Kontrolle (controllable by self) und negativ mit generellem Stressempfinden (stressfulness). Wie zu erwarten, kann ein empfundenes Kontrollgefühl mit positiver Stimmung und Zuversicht einhergehen. Die negative, eher schwächere Korrelation [28] von Freude und generellem Stressempfinden ist plausibel: Stress und Freude können nicht nebeneinander existieren. Ebenso lassen sich die signifikanten Korrelationen von Entspannung mit den einzelnen SAM-Skalen erklären. Entspannung korrelierte mit Bedrohung (threat) sowie mittelhoch mit generellem Stressempfinden (stressfulness). Wer nach dem Lesen der Geschichte entspannt bleibt, wird weder Bedrohung noch Anspannung empfinden. Die Höhe der Korrelationen mit den beiden Skalen Freude und Entspannung ist nicht so stark ausgeprägt wie in Hypothese 6 gefordert. Dennoch belegen die schwachen bis mittelhohen Korrelationswerte eine tendenziell gute diskriminante Validität. Freude und Entspannung entstehen nur in Abwesenheit von akutem Stress.

In der Gesamtschau werteten die Autoren Hypothese 6 aufgrund der Höhe und der Richtung der gefundenen Korrelationen weitestgehend als bestätigt. SAM korrelierte signifikant positiv und in den meisten Fällen in ausreichender Höhe mit inhaltlich nahen Konstrukten (z. B. STAI) und korrelierte negativ (wenn auch nicht immer in der geforderten Höhe) mit Maßen zur Bestimmung der Divergenz (Emotionsskala Freude bzw. Entspannung).

In der Gesamtschau aller Ergebnisse sollten sowohl die Ergebnisse von Peacock und Wong [6] als auch die Resultate von Roesch [32] und Rowley [33] sowie unsere Ergebnisse zum Anlass genommen werden, den theoretischen Bezug zum Transaktionalen Stressmodell von Lazarus und Folkman [6] zu hinterfragen. Darin wurde ursprünglich angenommen, dass Bedeutsamkeit (centrality) orthogonal zu Herausforderung (challenge) und Bedrohung (threat) angeordnet ist. Es wurde postuliert, dass eine Bedrohung (threat)

ebenso wie eine Herausforderung (challenge) nur dann Beachtung findet, wenn sie als bedeutsam (zentral) für das eigene Selbstkonzept (im Sinne von „Gefahr“ bzw. „selbstwertsteigernd“) wahrgenommen wird. Verschiedene Studienergebnisse (inklusive der Originalarbeit und unserer eigenen Studie) legen nun aber nahe, dass zumindest Bedrohung (threat) und Bedeutsamkeit (centrality) auf einem Faktor laden und somit nicht unabhängig sind. So sind Bedrohung (threat) und Bedeutsamkeit (centrality) enger miteinander verbunden als Herausforderung (challenge) und Bedeutsamkeit (centrality), bei denen die Orthogonalität bestehen bleibt. Dies würde eine Abkehr vom ursprünglichen Modell bedeuten. Eine Interpretation könnte etwa lauten, dass gefährvolle Ereignisse wesentlicher und zentraler sind als herausfordernde Ereignisse. In einer weiteren Studie [10] blieb die Annahme der Orthogonalität von Bedeutsamkeit (centrality) zu den beiden anderen Faktoren hingegen aufrechterhalten, obwohl Bedeutsamkeit (centrality) sowohl mit Bedrohung (threat) als auch mit Herausforderung (challenge), wenn auch nicht so hoch wie in unserer Studie, aber immer noch signifikant ( $r > 0,40$ ) korrelierte. Hier lag die höhere Korrelation sogar bei Herausforderung (challenge) und Bedeutsamkeit (centrality) ( $r = 0,49$ ). Eine weitere Unterstützung für die Beibehaltung der drei Faktoren zur Erfassung des primary appraisal liefern die Ergebnisse von Durak [42]. In einer Vergleichsstudie mit verschiedenen Instrumenten zur Erfassung von Stress und Coping lieferten die drei Variablen zum primary appraisal jeweils unterschiedliche Beiträge in Regressionsmodellen zur Vorhersage von anderen stressauslösenden bzw. stressverarbeitenden Variablen. Hingegen bestärkt dieselbe Studie [42] aber ebenso unsere Ergebnisse, indem zwischen Bedeutsamkeit (centrality) und Bedrohung (threat) eine hohe Korrelation gefunden wurde, die im Übrigen fast doppelt so hoch war wie der Zusammenhang zwischen Bedeutsamkeit (centrality) und Herausforderung (challenge) ( $r = 0,58$  bzw.  $r = 0,32$ ). Die zum Teil gegensätzliche Ergebnislage führt dazu, dass eine letztliche Entscheidung über die Aufrechterhaltung bzw. Verwerfung der Theorie von Lazarus und Folkman [7] offen bleiben und in weiteren Studien geklärt werden muss. Obwohl die Annahmen von Lazarus und Folkman [7] nicht vollkommen bestätigt werden konnten, belegen sowohl unsere statistischen Auswertungen als auch die zunehmende Verbreitung des SAM dessen Wert. Sicherlich wäre eine weitere Verkürzung oder Veränderung von Items zu überdenken, da gegenwärtig vor allem die Items zu Bedrohung (threat) und Bedeutsamkeit (centrality) Ähnliches messen. Trotz dieser Mängel sind die Autoren in der Gesamtschau vom sinnhaften Einsatz des SAM überzeugt. Aufgrund der eingeschränkten Repräsentativität der Stichprobe (Studenten) wäre als nächster Schritt zur Etablierung des SAM anzustreben, die Teilnehmerzahl zu erhöhen und ihn verschiedenen Populationen und Patientengruppen (z. B. Menschen mit Burnout-Symptomen, Tinnituspatienten oder Teilnehmern eines Meditationskurses etc.) zur weiteren Validierung vorzulegen. Schließlich könnte eine Normierung zur Aufstellung von Prozenträngen oder zu einer Stanineverteilung führen, um eine Einordnung hinsichtlich Art und Intensität des Stressempfindens vornehmen zu können.

### Ergänzendes Material

▼  
Anhang 1: Deutsche Übersetzung des SAM, Anhang 2: Stressinduktionsszenarium JOB und Anhang 3: Stressinduktionsszenarium HIV finden Sie online unter <http://dx.doi.org/10.1055/s-0034-1399727>

### Take Home Message

Die Erfassung von situationsabhängigem Stress kann in den verschiedensten Feldern von Grundlagenforschung und psychologischen Anwendungsfeldern (z. B. im klinischen, gesundheitspsychologischen und sportpsychologischen Bereich) wertvoll sein. Ziel der vorliegenden Arbeit war es, ein kurzes, aber dennoch ausreichend reliables und valides Erhebungsinstrument für akutes Stresserleben in den deutschen Sprachraum einzuführen. In unserer Untersuchung wurden neben der Übersetzung die Faktorenstruktur, die internen Konsistenzen sowie die konvergente Validität des SAM in deutscher Sprache in weiten Teilen nachgewiesen, sodass einem legitimen Einsatz des ins Deutsche übersetzten SAM grundsätzlich nichts mehr im Wege steht. Allerdings konnte besonders die Faktorenstruktur des fünften Faktors nur teilweise repliziert werden. Um die Ergebnisse zu erhärten, sollten noch weitere Validierungsstudien u. a. an verschiedenen klinischen Stichproben durchgeführt werden.

**Interessenkonflikt:** Die Autoren geben an, dass kein Interessenkonflikt besteht.

### Literatur

- 1 Kälén W. Deutsche Übersetzung des COPE von Carver, Scheier und Weintraub 1989. Institut für Psychologie: Universität Bern; 1994
- 2 Gaab J. PASA – Primary Appraisal Secondary Appraisal: ein Fragebogen zur Erfassung von situationsbezogenen kognitiven Bewertungen. Verhaltenstherapie 2009; 19: 114 – 115
- 3 Jack M. Fragebogen zur Erfassung von Ressourcen und Selbstmanagementfähigkeiten (FERUS). Göttingen: Hogrefe; 2007
- 4 Janke W, Erdmann G. Der Stressverarbeitungsfragebogen (SVF 120). Kurzbeschreibung und grundlegende Kennwerte. Hogrefe: Göttingen; 1997
- 5 Schulz P, Schlotz W, Becker P. Trierer Inventar zum chronischen Stress (TICS): TICS; Manual. Göttingen: Hogrefe; 2004
- 6 Peacock EJ, Wong PTP. The Stress Appraisal Measure (SAM) – a Multidimensional Approach to Cognitive Appraisal. Stress Medicine 1990; 6: 227 – 236
- 7 Lazarus RS, Folkman S. Stress, appraisal, and coping. New York: Springer; 1984
- 8 Kälén W. Deutsche 24-Item Kurzform des „Coping Inventory for Stressful Situations (CISS)“ von N. S. Endler & J. D. A. Parker basierend auf der Übersetzung von N. Semmer, F. Tschan & V. Schade (unveröffentlichter Fragebogen). Universität Bern: Institut für Psychologie; 1995
- 9 Meichenbaum D. Kognitive Verhaltensmodifikation. Weinheim: Beltz Psychologie Verlags Union; 1995
- 10 Levy A, Nicholls A, Polman R. Cognitive Appraisals in Sport: The Direct and Moderating Role of Mental Toughness. International Journal of Applied Psychology 2012; 2: 71 – 76
- 11 Antonovsky A. Health, stress, and coping: [new perspectives on mental and physical well-being]. The Jossey-Bass social and behavioral science series. San Francisco: Jossey-Bass; 1979
- 12 Antonovsky A. Unraveling the mystery of health: how people manage stress and stay well. The Jossey-Bass health series. San Francisco: Jossey-Bass; 1987
- 13 Kanfer FH, Reinecker H, Schmelzer D. Selbstmanagement-Therapie: Ein Lehrbuch für die klinische Praxis. 2. überarb. Aufl. Berlin: Springer; 1996
- 14 Bandura A. Self-efficacy: the exercise of control. New York: W. H. Freeman; 1997
- 15 Beck AT, Hautzinger M. Kognitive Therapie der Depression. 2. veränd. Aufl. Weinheim: Psychologie Verlags Union; 1986
- 16 Sommer G, Fydrich T. Soziale Unterstützung: Diagnostik, Konzepte, F-SOZU. Materialien. Deutsche Gesellschaft für Verhaltenstherapie. Tübingen: Deutsche Gesellschaft für Verhaltenstherapie; 1989

- 17 Buchwald P, Schwarzer C, Hobfoll SE. Stress gemeinsam bewältigen: Ressourcenmanagement und multiaxiales Coping. Hogrefe: Göttingen; 2004
- 18 Carver CS, Scheier MF. Attention and self-regulation: a control-theory approach to human behavior. Springer series in social psychology. Berlin: Springer; 1981
- 19 Graungaard AH, Andersen JS, Skov L. When resources get sparse: a longitudinal, qualitative study of emotions, coping and resource-creation when parenting a young child with severe disabilities. *Health* 2011; 15: 115–136
- 20 Ntoumanis N, Edmunds J, Duda JL. Understanding the coping process from a self-determination theory perspective. *Br J Health Psychol* 2009; 14: 249–260
- 21 Meichenbaum D. Kognitive Verhaltensmodifikation. München: U und S Psychologie; 1979
- 22 Garcia-Campayo J. Validation of the Spanish versions of the long (26 items) and short (12 items) forms of the Self-Compassion Scale (SCS). *Health and Quality of Life Outcome* 2014; 12
- 23 Kemper C, Lutz J, Neuser J. Konstruktion und Validierung einer Kurzform der Skala Angst vor negativer Bewertung (SANB-5). *Klinische Diagnostik und Evaluation* 2012; 4: 343–360
- 24 Wirtz PH, Thomas L, Domes G et al. Psychoendocrine validation of a short measure for assessment of perceived stress management skills in different non-clinical populations. *Psychoneuroendocrinology* 2013; 38: 572–586
- 25 Laux L. State-Trait-Angstinventar: theoretische Grundlagen und Handanweisung. Beltz Test. Testgesellschaft. 1 Testmappe. Weinheim: Beltz; 1981
- 26 Kelava A, Schermelleh-Engel K. Latent-State-Trait-Theorie (LST-Theorie). In: Moosbrugger H, Kelava A (ed) *Testtheorie und Fragebogenkonstruktion*. Berlin: Springer; 2008: 343–360
- 27 Association, A. A.P. APA: American Psychological Association. Available from: <http://www.apa.org/helpcenter/stress-kinds.aspx>
- 28 Brosius F. SPSS 16: das mitp-Standardwerk. Heidelberg: Redline; 2008
- 29 Team RC. A language and environment for statistical computing; 2013
- 30 Brosius F. SPSS 16: das mitp-Standardwerk. Redline: Heidelberg; 2008
- 31 Craney TA, Surlles JG. Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering* 2002; 14: 391–403
- 32 Roesch SC, Rowley AA. Evaluating and developing a multidimensional, dispositional measure of appraisal. *J Pers Assess* 2005; 85: 188–196
- 33 Rowley AA, Roesch SC, Jurica BJ et al. Developing and validating a stress appraisal measure for minority adolescents. *J Adolesc* 2005; 28: 547–557
- 34 Hattie J. Fig. 2.4p. 19. Visible learning a synthesis of over 800 meta-analyses relating to achievement. London: Routledge; 2009
- 35 Newby JM, Moulds ML. Intrusive memories of negative events in depression: Is the centrality of the event important? *Journal of Behavior Therapy and Experimental Psychiatry* 2011; 42: 277–283
- 36 Anshel MH, Robertson M, Caputi P. Sources of acute stress and their appraisals and reappraisals among Australian police as a function of previous experience. *Journal of Occupational and Organizational Psychology* 1997; 70: 337–356
- 37 Chesler MA, Barbarin OA. Difficulties of Providing Help in a Crisis – Relationships between Parents of Children with Cancer and Their Friends. *Journal of Social Issues* 1984; 40: 113–134
- 38 La Greca AM. Stress and coping in child health. *Advances in pediatric psychology*. New York: The Guilford Press; 1992
- 39 Reichman SRF, Miller AC, Gordon RM et al. Stress appraisal and coping in mothers of NICU infants. *Children's Health Care* 2000; 29: 279–293
- 40 Newby JM, Moulds ML. Characteristics of intrusive memories in a community sample of depressed, recovered depressed and never-depressed individuals. *Behaviour Research and Therapy* 2011; 49: 234–243
- 41 Besser A, Gordon I, Flett L et al. Perfectionism, Cognition, and Affect in Response to Performance. Failure vs. success. *Journal of Rational-Emotive & Cognitive-Behavior Therapy* 2004; 22: 301–328
- 42 Durak M. The Relationship between Cognitive Appraisal of Stress, Coping Strategies and Psychological Distress among Correctional Officers: Personal and Environmental Factors. *Middle East Technical University*; 2007