Diagnostik der Beschwerdenvalidität von psychischen Störungen in der sozialmedizinischen Begutachtung

Kumulative Dissertation

zur Erlangung der Doktorwürde

durch den Promotionsausschuss Dr. rer. nat.

am Fachbereich 11 Gesundheits- und Humanwissenschaften

der Universität Bremen

vorgelegt von

Dipl. Psych. Franziska Walter

Gutachterin: PD Dr. Monika Daseking

Gutachter: Prof. Dr. Dr. Uwe Koch

Danksagung

Zunächst möchte ich mich besonders bei meinem Doktorvater Prof. Dr. Franz Petermann bedanken, der mir die Chance gab, promovieren zu können. Das uneingeschränkte Engagement durch anregende Diskussionen und eine stets offene Tür haben diese Arbeit erst möglich gemacht.

In gleicher Weise gilt auch PD Dr. Axel Kobelt mein aufrichtiger Dank. Die ansteckende Begeisterung für die Wissenschaft, die ununterbrochene Unterstützung in Form von fachlichem Austausch sowie die stetige Wertschätzung meiner Arbeit waren für meine Dissertation unersetzlich.

Ebenso möchte ich mich herzlich bei PD Dr. Monika Daseking für ihren enthusiastischen Einsatz im Doktorandenkolleg und die vielen fachlichen Inspirationen bedanken.

Besonders danken möchte ich meinem Gutachter, Prof. Dr. Dr. Uwe Koch, für die Bereitschaft diese Arbeit zu bewerten.

Herzlich bedanken möchte ich mich beim gesamten Doktorandenkolleg und meinen Kollegen im ZKPR für jedwede Unterstützungen in Form von gemeinsamen Diskussionen und fachbezogenen Gesprächen.

Ein großer Dank gebührt meinen Kolleginnen Franziska Korsch und Dr. Ulrike de Vries, die mir vor allem in der letzten Phase ein besonderer Halt waren. Sie haben mit mir gelacht, gelitten und mich wieder aufgebaut.

Mein allerherzlichster Dank gilt meinen Eltern Christel und Detlef Walter, die mich immer uneingeschränkt unterstützen. Ich weiß, dass sie immer stolz auf mich sein werden und das unabhängig von erbrachten Erfolgen und Leistungen. Was wäre ich nur ohne euch!

Weiterhin möchte ich mich ganz besonders bei Juliane Walter bedanken. Ihre schwesterliche Freundschaft bedeutet mir unendlich viel. Danke für deinen Beistand, den du jederzeit bereit bist zu geben.

Mein allergrößter Dank gilt Christian Tiedt, der mich immer uneingeschränkt mit all seinen liebevollen, aufmunternden Worten und Gesten unterstützt. Danke für dein grenzenloses Verständnis und dein bedingungslos offenes Ohr.

Zusammenfassung

Fragestellung: Im Rahmen der vorliegenden Dissertation soll die Notwendigkeit der Beschwerdenvalidierung in der sozialmedizinischen Begutachtung überprüft sowie ein Screening für die Erfassung von negativen Antwortverzerrungen bei psychischen Störungen konzipiert und validiert werden. Darüber hinaus werden die Einflüsse von Instruktionen, die vor dem Ausfüllen eines Fragebogens gegeben werden und die Motivation zum Rentenerhalt bei Antragstellern einer Rente wegen verminderter Erwerbsfähigkeit untersucht.

Methode: Studie I ermittelt auf der Grundlage einer Literaturrecherche deutschsprachige Verfahren zur Beschwerdenvalidierung und prüft deren Eignung zur Erfassung von negativen Antwortverzerrungen bei psychischen Störungen. Ein Analogstudiendesign wurde genutzt, um etwaige Unterschiede in der Beschwerdendarstellung von instruierten Simulanten, einer Kontrollgruppe sowie depressiven Patienten und Schmerzpatienten mit einer depressiven Symptomatik in klinischen Symptomskalen zu ermitteln (vgl. Studie II). Auf der Basis dieses Designs wurde anschließend ein Beschwerdenvalidierungstest konstruiert und deren Güte bezüglich der Erfassung von negativen Antwortverzerrungen bei psychischen Störungen überprüft (vgl. Studie IV). Mithilfe eines randomisierten Kontrollgruppendesigns wurde bei Antragstellern einer Rente wegen verminderter Erwerbsfähigkeit aufgrund einer psychischen Störung untersucht, ob Hinweise und die Motivation zum Erhalt einer Rente zu einer verzerrten Beschwerdendarstellung führen können (vgl. Studie III).

Ergebnisse: Instruierte Simulanten zeigen in Symptomskalen Extremantworten und Übergeneralisierungen. Auf der Grundlage dieser Erkenntnise wurde ein neuer Beschwerdenvalidierungstest konzipiert, der hinreichend gute psychometrische Kennwerte aufweist. Die Analysen zeigen, dass Instruktionen zu einer verzerrten Beschwerdendarstellung führen. Darüber hinaus neigen Antragsteller zu negativen Antwortverzerrungen, wenn sie mit dem Rentenstatus die Chance, wieder gesund zu werden, verbinden.

Schlussfolgerung: Obwohl der entwickelte Beschwerdenvalidierungstest zufriedenstellende Kennwerte aufweist, sollte die Güte des Tests in weiteren praxisnäheren Studien beispielsweise in einem known-group-Design überprüft werden. Das Vorliegen negativer Antwortverzerrungen sollte auf mögliche Ursachen hin hinterfragt werden, da verzerrte Beschwerdendarstellungen durch Instruktionen und unterschiedliche Motive hervorgerufen werden können.

Abstract

Aims: The major aim of this thesis is to determine the need of symptom validation in the socio-medical assessment and to construct a screening for the detection of negative response bias by mental disorders. Additionally, the effects of being instructed before completing a questionnaire and the motivation for obtaining a pension are examined in claimants for disability pension.

Methods: On the basis of a systematic review German symptom validation questionnaires determined. Study I proved the suitability of the symptom validation tests for the detection of negative response bias in mental disorders. On the one hand an analogue study design is used, to detect any differences in the presentation of complaints between instructed malingerers, a control group, and inpatients with major depression or pain disorder with depressive symptoms in clinical symptom scales (see Study II). On the other hand, the analogue study design (see Study IV) is used to construct and validate the new screening for the detection of negative response bias by mental disorders. Using a randomized control group design, the effects of giving instructions in advance and the motivation obtaining a pension on symptom presentation are analyzed (see study III).

Results: Instructed malingerers tend to provide extreme responses and overgeneralization in symptom scales. Based on these findings a new symptom validation test was constructed, which has good psychometric characteristics for reliability and validity. Instructions produce negative response bias in symptom scales. In addition, claimants for disability pension tend to negative response bias when they think to recover through their pension.

Conclusion: Although the developed symptom validation test provides sufficient psychometric characteristics, the validation of the test should be continued in other more practically oriented in a known-group-Design, for example. The occurrence of any negative response bias should be scrutinized for possible causes since instructions and different motives can produce invalid symptoms.

Inhaltsverzeichnis

Tabellenverzeichnis **7**
Abbildungsverzeichnis **9**
Abkürzungsverzeichnis **10**
Publikationen **12**
Wissenschaftliches Tätigkeitsfeld **13**
Einleitung **14**

- Theoretischer Teil -

1 Beschwerdenvalidierung **16**
1.1 Begriffe und Definitionen 16
1.2 Studiendesign in der Erforschung der Beschwerdenvalidierung 17
1.3 Möglichkeiten der Erfassung von negativen Antwortverzerrungen 19
1.4 Zusammenfassung 24
2 Beschwerdenvalidierung im Kontext der sozialmedizinischen Begutachtung **25**
2.1 Die sozialmedizinische Begutachtung 25
2.2 Psychische Erkrankungen und die Rente wegen verminderter Erwerbsfähigkeit 27
2.3 Häufigkeit von negativen Antwortverzerrungen 29
2.4 Einflüsse der Beschwerdendarstellung 32
2.5 Ethische Aspekte und die Kontroverse der Beschwerdenvalidierung 34
2.6 Zusammenfassung 35
3 Fragestellung **36**

- Empirischer Teil -

4 Methode **39**
4.1 Testentwicklung 39
4.2 Stichprobe 41
4.3 Studienablauf 48
4.4 Messinstrumente 50
4.5 Datenanalyse 56
5 Ergebnisse **60**
5.1 Beschwerdenvalidierung in der sozialmedizinischen Begutachtung (Studie I) 60
5.2 Stichprobe 62
5.3 Entwicklung des BEVA (u.a. Studie II) 64

5.4 Konstruktion und Validierung des BEVA (Studie IV) 72
5.5 Einflüsse von Instruktionen und der Motivation auf die Beschwerdendarstellung (Studie III) 75
5.6 Fazit 78
6 Diskussion 80
6.1 Notwendigkeit der Entwicklung eines Beschwerdenvalidierungstests 80
6.2 Fragebogen BEVA 82
6.3 Einflüsse auf die Beschwerdendarstellung in der sozialmedizinischen Begutachtung 86
6.4 Limitationen 88
6.5 Implikationen für die Forschung und Praxis 92
6.6 Fazit 93
Literatur 95

Anhang A: Studie I

Anhang B: Studie II

Anhang C: Studie III

Anhang D: Studie IV

Anhang E: Fragebogen der Expertenbefragung

Anhang F: Materialien Studie II und IV

Anhang G: Materialien Studie III

Anhang H: Eigenleistung der Doktorandin

Anhang I: Eidesstattliche Erklärung

Tabellenverzeichnis

Tabelle 1: Differentialdiagnostische Unterscheidung der Kategorien Simulation/ Aggravation, artifizielle Störung und somatoforme Störung anhand der Dimensionen Beschwerdendarstellung und Motivation nach Merten (2005).....17

Tabelle 2: Diagnoseansätze sowie Beispiele für Verfahren mit kurzer Beschreibung..... 22

Tabelle 3: Anzahl der Rentenzugänge wegen verminderter Erwerbsfähigkeit von 2013 nach 1. Diagnose (psychische und Verhaltensstörungen) nach Kapitel V Bereich F des ICD-10 (Gesundheitsberichterstattung des Bundes, 2013).....29

Tabelle 4: Ermittelte Prävalenzen negativer Antwortverzerrungen für unterschiedliche Begutachtungskontexte und Diagnosegruppen nach einer Studie von Mittenberg, Patton, Canyock und Condit (2002).....30

Tabelle 5: Allgemeine Angaben der Expertenstichprobe..... 42

Tabelle 6: Ablaufschritte der Analogstudie 49

Tabelle 7: Ablaufschritte der Studie III..... 50

Tabelle 8: Einsatzbereich, gemessenes Konstrukt und etwaige Anmerkungen zu allen Verfahren, die in den Studien der vorliegenden Dissertation genutzt wurden 55

Tabelle 9: Auflistung und Erklärung der statistischen Analysen aller promotionsrelevanten Studien und der Dissertation 56

Tabelle 10: Personenbezogene Daten der Studierendenstichprobe der Prä-Studie 62

Tabelle 11: Endergebnis der qualitativen Analyse der Expertenbefragung 65

Tabelle 12: Beispielitems der Verfahren für die Itemgenerierung und Umformulierung des jeweiligen Items im BEVA.....66

Tabelle 13: Vergleich der mittleren Rohwertsummen zwischen den Untersuchungsgruppen für die Skalen des Moduls A der HEALTH-49 69

Tabelle 14: Vollständiger Itempool des BEVA 69

Tabelle 15: Anzahl negativer Antwortverzerrungen im SFSS (Cut-Off-Wert = 16) je Untersuchungsgruppe der Studie III.....77

Tabelle 16: Darlegung der Arbeitsanteile der Promovendin an allen promotionsrelevanten Publikationen ..... Anhang

Abbildungsverzeichnis

Abbildung 1: Untergliederung der Rente wegen verminderter Erwerbsfähigkeit nach Götz et al. (2011) 26

Abbildung 2: Anzahl der Rentenzugänge für die Jahre 2005 bis einschließlich 2013 wegen verminderter Erwerbsfähigkeit nach 1. Diagnose (Gesundheitsberichterstattung des Bundes, 2013) 28

Abbildung 3: Anzahl der Teilnehmer der Konstruktion des BEVA sowie der Studie II mit jeweiligen Ausschlusskriterien 43

Abbildung 4: Anzahl der Teilnehmer der Validierung des BEVA mit jeweiligen Ausschlusskriterium 44

Abbildung 5: Anzahl der Patienten der Konstruktion des BEVA sowie der Studie II mit jeweiligen Ausschlusskriterien 45

Abbildung 6: Anzahl der Patienten der Validierung des BEVA mit jeweiligen Ausschlusskriterien 46

Abbildung 7: Anzahl der Antragsteller der Studie III mit jeweiligen Ausschlusskriterien 47

Abbildung 8: Balkendiagramm der Mittelwerte und dazugehörigen Standardabweichungen für die Skalen des Moduls A der HEALTH-49 für die jeweiligen Untersuchungsgruppen der Studie II 67

Abbildung 9: Histogramm der Häufigkeiten der Summenwerte im BEVA je Untersuchungsgruppe in der Konstruktion 73

Abbildung 10: Histogramm der Häufigkeiten der Summenwerte im BEVA je Untersuchungsgruppe in der Validierung 74

Abbildung 11: Balkendiagramm der Mittelwerte und dazugehörigen Standardabweichungen für die Skalen des Moduls A der HEALTH-49 für die jeweiligen Gruppen der Studie III 76

Abbildung 12: Korrelation der Summenwerte im SFSS und dem Item „Die Rente bedeutet für mich, eine Chance, wieder gesund zu werden“ und dazugehöriger Regressionsgeraden in der Studie III 78

Abkürzungsverzeichnis

ABCN	American Board of Clinical Neuropsycchology
Abb.	Abbildung
AKGT	Amsterdamer Kurzzeitgedächtnistest
ANOVA	analysis of variance
AUC	area under the curve
BEVA	Beschwerdenvalidierungstest (neu entwickelter Test)
BDI-II	Beck Depressionsinventar-II
BVT	Beschwerdenvalidierungstest
bzw.	beziehungsweise
DGPPN	Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde
d.h.	das heißt
DP	depressive Patienten
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders-IV
EM	Expectation Maximization
FBS	MMPI-2**:** Fake-Bad Skala
FIT	15-Item-Test
FPI-R	Freiburger Persönlichkeitsinventar-Revision
HEALTH-49	Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis
ICD-10	Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme-10
IS	instruierte Simulanten
KG	Kontrollgruppe
LW-Skala	MMPI-2: Lachar und Wrobels kritische Items Skala
OR	Odds Ratio
MANOVA	multivariate analysis of variance
MENT	Morel Emotional Numbing Test
MMPI-2	Minnesota Multiphasic Personality Inventory-2
MSVT	Medical Symptom Validity Test
O-S Skala	MMPI-2: Skala der Differenz aus offensichtlichen und subtilen Symptomen
PSSI	Persönlichkeits-Stil und Störungs-Inventar

PPW	Positiv Prädiktiver Wert
ROC	Receiver Operating Characteristics
SFSS	Strukturierter Fragebogen Simulierter Symptome
SGB	Sozialgesetzbuch
SIRS	Structured Interview of Reported Symptoms
SKID	Strukturiertes Klinisches Interview für DSM-IV
SP	Schmerzpatienten
s.u.	siehe unten
Tab.	Tabelle
TBFN	Testbatterie zur Forensischen Neuropsychologie
TOMM	Test of Memory Malingering
u.a.	unter anderem
VEI	Verhaltens- und Erlebnisinventar
Vgl.	Vergleich
WMT	Word Memory Test
ZKPR	Zentrum für Klinische Psychologie und Rehabilitation

Publikationen

Die vorliegende kumulative Dissertation basiert auf den folgenden Publikationen, die in der Dissertation durch römische Ziffern gekennzeichnet werden.

Theoretische Arbeit

Studie I (vgl. Anhang A)

Walter, F., Petermann, F. & Kobelt, A. (2012). Beschwerdenvalidierung: Ein aktueller Überblick. *Die Rehabilitation, 51*, 342-348.

Empirische Arbeiten

Studie II (vgl. Anhang B)

Walter, F., Petermann, F., Dietrich, D. E. & Kobelt, A. (2013). Wie können Beschwerden im Rahmen der medizinischen Rehabilitation validiert werden? *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin, 23*, 334-340.

Studie III (vgl. Anhang C)

Walter, F., Geissler, N., Petermann, F. & Kobelt, A. (2015). Beeinflussen Instruktionen und die Motivation für eine Rente die Beschwerdedarstellung in Fragebogen? *Die Rehabilitation*. In Druck.

Studie IV (vgl. Anhang D)

Walter, F., Petermann, F. & Kobelt, A. (2015). Erfassung von negativen Antwortverzerrungen – Entwicklung und Validierung des BEVA. Eingereicht.

Wissenschaftliches Tätigkeitsfeld

Die vorliegende Dissertation wurde im Rahmen einer Kooperation mit der Deutschen Rentenversicherung Braunschweig-Hannover und dem Zentrum für Klinische Psychologie und Rehabilitation (ZKPR) an der Universität Bremen vom 1.1.2012 bis 31.12.2014 umgesetzt. Der Schwerpunkt der Arbeit liegt im Bereich der Beschwerdenvalidierung. Im Verlauf der Dissertation wurde ein Beschwerdenvalidierungstest für die sozialmedizinische Begutachtung entwickelt sowie unterschiedliche Faktoren für negative Antwortverzerrungen in Fragebogen analysiert.

Im Rahmen der sozialmedizinischen Begutachtung für die Beurteilung der Erwerbsunfähigkeit soll ermittelt werden, inwiefern die Leistungsfähigkeit des Versicherten durch eine psychische Erkrankung eingeschränkt ist. Mithilfe des in dieser Promotion entwickelten Verfahrens sollen die berichteten Beschwerden validiert werden. Hierfür wurden in Kooperation verschiedener Kliniken wie den AMEOS Kliniken in Hildesheim und Bremen sowie den Rehabilitationskliniken Oberharz und Bad Pyrmont Daten von depressiven Patienten erhoben. In der Rehabilitationsklinik Bad Eilsen konnten für die Datengewinnung Schmerzpatienten rekrutiert werden. Um mögliche Einflüsse der Beschwerdendarstellung in Fragebogen zu analysieren, wurden Antragsteller einer Rente wegen verminderter Erwerbsfähigkeit in den Gutachtenstellen der deutschen Rentenversicherung Braunschweig-Hannover in Brauschweig und Hannover rekrutiert.

Im Zeitraum der Dissertation wurden neben den promotionsrelevanten Studien an den folgenden Kongressbeiträgen mitgearbeitet:

Kobelt, A., Geissler N., Walter, F. & Petermann, F. (2014). Reicht es, im Rahmen der Begutachtung von Rentenantragstellern zu warnen, dass die Angaben überprüft werden, um weniger verzerrte Antwortergebnisse zu erhalten?. [Abstract] In Deutsche Rentenversicherung Bund (Hrsg.), *23. Rehabilitationswissenschaftliches Kolloquium – Arbeit – Gesundheit - Rehabilitation vom 10. bis 12. März 2014 in Karlsruhe* (DRV-Schriften, Bd. 103, S. 319-321). Berlin: Deutsche Rentenversicherung Bund.

Walter, F., Petermann, F., Kobelt, A. (2015). Entwicklung eines Screeningverfahrens für die Beschwerdenvalidierung von Erkrankungen mit depressiver Symptomatik. [Abstract] In Deutsche Rentenversicherung Bund (Hrsg.), *24. Rehabilitationswissenschaftliches Kolloquium – Psychische Störungen – Herausforderungen für Prävention und Rehabilitation vom 16. bis 18. März 2015 in Augsburg.* Berlin: Deutsche Rentenversicherung Bund. In Druck.

Einleitung

„Überzeugungen sind gefährlichere Feinde der Wahrheit als Lügen“

Friedrich Nietzsche (1879)

Um die Diagnose einer psychischen Erkrankung stellen zu können, ist der Kliniker, Diagnostiker bzw. Gutachter immer auf die Schilderungen seines Gegenübers angewiesen. Die Diagnostik soll beispielsweise durch Symptomfragebogen, Leistungstests und klinische Interviews objektiviert und standardisiert werden. Dies ist auch notwendig, da Studien belegen, dass sich Experten leichter täuschen lassen, wenn sie sich allein auf ihren klinischen Eindruck verlassen (Rosenhan, 1973; Vrij, 2000). Die Vorgehensweise der Diagnostik einer psychischen Erkrankung erfordert aber, dass sowohl der Kliniker als auch der Untersuchte dasselbe Ziel verfolgen – nämlich die Diagnostik einer psychischen Störung nach dem momentan wahrheitsgetreuen Ist-Zustand der Erkrankung. Jedoch können beispielsweise in der forensischen Begutachtung oder im zivil-, sozial- und verwaltungsrechtlichen Kontext eine Diagnose und daraus resultierende externe Vergünstigungen miteinander verknüpft sein. Das bedeutet, die Motivation Beschwerden vorzutäuschen oder zu übertreiben, könnte in dem Erhalt einer Rente oder von Schmerzensgeldern sowie dem Entgehen einer gerichtlichen Verfolgung begründet sein. Die üblicherweise eingesetzten Verfahren zur Validierung von Beschwerden sind allerdings sehr transparent und durchschaubar, das heißt der Zweck und die Zielrichtung ist für Laien schnell erkennbar, was dazu führt, dass Testergebnisse leicht zu manipulieren sind.

Daher erscheint es notwendig ein Screening zu entwickeln, dass etwaige negative Antwortverzerrungen erfassen kann, ohne dass der Zweck der Befragung sofort durchschaubar ist. Dafür ist es entscheidend, die Unterschiede des Antwortverhaltens von Simulanten und psychisch Erkrankten zu ermitteln. Darüber hinaus sollten in der sozialmedizinischen Begutachtung Faktoren untersucht werden, die wiederum die Beschwerdendarstellung beeinflussen und somit zu negativen Antwortverzerrungen führen können.

Die vorliegende Dissertation ist in zwei Hauptteile untergliedert. Im ersten Abschnitt - dem theoretischen Rahmen – wird im Kapitel 1 allgemein auf die Beschwerdenvalidierung Bezug genommen. Hierfür werden die Begriffe sowie Definitionen benannt und vorgestellt. Weiterhin wird auf

die Besonderheiten der Erforschung im Kontext der Beschwerdenvalidierung eingegangen. Darauf folgt im Kapitel 1.3 die Darstellung der Möglichkeiten der Erfassung von negativen Antwortverzerrungen. Dieser Abschnitt beschreibt Diagnoseansätze und ordnet ihnen konkrete Verfahrensbeispiele zu. Im Kapitel 2 erfolgt eine genauere Betrachtung der Beschwerdenvalidierung im Rahmen der sozialmedizinischen Begutachtung. Zunächst wird die sozialmedizinische Begutachtung losgelöst von der Beschwerdenvalidierung betrachtet und vor allem die Rente wegen verminderter Erwerbsfähigkeit als Beispiel beleuchtet. Danach gibt das Kapitel 2.2 einen Überblick über die am häufigsten auftretenden psychischen Störungen in der Begutachtung der Rente wegen verminderter Erwerbsfähigkeit. Es soll aufgezeigt werden, welche Störungen für die Entwicklung des Verfahrens relevant sind. Kapitel 2.3 stellt einen Überblick der Häufigkeiten negativer Antwortverzerrungen im Kontext der sozialmedizinischen Begutachtung dar. Daraufhin werden Einflüsse diskutiert, die in der sozialmedizinischen Begutachtung zu einer verzerrten Beschwerdendarstellung führen können. Schlussendlich wird auf die besondere ethische Herangehensweise der Diagnostik der Beschwerdenvalidität bei psychischen Erkrankungen eingegangen. Anhand der in Kapitel 1 und 2 dargestellten Theorie erfolgt die Herleitung von drei Themenschwerpunkten und dazugehörigen Fragestellungen, die im Rahmen der vorliegenden Dissertation bearbeitet wurden. Das Kapitel 3 bildet damit den Schluss des theoretischen Abschnitts.

Im empirischen Teil dieser Arbeit findet im Kapitel 4 zunächst die Betrachtung der Methodik der promotionsrelevanten Studien statt. Zunächst werden die einzelnen Schritte der Entwicklung des im Rahmen dieser Dissertation entstandenen Testverfahrens explizit beschrieben. Darüber hinaus erfolgt die Darstellung der Stichprobe, des Studienablaufs, der genutzten Erhebungsverfahren und der Datenanalyse für die empirischen sowie theoretischen promotionsrelevanten Publikationen. Weiterhin werden im Kapitel 5 die Ergebnisse der Studien sowie neue Analyseerkenntnisse geordnet nach den drei Themenschwerpunkten vorgestellt. Schlussendlich stellt das Kapitel 6 einen Überblick aller Resultate in Bezug auf die Bedeutung für die jeweilig bearbeiteten Fragestellungen und der Implikationen für die weitere Forschung und Praxis dar.

Im Anhang sind die promotionsrelevanten Publikationen, die Darstellung des Eigenanteils sowie die eidesstattliche Erklärung angefügt. Darüber hinaus können dem Anhang die eigens für die Promotion entwickelten Fragebogen entnommen werden.

1 Beschwerdenvalidierung

1.1 Begriffe und Definitionen

Bei der Beschäftigung mit der *Beschwerdenvalidierung* müssen zunächst einige Begrifflichkeiten geklärt und definiert werden, die mittelbar und unmittelbar mit der Thematik zusammenhängen. Nach Merten (2011, S.103) ist die *Beschwerdenvalidität*: „ … die Gültigkeit einer Beschwerdeschilderung oder Symptompräsentation durch einen Untersuchten." Als *Beschwerdenvalidierung* wird der diagnostische Prozess bezeichnet, der die beschriebenen Beschwerden hinsichtlich ihres Wahrheitsgehalts beurteilen soll. Mit den sogenannten *Beschwerdenvalidierungstests* (BVT) sollen die Gültigkeit der dargebotenen Beschwerden sowie etwaige Antwortverzerrungen erfasst werden. Im Allgemeinen erfolgt eine Unterscheidung zwischen kognitiven BVT und Verfahren, die psychische Beschwerden validieren sollen. *Antwortverzerrungen* zeichnen sich durch eine mangelnde Leistungsbereitschaft bzw. Testmotivation sowie fehlerhafte, ungenaue und unvollständige Antworten aus (Merten, Stevens & Blaskewitz, 2007). Es können sowohl *negative* als auch *positive Antwortverzerrungen* vorliegen. Letztere wird auch als *Dissimulation* bezeichnet und beschreibt reale Symptome, die durch einen Untersuchten verheimlicht bzw. verharmlost werden (Cima, Merckelbach, Hollnack, Butt, Kremer, Schellbach-Matties & Muris, 2003).

Im Rahmen der sozialmedizinischen Begutachtung spielen vor allem *negative Antwortverzerrungen* eine bedeutende Rolle. Hierbei wird grundsätzlich zwischen Simulation und Aggravation differenziert. Letztere liegt vor, wenn bei dem Betroffenen im Kern reale Symptome vorhanden sind, diese jedoch weitaus schwerwiegender präsentiert werden (Merten, 2011). Bei einer Simulation werden durch den Untersuchten Beschwerden oder eine Störung absichtlich vorgetäuscht. Differentialdiagnostisch müssen in diesem Zusammenhang folgende drei Kategorien betrachtet werden: Simulation/Aggravation, artifizielle Störung und somatoforme Störung. Wie in der Tabelle 1 zu erkennen, erfolgt die Unterscheidung nach Merten (2005) anhand zweier Dimensionen: einer *bewussten* vs. *unbewussten Symptompräsentation* sowie einer *bewussten* vs. *unbewussten zugrundeliegenden Motivation*. In der Literatur werden weitere Bedingungen diskutiert, die zu negativen Antwortverzerrungen führen können. Dazu zählen bspw. situative Faktoren sowie die Interaktion zwischen Gutachter und Betroffenem. Demnach können negative Antwortverzerrungen nicht bedingungslos einer Aggravation oder Simulation zugeschrieben werden. Es gilt, immer den Ursprung der beschriebenen Beschwerdendarstellung zu überprüfen.

Tabelle 1: Differentialdiagnostische Unterscheidung der Kategorien Simulation/Aggravation, artifizielle Störung und somatoforme Störung anhand der Dimensionen Beschwerdendarstellung und Motivation nach Merten (2005)

Diagnostische Kategorie	Beschwerdendarstellung	Motivation
Simulation/Aggravation	Bewusst und absichtlich	Reflektiert (bewusst)/ sekundärer Krankheitsgewinn
Artifizielle Störung	Bewusst und absichtlich	Unreflektiert (unbewusst)/ primärer Krankheitsgewinn
Somatoforme Störung	Unbewusst und nicht absichtlich	Unreflektiert (unbewusst)/ primärer Krankheitsgewinn

Zusätzlich stellt die Erforschung der Beschwerdenvalidierung und der mit ihr in Zusammenhang stehenden Faktoren eine große Herausforderung dar, da für die Erfassung von negativen Antwortverzerrungen kein Goldstandard vorliegt. Nachfolgend werden Studiendesigns vorgestellt, die versuchen, dieser Problematik durch verschiedene Ansätze entgegenzuwirken.

1.2 Studiendesign in der Erforschung der Beschwerdenvalidierung

Aus der Literatur sind vier Basisdesigns bekannt (Rogers, 2008), die im Wesentlichen für Studien der Beschwerdenvalidierung verwendet werden, wobei das Analogstudiendesign sowie das known-group-Design komplementär bezüglich ihrer jeweiligen Stärken sind. Da die Zuordnung zu den voneinander unabhängigen Gruppen (beispielsweise Simulant vs. Nicht-Simulant) im known-group-Design eine hohe Herausforderung darstellt, werden zwei zusätzliche Designs — das Prävalenzdesign sowie die Bootstrapping-Vergleiche — in der Erforschung der Beschwerdenvalidierung genutzt. Im Folgenden werden die jeweiligen Charakteristika und Stärken der vier Designs herausgestellt.

Analogstudiendesign. Für die Erforschung der Beschwerdenvalidierung und ihrer beeinflussenden Faktoren ist das Analogstudiendesign am weitesten verbreitet (Merten, Henry & Hilsabeck, 2004). Es ermöglicht eine sehr hohe interne Validität, da die Studienteilnehmer randomisiert verschiedenen, experimentellen Bedingungen zugeordnet und standardisierte Methoden genutzt werden. In den meisten Studien werden bspw. Personen der Allgemeinbevölkerung entweder dazu aufgefordert, eine Erkrankung vorzutäuschen („instruierte Simulanten“, nachfolgend: IS) oder sie werden einer Kontrollbedingung („gesunde Teilnehmer“) zugeordnet. Damit die vorgetäuschten Erkrankungen von wirklichen Symptomen einer Störung unterschieden werden können, werden

vor allem die Daten der instruierten Simulanten mit denen einer klinischen Stichprobe verglichen. Die Aufgabe der Kontrollgruppe (KG) bzw. der Teilnehmer der klinischen Gruppe ist es, die ihnen bereitgestellten diagnostischen Verfahren nach ihrem derzeitigen Krankheitsstatus wahrheitsgemäß zu beantworten. Der Studienablauf besteht dabei meist aus mehreren Bestandteilen. Zunächst erhalten die IS ein Szenario bzw. eine Falldarstellung eines fiktiven Patienten, auf das in den nachstehenden Fragebogen Bezug genommen werden soll. Anschließend folgt ein Fragebogen, der erfassen soll, ob der Text aufmerksam gelesen bzw. die Aufgabe verstanden wurde. Am Ende jeder Testung wird wiederum mit einem Fragebogen überprüft, ob die IS der Aufgabe Folge geleistet haben (Bianchini, Mathias & Greve, 2001). Der Nachteil dieses Studiendesigns besteht darin, dass die externe Validität sehr gering ist. Die IS täuschen unter anderen Umständen psychische Erkrankung vor, als Personen, die unter realen Bedingungen Beschwerden verzerrend darstellen.

Known-group-Design. Im Rahmen dieses Studiendesigns gibt es zwei unabhängige und diskrete Phasen. Zunächst werden Patienten durch unabhängige Experten bzw. Kliniker in real erkrankte Patienten und Personen, die ihre Beschwerden verzerrt darstellen, eingeteilt. Erst dann erfolgt eine systematische Analyse der Unterschiede zwischen diesen Gruppen. Dieses Design gewährt eine hohe externe Validität, da Betroffene unter realen diagnostischen und klinischen Bedingungen den Untersuchungsgruppen zugeordnet werden. Gleichzeitig ist die interne Validität sehr gering, da in der Regel keine standardisierten Bedingungen und kaum experimentelle Umstände gewährleistet werden können.

Prävalenz-Design. Bei Prävalenz-Designs werden zwei Stichproben miteinander verglichen, von denen bekannt ist, dass in einer dieser Stichproben mehr Antwortverzerrungen auftreten. Beispielsweise wird die Beschwerdendarstellung einer klinischen Stichprobe einer Begutachtungsstichprobe gegenübergestellt. Die Annahme besteht darin, dass in einer dieser Gruppen bei jedem Teilnehmer Antwortverzerrungen auftreten, hingegen ist dieses bei der zweiten Gruppe nicht der Fall. Die interne Validität des Prävalenz-Designs ist wiederum sehr gering, da keine standardisierten Bedingungen vorliegen. Die externe Validität ist mäßig, zwar sind die Teilnehmer im klinischen Kontext involviert, die Einteilung auf der Grundlage von Basisraten ist allerdings stark fehleranfällig.

Bootstrapping-Vergleiche. Bei diesem Design werden etablierte Kriterien oder Verfahren genutzt, um die Gruppen real erkrankte Patienten und Simulanten herauszufiltern (Slick, Sherman & Iverson, 1999). Durch die Verwendung von mehreren BVT soll die Spezifität maximiert werden.

Das bedeutet, dass die Wahrscheinlichkeit, dass ein Patient fälschlicherweise als Simulant klassifiziert wird, minimiert werden soll. Die interne Validität ist bei diesem Ansatz sehr gering, da keine standardisierten experimentellen Bedingungen vorliegen. Die Studienteilnehmer werden im Rahmen dieses Settings abermals unter klinischen Bedingungen rekrutiert. Die externe Validität wird durch Rogers (2008) als mittelstark eingeschätzt.

Die vorausgegangen Schilderungen verdeutlichen die Problematik in der Erforschung der Beschwerdenvalidierung. Es ist möglich, entweder eine hohe interne, oder externe Validität zu gewährleisten.

1.3 Möglichkeiten der Erfassung von negativen Antwortverzerrungen

Im Folgenden werden unterschiedliche Diagnose- und Forschungsstrategien beschrieben, die eingesetzt werden können, um etwaige BVT zu konzipieren. Die Tabelle 2 ordnet den möglichen Strategien Verfahrensbeispiele zu und erklärt diese kurz. Im Weiteren wird auf einen Teil dieser Verfahren immer wieder Bezug genommen, da sie in später zitierten Studien eingesetzt wurden.

Rogers (2008) beschreibt für den Nachweis von vorgetäuschten psychischen sowie neuropsychologischen Symptomen unterschiedliche Forschungs- und Diagnosestrategien. Ausgewählte Beispiele für besonders im deutschsprachigen Raum bedeutende Testverfahren, die auf den nachfolgenden beschriebenen sechs Ansätzen beruhen, sind in Tabelle 2 aufgelistet. Speziell für die neuropsychologische Beschwerdenvalidierung bewährt sich eine Einteilung von Rogers, Harrel und Liff (1993), die folgende sechs Strategien vorschlagen:

1. **Testdeckeneffekt oder Prinzip der verdeckten Leichtigkeit:**
   Das Prinzip beruht darauf, dass sehr einfache Aufgaben oder Fragen (Was ist größer: ein Hund oder ein Pferd?) nicht bewältigt werden können. Die Aufgaben sind jedoch so gestellt, dass sie sogar von stark beeinträchtigten hirngeschädigten Personen gelöst werden können.

2. **Leistungskurve:**
   Die Leistungskurve soll den Schwierigkeitsgrad unterschiedlicher Aufgaben plausibel abbilden. Bei dieser Strategie zur Aufdeckung von vorgetäuschten Symptomen wird davon ausgegangen, dass leichte Aufgaben nicht gelöst werden, schwierigere hingegen schon. Der Untersuchte weist in diesem Fall unübliche Schwierigkeitsrangreihen auf.

3. **Fehlergröße:**

Hierbei handelt es sich um ein Antwortverhalten, bei dem Personen regelhaft knapp von der richtigen Antwort abweichen. Es wird angenommen, dass die korrekten Antworten bekannt sind, jedoch verfälscht werden. Dieses Antwortverhalten kann beispielsweise bei Rechen- oder Orientierungsaufgaben beobachtet werden.

4. **Alternativwahlverfahren (BVT im eigentlichen Sinne):**

Dieses Prinzip gilt als die bisher am besten untersuchte und verbreiteteste Methode der Beschwerdenvalidierung. Die Alternativwahlverfahren haben den Begriff der BVT geprägt. Dennoch schließt der Begriff der BVT heute mehr Methoden sowie Strategien der Beschwerdenvalidierung ein und wird somit breiter genutzt. Die Methode gilt als einziger Ansatzpunkt, mit der definitive Beweise für vorgetäuschte kognitive Symptome erfasst werden können. Bei diesem Ansatz werden zwei Auswahlalternativen je Item dargeboten, sodass bei einer a-priori-Ratewahrscheinlichkeit von 50 % das Antwortverhalten durch eine zugrundegelegte Binomialverteilung analysiert werden kann.

5. **Inkonsistente oder atypische Leistungsprofile:**

Bei diesem Ansatz werden die Leistungsprofile der untersuchten Person auf ihre Konsistenzen und Plausibilitäten hin überprüft. Es sollen verschiedene (Unter-)Tests bzw. die Ergebnisse zweier Messzeitpunkte untersucht werden, um somit mögliche Differenzen zu ermitteln. Im Allgemeinen ist dieser Ansatz jedoch schwer zu operationalisieren und demnach schwieriger anzuwenden. Strauss, Hultsch, Hunter, Slick, Patry und Levy-Bencheton (1999) stellten jedoch heraus, dass die intraindividuelle Variabilität der Testleistung, gemessen über mehrere Zeitpunkte hinweg, eine gute Methode für die Erfassung von negativen Antwortverzerrungen darstellt.

6. **Darstellung von psychischen Symptomen:**

Psychische Erkrankungen wie Depressionen oder Angststörungen werden häufig mit kognitiven Defiziten assoziiert bzw. gelten auch als Folgeerkrankungen. Bei diesem Ansatz soll die Plausibilität von psychischen Symptomen überprüft werden.

Speziell für die Erfassung von vorgetäuschten psychischen Symptomen wurden durch Rogers, Sewell, Martin und Vitacco (2003) diagnostische Methoden und Strategien entwickelt. Für die

Konstruktion der Validitätsskalen des *Minnesota Multiphasic Personality Inventory-2* (MMPI-2; Hathaway & McKinley, 2000) wurden folgende Ansätze genutzt:

1. **Seltene Symptome**

   Bei diesem Ansatz werden Symptome beschrieben, die kaum bei wirklich psychisch erkrankten Patienten auftreten. Die Idee hinter diesem Ansatz ist, dass Simulanten nicht in der Lage sind, zwischen ungewöhnlichen und häufig auftretenden Symptomen zu differenzieren.

2. **Symptomschwere**

   Diese Methode berücksichtigt die Anzahl aller möglichen beeinträchtigenden Symptome einer psychischen Erkrankung. Sie basiert auf der Prämisse, dass Simulanten die Symptomschwere nicht beachten und eine sehr hohe Anzahl von Beschwerden angeben, die sie beeinträchtigen.

3. **Vergleich von offensichtlichen und subtilen Symptomen**

   Offensichtliche Symptome kennzeichnen sich dadurch, dass sie auch von Laien einfach einer Psychopathologie zugeordnet werden können. Wohingegen subtile Beschwerden meist nur von Experten erkannt werden. Die implizite Strategie bezieht sich darauf, dass Simulanten eher offensichtliche Symptome wiedererkennen und die Tendenz haben, diese anzugeben. Die Schwierigkeit dieses Ansatzes besteht darin, Symptome zu generieren, die bei einer psychischen Erkrankung auftreten, aber bei einem Laien den Anschein erwecken, als seien sie nicht zugehörig.

4. **Falsche Stereotypen**

   Dieser Ansatz beruht darauf, dass Simulanten nicht in der Lage sind, echte Psychopathologie von falschen Stereotypen zu unterscheiden. So werden Items konstruiert, die breit gefächerte Vorstellungen von psychischen Erkrankungen abbilden. Die Inhalte können bizarr sein und bestimmte Annahmen über die Affektivität, die somatischen Beschwerden, Konflikte in der Kindheit etc. bedienen.

Tabelle 2: Diagnoseansätze sowie Beispiele für Verfahren mit kurzer Beschreibung

Diagnoseansatz	Beispiel (Abkürzung)	Autor	Beschreibung
Testdeckeneffekt oder Prinzip der verdeckten Leichtigkeit	15-Item-Test (FIT)	Rey (1958)	Der Proband soll sich für 10 s 15 Zeichen auf einer Tafel einprägen und direkt im Anschluss wiedergeben.
Leistungskurve	Punkte-Zähl-Test	Rey (1941)	Es sollen Punkte auf unterschiedlichen Karten gezählt werden. Rey geht davon aus, dass gruppierte Punkte schneller gezählt werden können als nicht gruppierte Punkte. Der Gesamtscore wird folgendermaßen berechnet: $\text{Score} = \frac{\text{Zeit(gruppiert)} \times (\text{Anzahl Fehler} + 0.5)}{\text{Zeit (nicht gruppiert)}}$
Fehlergröße			Beispielsweise Orientierungsfragen oder Rechenaufgaben.
Alternativwahlverfahren (BVT im eigentlichen Sinne)	Amsterdamer Kurzzeitgedächtnistest (AKGT)	Schmand und Lindeboom (2005)	Das Testprinzip basiert auf einer Mehrfachauswahl von drei richtigen aus fünf möglichen Antworten. Statt einer Biniomialverteilung wird eine hypergeometrische Verteilung zugrunde gelegt.
	Word Memory Test (WMT)	Green (2003a)	Leicht assoziierbare Wortpaare sollen erlernt werden, die dann später im Zwangswahlverfahren wiedererkannt werden müssen.
	Medical Symptom Validity Test (MSVT)	Green (2003b)	Vereinfachte Version des WMT für demenzielle Erkrankungen.
	Test of Memory Malingering (TOMM)	Tombaugh (1996)	Es sollen im Alternativwahlprinzip Schwarz-Weiß-Zeichnungen von Objekten und Tieren wiedererkannt werden. Ausgewertet werden nur die Durchgänge zwei und drei.
Inkonsistente oder atypische Leistungsprofile	Testbatterie zur Forensischen Neuropsychologie (TBFN)	Heubrock und Peterman (2011)	Die TBFN umfasst insgesamt 23 Untertests, die vor allem für geltend gemachte neuropsychologische Erkrankungen konzipiert wurden.
Darstellung von psychischen Symptomen	Strukturierter Fragebogen Simulierter Symptome (SFSS)	(Cima, Hollnack, et al., 2003)	Es werden bizarre, atypische und unplausible Beschwerden für folgende Störungen erfasst: Psychosen, affektive Störungen, neurologische Beeinträchtigungen, mnestische Störungen & niedrige Intelligenz. Berechnung eines Gesamtwerts sowie Unterskalenwerte sind möglich.
	Minnesota Multiphasic Personality Inventory-2 (MMPI-2)	Engel (2000)	Umfassendes klinisches Inventar, das Validitätsskalen für die Erfassung der Beschwerdendarstellung enthält (s.u.).

Fortsetzung Tabelle 2

Diagnoseansatz	Beispiel (Abkürzung)	Autor	Beschreibung
	Verhaltens- und Erlebnisinventar (VEI)	Engel und Goves (2007)	Umfassendes klinisches Inventar, das Validitätsskalen für die Erfassung der Beschwerdendarstellung enthält (s.u.).
	Morel Emotional Numbing Test (MENT)	Morel (1998)	Konzipiert für Antwortmanipulationen bei der posttraumatischen Belastungsstörung, bezieht sich auf das Symptom der emotionalen Betäubung.
	Structured Interview of Reported Symptoms (SIRS)	Rogers, Sewell und Gillard (2010)	Fremdrating zur Erfassung von negativen Antwortverzerrungen.
Seltene Symptome	Skala „Ungünstiger Eindruck“ des VEI		Diese Skala besteht aus neun Items mit bizarren sowie ungewöhnlichen Symptomen.
	$F_{psychiatric}$-Skala des MMPI-2		Die Skala enthält 27 Items, die von weniger als 20 % der psychiatrischen Patienten bejaht wurde.
	Skala „Seltenheit“ des VEI		Die Skala enthält acht Items über Symptome, die von Gesunden und Patienten selten mit *„ja“* beantwortet werden.
Symptomschwere	LW-Skala (Lachar und Wrobels kritische Items) des MMPI-2	(Lachar & Wrobel, 1979)	Es wurden 130 kritische Items herauskristallisiert, die 14 Bereiche von psychologischen Symptomen repräsentieren.
Vergleich von offensichtlichen und subtilen Symptomen	O-S Skala des MMPI-2	(Wiener, 1948)	Differenz aller Subskalen der offensichtlichen Symptome und den Subskalen aller subtilen Symptome.
Falsche Stereotypen	Fake-Bad (FBS) Skala des MMPI-2; jetzt: Symptom-Validität-Skala	(Lees-Haley, English & Glenn, 1991)	Besteht aus 43 Items. Wurde ursprünglich für die Erfassung von Simulation von emotionalem Stress bei körperlichen Verletzungen konstruiert.

Slick-Kriterien. Eine bedeutende Arbeit für die Diagnostik der Beschwerdenvalidierung stellt immer noch ein Artikel von Slick et al. (1999) dar. Die Arbeitsgruppe entwickelte diagnostische Kriterien, die für die Diagnosestellung einer vorgetäuschten, neurokognitiven Störung elementar sind. Anhand dieser sogenannten Slick-Kriterien können verschiedene Grade diagnostischer Sicherheit für das Auftreten von simulativen Tendenzen abgeleitet werden. Weitere Modifikationen wurden durch Larrabee, Greiffenstein, Greve und Bianchini (2007) vorgeschlagen.

1.4 Zusammenfassung

Die Erfassung von negativen Antwortverzerrungen stellt sowohl in der Diagnostik als auch in der Forschung eine große Herausforderung dar. Für die Erforschung von bedingenden Einflussfaktoren sollten die Stärken und Schwächen eines jeden Designs abgewogen werden, um Aussagen zur Beschwerdenvalidität treffen zu können. In der Diagnostik muss darüber hinaus die zugrundeliegende Motivation für Antwortverzerrungen abgeklärt werden, um differentialdiagnostische Angaben machen zu können. Im nächsten Kapitel wird nun die Beschwerdenvalidierung im Rahmen der sozialmedizinischen Begutachtung erörtert. Näher betrachtet wird die Berentung wegen verminderter Erwerbsfähigkeit.

2 Beschwerdenvalidierung im Kontext der sozialmedizinischen Begutachtung

Im Rahmen dieses Kapitels soll die Einordnung der Beschwerdenvalidierung in die sozialmedizinische Begutachtung erfolgen. Dabei ist es zunächst wichtig, Aspekte der sozialmedizinischen Begutachtung und die am häufigsten auftretenden Erkrankungen losgelöst von der Beschwerdenvalidierung zu beleuchten. Im Weiteren werden die Auftretenswahrscheinlichkeit von negativen Antwortverzerrungen sowie besondere Einflüsse der Beschwerdendarstellung in diesem Kontext betrachtet. Abschließend erfolgt eine Betrachtung der Kontroverse der Beschwerdenvalidierung.

2.1 Die sozialmedizinische Begutachtung

Sozialrechtliche Gutachterfragestellungen können einerseits die Einschätzung der Behandlungsbedürftigkeit sowie die medizinische oder berufliche Rehabilitationsbedürftigkeit umfassen. Andererseits ist die Frage nach der Beurteilung der Leistungsfähigkeit Teil der sozialmedizinischen Begutachtung. Im Rahmen des Rechts der sozialen Entschädigung im Schwerbehindertenrecht und im Recht der gesetzlichen Unfallversicherung wird der Grad der Leistungsminderung überprüft sowie das ursächliche, schädigende Ereignis bewertet (Götz, Roth & Chojetzki, 2011). Dagegen ist in der Begutachtung der Kranken-, Arbeitslosen- und Rentenversicherungen die Beurteilung des verbliebenen Restleistungsvermögens und die daraus resultierende Einsatzfähigkeit im Erwerbsleben von zentraler Bedeutung.

Im Weiteren soll nun der Fokus auf die Rentenbegutachtung gerichtet sein. Durch eine Rente wegen verminderter Erwerbsfähigkeit (vgl. Abb. 1) aufgrund einer Erkrankung oder Behinderung soll der teilweise oder auch volle Verlust der Erwerbsfähigkeit finanziell kompensiert werden (Mehrhoff et al., 2012). Vor der Gewährung dieser Rente ist stets zu überprüfen, ob die Erwerbsfähigkeit durch eine medizinische Rehabilitation oder durch Leistungen zur Teilhabe am Arbeitsplatz wiederhergestellt oder verbessert werden kann. Dieser Grundsatz wird kurz gefasst als „Reha vor Rente“ betitelt. Die Beurteilung des Restleistungsvermögens im Rahmen der sozialmedizinischen Begutachtung wird anhand folgender drei Stufen differenziert (§ 43 Abs. 2 SGB VI):

1. **Volle Erwerbsminderung:** der Versicherte kann aufgrund einer Erkrankung oder Behinderung und unter den üblichen Bedingungen des Arbeitsmarktes nur 0 bis unter 3 Stunden erwerbstätig sein.

2. **Teilweise erwerbsgemindert:** der Versicherte kann aufgrund einer Erkrankung oder Behinderung und unter den üblichen Bedingungen des Arbeitsmarktes nur 3 bis unter 6 Stunden erwerbstätig sein.

3. **Nicht erwerbsgemindert:** der Versicherte kann aufgrund einer Erkrankung oder Behinderung und unter den üblichen Bedingungen des Arbeitsmarktes mindestens 6 Stunden erwerbstätig sein.

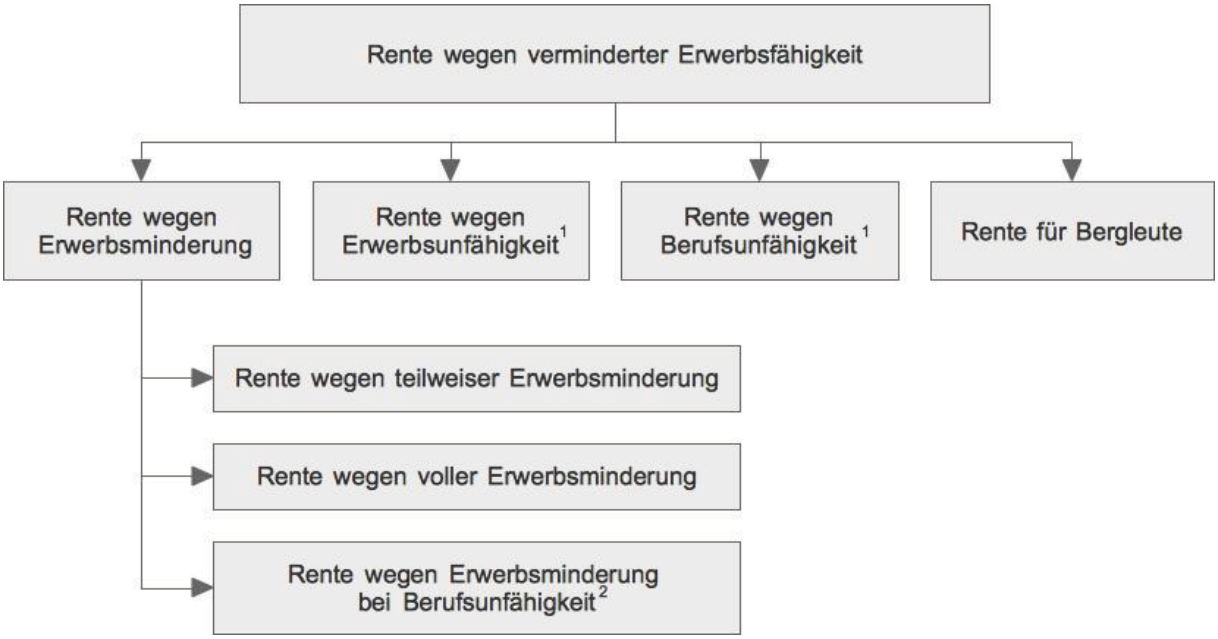

Abbildung 1: Untergliederung der Rente wegen verminderter Erwerbsfähigkeit nach Götz et al. (2011)
Anmerkungen: [1] nach altem Rentenrecht bis 31.12.2000: Besitzstandregelung; [2] Übergangsregelung für vor dem 2.1.1961 geborene Versicherte

Die Grundlagen der Begutachtung bzw. Beurteilung der körperlichen Leistungsfähigkeit sind körperliche Leistungsparameter (Schneider, Becker, Dohrenbusch, Freyberger, Gündel & Henningsen, 2010). Bei der psychosomatischen Begutachtung hingegen wird die Wechselwirkung aus sozialen, somatischen und psychischen Faktoren und den daraus resultierenden Auswirkungen auf die Leistungsfähigkeit beurteilt (Schneider et al., 2011). Daraus folgt, dass lediglich die psychischen Diagnosen keine Aussagekraft für die Leistungsbeurteilung haben. Für die Begutachtung wurden Leitlinien von verschiedenen Autoren vorgeschlagen (Dohrenbusch & Schneider, 2011; Schickel & Henningsen, 2010; Widder, Dertwinkel, Egle, Foerster & Schiltenwolf, 2007). Im Folgenden soll eine genauere Betrachtung der Bedeutung von psychischen Erkrankungen als Hauptdiagnose bzw. komorbide Störung im Rahmen der Rente wegen verminderter Erwerbsfähigkeit erfolgen.

2.2 Psychische Erkrankungen und die Rente wegen verminderter Erwerbsfähigkeit

Im Jahr 2013 lagen laut Statistik der Deutschen Rentenversicherung 175 135 neue Rentenzugänge wegen verminderter Erwerbsfähigkeit[1] vor (statistisches Bundesamt, 2013). Die häufigsten Ursachen für den Erhalt einer Rente wegen verminderter Erwerbsfähigkeit waren psychische und Verhaltensstörungen, die bei 43 % der Versicherten auftraten (Sperling, Biermann, Löwe, Kornhuber & Reulbach, 2010). Wenn die Rentenzugänge wegen verminderter Erwerbsfähigkeit von 2005 bis 2013 für die verschiedenen Diagnosegrundgruppen betrachtet werden, zeigt sich, dass die psychischen Erkrankungen stark angestiegen sind, wobei die Anzahl der übrigen Indikationen für die jährlichen Rentenzugänge relativ konstant blieben oder sich sogar leicht reduzierten (vgl. Abb. 2; Gesundheitsberichterstattung des Bundes, 2013) Durch den zahlenmäßigen Anstieg von psychischen Erkrankungen ergeben sich nicht nur Probleme in der Gesundheitsversorgung, sondern es resultieren unter anderem aufgrund der frühzeitigen Berentungen erhebliche volkswirtschaftliche Kosten (Schneider, 2007).

Bei differenzierter Betrachtung der psychischen Erkrankungen wird deutlich, dass affektive Störungen des Kapitels V F30-F39 der Internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme-10 (ICD-10; Weltgesundheitsorganisation, 2005) die häufigste Ursache im Jahr 2013 für eine Rente wegen verminderter Erwerbsfähigkeit darstellen (vgl. Tab. 3; Gesundheitsberichterstattung des Bundes, 2013. Es folgen neurotische, Belastungs- und somatoforme Störungen (F40-F48) sowie psychische und Verhaltensstörungen durch psychotrope Substanzen (F10-F19).Zudem zeigen Studien die Bedeutung von psychischen Belastungen bei primär somatischen Erkrankungen auf. So führen kardiovaskuläre sowie muskuloskelettale Erkrankungen eher zu Berufsunfähigkeit, wenn gleichzeitig psychische Beeinträchtigungen vorliegen (Gjesdal et al., 2009). Weiterhin weisen depressive Erkrankungen als komorbide Störungen zu primär organischen Erkrankungen eine Prävalenzrate von bis zu 70 % auf (Egede, 2005). In einer weiteren Studie (Schmidt, Bernert & Spyra, 2014), in der psychische Komorbiditäten bei Rückenschmerzpatienten in einer Reha-Kohorte von 2002-2009 betrachtet wurden, stellten die Autoren einen starken Einfluss von psychischen Störungen auf den Chronifizierungsprozess bei Rückenschmerzen fest. Die Häufigkeit von komorbiden Störungen beträgt im Rahmen der Studie zwischen 17.6 % und 20.9 %. Zudem steigt das Risiko einer Berentung wegen verminderter Erwerbsfähigkeit, falls bei den Patienten eine komorbide psychische Störung vorliegt.

[1] Ohne Rente für Bergleute wegen Vollendung des 50. Lebensjahres. Dies gilt auch für nachfolgende Angaben der Rentenzugänge wegen verminderter Erwerbsfähigkeit

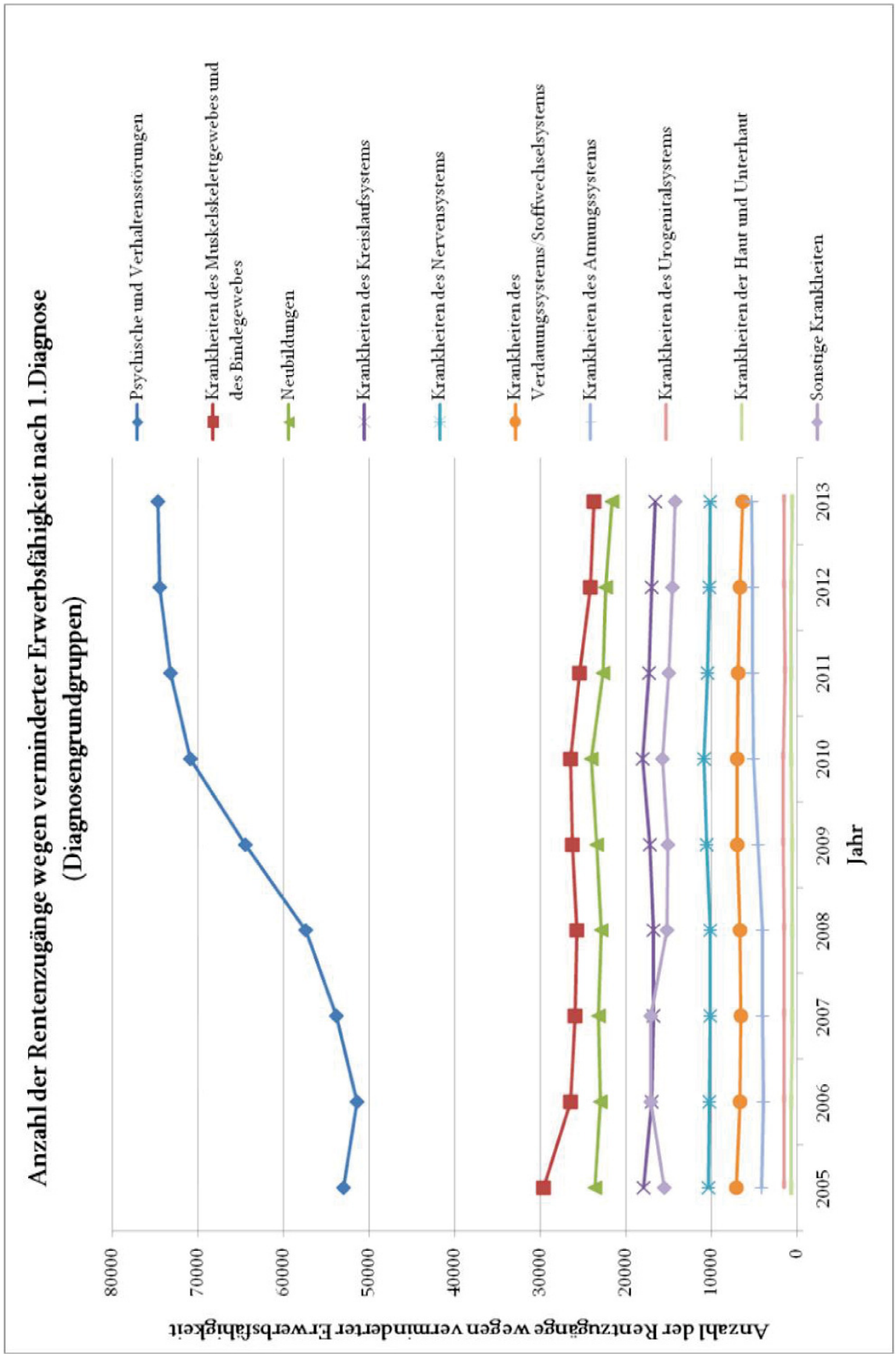

Abbildung 2: Anzahl der Rentenzugänge für die Jahre 2005 bis einschließlich 2013 wegen verminderter Erwerbsfähigkeit nach 1. Diagnose (Gesundheitsberichterstattung des Bundes, 2013)
Anmerkung: Ohne Rente für Bergleute wegen Vollendung des 50. Lebensjahres. Dies gilt auch für nachfolgende Angaben der Rentenzugänge wegen verminderter Erwerbsfähigkeit

Die vorausgegangen Daten belegen, dass psychische Erkrankungen nicht nur als Hauptdiagnose, sondern auch als komorbide Störung zu organischen Erkrankungen einen erheblichen Einfluss auf die Rente wegen verminderter Erwerbsfähigkeit haben (Mykletun et al., 2006).

Tabelle 3: Anzahl der Rentenzugänge wegen verminderter Erwerbsfähigkeit von 2013 nach 1. Diagnose (psychische und Verhaltensstörungen) nach Kapitel V Bereich F des ICD-10 (Gesundheitsberichterstattung des Bundes, 2013)

Diagnosegruppe	Anzahl
F00-F09 Organische, einschließlich symptomatischer psychischer Störungen	3013
F10-F19 Psychische und Verhaltensstörungen durch psychotrope Substanzen	8320
F20-F29 Schizophrenie, schizotype und wahnhafte Störungen	7430
F30-F39 affektive Störungen	32884
F40-F48 Neurotische, Belastungs- und somatoforme Störungen	15737
F40-F42 Phobische Störungen, andere Angststörungen und Zwangsstörungen	5519
F43 Reaktionen auf schwere Belastungen und Anpassungsstörungen	4671
F45 Somatoforme Störungen	4736
F48 Andere neurotische Störungen	442
F50-F59 Verhaltensauffälligkeiten mit körperlichen Störungen und Faktoren	401
F60-F69 Persönlichkeits- und Verhaltensstörungen	4199
F70-F79 Intelligenzminderung	2232
F80-F89 Entwicklungsstörungen	295
F90-F98 Verhaltens- und emotionale Störungen mit Beginn in der Kindheit und Jugend	169
F99 Nicht näher bezeichnete psychische Störungen	65

2.3 Häufigkeit von negativen Antwortverzerrungen

Um die Notwendigkeit von BVT für die sozialmedizinische Begutachtung zu verdeutlichen, sollten die Prävalenzen von negativen Antwortverzerrungen bei psychischen und psychosomatischen Erkrankungen betrachtet werden. Darüber hinaus werden die Basisraten von negativen Antwortverzerrungen für den rehabilitativen und klinischen Kontext dargestellt. Die Ermittlung der Häufigkeiten von negativen Antwortverzerrungen stellt ein großes methodisches Problem dar. Das einzig sichere Kriterium für die Erfassung von simulativen Tendenzen wäre das wahrheitsgemäße Eingeständnis eines Betroffenen, eine Erkrankung vorgetäuscht zu haben (Rosen, 2006). Aus diesem Grund basieren die meisten Schätzungen zu den Prävalenzen auf der Auswertung von auffälligen Testergebnissen in üblichen BVT.

Bei einer häufig zitierten multizentrischen Studie von Mittenberg, Patton, Canyock und Condit (2002) beruhten die ermittelten Häufigkeiten jedoch auf Schätzungen von 131 Mitgliedern des

American Board of Clinical Neuropsycchology (ABCN). Die Studie verdeutlichte, dass die Basisraten je nach Begutachtungskontext oder Diagnosegruppe stark schwanken (vgl. Tab. 4). Larrabee (2003) schlussfolgerte, dass die Prävalenz von negativen Antwortverzerrungen bei durchschnittlich 40 % liegt. Er bezog sich hierbei auf elf Studien, die im neuropsychologischen Kontext, bei denen ein sekundärer Krankheitsgewinn nicht auszuschließen war, Häufigkeiten von negativen Antwortverzerrungen ermittelten. Im Rahmen dieser elf Studien schwankten die Ergebnisse von 15 % bis 64 %. Es zeigt sich, dass die Angaben zur Schätzung der Prävalenzen nicht stabil sind und teilweise sogar mit über 50 % beziffert werden. Die Ergebnisse einer Erhebung in Großbritannien verdeutlichten (Gill, Green, Flaro & Pucci, 2007), dass 62 % der Personen, die aufgrund von Schmerzensgeldforderungen psychiatrisch untersucht wurden, auffällige Testergebnisse im MSVT aufwiesen. Ähnlich hohe Prävalenzen wurden in Studien mit geltend gemachten Schleudertraumata (Schmand, Lindeboom, Schagen, Heijt, Koene & Hamburger, 1998), leichten Schädel-Hirn-Verletzungen (Armistead-Jehle, 2010) sowie einer Studierendenstichprobe mit bestehendem Aufmerksamtkeits-Defizit-Syndrom (Sullivan, May & Galbally, 2007) ermittelt.

Tabelle 4: Ermittelte Prävalenzen negativer Antwortverzerrungen für unterschiedliche Begutachtungskontexte und Diagnosegruppen nach einer Studie von Mittenberg, Patton, Canyock und Condit (2002)

Begutachtungskontext	M (95 % CI)
Fälle mit Körperverletzung (Verletzungsfolgen)	28.66 (3.71)
Im Kontext von Arbeitsunfällen	30.12 (4.22)
Forensischer Kontext	19.25 (5.85)
Medizinische oder psychiatrische Fälle	8.11 (1.56)

Diagnosegruppe	M (95 % CI)
Leichtes Schädel-Hirn-Trauma	38.50 (7.06)
Fibromyalgie oder chronische Fatigue	34.74 (5.86)
Schmerz oder somatoforme Störungen	31.41 (5.67)
Neurotoxische Störungen	26.49 (5.98)
Verletzungen durch Strom	21.99 (6.02)
Depressive Störungen	14.95 (4.32)
Angststörungen	13.57 (4.26)
Dissoziative Störungen	10.51 (5.33)
Schlaganfall	9.35 (3.37)
Mittleres oder schweres Schädel-Hirn-Trauma	8.82 (2.00)
Vaskuläre Demenz	2.39 (1.02)

Anmerkung: M=berichteter Mittelwert in Prozent, 95 % CI= berichtetes 95 % Konfidenzintervall

Schätzungen der Basisrate liegen mittlerweile auch für deutsche Stichproben im Kontext der sozialmedizinischen Begutachtung vor (Stevens, Friedel, Mehren & Merten, 2008). Unter anderem zeigte sich, dass im Rahmen einer neurologischen Untersuchung 44.3 % von 235 Begutachteten auffällige Testergebnisse im WMT sowie MSVT aufwiesen. Bei einer Teilstichprobe von 47 Personen mit posttraumatischer Belastungsstörung lag die Basisrate mit 51 % sogar etwas höher (Merten, Friedel & Stevens, 2006). Eine zusätzliche Studie ermittelte, in der prospektiv Personen mit einer geltend gemachten posttraumatischen Belastungsstörung untersucht wurden, dass 70 % in einem von drei BVT (WMT, SFSS, MENT) auffällige Testergebnisse zeigten (Merten, Thies, Schneider & Stevens, 2009).

Der Großteil der bisher berichteten Ergebnisse zur Basisrate von negativen Antwortverzerrungen basiert auf Studien, die sich auf die sozialmedizinische Begutachtung bezogen. Diese Betrachtung macht dahingehend auch Sinn, da im gutachterlichen Kontext höchstwahrscheinlich mit einer erhöhten Häufigkeit von negativen Antwortverzerrungen zu rechnen ist, da besonders hier der sekundäre Krankheitsgewinn im Fokus steht (Johnson-Greene, Brooks & Ference, 2013). Jedoch zeigte eine Studie von Kobelt, Göbber, Bassler und Petermann (2012), dass außerdem im rehabilitativen Kontext mit Antwortverzerrungen zu rechnen ist. 40 % der untersuchten Patienten waren auffällig im SFSS und hatten gleichzeitig ein stärker ausgebildetes Rentenbegehren.

Die vorherigen Ausführungen belegen, dass negative Antwortverzerrungen in nicht unerheblichem Maße auftreten können. Besonders im Begutachtungskontext, bei dem ein sekundärer Krankheitsgewinn oftmals nicht auszuschließen ist, scheint die Basisrate hinreichend hoch zu sein (Dohrenbusch, 2009; Dohrenbusch, Henningsen & Merten, 2011). Die Schätzungen der Prävalenzen schwanken je nach Studie zwischen 20 und 60 %. Jedoch belegen die Untersuchungen immer das Auftreten von Antwortverzerrungen. Die Notwendigkeit einer Beschwerdenvalidierung im Begutachtungskontext wird deutlich. Dennoch sollte immer hinterfragt werden, aus welchen Bedingungen heraus negative Antwortverzerrungen resultieren, da sie nicht immer ein Resultat einer Simulation oder Aggravation sein müssen. Daher werden im folgenden Kapitel mögliche Einflüsse einer verzerrten Beschwerdendarstellung betrachtet.

2.4 Einflüsse der Beschwerdendarstellung

Warnungen und Coaching. Insbesondere durch das Internet haben Antragsteller der Rente wegen verminderter Erwerbsfähigkeit die Möglichkeit, sich über verfügbare Quellen auf die kommende sozialmedizinische Begutachtung vorzubereiten. Dies stellt gleichermaßen für den Gutachter sowie für den Testentwickler eines BVT eine große Herausforderung dar. Bisherige Studien in der sogenannten Coaching-Forschung erörterten vor allem die Fragestellung, inwiefern Verfahren resistent gegenüber solchen Instruktionen sind bzw. im welchem Ausmaß gezielt Anleitungen dargestellt werden müssen, um ein Testergebnis zu beeinflussen (Jelicic, Ceunen, Petersen & Merckelbach, 2011). Merten (2014) schlägt vier Stufen der Anleitung für die experimentelle Coaching-Forschung vor, die vorwiegend auf Analogstudien mit gesunden Teilnehmern zurückgreifen:

1. Naive Simulanten: Teilnehmer erhalten keinerlei Anleitung.
2. Symptom-Information oder Symptom-Coaching: Weitergabe von Informationen der Erkrankung.
3. Warning: Die Teilnehmer bzw. Patienten werden darüber informiert, dass in der nachfolgenden Untersuchung Verfahren eingesetzt werden, die negative Antwortverzerrungen erfassen können.
4. Coaching: Vermittlung von umfangreichen Strategien und Informationen über Verfahren, die negative Antwortverzerrung erfassen.

Im Rahmen der sozialmedizinischen Begutachtungen könnten Warnungen zudem eine Rolle spielen, wenn den Antragstellern erläutert wird, dass nachfolgende Verfahren eingesetzt werden, die ihre Beschwerdendarstellung validieren können. Bisherige Studien, die die Effekte von Warnungen untersucht haben, weisen heterogene Ergebnisse auf. Einerseits zeigen Studien, dass Anleitungen keinerlei Effekte auf die Beschwerdendarstellung haben und die eingesetzten Testverfahren relativ robust gegenüber Warnungen sind (Gunstad & Suhr, 2004; Johnson, Bellah, Dodge, Kelley & Livingston, 1998; Suhr, J., Gunstad, Greub & Barrash, 2004; Sullivan, K., Keane & Deffenti, 2001; Sullivan, K. & Richer, 2002; Wong, Lerner-Poppen & Durham, 1998). Andererseits verdeutlichen weitere Studien signifikante Effekte zwischen gewarnten und nicht gewarnten Teilnehmern (Dunn, Shear, Howe & Ris, 2003; Gunstad & Suhr, 2001; King & Sullivan, 2009; Schenk & Sullivan, 2010; Suhr, J. A. & Gunstad, 2000). Johnson und Lesniak-Karpiak (1997) schließen aus diesem Ergebnis, dass negative Antwortverzerrungen durch Warnungen reduziert

und dadurch die Antwortprofile der Teilnehmer valider werden. Diesen Ansatz kritisierten Youngjohn, Less-Haley und Binder (1999) und konstatierten, dass Warnungen die Validität von Verfahren verringern, wodurch Simulanten schlechter diagnostiziert werden können. Die Heterogenität der Studienergebnisse kann dadurch erklärt werden, dass es starke Differenzen in der methodischen Umsetzung des Analogstudiendesigns gibt. So schwankt beispielsweise die Darstellung der Szenarien stark. Weiterhin ist die Abgrenzung von Warnung und Coaching nicht immer eindeutig. Bisher fehlen weiterhin Studienergebnisse, die belegen, wie sich Instruktionen im *„realen"* klinischen Setting oder innerhalb einer Begutachtungssituation auswirken könnten.

Motivation. Das Auftreten von simulativen Tendenzen wird regelmäßig mit einem sekundären Krankheitsgewinn verknüpft (Sweet, Malina & Ecklund-Johnson, 2006). Hierbei steht das Motiv des persönlichen Gewinns im Vordergrund. Das bedeutet, dass Personen, die sich aufgrund ihrer Erkrankung beispielsweise eine Entschädigung oder finanzielle Vorteile erhoffen, eher dazu neigen, negative Antwortverzerrungen zu zeigen. Studien belegen, dass besonders das Motiv *„Aussicht auf eine Entschädigung"* dazu führt, dass negative Antwortverzerrungen dargeboten werden (Erdal, 2004). Auch in nicht-experimentellen Studien konnte verdeutlicht werden, dass Patienten mit einem geltend gemachten Schädelhirntrauma schlechter in neuropsychologischen Verfahren abschnitten, wenn sie gleichzeitig in ein Gerichtsverfahren involviert waren, in dem sie Schmerzensgeld forderten (Meyers & Volbrecht, 1998). Lanyon und Almer (2002) konnten dieses Ergebnis auch bei Klägern mit geltend gemachten psychischen Symptomen replizieren. Jedoch stellten sie weiterhin heraus, dass Patienten, die arbeitsunfähig waren, ihre Ansprüche jedoch nicht vor Gericht geltend machten, auch erhöhte Werte im MMPI-2 aufwiesen. Erdal (2009) schloss daraus, dass der alleinige Status der Arbeitsunfähigkeit bei den Patienten zur Emotionalität, Eigenfokussierung und einem Beachtungsbedürfnis führt. In ihrer Analogstudie verdeutlichte sie, dass die Teilnehmer der Motivationsbedingung *„Beachtungsbedürfnis"* genauso auffällig in BVT waren wie die Teilnehmer der Bedingung *„Aussicht auf eine Entschädigung"*. Einen Unterschied in ihrem Antwortverhalten gab es jedoch trotzdem. Probanden der Motivationsbedingung *„Beachtungsbedürfnis"* neigen dazu, viele falsche Antworten zu generieren, um auf ihre Beschwerden hinzuweisen. Wohingegen Teilnehmer mit dem Motiv *„Aussicht auf eine Entschädigung"* in neuropsychologischen Testverfahren viele Auslasser (Henry, 2005) zeigen.

Im vorliegenden Kapitel wurde eine Auswahl von Faktoren vorgestellt, die in der Literatur diskutiert werden und möglicherweise die Beschwerdendarstellung beeinflussen könnte. Studienergebnisse zu Warnungen bzw. Instruktionen sind sehr heterogen und wurden nicht im klinischen

Rahmen untersucht. Des Weiteren kann sowohl das Motiv *„Erhalt einer Entschädigung“* als auch *„Beachtungsbedürfnis“* zu einer verzerrten Beschwerdendarstellung führen. Ob der Einsatz von BVT in der sozialmedizinischen Begutachtung von psychischen Erkrankungen überhaupt notwendig und sogar ethisch vertretbar ist, wird im Folgenden eruiert.

2.5 Ethische Aspekte und die Kontroverse der Beschwerdenvalidierung

Bei der ethischen Beurteilung über den Einsatz von BVT sollte bedacht werden, dass es nicht möglich ist, ein Verfahren zu entwickeln, das eine 100 % Klassifikationsgenauigkeit ermöglicht. Eine — wenn auch geringe — Falsch-Positiv-Rate muss bei der Anwendung von BVT und Interpretation der daraus resultierenden Testergebnisse immer in Erwägung gezogen werden. Daraus ergeben sich verschiedene ethische Konflikte, die unterschiedlich bewertet werden. Die in Deutschland als ausreichend validiert erachteten und meist aus dem neuropsychologischen Bereich stammenden BVT weisen im Mittel eine Spezifität von 90 % auf (Widder, 2011). Das hat zur Folge, dass 10 % der Betroffenen fälschlicherweise als Simulanten diagnostiziert werden. Diese falsche Zuordnung kann zwangsweise schwerwiegende Konsequenzen für den Antragsteller haben und zu einer Stigmatisierung führen.

Die deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) sah sich in einer Stellungnahme (Dressing, Foerster, Widder, Schneider & Falkai, 2011) dazu veranlasst, den Einsatz von BVT kritisch zu hinterfragen. Die Autoren konstatieren, dass BVT vor allem für die Erfassung von negativen Antwortverzerrungen bei kognitiven Störungen konzipiert wurden und daher grundsätzlich bei der Begutachtung von psychiatrischen Störungen an ihre Grenzen stoßen. Sie stellen damit den Beitrag, den Neuropsychologen in der Begutachtung von psychischen Erkrankungen leisten können, infrage. Diese Kritik wird im Rahmen der Forschung aus verschiedenen Blickpunkten kontrovers diskutiert (Noeker & Petermann, 2011). Merten, Stevens und Blaskewitz (2007) diskutieren, dass neuropsychologische BVT auch bei psychischen sowie somatischen Erkrankungen Anwendung finden können, da bei diesen Störungen kognitive Einschränkungen zum Beschwerdenbild dazu gehören. Schmidt, Lanquillon und Ullmann (2011) betrachten diesen Einsatz kritisch und konstatieren, dass psychisch erkrankte Patienten nicht immer zugänglich für die Erfassung von kognitiven Einbußen sind, was nicht automatisch auf ein nicht-authentisches Antwortverhalten schließen lässt.

Zudem führt die DGPPN eine Studie von Foerster (1984) an, in der neurotische Rentenbewerber unabhängig vom Erhalt der Rente weiterhin Beeinträchtigungen in ihrer psychischen Gesundheit

zeigten. Sie schließen daraus, dass eine hohe Prävalenz von Simulation in der sozialmedizinischen Begutachtung von neurotischen Betroffenen nicht angenommen werden kann. In dieser Stellungnahme hinterfragen die Autoren außerdem, ob durch den alleinigen Einsatz von BVT Aggravation bzw. Simulation objektiv gemessen werden kann. Sie fordern ein kritisches Diskutieren der Methodik der BVT in der gutachterlichen Praxis, um einer „missbräuchlichen Anwendung" entgegenzuwirken.

2.6 Zusammenfassung

Studien zu den Häufigkeiten von negativen Antwortverzerrungen bzw. Simulationen zeigen, dass BVT vor allem im Kontext von Begutachtungen, bei denen ein sekundärer Krankheitsgewinn im Vordergrund steht, angebracht sind. Die Berentungen wegen verminderter Erwerbsfähigkeit aufgrund einer psychischen Erkrankung haben in den letzten fünf Jahren stark zugenommen, wodurch erhebliche volkswirtschaftliche Kosten verursacht werden können. Doch der Einsatz von BVT sollte stets ethisch hinterfragt werden, um auch der Stigmatisierung von Patienten als Simulanten entgegenzuwirken. Antwortverzerrungen sollten daher stets auf ihre zugrundeliegende Entstehung hin überprüft werden.

3 Fragestellung

Bei der Rente wegen verminderter Erwerbsfähigkeit spielen psychische Erkrankungen und vor allem affektive Störungen als Hauptdiagnose oder komorbide Störung bei vorwiegend somatischen Erkrankungen eine bedeutende Rolle. Zudem weisen Studien darauf hin, dass negative Antwortverzerrungen in nicht unerheblichem Maße auftreten und daher im Rahmen der sozialmedizinischen Begutachtung in Betracht gezogen werden sollten. Dabei sollte immer berücksichtigt werden, dass negative Antwortverzerrungen nicht ausnahmslos mit einer Simulation oder Aggravation gleichzustellen sind, sondern die zugrundeliegende Ursache des Antwortverhaltens betrachtet werden sollte. Bisher liegen besonders im neuropsychologischen Bereich Verfahren vor, die Simulation bzw. negative Antwortverzerrungen erfassen können.

In der vorliegenden Dissertation wird die Diagnostik der Beschwerdenvalidierung in der sozialmedizinischen Begutachtung beispielsweise bei der Berentung wegen verminderter Erwerbsfähigkeit aufgrund einer psychischen Störung fokussiert. Zunächst sollen hierbei bereits vorliegende Verfahren betrachtet werden und hinsichtlich ihrer Eignung für die sozialmedizinische Begutachtung von psychischen Störungen analysiert werden. Daraufhin soll ein geeigneter BVT konzipiert werden, dessen Güte auf der Grundlage eines Analogstudiendesigns überprüft wird. Darüber hinaus werden anhand einer klinischen Stichprobe etwaige Einflussfaktoren beleuchtet, die die Beschwerdendarstellung im Begutachtungskontext beeinflussen könnten. Auf der Grundlage dieser Überlegungen ergeben sich drei thematische Schwerpunkte mit dazugehörigen Fragestellungen.

Erster Schwerpunkt: Notwendigkeit des Einsatzes und der Neuentwicklung eines BVT für die sozialmedizinische Begutachtung

Im Kontext des ersten Schwerpunkts werden die bisherigen deutschsprachigen BVT gesichtet und deren Vor- und Nachteile für die sozialmedizinische Begutachtung von psychischen Störungen erörtert. Für die Beantwortung der folgenden Fragestellungen wurden die Ergebnisse der Studie I (vgl. Anhang A) herangezogen:

- Welche Verfahren liegen im deutschsprachigen Raum für die Beschwerdenvalidierung im Allgemeinen vor?
- Sind die vorhandenen Verfahren für den Einsatz in der sozialmedizinischen Begutachtung geeignet?

- Können die vorliegenden deutschsprachigen Verfahren Beschwerden bei psychischen Erkrankungen validieren?

Zweiter Schwerpunkt: Entwicklung und Validierung eines BVT für die Begutachtung der Rente wegen Erwerbsminderung aufgrund einer psychischen Störung

Der zweite inhaltliche Schwerpunkt ist dadurch gekennzeichnet, dass auf der Grundlage eines Analogstudiendesigns zunächst ein neuer BVT für psychische Erkrankungen entwickelt und darauffolgend validiert werden soll. Folgende Fragestellungen wurden mithilfe der Ergebnisse der Studien II (vgl. Anhang B) und IV (vgl. Anhang D) bearbeitet:

- Wie unterscheiden sich instruierte Simulanten, psychisch erkrankte Patienten und gesunde Studienteilnehmer in ihrem Antwortverhalten in Fragebogen?
- Trennen die entwickelten Items des neuen BVT zwischen den Gruppen instruierte Simulanten, psychisch erkrankte Patienten und gesunde Teilnehmer?
- Wie sind die Gütekriterien Sensitivität, Spezifität sowie prädiktive Werte für das neu entwickelte und validierte Verfahren einzuschätzen?

Dritter Schwerpunkt: Einflussfaktoren des Antwortverhaltens in Fragebogen bei Antragstellern der Rente wegen Erwerbsminderung aufgrund einer psychischen Störung

Im Rahmen des dritten inhaltlichen Schwerpunkts soll anhand der Antragsteller einer Rente wegen verminderter Erwerbsfähigkeit aufgrund einer psychischen Störung, Faktoren untersucht werden, die die Beschwerdendarstellung in Symptomfragebogen sowie BVT beeinflussen. Hierfür wurde einerseits der Einfluss von Instruktionen betrachtet, die die Antragsteller vor einer jeden Begutachtung erhalten. Andererseits wurden Motive für eine verzerrte Beschwerdendarstellung eruiert. Nachfolgende Fragestellungen wurden auf der Grundlage der Erkenntnisse der Studie III (vgl. Anhang C) untersucht:

- Welchen Einfluss haben Warnungen in der sozialmedizinischen Begutachtung im Rahmen eines Rentenantrags aufgrund einer psychischen Störung auf das Antwortverhalten der Antragsteller?

- Welche Motive können die Beschwerdendarstellung in der sozialmedizinischen Begutachtung im Rahmen eines Rentenantrags aufgrund einer psychischen Störung verändern?

4 Methode

Im Rahmen der vorliegenden Dissertation wurde für die Entwicklung des BEVA (Beschwerdenvalidierungstest) und die Beantwortung der damit in Zusammenhang stehenden Fragestellungen ein Analogstudiendesign genutzt (vgl. Studie II und IV). Dabei wurden Experten für die Itemgenerierung des BEVA befragt. Weiterhin wurde in einer Prä-Studie eine Studierendenstichprobe zur Generierung und Überprüfung von Items rekrutiert. Für die Konstruktion sowie Validierung des Screenings wurden Patienten mit depressiven Störungen sowie Schmerzerkrankungen, gesunde Teilnehmer und IS rekrutiert. Um etwaige Einflussfaktoren von negativen Antwortverzerrungen zu ermitteln, wurden zusätzlich die Daten von Antragstellern einer Rente wegen Erwerbsminderung ermittelt. Der Studie III liegt ein randomisiertes Kontrollgruppendesign zugrunde. Im Folgenden wird eine ausführliche Darstellung der Testentwicklung, der Stichprobe, der Studienabläufe, der verwendeten Messmethoden sowie der Datenanalyse erteilt.

4.1 Testentwicklung

Das Screening, das im Rahmen der vorliegenden Dissertation entwickelt wurde, basiert auf einer externalen Testkonstruktion (Amelang & Zielinski, 2004), da das Verfahren zwischen den Gruppen *real psychisch erkrankt* und *negative Antwortverzerrer* unterscheiden soll. Der Einsatzbereich des BEVA soll in erster Linie in der sozialmedizinischen Begutachtung von Personen mit einer depressiven Störung oder einer Schmerzerkrankung mit depressiver Symptomatik liegen. Um keinen Missmut bei den Antragstellern zu erwecken und dadurch eine erhöhte Abbruchrate auszulösen, wurden die Items nicht bizarr und absurd formuliert. Da das Bearbeiten des Fragebogens durch Antworttendenzen wie der Akquieszenz beeinflusst sein kann, wurden einige Aussagen positiv, also nicht zu einer Erkrankung gehörend, formuliert. Die Itemgenerierung beruhte auf folgenden Informationsquellen:

1. **Ergebnisse systematischer Literaturrecherche**

Auf der Grundlage einer systematischen Literaturrecherche mithilfe der Begriffe *„symptom validation“, „malingering“* und *„exageration“* wurden Studien ermittelt, die einen Beitrag zur Itemgenerierung leisten. Beispielweise zeigten Studien den Zusammenhang von Persönlichkeitsstilen und Simulation bzw. negativen Antwortverzerrungen (Cima, Pantus & Dams, 2007; Edens, Buffington & Tomicic, 2000; Grieve, 2012; Poythress, Edens & Watkins, 2001). Daher wurden Items zur Erfassung von antisozialer und negativistischer Persönlichkeitsausprägungen formuliert. Des Weiteren verdeutlichten Studien, dass Simulanten zu Extremantworten neigen (Giger,

Merten, Merckelbach & Oswald, 2010). Um dieses Antwortverhalten abbilden zu können, wurde eine vierstufige Skala gewählt. Anhand einer dichotomen Skala hätte diese Tendenz nicht ausreichend erfasst werden können, da die Varianz der Antwortmöglichkeiten eingeschränkt ist.

2. Expertenbefragung über Simulation, Aggravation bzw. negative Antwortverzerrungen

Vor der Beschreibung der Befragung, sei auf Kapitel 4.2 verwiesen, das eine ausführliche Darstellung der Zusammensetzung der Expertenstichprobe liefert. Die Experten erhielten per E-Mail einen Fragebogen, der aus allgemeinen Angaben sowie drei verschiedenen Modulen bestand. Anhand von offenen und geschlossenen Fragen wurden das Wissen und die Erfahrungen über Patienten mit negativen Antwortverzerrungen erfasst. Die Critical Incident Technique (CIT; Flanagan, 1954) wurde hierbei genutzt, um etwaige kritische Situationen, in denen Patienten negative Antwortverzerrungen zeigen, zu erläutern. Mithilfe dieser Technik konnten mögliche Einstellungen, Verhaltensweisen sowie Äußerungen von Patienten mit negativen Antwortverzerrungen erfasst und denkbare Messmethoden ermittelt werden. In einem weiteren Abschnitt sollten die Experten atypische Symptome für affektive und somatoforme Störungen generieren. Diese Aufgabe beruht auf der diagnostischen Strategie der seltenen Symptome (vgl. Kapitel 1.3). Auf der Grundlage dieser Angaben der Experten sollten Items generiert werden, die von Laien nur schwer als nicht zum Störungsbild zugehörige Symptome erkannt werden. In einem letzten Abschnitt sollten die Experten die bereits formulierten 26 Items auf ihre Eignung für die Erfassung von negativen Antwortverzerrungen beurteilen. Diese Einschätzung erfolgte dichotom in *geeignet* versus *ungeeignet*. Der vollständige Fragebogen ist dem Anhang E zu entnehmen.

Die Auswertung der offenen Fragen für den ersten und zweiten Abschnitt der Befragung erfolgte qualitativ. Zunächst wurden die Antworten aller Experten zu den Kategorien: allgemeine Angaben, affektive Störungen, somatoforme Störungen und Angststörungen zusammengefasst. Aus diesen Schlussfolgerungen wurde anschließend eine Tabelle ermittelt, die dazu diente, Items zu formulieren. Hierfür wurden inhaltlich äquivalente Äußerungen zusammengefasst oder redundant erscheinende Inhalte komplett entfernt. Beispielsweise wurden Formulierungen wie *„In der Regel neige ich dazu, das subjektive Leiden so zu beschreiben, wie der Patient das vorträgt.“* in der finalen Tabelle nicht aufgenommen, da es nicht möglich ist, ein äquivalentes Item für diese Äußerung zu formulieren, das etwaige Antwortverzerrungen erfassen würde. Die Auswertung des dritten Abschnitts des Expertenfragebogens sollte deskriptiv erfolgen, indem die Beurteilung der bereits generierten Items als *geeignet* oder *ungeeignet* von allen Experten ausgezählt wurden.

Jedoch lagen nur Antworten von fünf Experten vor. Dies führte dazu, dass die Auswertung ausblieb, da das Ergebnis als nicht aussagekräftig durch die Promovendin beurteilt wurde.

3. **Orientierung an bereits vorhandene Fragebogen und Skalen**

Für die Itemformulierung wurde sich unter anderem an bestehenden Items des SFSS orientiert, die für die Skala *„affektive Störung"* konstruiert wurden. Dabei konnten nicht alle Aussagen des SFSS genutzt werden, weil einige Items sehr bizarr verfasst wurden. Dies erscheint für die Anwendung in der sozialmedizinischen Begutachtung nicht geeignet. Weitere Testverfahren, die als Orientierung vor allem für die Konstruktion von Items der Persönlichkeitsstile dienten, waren der MMPI-2 sowie das Persönlichkeits-Stil und Störungsinventar (PSSI; Kuhl & Kazén, 2009).

Darüber hinaus wurde im Rahmen der Verfahrensentwicklung eine Prä-Testung mithilfe einer Studierendenstichprobe durchgeführt. Diese Studie diente vor allem dazu, den Ablauf des Analogstudiendesigns, die Verständlichkeit des Szenarios, der Items des BEVA sowie der Fragebogen *„Rollenkonformität"* und *„Rollenerfüllung"* zu überprüfen. Weiterhin wurden nach jeder Testung Anmerkungen und Anregungen zur Itemformulierung von den Studierenden eingeholt. Daraufhin wurde der Itempool des BEVA wiederum von 66 auf 76 Aussagen erweitert, deren Eignung im Rahmen der Konstruktion überprüft wurde.

4.2 Stichprobe

Im nachstehenden Abschnitt erfolgt eine Darstellung der Stichproben, die im Rahmen der promotionsrelevanten Studien rekrutiert wurden. Zunächst wird eine Beschreibung der Teilnehmer dargelegt (Expertenstichprobe und Prä-Test Stichprobe), deren Daten für die Testentwicklung von Bedeutung waren. Des Weiteren erfolgt eine gemeinsame Darstellung der Stichproben der Studien II und IV, da sich die Teilnehmer der Studie II aus einem Teil der Konstruktionsstichprobe der Studie IV ermittelten. Beide Stichproben waren Element eines Analogstudiendesigns und wurden für die Erhebungen der Konstruktion sowie Validierung des BEVA rekrutiert. Am Ende wird eine Erläuterung der Personen der Studie III gegeben, deren Daten für die Beantwortung der Frage nach möglichen Einflüssen in der Beschwerdendarstellung genutzt wurden.

4.2.1 Stichproben der Testentwicklung

Expertenstichprobe. Die Expertenstichprobe wurde über einen E-Mail-Verteiler der Deutschen Rentenversicherung Braunschweig-Hannover rekrutiert. In diesem Verteiler wurden alle Gutachter angesprochen, die für die Deutsche Rentenversicherung Braunschweig-Hannover sozialmedi-

zinische Begutachtungen von psychischen Erkrankungen durchführen. In der E-Mail wurde das Projekt beschrieben sowie darum gebeten an einer Expertenbefragung bezüglich der Entwicklung eines BVT teilzunehmen. Der Expertenfragebogen (vgl. Anhang E) war dem E-Mail Anhang zu entnehmen. Es wurden fünf ausgefüllte Exemplare von Gutachtern zurück gesendet. Allgemeine Angaben bezüglich der beruflichen Erfahrungen sowie die Expertise von negativen Antwortverzerrungen sind Tabelle 5 zu entnehmen.

Tabelle 5: Allgemeine Angaben der Expertenstichprobe

Allg. Angaben	Gutachter_1	Gutachterin_2	Gutachter_3	Gutachter_4	Gutachter_5
Facharzt für ...	Orthopädie	Psychiatrie	Neurologie, Psychiatrie	Psychiatrie, Psychotherapie, Sozialmedizin	Neurologie, Psychiatrie
Seit wann begutachten Sie?	1993	1984	1986	1991	1996
Wie viele Patienten begutachten Sie im Monat?	50	60	50	15	50
Wie oft hatten sie den Verdacht, dass Aggravationen bzw. Simulationstendenzen vorliegen?	40 von 50	30 von 60	10 von 50	2 von 15	50 von 60
Welche psychischen Erkrankungen treten am häufigsten in Ihrer Begutachtungspraxis auf?	Depressionen somatoforme Schmerzstörungen	Depressionen Suchterkrankungen somatoforme Störungen	Depressionen Schmerzen Tinnitus	Depressive Störungen	Abhängigkeitsstörungen

Prä-Test-Stichprobe. Die Prä-Test-Stichprobe setzte sich aus Studierenden der Psychologie des Bachelor- sowie des Masterstudiengangs der Universität Bremen des Wintersemesters 2012/2013 zusammen. Es konnten Daten von 87 Studierenden des 1. Mastersemesters sowie des 3. Mastersemesters gewonnen werden. Weiterhin wurden 40 Bachelor-Studierende des 5. Semesters rekrutiert. Dabei wurden insgesamt 84 Studierende instruiert, eine Depression oder eine somatoforme Störung in einem Fragebogen vorzutäuschen. 43 Personen haben den Fragebogen anhand ihres wirklichen derzeitigen Gesundheitszustands ausgefüllt.

4.2.2 Stichproben der Studien II und IV

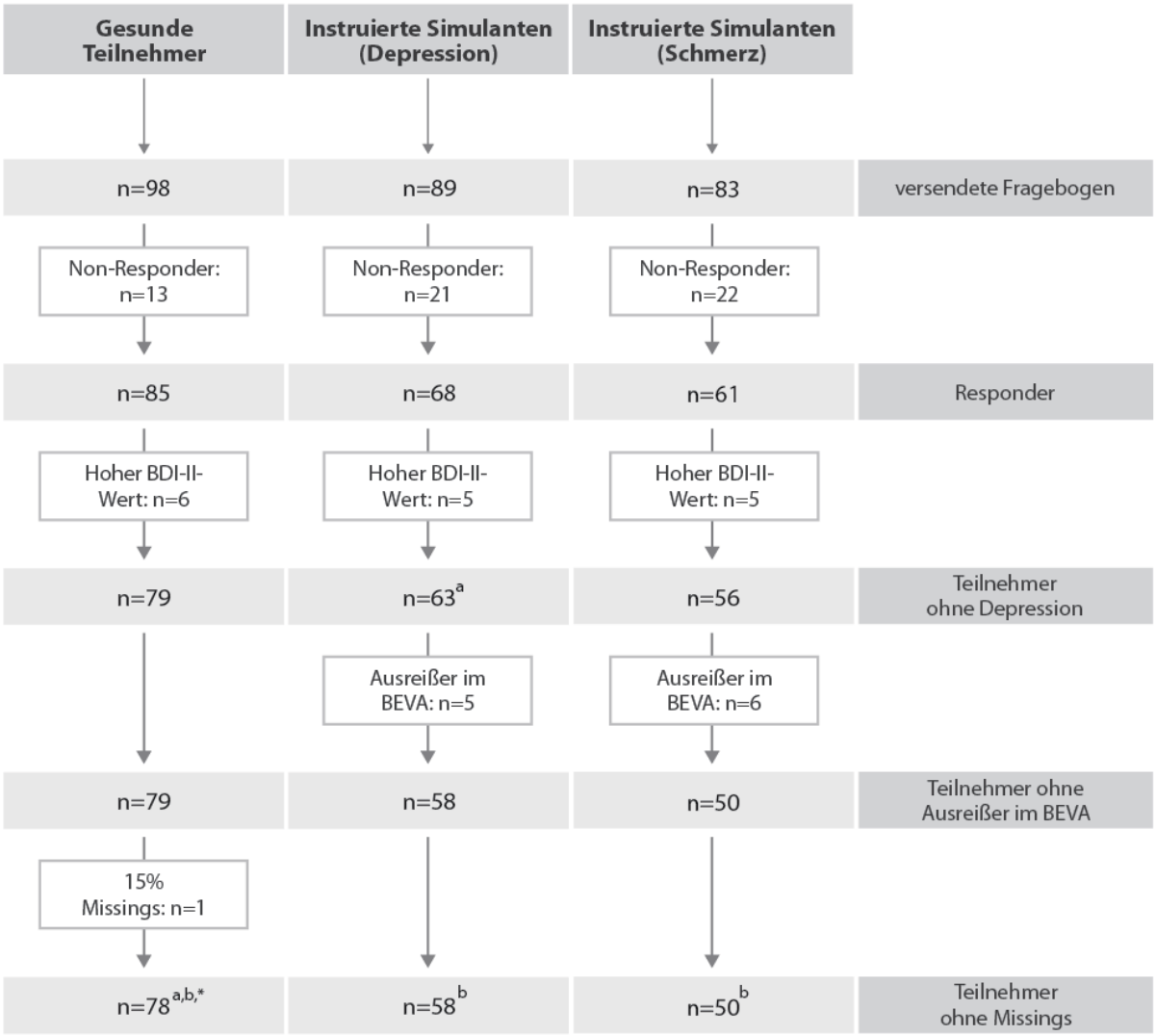

Abbildung 3: Anzahl der Teilnehmer der Konstruktion des BEVA sowie der Studie II mit jeweiligen Ausschlusskriterien
Anmerkung: BEVA= Beschwerdenvalidierungstest; BDI-II = Beck Depressionsinventar-II; hoher BDI-II Wert = Schweregrad vier und fünf; Ausreißer im BEVA = Summenwert liegt 20 Rohwerte unter dem Mittelwert der jeweiligen Gruppe liegt; Missings= fehlende Werte [a] Anzahl der Teilnehmer der Studie II; [b] Anzahl der Teilnehmer in der Konstruktion des BEVA; [*] von einem Teilnehmer sind die Daten nach Abschluss der Studie II eingetroffen

Instruierte Simulanten. Diese Teilstichprobe wurde mithilfe von Pressemitteilungen rekrutiert. Studieninteressierte wurden quasi-randomisiert der Instruktion „Depression“ bzw. „Schmerz“ zugeordnet. Gesucht wurden Personen, die aktuell keine depressive Symptomatik aufwiesen und im Alter von 40 Jahren (Konstruktion) bzw. 35 Jahren (Validierung) waren. In den Abbildungen 3 und 4 sind die endgültigen Fallzahlen der Studien II und IV sowie die Anzahl der ausgeschlossenen Teilnehmer bezüglich des jeweiligen Ausschlusskriteriums ersichtlich. Wenn Teilnehmer die

Aufgabenstellung nicht verstanden haben, in Beck-Depressionsinventar-II (BDI-II; Hautzinger, Keller & Kühner, 2009) einen Summenwert über 19 hatten oder 15 % Missings in den studienrelevanten Fragebogen aufwiesen, wurden sie aus der Studie ausgeschlossen. Das Aufgabenverständnis wurde anhand der Fragebogen Rollenkonformität und Rollenerfüllung gemessen. Wiederum wurden Personen aus der Datenanalyse ausgeschlossen, falls sie im BEVA einen Summenwert aufwiesen, der 20 Rohwerte unter dem Mittelwert der jeweiligen Gruppe lag.

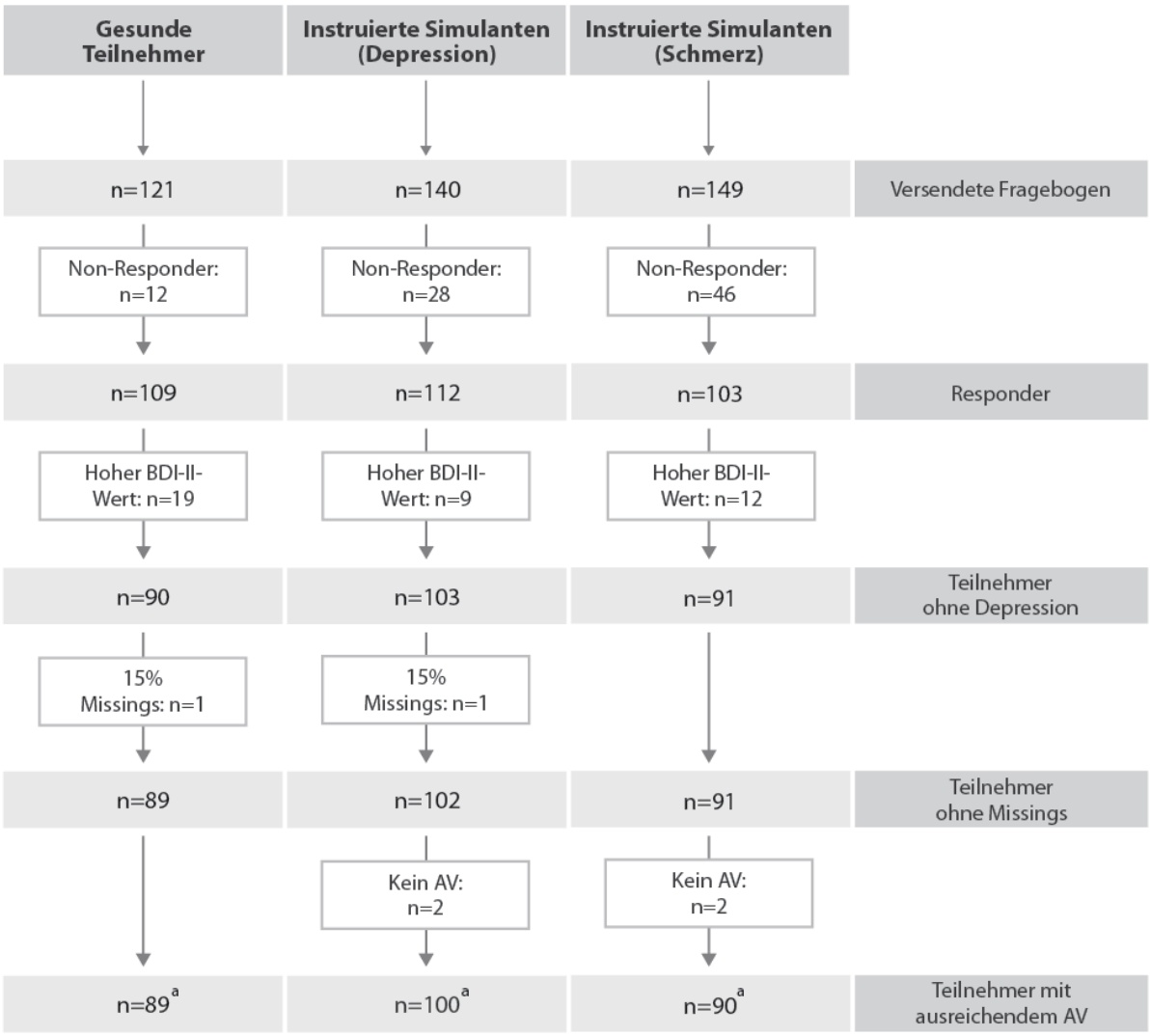

Abbildung 4: Anzahl der Teilnehmer der Validierung des BEVA mit jeweiligen Ausschlusskriterium
Anmerkung: BEVA= Beschwerdenvalidierungstest; BDI-II = Beck Depressionsinventar-II; hoher BDI-II Wert = Schweregrad vier und fünf; AV= Aufgabenverständnis; Missings= fehlende Werte [a] Anzahl der Teilnehmer der Validierung

Gesunde Teilnehmer. Die Kontrollgruppe wurde abermals über eine weitere Pressemitteilung gesucht (vgl. Abb. 3 und 4). Die Ein- und Ausschlusskriterien waren mit denen der IS identisch. Jedoch spielte in dieser Gruppe das Aufgabenverständnis keine bedeutsame Rolle, da die gesunden

Teilnehmer keine Erkrankung vortäuschen mussten. Wiederum wären Personen ausgeschlossen worden, falls ihr Rohwert im BEVA über 20 Wertpunkte unter dem Gruppenmittelwert gelegen hätte.

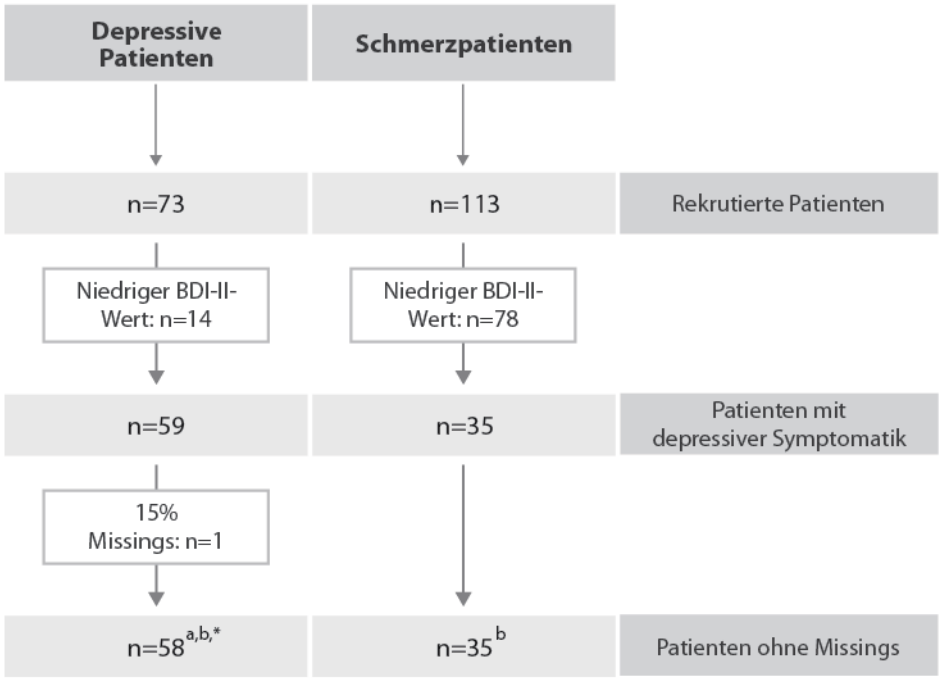

Abbildung 5: Anzahl der Patienten der Konstruktion des BEVA sowie der Studie II mit jeweiligen Ausschlusskriterien
Anmerkung: BEVA= Beschwerdenvalidierungstest; BDI-II = Beck Depressionsinventar-II; niedriger BDI-II Wert = Schweregrad eins bis drei (depressive Patienten); niedriger BDI-II Wert = Schweregrad eins und zwei (Schmerzpatienten); Missings= fehlende Werte [a] Anzahl der Teilnehmer der Studie II; [b] Anzahl der Teilnehmer in der Konstruktion des BEVA; [*] von vier Patienten sind die Daten nach Abschluss der Studie II eingetroffen

Depressive Patienten (DP). Die Abbildungen 5 und 6 zeigen die jeweiligen Fallzahlen der Studien II und IV. Aus ihnen wird ersichtlich, wie viele Patienten auf der Grundlage der Ein- und Ausschlusskriterien in die jeweiligen Studien integriert wurden. Die depressiven Patienten wiesen Diagnosen im Bereich F32-33 des Kapitels V des ICD-10 auf. Ausgeschlossen wurden Patienten mit folgenden komorbiden Störungen: organische, einschließlich symptomatischer psychischer Störungen (F00-F09), Schizophrenie, schizotype und wahnhafte Störungen (F20-F29), affektive Störungen mit manischen Episoden (F30), bipolare affektive Störungen (F31), dissoziative Störungen (F44) und Intelligenzminderung (F70-F79). Weitere Ausschlusskriterien waren: ein BDI-II Summenwert unter 20, 15 % Missings (fehlende Werte) in studienrelevanten Fragebogen sowie simulative Tendenzen (Angabe aus dem Therapeutenfragebogen). Die depressiven Patienten wurden mithilfe von Klinikpersonal und eingearbeiteten Forschungspraktikanten in den AMEOS Kli-

niken in Bremen und Hildesheim rekrutiert. Für die Validierung des BEVA wurden zusätzlich noch die Daten von Patienten in den Rehabilitationskliniken Oberharz und Bad Pyrmont erhoben. Alle Patienten wurden stationär behandelt und wiesen eine depressive Störung auf. Die Diagnose wurde von den behandelnden Ärzten bzw. Psychologen angegeben. Für die Rehabilitationspatienten lag sowohl die Einweisungsdiagnose als auch die Entlassungsdiagnose vor. Um zu gewährleisten, dass die Diagnosen vom Personal einer Klinik gestellt wurden, wurden ausschließlich die Entlassungsdiagnosen berücksichtigt.

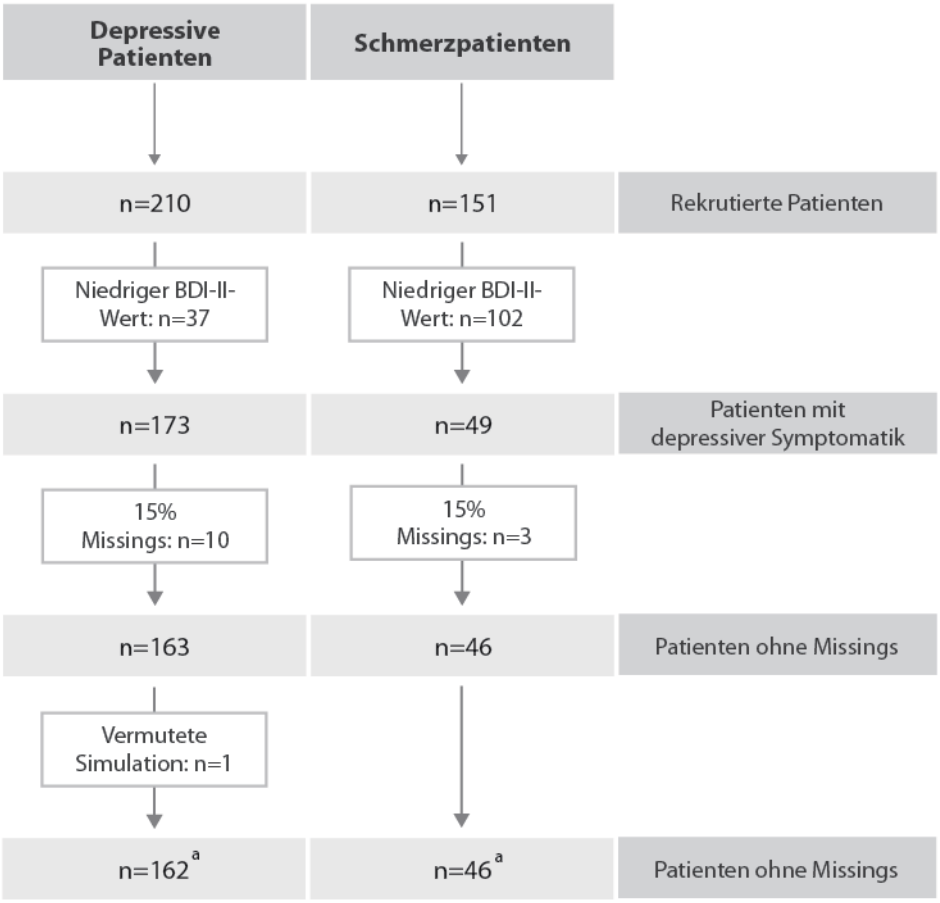

Abbildung 6: Anzahl der Patienten der Validierung des BEVA mit jeweiligen Ausschlusskriterien
Anmerkung: BEVA= Beschwerdenvalidierungstest; BDI-II= Beck Depressionsinventar-II ; niedriger BDI-II Wert = Schweregrad eins bis drei (depressive Patienten); niedriger BDI-II Wert = Schweregrad eins und zwei (Schmerzpatienten); Missings= fehlende Werte [a] Anzahl der Teilnehmer der Validierung des BEVA

Schmerzpatienten (SP). Diese Teilstichprobe wurde in der Rehabilitationsklinik Bad Eilsen rekrutiert. Die Patienten wiesen Diagnosen des Bereichs M (Krankheiten des Muskel-Skelett-Systems und Bindegewebes) des Kapitels XIII des ICD-10 auf sowie eine depressive Symptomatik, die anhand des BDI-II erfasst wurde. Für die Validierung des BEVA wurden zusätzlich Patienten mit einer somatoformen Schmerzstörung und depressiver Symptomatik eingeschlossen. Ebenso wird

aus den Abbildungen 5 und 6 ersichtlich wie viele Schmerzpatienten in den jeweiligen Studien rekrutiert wurden. Ausgeschlossen wurden Patienten mit einem Summenwert im BDI-II unter 14, 15 % fehlenden Werten in den studienrelevanten Fragebogen sowie simulativen Tendenzen (Angabe aus Therapeutenfragebogen).

Als Ergänzung zu den fehlenden Werte sei angemerkt, dass alle Teilnehmer der fünf Gruppen aus den jeweiligen Studien ausgeschlossen wurden, wenn mehr als ein Item je Skala der Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis (HEALTH-49; Rabung Harfst, Kawski, Koch, Wittchen & Schulz, 2009), drei Items des BDI-II sowie zehn (Konstruktion) bzw. acht (Validierung) Items des BEVA fehlten.

4.2.3 Stichprobe der Studie III

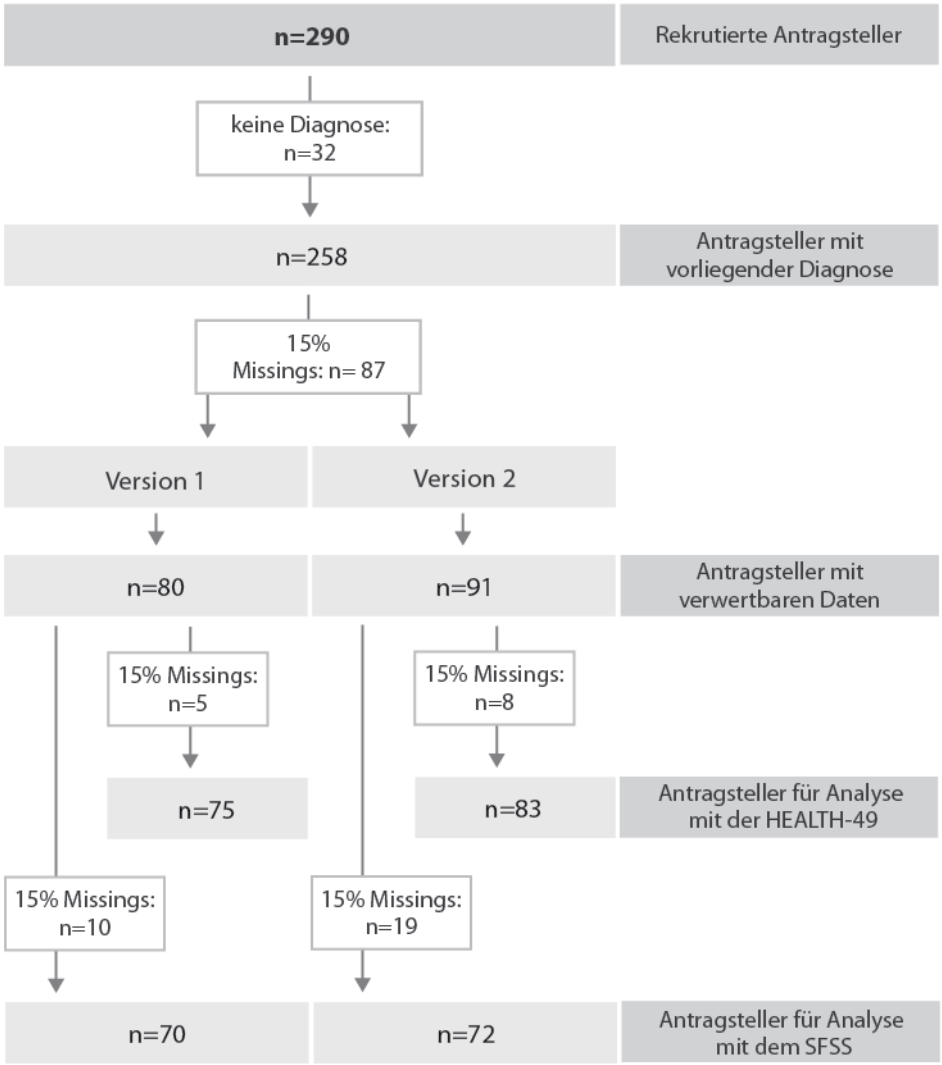

Abbildung 7: Anzahl der Antragsteller der Studie III mit jeweiligen Ausschlusskriterien
Anmerkung: HEALTH-49 = Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis; SFSS= Strukturierter Fragebogen Simulierter Symptome; Missings= fehlende Werte

Die Teilnehmer dieser Stichprobe wurden im Kontext der sozialmedizinischen Begutachtung der Rentenanträge wegen verminderter Erwerbsfähigkeit in den Begutachtungsstellen in Hannover und Braunschweig der Deutschen Rentenversicherung Braunschweig-Hannover rekrutiert. Die Erhebung fand im Zeitraum von Januar bis Juli 2013 statt. In den ungeraden Kalenderwochen dieses Zeitraums erhielten die Studienteilnehmer die Instruktionen mit Warnungshinweis. Wohingegen die Teilnehmer in den geraden Kalenderwochen des Studienzeitraums, die Instruktionen ohne Warnungshinweis bekamen. Die Antragsteller wurden demzufolge randomisiert den Bedingungen zugeordnet. In die Datenanalyse wurden Personen eingeschlossen, falls sie Diagnosen im Bereich F1, F3, F40-F42, F43.1, F43.2, F45, F60 oder F90 des Kapitels V des ICD-10 aufwiesen. In der Abbildung 7 ist die Anzahl der ausgeschlossenen Teilnehmer mit zugehörigem Kriterium dargestellt. Für die jeweiligen Teilfragestellungen ergeben sich immer wieder andere Fallzahlen, da Teilnehmer ausgeschlossen wurden, falls sie 15 % fehlende Werte in einem relevanten Fragebogen für die jeweilige Fragestellung zeigten. Das bedeutet, dass die Daten von Personen aus den Teilanalysen entfernt wurden, falls ein Item je Skala der HEALTH-49 sowie zehn Items des SFSS fehlten. Des Weiteren sind in Abbildung 7 die Daten von 142 Antragstellern dargestellt, die in die Analysen mit dem SFSS eingingen. Es wurde unter anderem eine Korrelation mit dem Summenwert des SFSS sowie verschiedenen Motivationsitems (vgl. Kapitel 4.4) gerechnet, für die sich folgende Fallzahlen ergaben: n= 112 *„finanzielle Sicherheit“*, n=114 *„Entlastung von beruflichen Anforderungen“*, n=106 *„nichts mehr wert zu sein“*, n=101 *„ausgeschlossen zu sein“* sowie n=115 *„eine Chance wieder gesund zu werden“.*

4.3 Studienablauf

Ablauf der Analogstudie. Das Ablaufschema für die Studien II und IV, in denen ein Analogstudiendesign genutzt wurde, ist in der nachfolgenden Tabelle 6 ersichtlich. Die eigens für die Studien entworfenen Fragebogen sind im Anhang F zu finden und werden im Kapitel 4.4 ausführlicher beschrieben. Die Datenerhebung der Konstruktion fand Juni bis August 2013 statt. Die Validierung dauerte von Januar bis August 2014.

Tabelle 6: Ablaufschritte der Analogstudie

	Instruierte Simulanten	**Kontrollgruppe**	**Patienten**
Anmerkungen	Quasi-Randomisierung zur der Gruppe „Depression“ oder „Schmerz“		Behandelndes Klinikpersonal erhielt Therapeutenfragebogen
1. Schritt	**Schriftliche Informationen und Ausfüllen der Einverständniserklärung**		
Anmerkungen	Postalischer Versand der Materialien inklusive Informationen zum Ablauf der Studie sowie Instruktionsbrief		Zusätzlich vom Klinikpersonal zum Ablauf der Studie instruiert
2. Schritt	**Soziodemografischer Fragebogen**		
Anmerkungen	Zusätzliche Frage nach der psychotherapeutischen Behandlung		
3. Schritt	**Beck Depressions-Inventar-II (BDI-II)**		
Nur von den instruierten Simulanten ausgefüllt	Szenario (Depression oder somatoforme Schmerzstörung)		
	Fragebogen zur Rollenkonformität		
4. Schritt	**Beschwerdenvalidierungstest (BEVA)**		
5. Schritt	**Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis (HEALTH-49)**		
Nur von den instruierten Simulanten ausgefüllt	Fragebogen zur Rollenerfüllung		

Bevor die IS eine psychische Erkrankung in den Fragebogen BEVA und HEALTH-49 vortäuschen sollten, erhielten sie ein Szenario, das es ihnen erleichtern sollte, sich in die kommende Situation einzufühlen. Das Szenario wurde wie folgt geschildert:

„Stellen Sie sich vor, Sie würden seit etwa 15 Jahren an immer wieder auftretenden Episoden einer Depression leiden. Aufgrund Ihrer psychischen Erkrankung befinden Sie sich zeitweilig in Behandlung. Sie haben verschiedene Therapieversuche unternommen und zuletzt nahmen Sie an einer medizinischen Rehabilitation teil. Allmählich verbessert sich Ihr Gesundheitszustand. Durch Ihre Erkrankung mussten Sie jedoch wiederholt Ihre Arbeit unterbrechen. Aufgrund der langen Abwesenheit von Ihrer Arbeitsstelle fühlen Sie sich dort sehr unbehaglich. Es kommt zu immer mehr Problemen. Umstrukturierungen in der Firma führen dazu, dass Sie zunehmend gestresst sind. Zudem fühlen Sie sich häufig von Ihren Arbeitskollegen missverstanden und zudem gemobbt. Es treten viele Konflikte auf. Ihr Vorgesetzter und Ihre Kollegen sehen Sie als eine Person mit Ecken und Kanten, die schwer zu integrieren sei.

Starke Leiden aufgrund Ihrer früheren Depression haben Sie nicht mehr, jedoch stresst Sie Ihre Arbeitsplatzsituation. Daher möchten Sie eine Erwerbsunfähigkeitsrente beantragen. Diese kann Ihnen aber nur aufgrund einer psychischen Erkrankung gewährleistet werden und nicht wegen der auftretenden Arbeitsplatzproblematik. Deshalb entscheiden Sie sich dafür, Ihr psychisches Wohlbefinden besonders problematisch darzustellen und eine massive Depression in der Begutachtung der Erwerbsunfähigkeit vorzutäuschen."

Ablauf Studie III. Der Ablauf der Studie III ist in der nachfolgenden Tabelle 7 dargestellt. Der in dieser Studie verwendete soziodemografische Fragebogen sowie die Patienteninformationen inklusive der Warnungs-Instruktionen sind wiederum dem Anhang G zu entnehmen. Die Materialien sind zusätzlich aufgeführt, da sie sich in einigen Aspekten von den in den anderen Studien dieser Dissertation verwendeten Materialien unterscheiden.

Tabelle 7: Ablaufschritte der Studie III

	Version I	Version II
Anmerkung	Die Antragsteller wurden randomisiert der jeweiligen Version zugeordnet.	
1. Schritt	Patienteninformation mit zusätzlicher Instruktion sowie Einverständniserklärung	Patienteninformation ohne zusätzlicher Instruktion sowie Einverständniserklärung
Anmerkung	Instruktion: „Mit einem Fragebogen haben Sie die Möglichkeit, Ihre psychischen Beschwerden genauer zu beschreiben, wobei die Testung gleichzeitig die Nachvollziehbarkeit der Angaben erlaubt."	
2. Schritt	Soziodemografischer Fragebogen	
Anmerkung	Der Fragebogen beinhaltet ein Item zur Motivation des Erhalts einer Erwerbsunfähigkeitsrente	
3. Schritt	Modul A der Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis (HEALTH-49)	
4. Schritt	Strukturierter Fragebogen Simulierter Symptome (SFSS)	

4.4 Messinstrumente

Das folgende Kapitel stellt die im Rahmen dieser Dissertation eingesetzten Verfahren vor. Es gliedert sich weiterhin in die Abschnitte *nicht standardisierte Fragebogen* und *standardisierte Fragebogen*. Darüber hinaus ist aus Tabelle 8 am Ende des Abschnitts erkennbar, welche gemessenen Konstrukte durch die Verfahren erfasst wurden und in welchen empirischen Studien ihre Ergebnisse einflossen. Weiterhin sind die selbst entwickelten Fragebogen, die in diesem Kapitel unter der Überschrift *nicht standardisierte Fragebogen* beschrieben werden, im Anhang F und G ein-

sehbar. Zusätzlich werden die Testverfahren vorgestellt, die für die Testentwicklung des BEVA hinzugezogen wurden. Diese sind im Abschnitt *standardisierte Fragebogen* ersichtlich.

4.4.1 Nicht standardisierte Fragebogen

Soziodemografischer Fragebogen. Dieser Fragenkatalog enthält übliche Items zur Person wie Alter, Geschlecht, Staatsangehörigkeit, Schulbildung, Berufsausbildung und Vorerkrankungen. Die gesunden Studienteilnehmer erhielten zusätzlich eine Frage zu einer möglichen derzeitigen bzw. früheren psychotherapeutischen Behandlung. Bei der Erhebung der Stichprobe zu Studie III wies der soziodemografische Fragebogen Abweichung auf, da die Studie in dem Kontext der sozialmedizinischen Begutachtung integriert war und weitere Fragestellungen beantwortet werden sollten. So wurde durch das zusätzliche Item *„Sind Sie selbst, Ihre Eltern, oder Ihre Großeltern in einem anderen Land als Deutschland geboren?“* nach der Definition des Statistischen Bundesamts (2013) der Migrationshintergrund erfasst. Weiterhin lagen keine Fragen zur Schulbildung sowie Berufsausbildung vor, sondern zur letzten Tätigkeit. Zudem war das Item bezüglich der Motivation (s.u.) integriert und wird nachfolgend separat beschrieben.

Motivation. Die Motivation für den Erhalt einer Rente wurde anhand eines Items aus dem soziodemografischen Fragebogen ermittelt. Mithilfe einer fünfstufigen Skala die von 1= *trifft voll zu* bis 5= *trifft gar nicht zu* reichte, sollten die Antragsteller folgende sechs Aussagen bewerten: Die Rente bedeutet für mich …

1. … finanzielle Sicherheit
2. … Entlastung von beruflichen Anforderungen
3. … nichts mehr wert zu sein
4. … kein Hartz-IV Empfänger zu sein
5. … ausgeschlossen zu sein
6. … eine Chance wieder gesund zu sein

Fragebogen zur Rollenkonformität. Bei der Entwicklung dieses Fragebogens wurde sich an übliche Analogstudien und insbesondere an deren Ausführungen über die Beschreibung von möglichen Items orientiert (Merten, Green, Henry, Blaskewitz & Brockhaus, 2005; Sullivan, K. A. & Elliott, 2012; Tydecks, Merten & Gubbay, 2006). Er sollte das Verständnis für die Aufgabe „Vortäuschen einer psychischen Störung“ der IS überprüfen. Ein Anliegen dieses Fragebogens lag darin, mit eigenen Worten die vorliegende Aufgabe des Vortäuschens einer Störung darzulegen. Zunächst sollten die Teilnehmer jedoch bewerten, ob sie mit der vorliegenden Aufgabe einverstanden sind

(Antwortmöglichkeit *„ja“* vs. *„nein“*). Anhand von drei Antwortalternativen sollten die IS zusätzlich vier Aussagen und Fragen zum Szenario sowie zur Aufgabe beantworten. Es lagen Fragebogen für die IS (Depression) und IS (Schmerz) vor. War ein Item von sechs auffällig beantwortet, so wurde davon ausgegangen, dass die Teilnehmer die Aufgabe nicht verstanden hatten bzw. nicht erfüllen wollten. Daraufhin wurden die Probanden aus der Studie ausgeschlossen.

Fragebogen zur Rollenerfüllung. Bei der Entwicklung dieses Fragebogens wurde sich an gängigen Analogstudien orientiert. Dieser Fragebogen sollte das Erfüllen der Aufgabe „*Vortäuschen einer psychischen Störung“* am Ende des Fragebogenkatalogs erfassen. Das Verfahren enthält sechs Items. Es lagen differenzierte Versionen für die IS (Depression) und die IS (Schmerz) vor. Zunächst sollten die Teilnehmer einschätzen, ob sie den Fragebogen ihrer Meinung nach aus der Perspektive eines Rentenantragstellers beantwortet haben (*„ja“ vs. „nein“*). Danach folgten vier Fragen zur Einschätzung der Rollenerfüllung, anhand derer beispielsweise die Schwierigkeit der Aufgabe bewerten werden sollte. Am Ende wurde sich durch eine offene Frage nach Strategien zur Aufgabenlösung erkundigt. Falls die Teilnehmer ein Item von sechs auffällig beantworteten, wurden sie aus der Studie ausgeschlossen.

Therapeutenfragebogen. Die behandelnden Ärzte bzw. Psychologen sowie sozialmedizinischen Gutachter sollten auf der Grundlage der Basisdokumentation die Hauptdiagnosen sowie die komorbiden Störungen der Patienten angeben. Zusätzlich sollten die Ärzte oder Psychologen anhand zweier dreistufigen Skalen (0=*liegt vor*; 1= *liegt vielleicht vor*, 2= *liegt nicht vor*) einstufen, ob beim Patienten Aggravations- bzw. Simulationstendenzen vorliegen. Patienten wurden ausgeschlossen, falls Simulationstendenzen vielleicht bzw. eindeutig vorlagen. Lagen Aggravationen vor wurden die Daten dieser Personen ebenfalls aus den Analysen entfernt.

4.4.2 Standardisierte Testverfahren

BDI-II. Das BDI-II gilt als ein Selbstbeurteilungsfragebogen zur Erfassung der Schwere einer depressiven Symptomatik bei Jugendlichen ab 13 Jahren und Erwachsenen. Die Autoren orientierten sich bei der Konstruktion des Verfahrens an den Kriterien einer Major Depression des Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV; APA, 2000). Die Personen sollen anhand einer vierstufigen Skala einschätzen, ob die beschriebene Symptomatik in den letzten zwei Wochen bei ihnen vorlag. Falls die Symptomatik gar nicht vorhanden ist, sollte eine null angekreuzt werden und eine drei, wenn eine schwerwiegende Symptomatik vorliegt. Durch die Addition aller angekreuzten Aussagen kann ein Gesamtwert berechnet werden, der zwischen null und ein-

schließlich 63 liegen kann. Anhand einer Receiver Operating Characteristics (ROC)-Kurve haben Beck, Steer und Brown (1996) Grenzwerte für den Schweregrad einer Depression bei Patienten mit einer diagnostizierten Depression ermittelt, die wie folgt definiert sind: 0-8= *keine Depression*, 9-13= *minimale Depression*, 14-19= *leichte Depression*, 20-28= *mittelschwere Depression* sowie 29-63= *schwere Depression*. Dabei geben die Autoren an, dass sie die Sensitivität als wichtiger erachtet haben als die Spezifität, daher dem Vorgehen der meisten Kliniker gefolgt sind und eine niedrige Schwelle für die Erklärung einer Depression akzeptiert haben. Damit soll die Wahrscheinlichkeit minimiert werden, eine Person mit einer depressiven Störung fälschlicherweise nicht zu erkennen. Das Verfahren wurde bisher international sowie national vielfach auf seine Gütekriterien hin untersucht. Es konnten interne Konsistenzen zwischen alpha= .84 und . 94 sowie Retest-Reliabilitäten von rtt= .93 belegt werden.

HEALTH-49. Die HEALTH-49 stellen ein Selbstbeurteilungsfragebogen dar, der die Aspekte von psychosozialer Gesundheit erfassen soll. Sein modularer Aufbau erlaubt es, den Fragebogen als Gesamtes oder auch nur einzelne Teile einzusetzen. Das vorliegende Instrument besteht aus sechs Modulen, die wie folgt beschrieben werden: Modul A psychische und somatoforme Beschwerden (18 Items); Modul B psychisches Wohlbefinden (5 Items); Modul C interaktionelle Schwierigkeiten (7 Items); Modul D Selbstwirksamkeit (5 Items); Modul E Aktivität und Partizipation (6 Items); Modul F soziale Unterstützung/Belastung (8 Items). Weiterhin können im Modul A ein Gesamtwert bestimmt sowie drei Subskalen gebildet werden, die folgendermaßen definiert sind: somatoforme Beschwerden (7 Items); Depressivität (6 Items) sowie phobische Ängste (5 Items). Die interne Konsistenz liegt je Skala zwischen alpha= .76 und .91. Für die HEALTH-49 liegen Prozentrang- und t-Werte für gesunde Personen und unterschiedliche Patientengruppen vor.

SFSS. Der SFSS stellt ein aus dem englischen adaptierter Selbstbeurteilungsfragebogen zur Erfassung von Simulation und simulierten Symptomen dar. Der Fragebogen besteht aus 5 Skalen je 15 Items, die folgendermaßen definiert sind: niedrige Intelligenz, affektive Störungen, neurologische Beeinträchtigungen, Psychosen und mnestische Störungen. Das Antwortformat ist dichotom (*„wahr“* vs. *„falsch“*). Die Items wurden auf der Grundlage verschiedener Quellen gebildet. Einerseits wurden Fragen aus vorhandenen Verfahren wie dem MMPI, SIRS oder Wechsler Adult Scale-IV (WAIS-IV; Petermann, 2012) umformuliert. Andererseits wurden Items auf der Basis früherer Arbeiten der Autoren über die Charakteristika von Simulanten konstruiert. Die Items repräsentieren bizarre Inhalte, unrealistische Symptome oder stellen bestimmte Syndrome dar, die nur von einem Experten als atypisch erkannt werden können. Es kann je Skala ein Wert sowie ein

Gesamtwert ermittelt werden. Für die deutsche Adaption konnte ein Cut-Off-Wert von 16 für den Gesamtwert sowie eine Sensitivität von .87 bei gleichzeitiger Spezifität von .87 ermittelt werden. Smith und Burger (1997) stellten heraus, dass der Gesamtwert der effizienteste Indikator für eine Simulation darstellt, da die Subskalen in ihren einzelnen Sensitivitäten stark variierten. Jedoch können sie aus Sicht der Autoren eine hinreichend qualitative Information über die simulierten Syndrome geben.

PSSI. Das PSSI stellt ein Selbstbeurteilungsverfahren zur Erfassung der Ausprägung von Persönlichkeitsstilen dar, die pathologisch nicht bedeutsam sind, sich jedoch an den Persönlichkeitsstörungen der ICD-10 sowie des DSM-IV orientieren. Der Fragebogen enthält 14 Skalen mit je zehn Items, die folgendermaßen benannt sind: eigenwillig-paranoid, zurückhaltend-schizoid, ahnungsvoll-schizotypisch, spontan-borderline, liebenswürdig-histrionisch, ehrgeizig-narzisstisch, selbstkritisch-selbstunsicher, loyal-abhängig, sorgfältig-zwanghaft, kritisch-negativistisch, still-depressiv, hilfsbereit-selbstlos, optimistisch-rhapsodisch und selbstbehauptend-antisozial. Anhand einer vierstufigen Skala von *„trifft gar nicht zu"* bis *„trifft ausgesprochen zu"* sollen Personen die 140 Aussagen bewerten. Cronbachs alpha schwankt je Skala von .73 bis .85. Die Retest-Reliabilität (sechs Wochen) liegt je Skala zwischen rtt= .68 und .83. Für das Inventar sind alters- und geschlechtsspezifische Prozentrang- und T-Werte vorhanden. Darüber hinaus liegen Normen für sechs verschiedene Patientengruppen vor.

MMPI-2. Das MMPI-2 ist ein adaptierter und neu normierter klinischer Persönlichkeitstest, der aus 567 Items besteht. Er enthält drei Gültigkeitsskalen und zehn Basisskalen, die aus 370 Items gebildet werden können und bereits in der Vorgängerversion des MMPI enthalten waren. Neu hinzugekommen sind Skalen für klinische relevante Bereiche wie Medikamentenmissbrauch oder Suizidtendenzen sowie weitere Validitätsskalen. Das MMPI ist ein Musterbeispiel für ein Verfahren, das auf der Grundlage einer rein empirischen Testkonstruktion Items zu entsprechenden Skalen zugeordnet hat. Das Verfahren wurde umfangreich an einer deutschen repräsentativen Stichprobe normiert und umfasst geschlechtsspezifische Normen. Die interne Konsistenz für die Basisskalen liegt zwischen alpha= .60 und .88 sowie die der rational/faktorenanalytischen Inhaltsskalen bei alpha= .71 und .87. Eine Retestreliabilität konnte für die Basisskalen zwischen rtt= . 62 sowie .92 und für die Inhaltsskalen zwischen rtt= .63 und .91 ermittelt werden.

Tabelle 8: Einsatzbereich, gemessenes Konstrukt und etwaige Anmerkungen zu allen Verfahren, die in den Studien der vorliegenden Dissertation genutzt wurden

Einsatzbereich	Anmerkungen	Gemessenes Konstrukt
Soziodemografischer Fragebogen		
Studie II - IV	In der Studie III etwas abgewandelter Fragebogen und enthält das Motivationsitem	Repräsentativität der Stichprobe in Bezug auf die Allgemeinbevölkerung
Studie III	Zuzügliches Item: „Sind Sie selbst, Ihre Eltern oder Ihre Großeltern in einem anderen Land als Deutschland geboren?“	Erfassung des Migrationshintergrunds
Motivation		
Studie III		Misst die Motivation für den Erhalt einer Rente
Fragebogen zur Rollenkonformität		
Studie II und IV	Erhielten nur die instruierten Simulanten	Erfasst das Verständnis der vorgegebenen Aufgabe „Vortäuschung einer psychischen Störung“
Fragebogen zur Rollenerfüllung		
Studie II und IV	Erhielten nur die instruierten Simulanten	Erfasst am Ende der Studie, ob die Aufgabe „Vortäuschen einer psychischen Störung“ durch den Teilnehmer konform erfüllt wurde
Therapeutenfragebogen		
Studie II - IV		Erfasst die Hauptdiagnosen und komorbiden Störungen der Patientengruppe sowie vermutete Aggravations- und Simulationstendenzen
Beck Depressions-Inventar (BDI-II)		
Studie II und IV		Erfasst den Schweregrad der depressiven Symptomatik (Ausschlusskriterium)
Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis (HEALTH-49)		
Studie II und III	Nur Modul A	Erfasst klinische Symptome wie Depressivität, somatoforme Beschwerden und phobische Ängste
Studie IV	Diente als „passives“ Verfahren im Analogstudiendesign	
Strukturierter Fragebogen Simulierter Symptome (SFSS)		
Studie III		Erfasst simulative Tendenzen
Studie IV	Ausschließlich Skala „Depression“	Itemgenerierung
Persönlichkeits-Stil und Störungs-Inventar (PSSI)		

Fortsetzung Tabelle 8

Studie IV	Nur die Skalen kritisch-negativistisch, selbstbehauptend-antisozial	Itemgenerierung
Minnesota Multiphasic Personality Inventory-II (MMPI-II)		
Studie IV	Nur die Skalen soziale Verantwortung, antisoziale Einstellungen, schlechte Meinung, Fake-Bad Scale und Moral	Itemgenerierung

4.5 Datenanalyse

Alle statistischen Analysen erfolgten mit IBM SPSS 20.0 bzw. dem Nachfolgeprogramm IBM SPSS 22.0. Innerhalb der promotionsrelevanten Studien kamen verschiedene statistische Verfahren zum Einsatz, die im Überblick in der Tabelle 9 aufgelistet sind. Genauere Beschreibungen können den jeweiligen Publikationen entnommen werden.

Tabelle 9: Auflistung und Erklärung der statistischen Analysen aller promotionsrelevanten Studien und der Dissertation

Studie II	
Statistisches Analyseverfahren	**Erklärung**
Kolmogorov-Smirnov-Test	Überprüfung der Normalverteilungsannahme für die Daten der einzelnen Gruppen in allen studienrelevanten Skalen
Mann Whitney U-Test	Zusätzliche Überprüfung der Gruppenunterschiede für non-parametrische Daten bei Verletzung der Normalverteilungsannahme.
Chi^2-Statistik nach Roderick J.A. Little	Analyse von fehlenden Werten in allen studienrelevanten Skalen
Expectation-Maximization-Algorithmus (EM-Algorithmus)	Diente als Verfahren für die Datenersetzung fehlender Werte bei allen studienrelevanten Skalen
Univariate Varianzanalysen (ANOVAs)	Überprüfung der Gruppenunterschiede in den studienrelevanten Skalen sowie für das Alter
Post hoc t-Tests	Überprüfung der Gruppenunterschiede in den jeweiligen studienrelevanten Skalen
Effektgröße Cohens d	Bestimmung der Größe des statistischen Effekts der Gruppenunterschiede in den studienrelevanten Skalen
Chi^2-Test	Überprüfung der Gruppenunterschiede für non-parametrische Daten in den soziodemografischen Variablen (ausgenommen Alter)

Studie III	
Statistisches Analyseverfahren	**Erklärung**

Fortsetzung Tabelle 9

Kolmogorov-Smirnov-Test	Überprüfung der Normalverteilung je Gruppe in allen studienrelevanten Skalen
Chi^2-Statistik nach Roderick J.A. Little	Analyse von fehlenden Werten bei allen studienrelevanten Skalen
Expectation-Maximization- (EM-Algorithmus)	Diente als Verfahren für die Datenersetzung bei allen studienrelevanten Skalen
Univariate Varianzanalysen (ANOVAs)	Überprüfung der Gruppenunterschiede in den studienrelevanten Skalen sowie für das Alter
Effektgröße Cohens d	Statistisches Maß, zur Bestimmung der Größe des statistischen Effekts der Gruppenunterschiede in den studienrelevanten Skalen
Chi^2-Test	Überprüfung der Gruppenunterschiede in den personenbezogenen Variablen (ausgenommen Alter) sowie in der Anzahl der negativen Antwortverzerrungen
Risikomaß Odds Ratio (OR)	Berechnung der Wahrscheinlichkeit negative Antwortverzerrungen zu zeigen, falls die Personen gewarnt wurden
Korrelationskoeffizient Spearmans-Rho	Darstellung eines linearen Zusammenhangs der Variablen Motivation und Summenwert im SFSS
Fishers z-Transformation	Überprüfung der Unterschiede zwischen den berechneten Korrelationskoeffizienten für die jeweiligen Gruppen
Lineare Regression	Analyse, um etwaige Motive als Prädiktoren zu ermitteln, die das Kriterium Summenwert im SFSS voraussagen können

Studie IV	
Statistisches Analyseverfahren	**Erklärung**
Chi^2-Statistik nach Roderick J.A. Little	Analyse von fehlenden Werten für den BEVA Skalen
Mann Whitney U-Test	Überprüfung der Gruppenunterschiede je Item des BEVA für die Konstruktion sowie Validierung
Kolmogorov-Smirnov-Test	Überprüfung der Normalverteilung je Gruppe für die Variable Alter sowie Summenwert des BEVA
Univariate Varianzanalysen (ANOVA)	Überprüfung der Gruppenunterschiede für die Variable Alter sowie den Summenwert des BEVA
Post hoc t-Tests	Überprüfung der Gruppenunterschiede des Summenwerts im BEVA
Effektgröße Cohens d	Statistisches Maß, zur Bestimmung der Größe des statistischen Effekts der Gruppenunterschiede in den studienrelevanten Skalen
Chi^2-Test	Überprüfung der Gruppenunterschiede in den soziodemografischen Variablen (ausgenommen Alter)

Fortsetzung Tabelle 9

Receiver operating charateristics curve (ROC)	Ermittlung der Cut-Off-Werte im BEVA in der Konstruktionsstichprobe
Kontingenztafel	Ermittlung von Spezifität, Sensitivität, Negativ Prädiktiven Wert (NPW) und Positiv Prädiktiven Wert (PPW)
Cronbachs alpha	Ermittlung der internen Konsistenz des BEVA

Dissertation	
Statistisches Analyseverfahren	**Erklärung**
Kolmogorov-Smirnov-Test	Überprüfung der Normalverteilung für die Gruppen IS (Schmerz) und Schmerzpatienten in den Skalen der HEALTH-49 des Moduls A
Mann Whitney U-Test	Zusätzliche Überprüfung der Gruppenunterschiede bei Verletzung der Normalverteilungs-Voraussetzung
Univariate Varianzanalysen (ANOVAs)	Überprüfung der Gruppenunterschiede in den Skalen der HEALTH-49 des Moduls A sowie für das Alter
Post hoc t-Tests	Überprüfung der Gruppenunterschiede in den Skalen der HEALTH-49 des Moduls A
Effektgröße Cohens d	Statistisches Maß, zur Bestimmung der Größe des statistischen Effekts der Gruppenunterschiede in den Skalen der HEALTH-49 des Moduls A
Chi^2-Test	Überprüfung der Gruppenunterschieden in den soziodemografischen Variablen (ausgenommen Alter)

Anmerkung: BEVA= Beschwerdenvalidierungstest; SFSS= Strukturierter Fragebogen Simulierter Symptome; IS= instruierte Simulanten; HEALTH-49 = Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis

Weiterhin sollte anhand einer *theoretischen Überblicksarbeit* (vgl. Studie 1) ermittelt werden, inwiefern derzeitige deutschsprachige BVT dafür geeignet sind, um bei Patienten mit psychischen Erkrankungen in einer sozialmedizinischen Begutachtung negative Antwortverzerrungen bzw. Simulation oder Aggravation zu erfassen. Darüber hinaus sollte die Notwendigkeit einer Verfahrenskonstruktion erörtert werden. Hierfür wurden beispielhaft gängige Diagnoseansätze und übliche deutschsprachige Verfahren einbezogen und bezüglich ihrer Eignung überprüft. Anhand einer Literaturrecherche über die Datenbanken „PSYNDEX“ und „web of knowledge“ wurden Publikationen mithilfe der Stichwörter *„malingering“* und *„symptom validation“* gesucht. Darauffolgend wurden die so ermittelten BVT als Schlüsselwörter für die Suche in den bereits genannten Datenbanken gewählt, um zu weiteren Publikationen für die Überblicksarbeit zu gelangen. Im Weiteren

sollen nun die Ergebnisse dieser Literaturrecherche sowie der statistischen Analysen der empirischen Arbeiten beschrieben und eingeordnet werden.

5 Ergebnisse

Im vorliegenden Kapitel werden die Ergebnisse der promotionsrelevanten Publikationen dargelegt. Es erfolgt eine Zuordnung der Studien, zu den bereits vorgestellten aufeinander aufbauenden drei Schwerpunkten dieser Dissertation. Die Ergebnisse der Studie I (vgl. Anhang A) werden genutzt, um die Fragestellungen des ersten Schwerpunkts *„Notwendigkeit eines BVT in der sozialmedizinischen Begutachtung"* zu beantworten. Für die Bearbeitung der Fragestellungen des zweiten Schwerpunkts *„Entwicklung und Validierung eines BVT für die sozialmedizinische Begutachtung"* werden die Studien II (vgl. Anhang B) und IV (vgl. Anhang D) hinzugezogen. Im Rahmen der Darstellungen des dritten Schwerpunkts *„Einflüsse auf die Beschwerdendarstellung in der sozialmedizinischen Begutachtung"* werden die Ergebnisse der Studie III (vgl. Anhang C) präsentiert. Das Kapitel 5.2 gibt die Analysen der Stichprobenmerkmale wieder. Ergebnisse, die bisher in keiner Publikation Erwähnung fanden werden ausführlicher dargestellt.

5.1 Beschwerdenvalidierung in der sozialmedizinischen Begutachtung (Studie I)

Im Rahmen der Studie I wurde die Frage nach der Notwendigkeit der Konstruktion eines BVT für die sozialmedizinische Begutachtung von psychischen Erkrankungen erörtert. Auf der Grundlage einer Literaturrecherche wurden folgende Möglichkeiten der Erfassung von negativen Antwortverzerrungen herausgestellt: Kriterien für den Verdacht der Simulation, Validitätsparameter, Profilanalysen zur Validitätsbeurteilung, BVT sowie Selbst- und Fremdbeurteilungsverfahren. Weiterhin wurden zu jedem Ansatz Verfahrensbeispiele vorgestellt und nach ihrer Eignung zur Erfassung negativer Antwortverzerrungen bzw. Simulation und Aggravation bei Patienten mit psychischen Erkrankungen überprüft.

Dabei stellte sich heraus, dass die vorgestellten Kriterien für den Verdacht von Simulation unzureichende Sensitivitäten und Spezifitäten aufweisen (Rogers et al., 1993) und somit sowohl bei Patienten mit geltend gemachten kognitiven als auch psychischen Störungen keine Anwendung finden sollten. Ein weiteres Ergebnis aus Studie I ist, dass ein Großteil der im deutschsprachigen Raum etablierten Verfahren zur Beschwerdenvalidierung für neuropsychologische Erkrankungen konzipiert wurde. In der Studie I wird der Diskurs über den Einsatz von kognitiven Verfahren bei der Begutachtung von ausschließlich psychischen Erkrankungen betrachtet. Einige Autoren schlussfolgern, dass die Anwendung Schwierigkeiten aufweisen kann, da durch die kognitiven BVT meist die Leistungsmotivation gemessen wird, die bei Patienten mit einer depressiven Symptomatik stark eingeschränkt ist. Somit ist es möglich,

dass Patienten mit psychischen Störungen auffällige Testergebnisse erzielen, da kognitive BVT die Leistungsmotivation erfassen und nicht die invalide Symptomdarstellung. Weiterhin gilt es die Selbst- und Fremdbeurteilungsverfahren für die Erfassung von negativen Antwortverzerrungen, die zusätzlich für psychische Symptome geeignet scheinen, zu betrachten. Fremdbeurteilungsverfahren wie beispielsweise das SIRS liegen bis zum Zeitpunkt der Studie I für den deutschsprachigen Raum nicht vor. Allerdings zeigt die Publikation I, dass Kontrollskalen für die Validierung von Beschwerden in Selbstbeurteilungsverfahren vorliegen. Dennoch werden beispielsweise anhand der Offenheitsskala des Freiburger Persönlichkeits-Inventars (FPI-R; Fahrenberg, Hampel & Selg, 2010) anscheinend keine negativen Antwortverzerrungen gemessen, sondern das Konstrukt der Dissimulation (Merten, Friedel, Mehren & Stevens, 2007). Das ebenfalls Kontrollskalen beinhaltende MMPI-2 ist mit 567 Items ein sehr umfangreiches und zeitaufwendiges Verfahren. Zudem wurde der Einsatz des Verfahrens in Deutschland aufgrund seiner teilweise unzureichenden psychometrischen Qualität lange Zeit abgelehnt (Hank & Schwenkmezger, 2003). Das einzige Selbstbeurteilungsverfahren, welches speziell für die Erfassung negativer Antwortverzerrungen bzw. Simulation sowie für den deutschsprachigen Raum geeignet scheint, ist der SFSS. Wie Studie I verdeutlicht, wurde dieser Fragebogen bisher jedoch nur an einer forensischen Stichprobe sowie Studierendenstichprobe validiert und erscheint daher wenig geeignet für die sozialmedizinische Begutachtung. Darüber hinaus stößt der SFSS vermutlich durch die teilweise bizarr formulierten Items bei vielen Antragstellern der Rente wegen verminderter Erwerbsfähigkeit aufgrund einer psychischen Störung auf Ablehnung. Diese Vermutung konnte im weiteren Projektverlauf teilweise auch durch Studie III bestätigt werden, da nur 54 Personen den Fragebogen vollständig ausfüllten. Wohingegen 133 Teilnehmer das Modul A der HEALTH-49 vollständig bearbeiteten.

Die Ergebnisse der Studie I verdeutlichen, dass für die Erfassung von negativen Antwortverzerrungen und Simulation bei kognitiven Beeinträchtigungen auch im deutschsprachigen Raum viele gut validierte Verfahren vorliegen. Für den Bereich der psychischen Erkrankungen sind allerdings nur wenige Verfahren vorhanden, die auch nicht speziell für die Gruppe der Antragsteller einer Rente wegen Erwerbsminderung geeignet ist. Es ist daher notwendig für diesen klinischen Bereich einen Fragebogen für die Erfassung von negativen Antwortverzerrungen neu zu entwickeln und zu validieren.

5.2 Stichprobe

Nachstehend sollen die Analysen der Teilstichproben präsentiert werden. Dabei werden die Ergebnisse aus den Studien II-IV zusammenfassend dargestellt, da eine ausführliche Betrachtung bereits in den jeweiligen Publikationen vorgenommen wurde, die im Anhang B, C und D einsehbar sind. Die Ergebnisse der Studierendenstichprobe der Prä-Studie werden hingegen vollständig abgebildet, da es sich um bisher unveröffentlichte Ergebnisse handelt.

Prä-Studie. Die Tabelle 10 bietet eine ausführliche Darstellung der personenbezogenen Daten der Studierenden je Untersuchungsgruppe dar. Die drei Gruppen unterscheiden sich nicht signifikant in den personenbezogen Daten Alter, Geschlecht, Staatsangehörigkeit sowie Familienstand. Im Mittel sind die Studierenden etwa 25 Jahre alt, die meisten von ihnen sind weiblich, ledig und haben die deutsche Staatsangehörigkeit. Vier Personen in der Gruppe IS (Depression) haben eine psychische Erkrankung sowie zwei der Gruppe gesunde Studierende leiden unter einer chronischen Schmerzerkrankung. Der Großteil der Studierenden weist jedoch keine Erkrankung auf.

Tabelle 10: Personenbezogene Daten der Studierendenstichprobe der Prä-Studie

Personenbezogene Daten		Gesunde Studierende (n=43)	IS (Depression) (n=42)	IS (Schmerz) (n=42)	Statistische Analysen
Alter	M(SD)	24.30 (3.60)	25.81 (5.42)	24.43 (3.28)	$F_{2,124}$ = 1.673; *n.s.*
Geschlecht	Weiblich	81.4 % (35)	83.3 % (35)	85.7 % (36)	X^2= 0.288; df=2; n.s.
	Männlich	8.6 % (8)	6,7 % (7)	4.3 % (6)	
Staatsangehörigkeit	Deutsch	95.3 % (41)	90.5 % (38)	95.2 % (40)	X^2= 1.106; df=2; n.s.
	Andere	4.7 % (2)	9.5 % (4)	4.8 % (2)	
Familienstand	Ledig	97.7 % (42)	92.9 % (39)	95.2 % (40)	X^2= 3.714; df=4; n.s.
	Verheiratet	2.3 % (1)	7.1 % (3)	2.4 % (1)	
	Geschieden	0 % (0)	0 % (0)	2.4 % (1)	
	Verwitwet	0 % (0)	0 % (0)	0 % (0)	
Erkrankungen	Keine	95.3 % (41)	90.5 % (38)	100 % (42)	X^2= 12.202; df=4; p= .016
	Psychische Erkrankung	0 % (0)	9.5 % (4)	0 % (0)	
	Chronische Schmerzerkrankung	4.7 % (2)	0 % (0)	0 % (0)	

Anmerkungen: M= Mittelwert; SD=Standardabweichung; n=Teilstichprobengröße

Studie II. Die Stichprobe der Studie II ist aus einem Teil der Konstruktionsstichprobe des BEVA gebildet worden. Die rekrutierten Personen sollten sich in Teilen in den soziodemografischen Variablen mit den Antragstellern einer Rente wegen Erwerbsminderung vergleichen lassen. Beispielsweise waren die Teilnehmer der Kontrollgruppe durchschnittlich 50.32 (SD= 5.06), der IS

(Depression) 49.98 (SD=6.09) und die depressiven Patienten 50.07 (SD=5.33) Jahre alt. Sowohl in der Variablen Alter ($F_{2,191}$ = 0.152; n.s.) als auch im Geschlecht (X^2=1.924; df=2; n.s) und der Staatsangehörigkeit (X^2=2.343; df=2; n.s.) konnten keine signifikanten Effekte ermittelt werden. Da sich jedoch ein signifikanter Effekt zwischen den drei Gruppen in der Variablen *„Schulbildung"* zeigte, wurde diese als fester Faktor in die MANOVA einbezogen. Als weiterer fester Faktor fungierte die Gruppenzugehörigkeit und als abhängige Variablen galten die Summenwerte der Skalen *„Depressivität", „somatoforme Beschwerden"* sowie *„phobische Ängste"* der HEALTH-49. Die Analysen verdeutlichten keinen signifikanten multivariaten Effekt der *„Schulbildung"* für die Skalen des HEALTH-49 (Wilks λ = .902; $F_{12, 479}$ = 1.593; p = .090) sowie keinen multivariaten Interaktionseffekt für die Gruppenzugehörigkeit und Schulbildung (Wilks λ = .938; $F_{12, 479}$ = 0.980; p = .467). Die gleichen Ergebnisse konnten auch für die univariaten Analysen der jeweiligen Skalen des Moduls A der HEALTH-49 gezeigt werden. Weiterhin stellte sich ein signifikanter Effekt zwischen den Gruppen für die soziodemografische Variable „Familienstand" heraus (X^2=14.887; df=6; p=.016). Der Familienstand wurde dennoch nicht als eigenständiger Faktor in die MANOVA integriert, da in der Literatur keinerlei Hinweise dafür vorliegen, dass diese Variable einen Einfluss auf das Antwortverhalten haben könnte.

Studie III. Für alle personenbezogenen Daten wie Alter (t_{169} =0.318; n.s.), Geschlecht (X^2=0.542; df=1; n.s.), Staatsangehörigkeit (X^2=0.008; df=1; n.s.), Migrationshintergrund (X^2=0.176; df=1; n.s.), Tätigkeit (X^2=3.948; df=4; n.s.) sowie Diagnose (X^2=8.508; df=7; n.s.) konnten keine signifikanten Effekte zwischen den Gruppen „gewarnte Stichprobe" vs. „nicht gewarnte Stichprobe" nachgewiesen werden. Das Ergebnis verdeutlicht, dass durch das randomisierte Kontrollgruppendesign zwei in ihren personenbezogenen Daten vergleichbare Gruppen gebildet werden konnten und somit mögliche Störvariablen reduziert wurden.

Studie IV. In der Konstruktionsstichprobe zeigten sich für die soziodemografischen Variablen Alter ($F_{4,274}$ = 3.053; p = .017) sowie Schulbildung (X^2= 88.180; df=16; p = .000) zwischen allen Untersuchungsgruppen signifikante Effekte. Der Mittelwert für das Alter liegt bei den fünf Untersuchungsgruppen zwischen 46.48 (SD=6.06) für die IS (Schmerz) sowie den höchsten Mittelwert 50.11 (SD=8.13) für die Schmerzpatienten, sodass davon ausgegangen werden kann, dass alle Teilnehmer trotz des signifikanten Gruppenunterschieds sich in etwa in einer Altersgruppe befinden. Bezüglich der Variablen *„Schulbildung"* ist auffällig, dass viele Patienten einen Hauptschulabschluss oder die mittlere Reife besitzen. In den drei anderen Gruppen weisen hingegen sehr wenige Probanden einen Hauptschulabschluss auf, wohingegen die mittlere Reife

bzw. das Abitur oder die Fachhochschulreife stärker vertreten sind. Die Analysen zwischen den Gruppen der anderen soziodemografischen Variablen Staatsangehörigkeit (X^2= 4.630; df=4), Familienstand (X^2= 15.712; df=12; n.s.) und Geschlecht (X^2= 3.931; df=4; n.s.) wiesen keine Signifikanzen auf.

In der Validierungsstichprobe zeigte sich zwischen den Gruppen nur für die Variable Staatsangehörigkeit (X^2= 7.171; df=4; n.s.) kein signifikanter Effekt. Für die soziodemografische Variable Alter ($F_{4,482}$ = 3.528; p = .007) kann jedoch wiederum festgestellt werden, dass sich die Teilnehmer in einer Altersgruppe befinden. Die Gruppe der IS (Depression) hatte den geringsten Mittelwert von 44,13 Lebensjahren (SD=7.80) und die Schmerzpatienten wiederum den höchsten von 48,83 (SD=8.23). In der Kontrollgruppe sowie den Gruppen der IS waren 2/3 der Teilnehmer Frauen. Wohingegen bei den Patientengruppen das Geschlechterverhältnis ausgeglichen war. Dies zeigte sich wiederum in einem signifikanten Effekt (X^2= 19.548; df=4; p= .001). Ein weiterer signifikanter Effekt lag zwischen den Gruppen für die Variablen Familienstand (X^2= 36.244; df=12; p= .000) und Schulbildung (X^2= 111.245; df=16; p = .000) vor.

5.3 Entwicklung des BEVA (u.a. Studie II)

In diesem Kapitel werden zunächst die bisher unveröffentlichten Ergebnisse der Itemgenerierung sowie der Prä-Studie dargestellt. Darauffolgend werden zusammengefasst die Erkenntnisse der Studie II beschrieben, die zur Entwicklung des BEVA beigetragen haben.

Ergebnisse der Itemgenerierung. Wie im Kapitel 4.2 beschrieben, wurden im Verlauf der Itemgenerierung Informationen zur Formulierung etwaiger Items anhand einer Expertenbefragung erhoben. In der Tabelle 11 sind die Endergebnisse dieser qualitativen Analyse dargestellt. Darin sind Angaben zu Angststörungen nicht enthalten, da hierfür keine Aussagen von den Experten vorlagen. Auf der Grundlage der Angaben zu den somatoformen Störungen wurden beispielsweise folgende Items generiert: „*Ich habe so starke Schmerzen, dass ich sie nicht einmal mehr genau beschreiben kann.“* und „*Meine andauernden Schmerzen verbessern sich zu keiner Tageszeit.“*. Weiterhin ergab die Analyse, dass wiederholt auf das Freizeitverhalten und Hobbys durch die Experten Bezug genommen wurde, weshalb beispielsweise das nachstehende Item generiert wurde: *„Ich kann mich durch Sport und andere körperliche Aktivitäten von meinen Problemen ablenken.“*

Tabelle 11: Endergebnis der qualitativen Analyse der Expertenbefragung

Allgemeine Angaben	Affektive Störung	Somatoforme Störung
Erlebens- und Gestaltensfähigkeit im Alltag nicht eingeschränkt	Psychomotorik	Stark übersteigende Schmerzäußerungen
Fragen nach dem Ziel oder Zweck des Pat., ob er Entschädigung wünscht, Höherbewertung des GdB, gesellschaftliche Anerkennung	durchgehend schwunglos, auch rückblickend negative Kognitionen, (vs. retrospektive Hypernormalität) Schuldgefühle, fühlt sich nicht wert, Leistungen zu erhalten, ist geduldig, nicht fordernd/aggressiv, demütig, leise, wenig gepflegt	starke Betonung somatischer Beschwerden oft sehr exorbitant geäußert und im Vergleich zu wirklich sehr kranken somatischen Pat. z.B. im Akuthaus in keinerlei Korrelation stehend
Kein "Leidensdruck"	Betonung einer depressiven Symptomatik, wobei draußen dann gelacht wird	
Angaben zu Beruf- und Freizeitverhalten, die den anderen Angaben nicht entsprechen	stelle bei affektiven Störungen aber noch weitere Fragen, z.B. zur Alltagsbewältigung, Freunden, Hobbies etc.	
Aggravation als "Übertreibung" (nach Silomon) nur selten erlebt (am ehesten in Richtung einer besonderen "Erschöpfung")		
Offensichtlich geringer Leidensdruck trotz intensiv geschilderter Beschwerden		
Spontan diffuse Beschwerden, Bejahungstendenz, unzureichende Kooperation, Waddell-Zeichen positiv		
Aggressives Auftreten		
Leidendes Gebaren		

Anmerkung: In der Tabelle sind die Aussagen der Gutachter wortwörtlich aufgelistet

Als weitere Informationsquelle dienten bereits vorliegende Verfahren und Kontrollskalen. Aus der Tabelle 12 sind Beispiele der jeweiligen Originalitems der genutzten Verfahren und den dazugehörigen Umformulierungen im BEVA ersichtlich.

Tabelle 12: Beispielitems der Verfahren für die Itemgenerierung und Umformulierung des jeweiligen Items im BEVA

Originalfragebogen	Originalitem	Umformulierung im BEVA
SFSS; Skala „Depression“	Am besten fühle ich mich morgens nach einem guten Schlaf, obwohl ich die meiste Zeit depressiv bin.	Je länger der Tag andauert, desto schlechter fühle ich mich.
SFSS; Skala „Depression“	Wenn meine Depression zu stark wird, mache ich lange Sparziergänge oder Sport, um die Spannung zu vermindern.	Wenn ich besonders psychisch belastet bin, dann kann ich durch Sport meine Anspannung reduzieren.
MMPI-2; Skala „soziale Verantwortung“	Wenn mir jemand Unrecht tut, habe ich das Gefühl, dass ich es ihm heimzahlen müsste, wenn ich kann, schon um des Prinzips willen.	Wenn mir die Menschen Unrecht tun, dann habe ich das Gefühl, mich rächen zu müssen.
PSSI; Skala kritisch-negativistisch	Ich bin in meinem Leben oft ungerecht behandelt worden.	Meiner Meinung nach wurde ich oft zu Unrecht bestraft.
PSSI; Skala selbstbehauptend-antisozial	Meine Bedürfnisse lebe ich aus, auch wenn andere zurückstecken müssen.	Meine Anliegen kommen an erster Stelle, auch wenn andere sich dafür zurücknehmen müssen.

Anmerkung: BEVA=Beschwerdenvalidierungstest; SFSS=Strukturierter Fragebogen Simulierter Symptome; MMPI-2=Minnesota Multiphasic Personality Inventory-2; PSSI=Persönlichkeits-Stil und Störungs-Inventar

Ergebnisse der Prä-Studie. Anhand der Durchführung der Prä-Studie konnte festgestellt werden, dass der Ablauf der Studie, die Instruktionen für die Simulanten sowie die Verständlichkeit des Szenarios als ausreichend gut betrachtet werden können. Dies zeigte sich unter anderem dadurch, dass keiner der Studierenden aufgrund fehlenden Aufgabenverständnisses aus der Studie ausgeschlossen werden musste. Erfasst wurde das Aufgabenverständnis anhand der Fragebogen Rollenerfüllung bzw. Rollenkonformität. Darüber hinaus gab es bezüglich des Ablaufs der Studie keine negativen schriftlichen sowie mündlichen Anmerkungen der Studierenden. Jedoch trugen die Ergebnisse der Prä-Studie zu Veränderungen in der Itemformulierung bei. Bei 38 von 66 Items lagen fehlende Werte vor, wobei die höchste Anzahl an Auslassungen bei neun lag. Dies bedeutet, dass jedes Item von mindestens 93 % der Teilnehmer bewertet wurde. Dabei war auffällig, dass die meisten fehlenden Werte in der Kontrollgruppe auftraten, da in dieser Gruppe nur 29 der 43 Teilnehmer den Fragebogen vollständig ausfüllten. In den jeweiligen anderen Gruppen füllten 40 (IS „Schmerz“) bzw. 39 (IS „Depression“) den Fragebogen komplett aus. Bei der Analyse der Auslasser konnte ermittelt werden, dass dies vor allem Items betraf, die eine direkte Formulierung zu einer Erkrankung aufwiesen beispielsweise *„in schlimmen depressiven Phasen“* oder *„meine psychische Erkrankung“*. Diese Items können demzufolge nur dann beantwortet werden, wenn eine entsprechende Erkrankung vorliegt. Daher wurden für den später eingesetzten Fragebogen

andere Umschreibungen genutzt wie *„psychische Belastung“*. Weiterhin war die Verständlichkeit bei sehr langen Items eingeschränkt. Daraufhin wurden die Aussagen meist in zwei neu formulierten Items aufgeteilt, sodass schlussendlich ein Fragebogen mit 76 Items in der Konstruktionsphase eingesetzt wurde.

Ergebnisse von Studie II. Die paarweisen post-hoc-t-Tests zeigten, dass sich die IS signifikant sowohl von der Kontrollgruppe als auch von den depressiven Patienten in allen untersuchten Skalen *„Depressivität“* (**IS-KG**: t_{138}=25.336; p= .000; d=4.62; **IS-DP**: t_{115}= 15.450; p= .000; d= 0.82), *„phobische Ängste“* (**IS-KG**: t_{138}=15.450; p= .000; d=2.78; **IS-DP**: t_{115}= 3.951; p= .000; d= 0.74) sowie *„somatoforme Beschwerden“* (**IS-KG**: t_{138}=18.186; p= .000; d=3.41; **IS-DP**: t_{115}= 5.664; p= .000; d= 1.06) der HEALTH-49 unterscheiden. Die ermittelten Effekte aller paarweisen Vergleiche der Studie II sind nach Cohen (1988) als groß zu definieren. In der Abbildung 8 werden die Mittelwerte der drei Gruppen mit ihren jeweiligen Standardabweichungen für die untersuchten klinischen Skalen der HEALTH-49 ersichtlich.

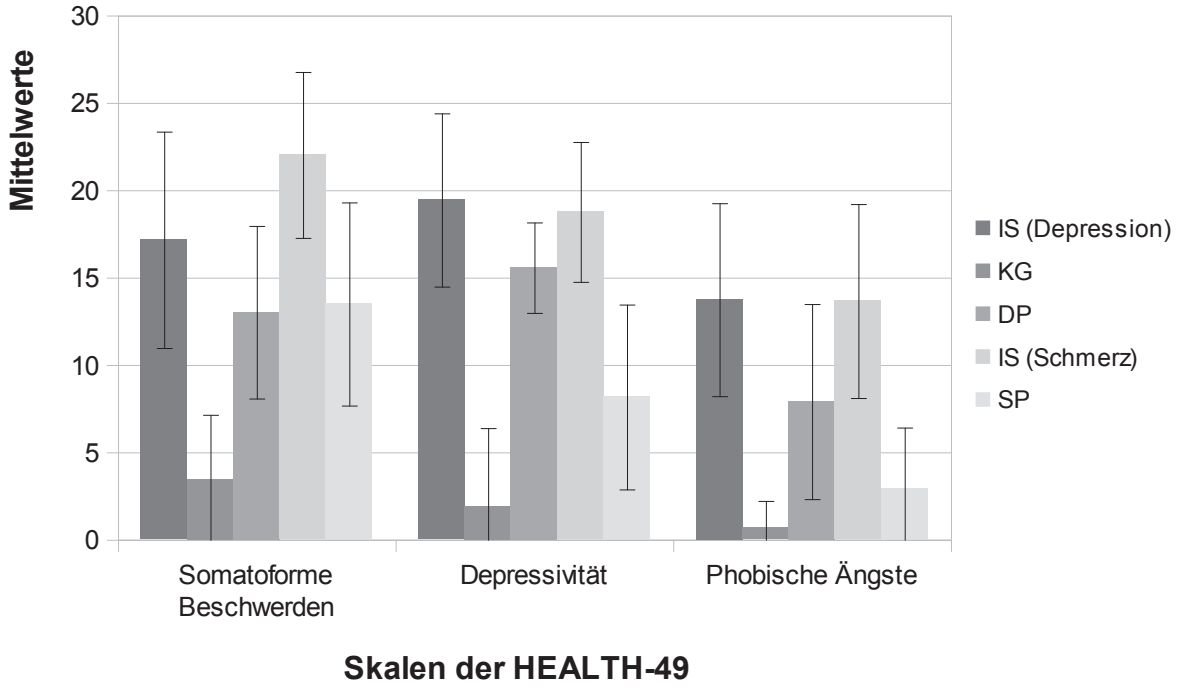

Abbildung 8: Balkendiagramm der Mittelwerte und dazugehörigen Standardabweichungen für die Skalen des Moduls A der HEALTH-49 für die jeweiligen Untersuchungsgruppen der Studie II
Anmerkung: HEALTH-49= Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis; IS= instruierte Simulanten; KG= Kontrollgruppe; DP= depressive Patienten; SP= Schmerzpatienten

Dieses Ergebnis verdeutlicht, dass sich IS im Vergleich zu depressiven Patienten durch Extremantworten und Übergeneralisierungen unterscheiden. Es bestätigt wiederum, dass sich das

Antwortverhalten von IS im Vergleich zu wirklich depressiven Patienten in einem Fragebogen abbilden lässt. Die Formulierung von Items des BEVA, die nicht ausschließlich depressive Symptome erfassen, sondern auch andere Erkrankungen wie Angststörungen oder Schmerzerkrankungen berücksichtigen, kann durch das Ergebnis der Übergeneralisierung gerechtfertigt werden. Zusätzlich begründet das Ergebnis der Extremantworten den Einsatz einer vierstufigen Skala, da eine derartige Schwankung im Rahmen eines dichotomen Antwortformats nicht abgebildet werden könnte. Darüber hinaus verdeutlicht Studie II, dass IS zwar Extremantworten zeigen, jedoch eine Varianz in ihrem Antwortverhalten erkennen lassen. Hinweise darauf lieferten die Analysen zur Überprüfung der Normalverteilung anhand des Kolmogorov-Smirnov-Tests für die IS in den Skalen *„Depressivität“* (D(63)=0.163; p= .069), *„phobische Ängste“* (D(63)=0.129; p= .249) sowie *„somatoforme Beschwerden“* (D(63)=0.078; p= .839).

In den Berechnungen für die Studie II wurden lediglich die Gruppen IS (Depression), depressive Patienten sowie die Kontrollgruppe einbezogen. Daher wurden für die vorliegende Dissertation weitere Rechnungen durchgeführt, um Ergebnisse der Untersuchungsgruppe „Schmerz“ für die Beantwortung der dargebotenen Fragestellungen beizutragen. Die Resultate dieser zusätzlichen Analysen sollen im Folgenden dargestellt werden.

Für die nachfolgenden Analysen wurde die Teilstichprobe aus der Stichprobe Konstruktion des BEVA gewonnen. Dies führt dazu, dass sich diese Kontrollgruppe im Vergleich zur Studie II um eine Person erhöht. Aus der Tabelle 13 ist ersichtlich, dass sich auch die IS (Schmerz) im Mittel signifikant von der Kontrollgruppe und den Schmerzpatienten unterscheiden. Dies bestätigt die Annahme, dass sich IS unabhängig von den simulierten Diagnosen im Allgemeinen durch Extremantworten und Übergeneralisierungen auszeichnen. In der Abbildung 8 sind darüber hinaus alle Untersuchungsgruppen mit Mittelwerten und zugehörigen Standardabweichungen abgebildet. Die Daten der Kontrollgruppe wurden aus der Studie II entnommen.

Weiterhin wurde auf der Grundlage der Kolmogorov-Smirnov-Tests geprüft, ob die Summenwerte für die klinischen Skalen der HEALTH-49 in den Gruppen Schmerzpatienten und IS (Schmerz) normalverteilt sind. Sie zeigten, dass die Werte der Gruppe IS (Schmerz) für die Skalen „Depressivität“ (D(50)=0.182; p= .000), „phobische Ängste“ (D(50)=0.127; p= .044) sowie „somatoforme Beschwerden“ (D(50)=0.142; p= .014) nicht normalverteilt sind. Verglichen mit der Gruppe der IS (Depression) zeigen die IS (Schmerz) keine starke Varianz in ihrem Antwortverhalten.

Tabelle 13: Vergleich der mittleren Rohwertsummen zwischen den Untersuchungsgruppen für die Skalen des Moduls A der HEALTH-49

Skalen des HEALTH-49	KG (1) n=78	SP (2) n=35	IS (3) n=50	Teststatistik						
				ANOVA		Post-hoc Analyse				Effekt-größe
	M (SD)	*M (SD)*	*M (SD)*	*F (df_1/df_2)*	*p*	Vergleich	*t*	α_{adjust}	*p*	*d*
Depressivität	1.85 (2.59)	8.17 (5.29)	18.76 (4.00)	309.412 (2 /160)	.000	(1) - (2)***	-6.722	.016	.000	1.75
						(1) - (3)***	-26.521	.016	.000	5.31
						(2) - (3)***	-10.505	.016	.000	2.34
Somatoforme Beschwerden	3.44 (3.66)	13.49 (5.81)	22.02 (4.75)	261.327 (2 /160)	.000	(1) - (2)***	-9.423	.016	.000	2.29
						(1) - (3)***	-23.544	.016	.000	4.55
						(2) - (3)***	-7.431	.016	.000	1.66
Phobische Ängste	0.67 (1.52)	2.91 (3.51)	13.66 (5.55)	203.625 (2 /160)	.000	(1) - (2)**	-3.638	.016	.001	0.97
						(1) - (3)***	-16.168	.016	.000	3.58
						(2) - (3)***	-10.922	.016	.000	2.26

Anmerkungen: HEALTH-49= Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis; KG = Kontrollgruppe; SP= Schmerzpatienten; IS = instruierte Simulanten; (1) - (2) = Vergleich zwischen den Mittelwerten von KG und SP; (1) - (3) = Vergleich zwischen den Mittelwerten von KG und IS; (2) - (3) = Vergleich zwischen den Mittelwerten von SP und IS; α_{adjust} = adjustiertes α-Niveau nach Bonferroni. **statistische Signifikanz bei $p < .01$.***statistische Signifikanz bei $p < .001$.

Die Summenwerte der Gruppe der Schmerzpatienten waren ausschließlich für die Skala *„somatoforme Beschwerden"* (D(35)=0.126; p= .178) normalverteilt. Daraus lässt sich schließen, dass die Werte der Skala *„phobische Ängste"* (D(35)=0.205; p= .001) sowie *„Depressivität"* (D(35)=0.216; p= .000) für die Schmerzpatienten wiederum nicht normalverteilt waren. Wie in der Studie II bereits belegt, waren auch die Summenwerte der Kontrollgruppe für alle Skalen der HEALTH-49 nicht normalverteilt. Daher wurden alle Ergebnisse, die sich durch parametrische Tests ergaben anhand von des Mann-Whitney U-Tests zusätzlich geprüft. Es zeigten sich keine abweichenden Erkenntnisse, daher werden nur die Ergebnisse der parametrischen Testung ausführlich berichtet.

Auf der Grundlage aller Feststellungen des Kapitels 5.3 wurde ein Fragebogen mit 76 Items entwickelt, der in der Tabelle 14 ersichtlich ist und in dieser Version in die nächste Studienphase der „Konstruktion" überging.

Tabelle 14: Vollständiger Itempool des BEVA

1.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
2.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
3.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.

Fortsetzung Tabelle 14

4.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
5.	Ich denke, dass sich viele Leute krank stellen, um unangenehme Dinge zu vermeiden.
6.	Ich rede häufig pausenlos.
7.	Ich habe psychische Probleme, mit denen ich allein nicht fertig werde.
8.	Ich sehe mit Freuden in die Zukunft.
9.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
10.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
11.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
12.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
13.	Bestimmte Körperhaltungen lindern meine Schmerzen.
14.	Psychopharmaka machen abhängig.
15.	Durch meine Erkrankung habe ich Einschränkungen im Beruf.
16.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
17.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
18.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
19.	Obwohl ich mich von einigen Tätigkeiten zurückgezogen habe, bemühe ich mich, meine Angehörigen durch Tipps (z.B. in der Haushaltsführung) zu unterstützen.
20.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
21.	Bei meinen Erkrankungen kann mir nur meine Homöopathin/mein Homöopath helfen.
22.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
23.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
24.	Wenn ich merke, dass meine Schmerzen stärker werden, treffe ich genaue Vorsorgen, um diese zu ertragen.
25.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
26.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
27.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
28.	Wenn ich traurig bin, kann ich mir besonders gut Dinge merken.
29.	Ich führe regelmäßig Gespräche mit einer/einem PsychotherapeutIn bzw. PsychiaterIn.
30.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
31.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
32.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
33.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
34.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
35.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
36.	Ich erlebe keine Einschränkungen in meinem Sozial- und Freizeitleben.
37.	Ich nehme regelmäßig Medikamente zur Verbesserung meiner Stimmungslage.
38.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
39.	Meist tue ich das, was mir Freude bereitet, auch wenn andere dabei etwas zurückstecken müssen.
40.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
41.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
42.	Es ist keinem ein Vorwurf zu machen, der andere benachteiligt, wenn ihm eine Gelegenheit dazu geboten wird.

Fortsetzung Tabelle 14

43.	Im Allgemeinen bin ich mit meinem Leben zufrieden.
44.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
45.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
46.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
47.	Solange eine Regel nicht gebrochen wird, ist es in Ordnung sie zu umgehen.
48.	Ich leide unter Schuldgefühlen.
49.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
50.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
51.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
52.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
53.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
54.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
55.	Mein Gewicht bleibt auch in belastenden Phasen konstant.
56.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
57.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
58.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
59.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
60.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
61.	Viele Leute sehen mein Durchsetzungsvermögen als rücksichtslos an.
62.	Meiner Meinung nach übertreiben viele Menschen ihren Kummer, um die Anteilnahme ihrer Mitmenschen zu steigern.
63.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
64.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
65.	Meine Anliegen kommen an erster Stelle, auch wenn andere sich dafür zurücknehmen müssen.
66.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
67.	Wenn ich besonders psychisch belastet bin, widme ich mich vermehrt meinen Hobbys, um mein Leben erträglicher zu gestalten.
68.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
69.	Wenn mir etwas Gutes widerfährt, ist es mein Verdienst.
70.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
71.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
72.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
73.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
74.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
75.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.
76.	Aus verlagsrechtlichen Gründen kann das Item an dieser Stelle leider nicht abgedruckt werden.

Anmerkungen: BEVA= Beschwerdenvalidierungstest; Die Teilnehmer mussten anhand einer vierstufigen Skala „stimmt“, „stimmt eher“, „stimmt eher nicht“, „stimmt nicht“ die vorliegenden 76 Aussagen bewerten. entfernte Items in der Konstruktion; entfernte Items in der Validierung; Items des BEVA

5.4 Konstruktion und Validierung des BEVA (Studie IV)

Ergebnisse der Konstruktion. Zunächst erfolgte eine Analyse, bei der die paarweisen Zusammenhänge je Item auf signifikante Effekte geprüft wurden. Auf der Basis der Daten der Konstruktionsphase wurden anhand von Mann-Whitney-U Tests Items aus dem BEVA eliminiert, die in der Tabelle 14 hellgrau hinterlegt sind. Zwei weitere Items wurden wegen der Unterschiedsberechnungen in der Validierung aus dem BEVA gelöscht. Diese Items sind wiederum in der Tabelle 14 in mittelgrau hervorgehoben. Es wurden Items ausgeschlossen, falls keine signifikanten Unterschiede zwischen den Gruppen IS und Patienten bzw. IS und Kontrollgruppe vorlagen. Für die Analysen mit dem Mann-Whitney-U-Test wurden nur die Daten verwendet, bei denen keine Ersetzung aufgrund von fehlenden Werten erfolgt war.

Daraufhin wurde mithilfe der 51 Items der Cut-Off-Wert des BEVA für die Untersuchungsgruppen *depressive Störungen* und *Schmerz mit depressiver Symptomatik* ermittelt. Der BEVA ist ein Screening, das negative Antwortverzerrungen erfassen soll. Daher weisen Summenwerte, die über dem Cut-Off-Wert liegen, auf ein invalides Testergebnis. Es wurde für die Bestimmung des Cut-Off-Werts aus ethischen Gesichtspunkten eine maximale Spezifität gewählt. Das bedeutet, dass möglichst alle Patienten, die keine negativen Antwortverzerrungen aufweisen, einen Summenwert zeigen, der unter dem Cut-Off-Wert liegt. Die Festlegung einer möglichst maximalen Spezifität kann zu einer geringeren Sensitivität führen. Es ist daher möglich, dass instruierte Simulanten mit ihrem Gesamtwert unter dem Cut-Off-Wert liegen und daher als Personen mit validen Testergebnis klassifiziert werden.

Die Abbildung 9 zeigt, dass keiner der depressiven Patienten im BEVA einen Summenwert höher als 107 erreicht hat, weshalb dieser Wert als Cut-Off ermittelt wurde. In der Gruppe der IS (Depression) hatten 45 Personen einen Summenwert, der über diesem Cut-Off lag. Bei einem Summenwert von 107 kann eine Sensitivität von .78 bei gleichzeitiger Spezifität von 1 berechnet werden. Der niedrigste Summenwert, der in der Gruppe IS (Depression) erreicht wurde liegt bei 84. 12 depressive Patienten haben einen Summenwert der größer als bzw. gleich 84 ist, daher kann für diesen Wert eine Sensitivität von 1 bei gleichzeitiger Spezifität von .79 berechnet werden. Wobei die area under the Curve (AUC) .987 beträgt. Da der Wert höchstens 1 erreichen kann, ist davon auszugehen, dass die Trefferquote zur Erkennung von negativen Antwortverzerrungen bei depressiven Patienten im BEVA sehr hoch ist und die Fehlerrate als sehr niedrig anzusehen ist. Weiterhin ist aus der Abbildung 9 ersichtlich, dass die Schmerzpatienten einen Summenwert im BEVA von maximal 90 erreichen. Im Vergleich dazu

haben 44 IS (Schmerz) im BEVA einen Summenwert, der über 90 liegt. Bei diesem Wert kann eine Spezifität von 1 bei gleichzeitiger Sensitivität von .88 berechnet werden. Es zeigt sich, dass der minimalste Summenwert im BEVA für die Gruppe IS (Schmerz) 82 beträgt. Da 4 Schmerzpatienten über diesen Wert lagen, ergibt sich bei einer Sensitivität von 1 eine gleichzeitige Spezifität von .89. Die AUC liegt für diese Untersuchungsgruppe bei .991, was wiederum für eine sehr gute Trefferquote und geringe Fehlerrate spricht.

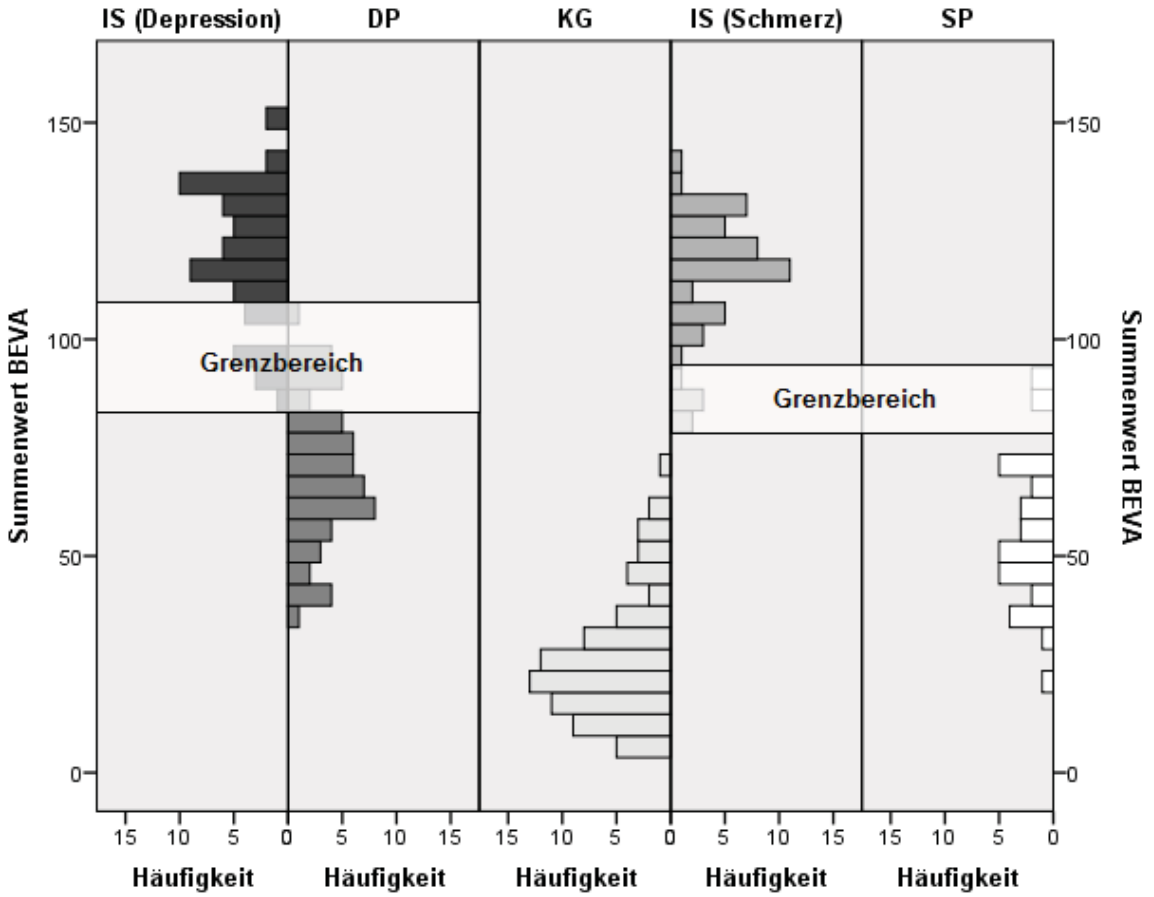

Abbildung 9: Histogramm der Häufigkeiten der Summenwerte im BEVA je Untersuchungsgruppe in der Konstruktion
Anmerkung: BEVA= Beschwerdenvalidierungstest; IS= instruierte Simulanten; DP= depressive Patienten; KG=Kontrollgruppe; SP= Schmerzpatienten; Grenzbereich= Bereich zwischen höchsten Summenwert der Patientengruppe sowie niedrigsten Summenwert der IS

Ergebnisse der Validierung. Für die Validierung des Screenings BEVA wurde für die Untersuchungsgruppe „Depression“ der Cut-Off-Wert von 107 genutzt, da aus ethischen Gesichtspunkten eine maximale Spezifität für das Verfahren angestrebt wird. Wie in der Abbildung 10 zu erkennen ist, wurden 75 von 100 IS (Depression) durch das Screening auch als

Simulanten erkannt. Weiterhin konnten anhand des Summenwerts des BEVA 154 der 162 depressive Patienten korrekt als Angehörige der tatsächlichen Patientengruppe ermittelt werden. Dadurch ergeben sich eine Sensitivität von .75, eine Spezifität von .95, ein PPW von .90 sowie ein NPW von .86. Unabhängig davon, ob die Teilnehmer IS (Depression) oder depressive Patienten sind, klassifiziert der Cut-Off-Wert des BEVA zu 87.4 % korrekt. Anhand der Analysen der Konstruktion konnte der niedrigste Summenwert für die IS (Depression) von 84 ermittelt werden. In der Validierungsstichprobe haben 99 von 100 IS (Depression) einen Summenwert, der größer ist als 84. In der Gruppe der depressiven Patienten hatten 129 von 162 einen geringeren Summenwert im BEVA als 84. Daher ergibt sich eine Sensitivität von .99 bei gleichzeitiger Spezifität von .80.

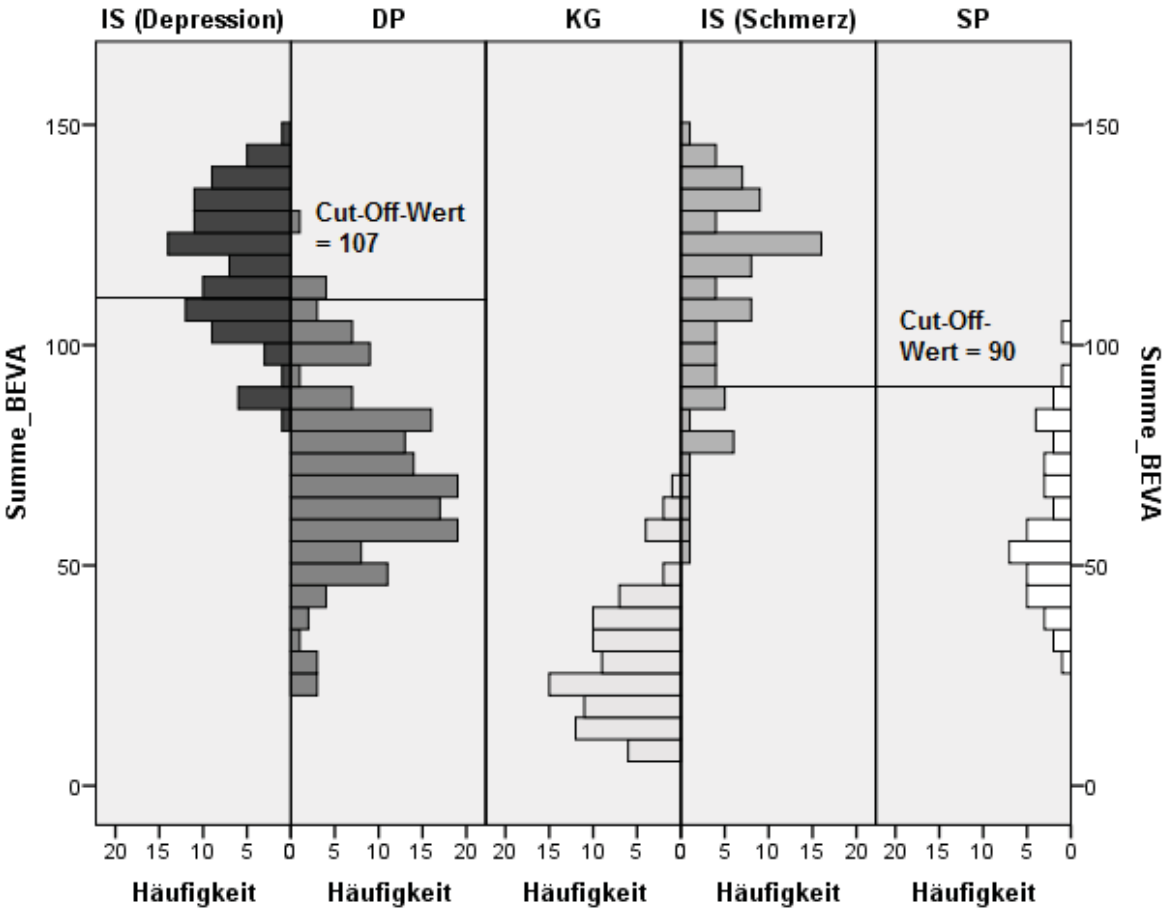

Abbildung 10: Histogramm der Häufigkeiten der Summenwerte im BEVA je Untersuchungsgruppe in der Validierung
Anmerkung: BEVA= Beschwerdenvalidierungstest; IS= instruierte Simulanten; DP= depressive Patienten; KG=Kontrollgruppe; SP= Schmerzpatienten; gemeint ist hier der Cut-Off-Wert im BEVA

Für die Untersuchungsgruppe *„Schmerz“* konnte anhand der Konstruktionsstichprobe ein Cut-Off-Wert von 90 ermittelt werden. Weiterhin zeigt die Abbildung 10, dass bei diesem Wert in der Validierungsstichprobe von 90 IS (Schmerz) 73 durch den BEVA korrekt als Simulanten ermittelt wurden. Weiterhin wurden durch das Screening mit der BEVA 44 von 46 Schmerzpatienten auch richtig als solche erkannt. Damit ergeben sich für diese Untersuchungsgruppe eine Sensitivität von .81, eine Spezifität von .96, ein PPW von .97 sowie eine NPW von .72. Unabhängig von der Gruppenzugehörigkeit liegt die korrekte Klassifikation durch den Cut-Off des BEVA bei 86 %. Für den niedrigsten berechneten Summenwert der Konstruktion für die IS (Schmerz), wiesen die Ergebnisse der Validierung eine Sensitivität von .87 bei gleichzeitiger Spezifität von .87 auf. Von 90 IS lagen 78 über sowie 40 von 46 Schmerpatienten unter dem Summenwert von 82.

Zuzüglich wurde anhand der Gesamtstichprobe der Validierung (N=487) die interne Konsistenz anhand von Cronbachs alpha ermittelt, die .976 beträgt.

5.5 Einflüsse von Instruktionen und der Motivation auf die Beschwerdendarstellung (Studie III)

Instruktion. Anhand der Ergebnisse der Studie III ist abzulesen, dass Warnungen sowohl einen Einfluss auf die Beschwerdendarstellung in einem BVT als auch in klinischen Symptomfragebogen haben. Dies zeigte sich darin, dass der Hinweis *„Mit einem Fragebogen haben Sie die Möglichkeit, Ihre psychischen Beschwerden genauer zu beschreiben, wobei die Testung gleichzeitig die Nachvollziehbarkeit der Angaben erlaubt."* dazu führte, dass die gewarnten Teilnehmer (Version 1) in den klinischen Skalen *„Depressivität“* ($F_{1,156}$ = 6.760; p= .010; d= 0.42), *„somatoforme Beschwerden“* ($F_{1,156}$ = 5.669; p= .018; d= 0.38) und *„phobische Ängste“* ($F_{1,156}$ = 11.367; p= .001; d= 0.54) der HEALTH-49 signifikant höhere Summenwerte zeigten als die nicht gewarnte Gruppe. Zusätzlich deutet dieses Ergebnis daraufhin, dass negative Antwortverzerrungen trotz der heterogenen Diagnosen in den jeweiligen Gruppen durch Extremantworten und Übergeneralisierungen gekennzeichnet sind. In der Abbildung 11 sind für die beiden Untersuchungsgruppen die Mittelwerte mit den dazugehörigen Standardabweichungen dargelegt.

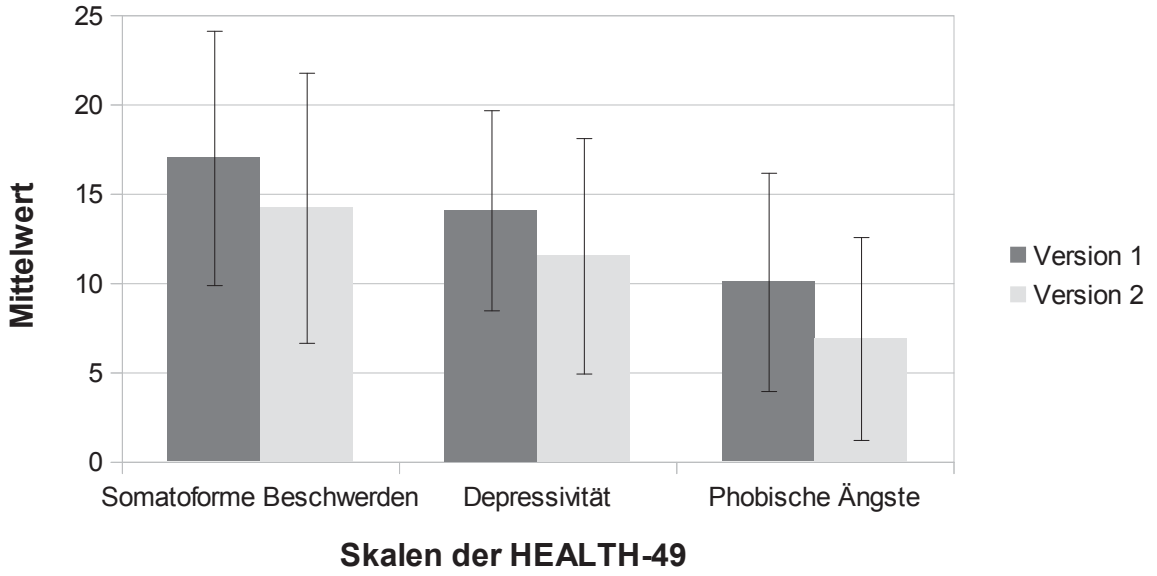

Abbildung 11: Balkendiagramm der Mittelwerte und dazugehörigen Standardabweichungen für die Skalen des Moduls A der HEALTH-49 für die jeweiligen Gruppen der Studie III
Anmerkungen: HEALTH-49= Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis; Version 1= gewarnte Gruppe; Version 2= nicht gewarnte Gruppe

Auch im Summenwert des SFSS spiegelt sich der Einfluss der Warnung wieder, da die gewarnte Gruppe (M = 20.16, SD = 9.53) durchschnittlich einen signifikant höheren Summenwert ($F_{1,140}$ = 5.854; p= .017; d = 0.41) aufwies als die nicht gewarnte Gruppe (M= 16.2; SD= 9.84). Weiterhin wurden die Teilnehmer der jeweiligen Gruppen auf der Grundlage des Cut-Off-Werts des SFSS in *auffällig* und *unauffällig* eingeteilt. Auffällig bedeutet hierbei, dass die Personen negative Antwortverzerrungen zeigen (vgl. Tab. 15). Es zeigte sich, dass die Personen der Version 1 nicht signifikant häufiger den Cut-Off-Wert von 16 überschritten haben. Jedoch verdeutlichte das OR, dass die Wahrscheinlichkeit um das fast 2-fache steigt ein auffälliges Ergebnis im SFSS aufzuweisen, wenn Personen den Hinweis aus der Version 1 erhielten (X^2= 3.389; df=1; p= .066; OR= 1.87). Dieses Ergebnis konkretisiert, dass der Cut-Off-Wert des SFSS nicht geeignet scheint, um eine korrekte Einteilung von Antragstellern in einer sozialmedizinischen Begutachtung bezüglich ihrer negativen Antwortverzerrungen zu gewährleisten. Darüber hinaus ist aus der Tabelle 15 ersichtlich, dass die Prävalenz gemessen mit dem SFSS für negative Antwortverzerrungen in etwa bei 48 % liegt. Auf Gruppenebene ergeben sich Basisraten von 56 % (Version 1) bzw. 40 % (Version 2).

Tabelle 15: Anzahl negativer Antwortverzerrungen im SFSS (Cut-Off-Wert = 16) je Untersuchungsgruppe der Studie III

		SFSS-Codierung		Gesamt
		0	1	
Version	1	31	39	70
	2	43	29	72
Gesamt		74	68	142

Anmerkungen: SFSS= Strukturierter Fragebogen Simulierter Symptome; Version 1 = gewarnte Gruppe; Version 2 = nicht gewarnte Gruppe. SFSS-Codierung 0 = Summenwert im SFSS liegt unter dem Cut-Off-Wert; SFSS-Codierung 1 = Summenwert im SFSS liegt über dem Cut-Off-Wert

Motivation. Für die Beantwortung der Fragestellung, ob eine zugrundliegende Motivation bezüglich des Erhalts einer Rente die Beschwerdendarstellung beeinflusst, wurden die Korrelationen zwischen Summenwert des SFSS und jedes einzelnen Motivationsitem betrachtet. Diese Darstellung konnte für die gesamte Stichprobe erfolgen, da sich keine signifikanten Effekte zwischen den Korrelationskoeffizienten der jeweiligen Gruppen der Studie III mittels Fishers z-Transformationen zeigten. Da im Rahmen der Studie mitunter fehlende Werte auftraten, ist die Teilstichprobengröße für die Betrachtung der einzelnen Items unterschiedlich. Die Studie III verdeutlicht, dass das Vorliegen des Motivs *„Die Rente bedeutet für mich, eine Chance wieder gesund zu werden"* anscheinend dazu führt, vermehrt negative Antwortverzerrungen zu zeigen (vgl Abb. 12). Es zeigt, dass Personen, die dem oben beschriebenen Motiv voll zustimmen, einen höheren Summenwert im SFSS erreicht. In einer linearen Regressionsanalyse konnte das genannte Motiv als einziger Prädiktor das Kriterium Summenwert des SFSS vorhersagen ($p = .008$). In diesem Modell waren neben dem genannten Motiv zusätzlich noch die Prädiktoren *„finanzielle Sicherheit", „Entlastung von beruflichen Anforderungen", „nichts mehr wert zu sein", „kein Hartz-IV-Empfänger zu sein"* sowie *„ausgeschlossen zu sein"*. Insgesamt zeigte die Modellprüfung statistische Signifikanz ($F_{6, 86} = 3.375$; $p= .005$; $R^2= .191$). In der Abbildung 12 ist das R^2 für ein zweites Modell angegeben, in dem das Motiv „eine Chance wieder gesund zu werden" als einziger Prädiktor eingefügt wurde ($F_{1;113}= 14.639$; $p= .000$; $R^2= .115$). Die Korrelationskoeffizienten sowie Kennwerte der linearen Regression können der Studie III im Anhang C entnommen werden.

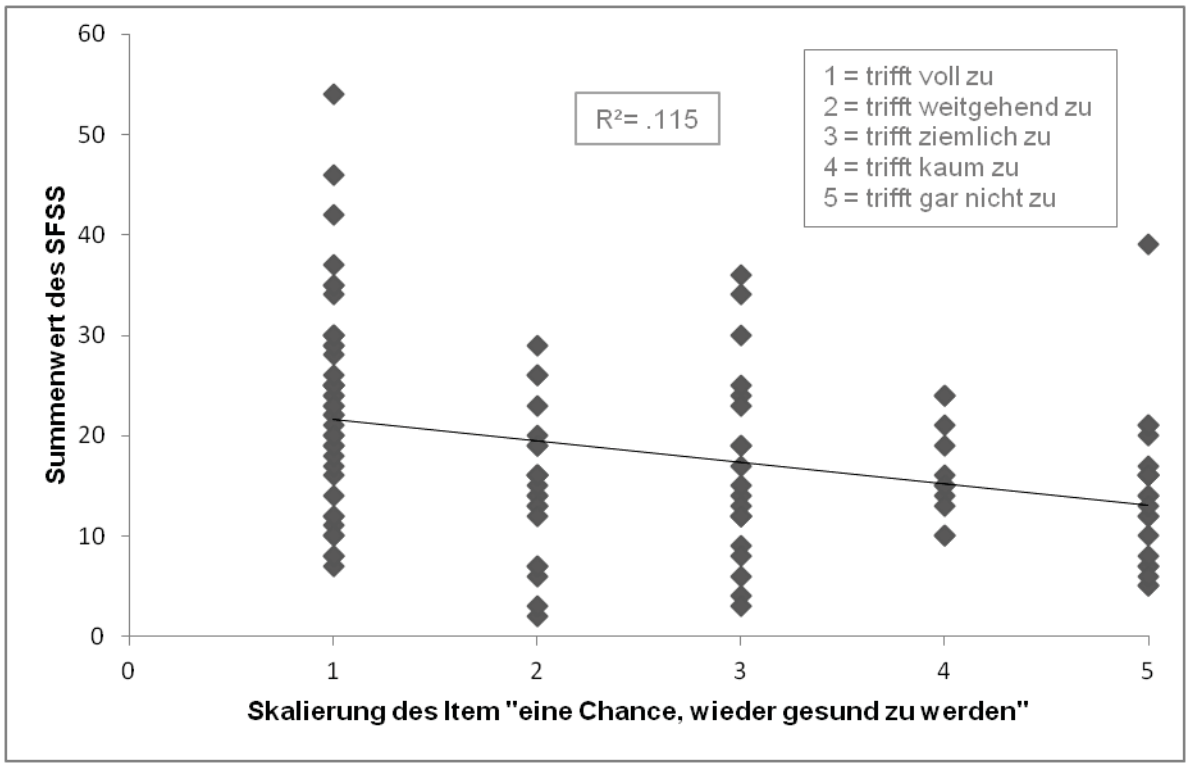

Abbildung 12: Korrelation der Summenwerte im SFSS und dem Item „Die Rente bedeutet für mich, eine Chance, wieder gesund zu werden" und dazugehöriger Regressionsgeraden in der Studie III
Anmerkung: SFSS= Strukturierter Fragebogen Simulierter Symptome

5.6 Fazit

Im Kapitel 5 wurden die Ergebnisse der drei Themenschwerpunkten dieser Dissertation präsentiert. Zum ersten Schwerpunkt kann anhand der Literaturanalyse konstatiert werden, dass im deutschsprachigen Raum gut validierte neuropsychologische BVT zur Erfassung der Leistungsbereitschaft und der Motivation bei kognitiven Beeinträchtigungen vorliegen. Über den Einsatz dieser Verfahren für die Erfassung von negativen Antwortverzerrungen bzw. Simulation oder Aggravation wird allerdings zwiespältig diskutiert. National gesehen, gibt es nur wenige BVT bzw. Validitätsskalen für geltend gemachte psychische Erkrankungen, die zudem meist auch nur an einer forensischen bzw. Studierendenstichprobe validiert wurden. Weiterhin liegen Validitätsskalen vor, die jedoch in sehr umfangreichen klinischen Fragebogen eingebettet sind. Daher war es notwendig einen Fragebogen zur Erfassung negativer Antwortverzerrungen bei psychischen Erkrankungen zu entwickeln und zu validieren.

Die Analysen im Rahmen des zweiten Schwerpunkts zeigen, dass sich IS im Mittel signifikant von Patienten sowie einer Kontrollgruppe in klinischen Skalen der HEALTH-49 unterscheiden. Die IS weisen im Mittel höhere Summenwerte auf als die anderen Untersuchungsgruppen. Sie zeigen

demnach in klinischen Fragebogen Extremantworten sowie Übergeneralisierungen. Im Gegensatz zu den IS (Schmerz) kennzeichnen sich die IS (Depression) dadurch, dass die Summenwerte der jeweiligen klinischen Skalen der HEALTH-49 normalverteilt sind. Der aus diesen Erkenntnissen sowie den Ergebnissen der Expertenbefragung und Prä-Studie entwickelte BEVA weist eine sehr gute Validität sowie Reliabilität auf. Für die Untersuchungsgruppe *„Schmerz"* wurde ein Cut-Off-Wert von 90 berechnet. Für die Untersuchungsgruppe *„Depression"* lag der ermittelte Cut-Off-Wert bei 107. Bei der Entwicklung wurde auf eine möglichst maximale Spezifität Wert gelegt. Durch den BEVA wurde ein Fragebogen mit zufriedenstellenden psychometrischen Kennwerten zur Erfassung von negativen Antwortverzerrungen bei Patienten mit depressiver Symptomatik entwickelt und validiert.

Die Ergebnisse des dritten Schwerpunkts verdeutlichen, dass Instruktionen einen Einfluss auf die Beschwerdendarstellung sowohl in den klinischen Skalen des HEALTH-49 haben als auch auf den Summenwert des SFSS. Gewarnte Personen zeigten im Mittel signifikante höhere Summenwerte in den Skalen *„Depressivität", „somatoforme Beschwerden"* sowie *„phobische Ängste"* und dem Summenwert im SFSS. Die Teilnehmer der zwei Untersuchungsgruppen wurden jeweils anhand des Cut-Off-Werts von 16 in die Gruppen mit *auffälligem Antwortverhalten* vs. *nicht auffälligem Antwortverhalten* eingeteilt. Die Chi^2-Statistik zeigte keinen signifikanten Effekt. Jedoch verdeutlichte das OR, dass die Wahrscheinlichkeit, ein auffälliges Testergebnis zu erhalten fast um das Zweifache erhöht ist, wenn die Personen vor der Bearbeitung des Fragebogens bezüglich seines Verwendungszwecks gewarnt wurden. Das Modell der linearen Regression lässt erkennen, dass der Prädiktor „die Rente bedeutet für mich, eine Chance wieder gesund zu werden" das Kriterium Summenwert des SFSS vorhersagen kann. Es konnte damit gezeigt werden, dass Instruktionen, die vor dem Ausfüllen eines Fragebogens schriftlich erteilt werden, sowie die Verknüpfung des Rentenstatus und die Möglichkeit der Gesundung zu negativen Antwortverzerrungen in Fragebogen führen können.

6 Diskussion

Die vorliegende Dissertation betrachtet auf der Grundlage von drei Themenschwerpunkten die Beschwerdenvalidierung im Kontext der sozialmedizinischen Begutachtung. Beginnend mit der Darstellung der Notwendigkeit, einen neuen BVT für die sozialmedizinische Begutachtung zu entwickeln, wird die folgende Diskussion eingeleitet. Hierfür werden über die Ergebnisse der Studie I (vgl. Anhang A) hinaus, Erkenntnisse der Studien II (Anhang B) und III (Anhang C) genutzt. Im Rahmen des zweiten Themenschwerpunkts wurde der BEVA für die Erfassung von negativen Antwortverzerrungen entwickelt. Im Kapitel 6.2 wird dessen Güte und Einsatz in der sozialmedizinischen Begutachtung erörtert. Es folgt eine Ausführung der Einflussfaktoren der Beschwerdendarstellung in der Begutachtung der Rente wegen verminderter Erwerbsfähigkeit, der dem dritten Themenschwerpunkt dieser Dissertation zuzuordnen ist. Im nachfolgenden Abschnitt werden die Limitationen der Studien betrachtet und Lösungsvorschläge dargelegt. Zusammenfassend werden anhand der Betrachtungen Implikationen für die Forschung und Praxis abgeleitet, bevor ein Fazit aus allen gesammelten Erkenntnissen dieser Dissertation gezogen wird.

6.1 Notwendigkeit der Entwicklung eines Beschwerdenvalidierungstests

Der theoretische Überblicksartikel verdeutlicht, dass die Entwicklung, Konstruktion und Validierung eines neuen BVT notwendig ist, da bisher keine ausreichend gut validierten und evaluierten deutschsprachigen Verfahren für die sozialmedizinische Begutachtung von psychischen Störungen zur Verfügung stehen. Ausreichend geprüfte BVT liegen bisher nur für kognitive Defizite vor, die auch ein Teilsymptom affektiver Störungen sein können (Cueni, Abbruzzese, Brühl & Herwig, 2011). Sie bilden jedoch nicht vollständig das Erscheinungsbild einer affektiven Störung ab. Weiterhin erfassen neuropsychologische BVT oftmals nur die Testmotivation. Patienten mit affektiven Störungen leiden jedoch häufig unter Antriebslosigkeit, die für negative Antwortverzerrungen in neuropsychologischen BVT verantwortlich gemacht werden kann. Weiterhin stellen Greiffenstein Baker, Gola, Donders und Miller (2002, S.1599) fest, dass: „… malingering in one domain does not automatically imply malingering in another domain." Diese Aussage konnte in einer Studie von Nelson und Sweet (2010) bestätigt werden. Sie zeigten, dass auffällige Testergebnisse in kognitiven Tests nicht zwangsläufig zu negativen Antwortverzerrungen in den Validitätsskalen des MMPI-2 führen. In dieser Untersuchung wiesen die Antragsteller extrem hohe Summenwerte in den Skalen des MMPI-2 auf und zeigten gleichzeitig sehr gute kognitive Leistungen, oder sie präsentierten klare kognitive Defizite und waren zugleich nicht auffällig in den Validitätsskalen. Die Ausfüh-

rungen verdeutlichen die Schwierigkeit des Einsatzes von neuropsychologischen BVT bei der Diagnostik von negativen Antwortverzerrungen bei psychischen Störungen. Möglicherweise führen die kognitiven BVT zu einem invaliden Testergebnis, obwohl keine negativen Antwortverzerrungen vorliegen oder es sind simulative Tendenzen vorhanden, die jedoch nicht in kognitiven BVT gezeigt werden. Dennoch können Validitätsparameter aus Leistungstests oder neuropsychologische BVT für ein umfassenderes Bild der Beschwerdenvalidierung in der Diagnostik bei psychischen Störungen hinzugezogen werden (Nelson, Sweet, Berry, Bryant & Granacher, 2007).

Das einzige Screening zur Beschwerdenvalidität bei psychischen Störungen stellt der SFSS dar. Dieser umfasst nicht nur Items für die Validierung von psychischen Störungen wie Depressionen oder Psychosen, sondern enthält zusätzlich Skalen für mnestische sowie neurologische Störungen und niedrige Intelligenz. Items wie *„Die Woche hat 6 Tage"* können somit schnell zu Irritationen und Ablehnungen bei den Patienten mit affektiven Störungen führen. Die Validierungsstudie von Cima et al. (2003) zeigt zudem, dass die jeweiligen Unterskalen im Gegensatz zur Gesamtskala nicht ausreichend gute psychometrische Kennwerte aufweisen. Daher erscheint es nicht sinnvoll, die Skala *„affektive Störungen"* losgelöst zur Erfassung von negativen Antwortverzerrung zu nutzen. International betrachtet, ist der SFSS (engl: Structured Inventory of Simulated Symptoms (Smith & Burger, 1997) umfangreich untersucht (Rogers, Robinson & Gillard, 2014). Jedoch fehlen im deutschsprachigen Raum noch Studien, die den Einsatz des SFSS in Kontexten außerhalb der Forensik beleuchten (van Impelen, Merckelbach, Jelicic & Merten, 2014).

Ergebnisse der Studie III weisen darauf hin, dass der SFSS möglicherweise nicht geeignet ist, um negative Antwortverzerrungen in der sozialmedizinischen Begutachtung zu erfassen. Die Ergebnisse der Studie III verdeutlichten, dass Instruktionen einen Einfluss auf die Beschwerdendarstellung haben. Es zeigte sich, dass sich die Untersuchungsgruppen im Mittel im Summenwert des SFSS signifikant unterschieden. Anhand des Cut-Off-Werts wurden die Testergebnisse je Gruppe in *auffällig* und *nicht auffällig* eingeteilt. Jedoch zeigte der Chi^2 -Test keinen signifikanten Unterschied. Auf der Grundlage der Ergebnisse des SFSS ergab sich eine Prävalenz von negativen Antwortverzerrungen in dieser Stichprobe je nach Betrachtungsweise der Gruppenzusammenstellung von 40 – 56 %. Ähnliche Ergebnisse konnten wiederum durch Kobelt, Göbber, Bassler und Petermann (2012) im rehabilitativen Kontext belegt werden, die auch den SFSS zur Schätzung hinzuzogen. Jedoch ist diese Basisrate von negativen Antwortverzerrungen für eine klinische Stichprobe sehr hoch. Andere Studien zeigten im Vergleich Prävalenzen von etwa 4-24 % (Locke, Smigielski, Powell & Stevens, 2008; Sharland &

Gfeller, 2007). Diese Studien geben Hinweise darauf, dass der SFSS wenig spezifisch hinsichtlich der Diagnostik von negativen Antwortverzerrungen ist und vermutlich das Auftreten von simulativen Tendenzen überschätzt.

Auch andere Verfahren, wie das MMPI-2 oder das VEI, enthalten Validitätsskalen zur Beschwerdenvalidierung. Einerseits weisen diese Skalen psychometrische Schwierigkeiten auf (Hank & Schwenkmezger, 2003; Vossler-Thies, Stevens, Engel & Licha, 2013), andererseits sind sie in sehr umfangreichen Verfahren eingebettet, deren Bearbeitung bis zu einer Stunde in Anspruch nehmen kann. Bei Patienten mit affektiven Störungen kann sich diese Bearbeitungszeit sogar noch deutlich verlängern. Für ein Screening scheinen demnach das MMPI-2 und das VEI nicht geeignet zu sein.

Um auf die Notwendigkeit der Entwicklung eines neuen BVT hinzuweisen, können die Ergebnisse der Studie II hinzugezogen werden. Es konnte belegt werden, dass sich das Antwortverhalten von IS und Patienten in klinischen Fragebogen unterscheidet und durch Testverfahren abgebildet werden kann. Jedoch weist die Studie II daraufhin, dass negative Antwortverzerrungen bzw. Simulation sowie Aggravation nur mit geeigneten Kontrollskalen und dazugehörigen Cut-Off-Werten erfasst werden können. Die Analysen verdeutlichen weiterhin, dass aufgrund hoher Summenwerte in klinischen Skalen nicht notwendigerweise auf eine schwere Psychopathologie geschlossen werden kann, sondern dass es auch ein Zeichen von negativen Antwortverzerrungen sein kann. Ein genaueres Bild hierzu kann lediglich ein speziell dafür entwickeltes Verfahren leisten.

Zusammenfassend kann konstatiert werden, dass die Ergebnisse der Studien I–III darauf hindeuten, dass es dringend erforderlich ist, ein zusätzliches Screening zur Beschwerdenvalidierung für die sozialmedizinische Begutachtung zu konzipieren. Die Ergebnisse der Konstruktion und Validierung des BEVA sollen nachfolgend im Kapitel 6.2 diskutiert werden.

6.2 Fragebogen BEVA

Die Validitätsmaße eines Screenings, wie Sensitivität und Spezifität, sind schwer zu schätzen, da das externe Kriterium schwierig festzulegen ist. Die Bestimmung dieses Kriteriums ist in der Forschung der Beschwerdenvalidierung zusätzlich kompliziert, da für Konstrukte wie Simulation, Aggravation oder negative Antwortverzerrungen kein Goldstandard vorliegt. Zudem sind Studien im Bereich der Beschwerdenvalidierung in der sozialmedizinischen Begutachtung im deutschsprachigen sowie internationalen Raum rar, sodass dies zusätzlich die Erforschung eines

neuen BVT erschwert. Werden falsche externe Kriterien festgelegt, so ist die Abbildung von korrekten Validitätsmaßen kaum möglich. Im Rahmen der Entwicklung des BEVA wurde ein Analogstudiendesign verwendet, welches den Vorteil hat, dass durch die Gruppenbildung *Antwortverzerrer* und *Nicht-Antwortverzerrer* klar definiert sind.

Im Rahmen der Konstruktion sowie später bei der Validierung des BEVA wurde auf eine studentische Stichprobe verzichtet und gesunde Personen im Alter von etwa 35 bis 60 Jahren über Pressemitteilungen für die Kontrollgruppe sowie die Gruppe der IS rekrutiert. Dieses Vorgehen ermöglicht eine bessere Vergleichbarkeit mit der Allgemeinbevölkerung bzw. den Antragstellern einer Rente wegen verminderter Erwerbsfähigkeit aufgrund einer psychischen Störung. Teilnehmer in derselben Alltagsspanne können sich möglicherweise besser in die Gefühlslage einer Person hineinversetzen, die aus verschiedenen Gründen eine Rente erhalten möchte. Diese Tatsache könnte Auswirkungen auf das Antwortverhalten in Fragebogen haben. Weiterhin war es durch die Rekrutierung über Pressemitteilungen möglich, das Merkmal *„Schulbildung“* sowie *„Berufsausbildung“* varianzreicher abzubilden. Durch den Verzicht einer studentischen Stichprobe bei der Konstruktion und Validierung des BEVA sollte trotz des Gebrauchs eines Analogstudiendesigns ein realitätsnäherer Bezug erreicht werden.

Da sich die Stichprobe der Studie II aus einem Teil der Konstruktionsstichprobe des BEVA zusammensetzt, können die bereits beschriebenen Vorteile des Analogstudiendesigns sowie der Vorteil des Verzichts einer studentischen Stichprobe auf die Studie II übertragen werden. Die Erkenntnisse, die aus der Studie II gewonnen wurden, trugen zur Entwicklung des BEVA bei. So zeigten die Analysen, dass sich Unterschiede von IS, depressiven Patienten und gesunden Personen in Symptomfragebogen abbilden lassen. In allen klinischen Skalen der HEALTH-49 wiesen die IS beider Untersuchungsgruppen signifikant höhere Mittelwerte auf. Dies bestätigt die Annahme, dass in einem BVT für psychische Erkrankungen Items enthalten sein sollten, die nicht nur auf seltene Symptome deuten, sondern übliche Beschwerden abbilden. Der Unterschied zu psychisch Erkrankten zeichnet sich dadurch aus, dass die IS *extrem* antworten. Das bedeutet, dass sie besonders viele Symptome mit einer gleichzeitigen schwerwiegenden Ausprägung aufweisen. Dieses Verhalten zeigen die IS nicht ausschließlich für die geltend gemachte Erkrankung, sondern sie zeigen ein übergeneralisierendes Antwortverhalten auf. Das heißt, dass übliche Symptome bekannter psychischer Erkrankung als eigene Beschwerden dargestellt werden. Daher berechnet sich der Summenwert für die Untersuchungsgruppe „Depression“ aus Items, die Symptome einer Schmerzerkrankung oder Phobie berücksichtigen. Ähnliches gilt für die Gruppe „Schmerz“.

Hierfür sind Items, die eine depressive Störung oder eine Phobie beschreiben im Gesamtwert enthalten. Damit die Tendenz der Extremantworten und Übergeneralisierungen ausreichend abgebildet werden kann, wurde eine vierstufige statt einer dichotomen Antwortskala gewählt. Das Antwortprofil der IS (Depression) ist dadurch gekennzeichnet, dass nicht bei jedem Item der extremste Wert angekreuzt wurde. Das wird dadurch ersichtlich, dass die Summenwerte der IS (Depression) normalverteilt sind und demnach eine Varianz aufweisen. Dies spricht wiederum dafür, dass die Diagnostik von negativen Antwortverzerrungen ohne ausreichende BVT oder Kontrollskalen durch klinische Skalen nicht ausreichend möglich ist.

Der BEVA weist für die beiden Untersuchungsgruppen „Depression“ und „Schmerz“ sehr gute psychometrische Kennwerte auf. Dies zeigt unter anderem die AUC der Konstruktion, die für die Gruppe *„Depression“* mit .987 und für die Untersuchungsgruppe *„Schmerz“* mit .991 berechnet wurde. Die AUC kann maximal den Wert 1 annehmen, der einen perfekten Zusammenhang beschreiben würde. Das würde bedeuten, dass alle Patienten unter dem ermittelten Cut-Off liegen würden und dementsprechend alle IS über dem Cut-Off. Da die Analysen auf eine AUC nahe des Wertes 1 hinweisen, kann daraus geschlossen werden, dass die Fehlerquote des BEVA im Allgemeinen als gering zu bewerten ist.

In der Validierungsstichprobe konnten die sehr guten Ergebnisse für beide Untersuchungsgruppen bestätigt werden. Dabei ist zu bemerken, dass die Vergleichsgruppe *„depressive Patienten“* ausschließlich aus Personen bestand, die eine sehr starke depressive Symptomatik aufwiesen. Das bedeutet, dass nicht ausschließlich Daten von Patienten mit einer Diagnose im Bereich des Kapitels V F30-39 in die Analysen eingingen, sondern dass im Rahmen der Studie mithilfe des BDI-II der Schweregrad der Depression bestimmt wurde. Dadurch waren nur Patienten mit einer mittelschweren bzw. schweren Symptomatik einbezogen. Daraus resultiert, dass der BEVA zwischen IS und Patienten mit einer schweren Psychopathologie unterscheiden kann. Ein ähnliches Vorgehen existierte bei den Schmerzpatienten. Da in dieser Gruppe somatische Beschwerden zunächst vordergründig sind, wurden für die Berechnungen auch die Daten von Personen mit einer leichten depressiven Symptomatik einbezogen. Die Analysen verdeutlichen, dass durch den Cut-Off-Wert des BEVA die Summenwerte der Schmerzpatienten und IS (Schmerz) ausreichend gut getrennt werden.

Weiterhin wurden die Validitätsmaße für den Wert bestimmt, bei dem die jeweils geringsten Summenwerte für die IS in der Konstruktionsstichprobe berechnet wurden. Für diesen niedrigsten Wert wurde eine Sensitivität von 1 ermittelt. Die Cut-Off-Werte wurden demgegenüber bei einer

Spezifität von 1 berechnet. Demzufolge ergibt sich zwischen den jeweiligen beschriebenen Werten ein Grenzbereich. Hierbei überschneidet sich eine gewisse Anzahl an Summenwerten der Patienten und der IS in der Konstruktionsstichprobe. In der Praxis kann der Grenzbereich dahingehend interpretiert werden, dass durch das Screening nicht vollständig festgestellt werden konnte, ob ein *auffälliges* oder *unauffälliges* Testergebnis vorliegt. Summenwerte in diesem Bereich sollten besonders kritisch hinterfragt werden. Dies bedeutet jedoch nicht, dass alle Werte, die außerhalb des Intervalls liegen, bedenkenlos hingenommen werden können, da der BEVA ausschließlich ein Screening darstellt, welches erste Anzeichen und Hinweise auf ein mögliches invalides Testergebnis gibt. Weiterhin sollte beachtet werden, dass ein auffälliges Testergebnis nicht automatisch ein Anzeichen für Simulation darstellt, sondern lediglich auf negative Antwortverzerrungen hinweist. Die Entwicklung eines Testverfahrens, welches eindeutig auf ein Vortäuschen einer psychischen Erkrankung hindeutet, stellt methodisch eine große Schwierigkeit dar. Testergebnisse, die auf negative Antwortverzerrungen hindeuten, müssen jederzeit kritisch hinterfragt werden, weil die Quelle der Beschwerdendarstellung nicht immer eindeutig ist. Für eine ausführlichere Darstellung ist hierzu auf Kapitel 6.3 zu verweisen.

Die Durchführbarkeit des BEVA im klinischen Setting kann als gut bewertet werden. Die Mehrzahl der Patienten, die sich für die Studie entschieden haben, füllten den soziodemografischen Fragebogen, BDI-II, BEVA und HEALTH-49 mit weniger als 15 % fehlenden Werten aus. Dies spricht dafür, dass alle Fragebogen auf wenig Ablehnung stießen und Items als angemessen formuliert angesehen wurden. Im Rahmen der Studie wurde jedoch nicht erhoben, warum Patienten der Teilnahme nicht zustimmten bzw. etwaige Fragebogen nicht ausgefüllt zurückgaben. Die Anwenderfreundlichkeit kann also nur auf der Grundlage der abgebenden Fragebogen getroffen werden. Weiterhin ist es möglich, dass der BEVA durch das Fehlen von bizarr formulierten Items nicht als ein BVT erkannt wurde, sondern eher als ein Fragebogen, der Symptome und Beschwerden erfasst (Merten, Diederich & Stevens, 2008). Dies kann dazu führen, dass die Beschwerdendarstellung durch etwaiges Misstrauen beim Ausfüllen weniger beeinflusst wird.

Der BEVA ist ein Screening, in dessen Konstruktion und Validierung Patienten mit einer mittelschweren und schweren depressiven Symptomatik berücksichtigt wurden. Der Schweregrad der Symptomatik wurde anhand des BDI-II erfasst. Dennoch ist der Einsatzbereich des BEVA vielfältig, da in den Analysen zusätzlich Schmerzpatienten inbegriffen waren. Somit ist das Screening auch bei der Beschwerdenvalidierung von psychosomatischen Erkrankungen anwend-

bar. Dieser Einsatz erscheint notwendig, wenn die Studie von Schmidt, Bernert und Spyra (2014) betrachtet wird. Sie zeigen, dass psychische Komorbiditäten bei Schmerzerkankungen häufig auftreten und das Risiko erhöhen, eine Rente wegen verminderter Erwerbsfähigkeit zu erhalten. Das führt dazu, dass die sozialmedizinische Begutachtung auf neue Herausforderungen trifft und psychische Komorbiditäten bei Schmerzerkrankungen beachtet werden sollten. Darüber hinaus kann daraus geschlossen werden, dass die Diagnostik der Beschwerdenvalidität bei Schmerzpatienten Bezug auf die psychische Symptomatik nehmen sollte. Die Studie IV verdeutlicht wiederum diese Problematik. Etwa 30 % der Schmerzpatienten in der Rehabilitationsklinik wiesen eine leichte bis schwere depressive Symptomatik auf.

6.3 Einflüsse auf die Beschwerdendarstellung in der sozialmedizinischen Begutachtung

Die Studie III liefert im deutschsprachigen Raum erste Indizien für den Einfluss von Instruktionen bzw. Warnungen auf die Beschwerdendarstellung im Kontext der sozialmedizinischen Begutachtung. Wie im Kapitel 2.4 dargestellt, liegen keine homogenen Ergebnisse in der Coaching-Forschung vor. Es zeigt sich, dass nur sehr wenige Studien den Einfluss von Warnungen im klinischen und nicht experimentellen Kontext untersucht haben.

Die vorliegende Studie III verdeutlicht, dass warnende Instruktionen vielfältige Auswirkungen auf das Antwortverhalten haben können. Die Warnung, dass nachfolgende Fragebogen die Nachvollziehbarkeit der berichteten Beschwerden erfassen können, kann missverstanden werden. Die Teilnehmer können diesen Hinweis als Chance verstehen, ihre möglichen Beschwerden wirkungsreich darzulegen. Gegebenenfalls führt die Instruktion dazu, dass die Antragsteller das Gefühl haben, durch die Befragung besondere Aufmerksamkeit zu genießen und daher ihre Symptome extremer darstellen. Dadurch kann interpretiert werden, dass die Teilnehmer einerseits ihr Antwortverhalten nicht als unehrlich wahrnehmen, andererseits zeigen Studien (Mazar & Ariely, 2006), dass Personen, trotz des Risikos einer Lüge überführt zu werden, ihre Unehrlichkeit nicht zügeln.

Der Entscheidungsprozess für Unehrlichkeit unterliegt auch einer Kosten-Nutzen- Analyse und moralischen Erwägungen auf Seiten der Teilnehmer (Shalvi, Dana, Handgraaf & De Dreu, 2011). Es ist möglich, dass die Instruktion den Teilnehmern verdeutlicht, dass sie die Möglichkeit haben, ihre Beschwerden darzustellen. Jedoch nehmen die Teilnehmer das Risiko als sehr gering wahr, dass verzerrtes Antwortverhalten durch den Fragebogen erfasst werden könnte (Loewen, Dawes, Mazar, Johannesson, Koellinger & Magnusson, 2013). Es kann angenommen werden, dass

unehrliches Verhalten von der Selbstregulation abhängig ist (Higgins, 1998). Das Cohens d zeigt, dass kleine Effekte für die Skalen *„Depressivität“* sowie *„somatoforme Störungen“* und ein mittlerer Effekt für die Skala *„phobische Ängste“* angenommen werden können. Möglicherweise ist es auch ein Ergebnis eines Abwägungsprozesses, dass geringere Verzerrungen noch unentdeckt bleiben.

Dies unterstützen die Ergebnisse der Studie II, in der keine warnenden Instruktionen gegeben wurden. Dort belegte das Cohens d zwischen den IS und den Patienten große Effekte in allen klinischen Skalen der HEALTH-49. Der Effekt der Instruktionen wurde weiterhin im SFSS deutlich. Die gewarnten Teilnehmer zeigten im Durchschnitt signifikant höhere Summenwerte als die nicht gewarnten Personen. Eine höhere Ausprägung spricht in diesem Fall für negative Antwortverzerrungen. Das OR verwies auf eine fast zweifach erhöhte Chance über den Cut-Off-Wert von 16 im SFSS zu liegen, wenn die Teilnehmer in der gewarnten Gruppe waren. Jedoch zeigte die Chi^2-Statistik keinen signifikanten Unterschied. Dieses Ergebnis könnte darauf hinweisen, dass der Cut-Off-Wert des SFSS weniger geeignet scheint, das Antwortverhalten korrekt in *auffällig* und *nicht auffällig* zu klassifizieren.

Als weiterer Einflussfaktor wurde die Motivation für den Erhalt einer Rente in der Studie III untersucht. Vielfach wurde in der Literatur (Heubrock & Petermann, 1998) ein sekundärer Krankheitsgewinn angenommen, der zu negativen Antwortverzerrungen führen kann. Diese These konnte im Rahmen der vorliegenden Studie nicht belegt werden. Die Variable „Die Rente bedeutet für mich finanzielle Sicherheit.“ zeigte keinen korrelativen, signifikanten Zusammenhang mit dem Summenwert des SFSS. Jedoch unterliegt dieses Item stark dem Störfaktor der sozialen Erwünschtheit. Es ist fraglich, ob Personen, auch wenn sie sich durch die Rente vor allem finanzielle Vorteile erhoffen, dieses direkt in einem Fragebogen im Rahmen einer sozialmedizinischen Begutachtung angeben.

Lediglich die Korrelation der Variablen „Die Rente bedeutet für mich, kein Hartz-IV-Empfänger zu sein.“ und des Summenwerts des SFSS zeigte eine statistische Signifikanz. Das bedeutet, dass Personen, die der beschriebenen Aussage voll zustimmten, gleichzeitig einen hohen Summenwert im SFSS zeigten. Es wird angenommen, dass je höher der Summenwert ist, desto mehr negative Antwortverzerrungen werden gezeigt. Weiterhin kann bei dieser Variablen interpretiert werden, dass die Antragsteller sich dadurch keinen finanziellen Zugewinn erhoffen, sondern lediglich nicht stigmatisiert werden wollen. Gleichwohl verdeutlichte das Modell der linearen Regression, dass nur der Prädiktor *„Die Rente bedeutet für mich, eine Chance wieder gesund zu werden.“* das

Kriterium Summenwert des SFSS vorhersagen kann. Personen, die die Heilung ihrer Krankheit einzig darin sehen, dass sie eine Rente erhalten, kennzeichnen dies vermutlich dadurch, dass sie vermehrt negative Antwortverzerrungen aufweisen. Jedoch legen Studien dar, dass eine Rente nicht zwangsläufig dazu führt, dass Frührentner wieder ins Erwerbsleben zurückkehren. Das bedeutet, dass sie nicht notwendigerweise durch eine Rente wieder gesund werden (Kobelt, Grosch, Hesse, Gebauer & Gutenbrunner, 2009).

Die Studie III zeigt eindrucksvoll, dass sich Gutachter sowie Kliniker den Einfluss von Instruktionen und der Motivation für den Erhalt einer Rente bewusst machen sollten, da die im Rahmen der Studie untersuchten Faktoren negative Antwortverzerrungen in Symptomskalen und BVT hervorbringen können. Gerade in der sozialmedizinischen Begutachtung können weitere externale Umstände zu einer verzerrten Beschwerdendarstellung führen, weil die Antragsteller für eine Rente wegen Erwerbsunfähigkeit stark vom Ergebnis der Begutachtung für ihre weitere Zukunft abhängig sind. Demnach bewirken nicht zwangsläufig bewusste Täuschungen eine Verzerrung, sondern auch Erkrankungen, Hinweise der Gutachter oder eine zugrundeliegende Motivation. Die Ursachen sowie Erklärungen für negative Antwortverzerrungen bzw. Simulationen oder Aggravationen sind folglich vielfältig.

6.4 Limitationen

Im folgenden Kapitel sollen die Grenzen der Studien dieser Dissertation aufgezeigt werden. Beginnend mit einer kritischen Betrachtung des Analogstudiendesigns, das in den Studien II und IV eingesetzt wurde. Darüber hinaus werden die Grenzen der Stichprobenerhebung und der Klassifikation der Diagnosen betrachtet. Separat werden die Limitationen der Studie III aufgeführt.

Analogstudiendesign. Im Kontext der Studien II und IV wurde ein Analogstudiendesign verwendet, dessen methodische Problematik (z.B. geringe externe Validität) teilweise im Kapitel 1.2 angesprochen wurde. Da für die Erfassung von negativen Antwortverzerrungen bzw. Simulation sowie Aggravation kein Goldstandard vorliegt, muss dem anhand von experimentellen Bedingungen entgegengewirkt werden. Durch das verwendete Design konnte zwar klar eingegrenzt werden, in welcher der Gruppen Personen sind, die eine psychische Erkrankung vortäuschen, jedoch kann ausgeschlossen werden, dass die Teilnehmer der Gruppe IS nicht die gleichen Motive und Beweggründe haben wie beispielsweise die Antragsteller einer Rente wegen verminderter Erwerbsfähigkeit. Zudem hatte ein Großteil der IS vermutlich keine Erfahrungen bzw. wenig Expertise im Zusammenhang mit psychischen Erkrankungen wie depressiven oder

somatoformen Störungen. Somit ist es schwieriger, sich wirklich in das Krankheitsbild hineinzuversetzen. Das mangelnde Wissen über psychische Erkrankungen kann wiederum dazu führen, dass extreme Übertreibungen auftreten. Im Rahmen der vorliegenden Analogstudien fehlte der Hinweis, die Beschwerden in den Fragebogen nicht übertrieben stark darzustellen bzw. die psychische Erkrankung so vorzutäuschen, dass die dahinterliegende Absicht nicht leicht erkannt wird. In folgenden Studien sollten diese Instruktionen hinzugefügt werden, um zu untersuchen, ob sich die Beschwerdendarstellung der IS verändert. Diese starke Überspitzung kann eine Ursache für die sehr gut ausfallenden psychometrischen Kennwerte des BEVA sein. Es ist möglich, dass hierdurch ein unrealistisch hoher Summenwert durch die IS erzielt wurde und das Screening daher gut zwischen psychisch Erkrankten und den IS trennt. Eine endgültige Aussage über die Güte und Qualität des BEVA kann nur anhand weiterer Studien erfolgen.

Durch den Einsatz eines known-group-Designs wäre eine realitätsnähere Umsetzung möglich. Jedoch ist hierdurch die Erfassung der Simulation schwieriger, wodurch eine falsche Gruppenzuordnung in *negative Antwortverzerrung* vs. *unauffälliges Antwortverhalten* nicht ausgeschlossen werden kann. Dieses Forschungsdesign weist also ein hohes Maß an externer Validität und ein geringes Maß an interner Validität auf. Es ist erforderlich, diese beiden Untersuchungsformen miteinander zu verknüpfen. Demzufolge ist es notwendig, dass das im Rahmen dieser Dissertation entwickelte Screening erneut im klinischen sowie gutachterlichen Kontext mit einem known-group-Design zu überprüfen, um die Validität sowie Reliabilität erneut berechnen zu können.

Stichprobe. Zunächst sollen die Grenzen der Untersuchungsbedingungen für die IS betrachtet werden. Alle gesunden Teilnehmer erhielten ihre Studienmaterialien postalisch. Daher ließ sich nicht ausreichend kontrollieren, unter welchen Bedingungen die Fragebogen ausgefüllt wurden. Es ist nicht eindeutig, ob Teilnehmer zusätzliche Unterstützung bzw. Informationen einholten oder sie durch andere Faktoren in der Beantwortung der Fragebogen gestört wurden. Jedoch kann das Wissen über psychische Erkrankungen oder Testverfahren auch durch kontrollierte Untersuchungsbedingungen nicht gänzlich ausgeschlossen werden.

Weiterhin sind die Unterschiede in den soziodemografischen Daten der Konstruktions- sowie Validierungsstichprobe kritisch zu bewerten. Hier zeigen sich in mehreren Variablen signifikante Effekte zwischen den Gruppen. Dies trifft unter anderem auf das Alter zu. Bei näherer Betrachtung der deskriptiven Werte deutet sich jedoch an, dass sich die Personen der jeweiligen

Gruppen trotz der signifikanten Effekte in etwa einer Altersrange befinden. Im Mittel sind alle Studienteilnehmer zwischen 44 und 50 Jahren alt.

Darüber hinaus liegt ein signifikanter Effekt für die Variable *„Schulbildung“* vor. Dies kann zum Teil dadurch begründet sein, dass die Pressemitteilungen auch über einen E-Mail-Server der Universität Bremen verschickt wurden und sich daher viele Akademiker für die Studie meldeten. Bisher liegen keine bekannten Studienergebnisse vor, die verdeutlichen, dass der Bildungsgrad einen Einfluss auf die Empfindungen und Darstellung von (vorgetäuschten) Beschwerden hat. Im Rahmen der Studie II, deren Stichprobe einen Teil der Konstruktionsstichprobe darstellt, wurde dieser mögliche Effekt auf der Grundlage einer MANOVA überprüft. Es konnte dargelegt werden, dass kein multivariater bzw. univariater Effekt für die Schulbildung auf die klinischen Skalen des HEALTH-49 vorlag. Darüber hinaus verdeutlichten die Analysen keinen signifikanten Interaktionseffekt für die Gruppenzugehörigkeit und die Schulbildung. Dieses Ergebnis weist darauf hin, dass es keinen Einfluss der Schulbildung auf das Antwortverhalten in den klinischen Skalen der HEALTH-49 gab. Dennoch ist zu bedenken, dass die Aufgabe des Simulierens einer psychischen Störung ein hohes Maß an Abstraktions- und Vorstellungsvermögen verlangt (Baron-Cohen, Tager-Flusberg & Cohen, 2000). Daraus lässt sich ableiten, dass die Aufgabe der IS von kognitiven Fähigkeiten abhängig ist. Ob sich Unterschiede abbilden lassen, wenn ein Großteil der IS einen anderen Bildungsabschluss hätte, muss in weiteren Studien überprüft werden.

Darüber hinaus zeigten sich in der Validierungsstichprobe signifikante Effekte für die soziodemografischen Variablen *„Familienstand“* und *„Geschlecht“*. Jedoch ist es aufgrund von fehlenden Studienergebnissen auszuschließen, dass der Familienstand einen Einfluss auf die etwaige Symptomdarstellung hat. Weiterhin liegen im Forschungsbereich der Beschwerdenvalidierung keine Studien vor, die Geschlechtseffekte bei negativen Antwortverzerrungen thematisieren. Dies gilt es zu überprüfen, um eine Aussage über den Einfluss des Geschlechts auf eine etwaig verzerrte Beschwerdendarstellung zu treffen.

Es kann geschlussfolgert werden, dass eine Konfundierung personenbezogener Daten auf die Testergebnisse und das Antwortverhalten nur auszuschließen ist, wenn ein matching durchgeführt worden wäre. Hierbei werden die jeweiligen Untersuchungsgruppen so zusammengestellt, dass sie in Bezug auf ihre personenbezogenen Daten übereinstimmen. Über die soziodemografischen Variablen hinaus soll im Folgenden spezifischer auf die Grenzen der Klassifikation der Diagnosen eingegangen werden.

Diagnosen. Die Diagnosen im Rahmen aller Studien erfolgten auf der Grundlage der Basisdokumentation der Kliniken bzw. der sozialmedizinischen Begutachtung. Damit konnte keine einheitliche Diagnostik gewährleistet werden, wie es beispielsweise bei der Durchführung des *„Strukturiertes Klinisches Interview für DSM-IV“* (SKID; Wittchen, Zaudig & Fydrich, 1997) der Fall gewesen wäre. Bei der Einteilung der Schmerzpatienten wurden Personen der Gruppe zugeordnet, wenn eine Diagnose des Kapitels XII M der ICD-10 vorlag. Somit ergab sich auch hier ein heterogenes Diagnosebild. Im Kontext der Studien II und IV wurde anhand des BDI-II die depressive Symptomatik gemessen, um ein gruppenübergreifendes Bild zu erhalten. Durch das Screening kann eine Diagnose zwar nicht gestellt werden. Es ermöglicht jedoch eine Aussage über den Schweregrad der depressiven Symptomatik.

Studie III. Im Kontext der Studie III wiesen die Teilnehmer auffällig viele fehlende Werte auf. Etwaige Gründe für das unvollständige Ausfüllen der Fragebogen konnten nicht konkret erfasst werden. Auffällig ist, dass den SFSS nur 54 Personen vollständig beantworteten, hingegen 133 Teilnehmer das Modul der HEALTH-49 komplett ausfüllten. Es scheint, dass das Ausfüllen des SFSS Unbehagen bei den Antragstellern auslöste, da gerade dieser Fragebogen viele bizarre Items enthält und ein sehr breites Spektrum von Diagnosen abbildet. Eine engmaschige Betreuung beim Ausfüllen der Fragebogen ist eventuell nötig, damit die Teilnehmer bei möglichen Problemen gezielt nachfragen können. Obwohl ausdrücklich nur Personen eingeschlossen wurden, die ihr Einverständnis zur vorliegenden Studie gaben, erweckt das Beantworten von Fragebogen im Rahmen einer sozialmedizinischen Begutachtung eventuell Misstrauen, was dazu führen kann, dass einige Items nicht beantwortet werden. In der Studie III wurden die Werte konservativ (SFSS) bzw. anhand von etablierten Imputationsverfahren ersetzt (Musil, Warner, Yobas & Jones, 2002). Jedoch leistet dieser Ersatz nur eine Schätzung und kann nicht perfekt das tatsächliche Antwortverhalten abbilden.

Weiterhin zeigen die Analysen zum Vergleich der personenbezogenen Daten zwischen den beiden Untersuchungsgruppen keine signifikanten Unterschiede. Die jeweiligen Gruppen sind in Bezug auf ihre Diagnosen sehr heterogen. Es ist nicht vollkommen auszuschließen, dass etwaige Effekte durch die Gruppenzusammensetzung beeinflusst wurden. Dennoch liegen bisher kaum Studienergebnisse vor, die negative Antwortverzerrungen im Zusammenhang mit einzelnen psychischen Störungen besonders betonen. Ausgenommen sind hier artifizielle Störungen, die jedoch in dieser Studie nicht auftraten.

6.5 Implikationen für die Forschung und Praxis

Als Ergebnis der einzelnen Studien dieser Dissertation ging ein neuer BVT hervor, der vor allem für die Beschwerdenvalidierung psychischer Störungen entwickelt wurde. Es ist notwendig, die Anwendbarkeit des Fragebogens sowie seine Validität und Reliabilität in einem praxisnahen Setting zu überprüfen. Dazu sollte die Übereinstimmung der Ärzte in der sozialmedizinischen Begutachtungen und das Testergebnis des BEVA untersucht werden. Zusätzlich könnte die Konstruktvalidität anhand der Korrelationen mit einem anderen BVT, beispielsweise dem SFSS oder den Validitätsskalen des MMPI-2 oder VEI, erfolgen. Dies würde zu einem umfassenderen Bild der Güte des BEVA beitragen. Des Weiteren ist der BEVA bisher nur bei Patienten mit depressiven Störungen und Schmerz mit depressiver Symptomatik untersucht worden. Es gilt, das Screening in weiteren Diagnosegruppen wie beispielsweise Angststörungen, Anpassungsstörungen oder posttraumatischen Belastungsstörungen zu testen, um den BEVA vielfältig in der sozialmedizinischen Begutachtung oder im klinischen Kontext anwenden zu können. Darüber hinaus erscheint es notwendig, mögliche Effekte des Geschlechts oder des Migrationshintergrunds auf das Antwortverhalten im BEVA zu überprüfen (Lilienfeld, Thames & Watts, 2013). Studien belegen, dass 50 % der Personen mit einem Migrationshintergrund im SFSS eine invalide Symptomdarstellung zeigen. Hingegen wiesen nur 33 % der Patienten ohne Migrationshintergrund negative Antwortverzerrungen auf (Göbber, Petermann, Piegza & Kobelt, 2012). Jedoch wurde im Rahmen der Studie auch gezeigt, dass sich die signifikanten Unterschiede vermutlich größtenteils durch Schicht- und Geschlechtszugehörigkeit erklären lassen. Die uneindeutigen Ergebnisse weisen darauf hin, dass ein dringender Forschungsbedarf über die Einflüsse von personenbezogenen Variablen auf die Beschwerdendarstellung vorliegt.

Vor allem die Ergebnisse der Studie III weisen darauf hin, dass der Ablauf von Begutachtungen möglicherweise weiter standardisiert werden sollte. Die Analysen zeigen, dass gewisse Instruktionen oder Hinweise die Beschwerdendarstellung in Fragebogen manipulieren und dies zu negativen Antwortverzerrungen führen kann. Bei Nichtbeachtung dieses Einflusses könnten falsche Schlussfolgerungen durch den Kliniker oder Gutachter gezogen werden. Es gilt, die Quelle des Antwortverhaltens stets zu überprüfen und mögliche Motive für Verzerrungen zu hinterfragen. Dies zeigen auch die Ergebnisse des zweiten Teils der Studie III, in der Motive im Zusammenhang mit auffälligen Testergebnissen verglichen wurden. Auch hier gilt, es im Kontext der Begutachtung weitere Fragen zu etwaigen Gründen für beispielsweise eine Rente umfassend zu hinterfragen. Eine weitere Implikation der Forschung besteht darin, zu überprüfen, ob sich die

abgebildeten Effekte der beschriebenen Studie im Rahmen einer Gruppe mit homogenen Diagnosen replizieren lassen. Zusätzlich scheint es interessant, ob das Antwortverhalten durch die Instruktionen sowie Motive auch in anderen Settings beeinflusst werden kann. Es ist möglich, dass diese Hinweise im klinisch therapeutischen Kontext durch den Patienten anders interpretiert werden als von Antragstellern einer Rente.

Weiterhin sollte in Betracht gezogen werden, den BEVA oder auch andere BVT im klinischen oder rehabilitativen Kontext anzuwenden. Feldmann (2004) beschreibt beispielsweise die besondere Problematik von Patienten mit artifiziellen Störungen. Das Krankheitsbild ist dadurch gekennzeichnet, dass die Betroffenen medizinische Zuwendung erhalten möchten und sich beispielsweise selbst verletzen oder vergiften. Zusätzlich fordern sie immer mehr medizinische Untersuchungen, die ihren Krankheitsstatus belegen sollen. Der Einsatz eines BVT könnte die Diagnostik einer artifiziellen Störung unterstützen.

Weiterhin sollte der BEVA in den klinisch-therapeutischen Alltag integriert werden. Liegen bei einem Patienten negative Antwortverzerrungen vor, kann dies einen erheblichen Einfluss auf seinen Therapieerfolg haben. Mögliche Motive für eventuell verzerrte Beschwerdendarstellungen sollten vom Kliniker hinterfragt werden, um somit möglicherweise den weiteren Therapieverlauf unterstützen zu können.

6.6 Fazit

Studien belegen, dass die Prävalenz von verzerrten Antwortprofilen im Bereich der sozialmedizinischen Begutachtung nicht unerheblich ist. Trotz des Wissens möglicher invalider Testergebnisse, liegen bisher wenige Publikationen vor, die sich mit der Beschwerdenvalidierung in der sozialmedizinischen Begutachtung beschäftigen. Zudem wurden psychische Erkrankungen wie affektive Störungen in der Forschung der Beschwerdenvalidierung meist ausgeklammert.

Im Rahmen der vorliegenden Dissertation konnte auf der Basis der promotionsrelevanten Studien verdeutlicht werden, dass gerade für die Erfassung negativer Antwortverzerrungen bei psychischen Erkrankungen kein ausreichend gut entwickeltes und validiertes Screening vorliegt. Es zeigte sich, dass die wenigen deutschsprachigen Verfahren für den besonderen Kontext der sozialmedizinischen Begutachtung nicht geeignet sind.

Es wurde auf den Erkenntnissen unterschiedlicher Informationsquellen der BEVA entwickelt, der für die sozialmedizinische Begutachtung von psychischen Störungen konzipiert wurde und zufrie-

denstellende psychometrische Kennwerte aufweist. Es ist jedoch darauf hinzuweisen, dass die Qualität des BEVA in künftigen Studien in einem known-group-Design überprüft werden sollte, da der Fragebogen bisher nur anhand eines Analogstudiendesigns validiert wurde. Um einen übergreifenden Einsatz des BEVA zu gewährleisten, sollte der Fragebogen in unterschiedlichen klinischen Kontexten und anhand einer Patientenstichprobe mit weiteren psychischen Erkrankungen eingesetzt werden.

Der BEVA ist ein Screening, das die Diagnostik in der sozialmedizinischen Begutachtung verbessern kann. Wichtig ist dabei, die Ursache für etwaige negative Antwortverzerrungen zu hinterfragen. Die Ergebnisse der vorliegenden Dissertation zeigen, dass ein mögliches verzerrtes Antwortverhalten in klinischen Fragebogen durch verschiedene Faktoren herbeigeführt werden kann. So führen mögliche Instruktionen, die den Antragstellern in Begutachtungssituationen gegeben werden, zu einem invaliden Testergebnis.

Im Übrigen kann der starke Wunsch, wieder gesund zu werden, zusätzlich die Beschwerdendarstellung in einem Fragebogen beeinflussen. Mögliche auffällige Testergebnisse sollten demnach nicht durch den Gutachter als gegeben hingenommen, sondern auf ihre möglichen Beweggründe hin überprüft werden. Die vorliegende Dissertation konnte hierbei nicht alle Gesichtspunkte von möglichen Einflussfaktoren analysieren. Für die weitere Forschung ist es wichtig, zusätzliche Variablen wie „coaching“ oder „das Wissen über psychische Erkrankungen“ auf den Einfluss des Antwortverhaltens in Fragebogen zu untersuchen.

Literatur

Amelang, M. & Zielinski, W. (2004). *Psychologische Diagnostik und Intervention* (3. Auflage). Heidelberg: Springer.

American Psychiatric Association (APA) (2000). *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR*. Washington, DC: American Psychiatric Association.

Armistead-Jehle, P. (2010). Symptom Validity Test performance in U.S. Veterans referred for evaluation of mild TBI. *Applied Neuropsychology, 17*, 52-59.

Baron-Cohen, S., Tager-Flusberg, H. & Cohen, D. J. (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd edition). New York: Oxford University Press.

Beck, A. T., Steer, R. A. & Brown, G. K. (1996). *Beck Depression Inventory - Second Edition. Manual.* San Antonio, TX: The Psychological Corporation.

Bianchini, K. J., Mathias, C. W. & Greve, K. W. (2001). Symptom Validity Testing: A critical review. *The Clinical Neuropsychologist, 15*, 19-45.

Blaskewitz, N., Gorny, I., Merten, T. & Kathmann, N. (2007). Anleitung zum nicht authentischen Testverhalten: Coaching-Praktiken und ihre Folgen. *Praxis der Rechtspsychologie, 17*, 108-124.

Cima, M., Hollnack, S., Kremer, K., Knauer, E., Schellbach-Matties, R., Klein, B. & Merckelbach, H. (2003). Strukturierter Fragebogen Simulierter Symptome. Der *Nervenarzt, 74*, 977-986.

Cima, M., Merckelbach, H., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R. & Muris, P. (2003). The other side of malingering: supernormality. *The Clinical Neuropsychologist, 17*, 235-243.

Cima, M., Pantus, M. & Dams, L. (2007). Simulation und Dissimulation in Abhängigkeit vom strafrechtlichen Kontext und der Persönlichkeit. *Praxis der Rechtspsychologie, 17*, 47-62.

Cohen, J. (1988). *Statistical power analysis for the behhavioral sciences*. New York: Erlbaum.

Cueni, C., Abbruzzese, E. A., Brühl, A. B. & Herwig, U. (2011). Neuropsychologische Aspekte der Depression. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 59*, 103-114.

Dohrenbusch, R. (2009). Symptom- und Beschwerdevalidierung chronifizierter Schmerzen in sozialmedizinischer Begutachtung. *Der Schmerz, 23*, 231-240.

Dohrenbusch, R., Henningsen, P. & Merten, T. (2011). Die Beurteilung von Aggravation und Dissimulation in der Begutachtung psychischer und psychosomatischer Störungen. *Versicherungsmedizin, 63*, 81-85.

Dohrenbusch, R. & Schneider, W. (2011). Integration psychologischer Testergebnisse in die Begutachtung psychischer Erkrankungen. *Versicherungsmedizin, 63*, 76-81.

Dressing, H., Foerster, K., Widder, B., Schneider, F. & Falkai, P. (2011). Zur Anwendung von Beschwerdenvalidierungstests in der psychiatrischen Begutachtung. Der *Nervenarzt, 82*, 388-390.

Dunn, T. M., Shear, P. K., Howe, S. & Ris, M. D. (2003). Detecting neuropsychological malingering: effects of coaching and information. *Archives of Clinical Neuropsychology, 18*, 121-134.

Edens, J. F., Buffington, J. K. & Tomicic, T. L. (2000). An investigation of the relationship between psychopathic traits and malingering on the Psychopathic Personality Inventory. *Assessment, 7*, 281-296.

Egede, L. E. (2005). Effect of comorbid chronic diseases on prevalence and odds of depression in adults with diabetes. *Psychosomatic Medicine, 67*, 46-51.

Engel, R. R. (2000). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2)*. Bern: Huber.

Erdal, K. (2004). The effects of motivation, coaching, and knowledges of neuropsychology on the simulated malingering of head injury. *Archives of Clinical Neuropsychology, 19*, 73-88.

Erdal, K. (2009). Why one fakes a head injury affects how one fakes a head injury. *Applied Neuropsychology, 16*, 42-48.

Fahrenberg, J., Hampel, R. & Selg, H. (2010). *Freiburger Persönlichkeitsinventar (FPI-R). Manual.* Göttingen: Hogrefe.

Feldmann, M. D. (2004). *Playing sick. Untangling the web of Munchausen syndrome, Munchausen by proxy, malingering & factitious disorder*. New York: Routledge.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

Foerster, K. (1984). *Neurotische Rentenbewerber*. Stuttgart: Enke.

Gesundheitsberichtserstattung des Bundes (2013). *Rentenzugänge wegen verminderter Erwerbsfähigkeit in der Gesetzlichen Rentenversicherung im Laufe des Berichtsjahres (Anzahl und je 100.000 aktiv Versicherte). Gliederungsmerkmale: Jahre, Region, Zugangsalter, Geschlecht, 1. Diagnose (ICD-10)* Verfügbar unter: https://www.gbe-bund.de/oowa921-install/servlet/oowa/aw92/WS0100/_XWD_PROC?_XWD_336/1/XWD_CUBE.DRILL/_XWD_364/D.946/14493 [30.01.2015].

Giger, P., Merten, T., Merckelbach, H. & Oswald, M. (2010). Willentliche Testverfälschung bei Verfahren zur Erfassung von Dissoziation: Ergebnisse einer Begutachtungsstudie. *Praxis der Rechtspsychologie, 20*, 131-147.

Gill, D., Green, P., Flaro, L. & Pucci, T. (2007). The role of effort testing in independent medical examinations. *The Medico-Legal Journal, 75*, 64-71.

Gjesdal, S., Haug, K., Ringdal, P., Mæland, J. G., Hagberg, J., Røraas, T. … Alexanderson, K. (2009). Sickness absence with musculoskeletal or mental diagnoses, transition into disability pension and all-cause mortality: A 9-year prospective cohort study. *Scandinavian Journal of Public Health, 37*, 387-394.

Göbber, J., Petermann, F., Piegza, M. & Kobelt, A. (2012). Beschwerdenvalidierung bei Rehabilitanden mit Migrationshintergrund in der Psychosomatik. *Die Rehabilitation, 51*, 356-364.

Götz, M., Roth, S. & Chojetzki, U. (2011). Rechtliche Grundlagen für Leistungen der gesetzlichen Rentenversicherung. In D. R. Bund (Hrsg.), *Sozialmedizinische Begutachtung für die gesetzliche Rentenversicherung* (S. 3-27). Heidelberg: Springer.

Green, P. (2003a). *Green´s Word Memory Test*. Edmonton: Green´s Publishing.

Green, P. (2003b). *Medical Symptom Validity Test: User's manual*. Edmonton: Green's Publishing.

Grieve, R. (2012). The role of personality, psychopathy, and previous experience with assessment in intentions to fake in psychological testing. *Current Psychology, 31*, 414-422.

Greiffenstein, M.F., Baker, W.J., Gola, T., Donders, J., & Miller, L. (2002). The fake bad scale in atypical and severe closed head injury litigants. *Journal of Clinical Psychology*, *58*, 1591–1600.

Groves, J. A. & Engel, R. R. (2007). The german adaptation and standardization of the Personality Assessment Inventory (PAI). *Journal of Personality Assessment, 88*, 49-56.

Gunstad, J. & Suhr, J. A. (2001). Efficacy of the full and abbreviated forms of the Portland Digit Recognition Test: vulnerability to coaching. *The Clinical Neuropsychologist, 15*, 397-404.

Gunstad, J. & Suhr, J. A. (2004). Use of the abbreviated Portland Digit Recognition Test in simulated malingering and neurological groups. *Journal of Forensic Neuropsychology, 4*, 33-47.

Hank, P. & Schwenkmezger, P. (2003). Das Minnesota Personality Inventory-2 (MMPI-2) Testbesprechung im Auftrag des Testkuratoriums. *Report Psychologie, 28*, 294-306.

Hathaway, S. R. & McKinley, J. C. (2000) *MMPI-2. Minnesota Multiphasic Personality Inventory 2.* Minneapolis: University of Minnesota Press.

Hautzinger, M., Keller, F. & Kühner, C. (2009). *BDI-II. Beck-Depressions-Inventar. Revision* (2. Auflage). Frankfurt: Pearson Assessment.

Henry, G. K. (2005). Probable malingering and performance on the test of variables of attention. *The Clinical Neuropsychologist, 19*, 121-129.

Heubrock, D. & Petermann, F. (1998). Neuropsychological assessment of suspected malingering: research results, evaluation techniques, and further directions of research and application. *European Journal of Psychological Assessment, 14*, 211-225.

Heubrock, D. & Petermann, F. (2011). *Testbatterie zur Forensischen Neuropsychologie, TBFN. Testmanual. Neuropsychologische Diagnostik bei Simulationsverdacht* (3. Auflage). Frankfurt: Pearson Assessment.

Higgins, E. T. (1998). *Promotion and prevention: Regulatory focus as a motivational principle*. New York: Academic Press.

Jelicic, M., Ceunen, E., Petersen, M. J. V. & Merckelbach, H. (2011). Detecting coached feigning using the Test of Malingering (TOMM) and the Structured Inventory of Malingered Symtomatology (SIMS). *Journal of Clinical Psychology, 67*, 850-855.

Johnson-Greene, D., Brooks, L. & Ference, T. (2013). Relationship between performance validity testing, disability status, and somatic complaints in patients with fibromyalgia. *Clinical Neuropsychologist, 27*, 148-158.

Johnson, J. L., Bellah, C. G., Dodge, T. I. M., Kelley, W. & Livingston, M. M. (1998). Effect of warning on feigned malingering on the WAIS-R in college samples. *Perceptual and Motor Skills, 87*, 152-154.

Johnson, J. L. & Lesniak-Karpiak, K. (1997). The effect of warning on malingering on memory and motor tasks in college samples. *Archives of Clinical Neuropsychology, 12*, 231-238.

King, J. & Sullivan, K. A. (2009). Deterring malingered psychopathology: The effect of warning simulating malingerers. *Behavioral Sciences & the Law, 27*, 35-49.

Kobelt, A., Göbber, J., Bassler, M. & Petermann, F. (2012). Beschwerdenvalidität im Rahmen stationärer psychosomatischer Rehabilitation. *Die Rehabilitation, 51*, 349-355.

Kobelt, A., Grosch, E., Hesse, B., Gebauer, E. & Gutenbrunner, C. (2009). Wollen psychisch erkrankte Versicherte, die eine befristete Rente wegen voller Erwerbsminderung beziehen, wieder ins Erwerbsleben eingegliedert werden? *Psychotherapie, Psychsosomatik, Medizinische Psychologie, 59*, 273-280.

Kobelt, A., Geissler N., Walter, F. & Petermann, F. (2014). Reicht es, im Rahmen der Begutachtung von Rentenantragstellern zu warnen, dass die Angaben überprüft werden, um weniger verzerrte Antwortergebnisse zu erhalten?. [Abstract] In Deutsche Rentenversicherung Bund (Hrsg.), *23. Rehabilitationswissenschaftliches Kolloquium – Arbeit – Gesundheit - Rehabilitation vom 10. bis 12. März 2014 in Karlsruhe* (DRV-Schriften, Bd. 103, S. 319-321). Berlin: Deutsche Rentenversicherung Bund.

Kuhl, J. & Kazén, M. (2009). *Persönlichkeits-Stil- und Störungs-Inventar* (2. Auflage). Göttingen: Hogrefe.

Lachar, D. & Wrobel, T. A. (1979). Validating clinicians' hunches: Construction of a new MMPI critical item set. *Journal of Consulting and Clinical Psychology, 47*, 277-284.

Lanyon, R. I. & Almer, E. R. (2002). Characteristics of compensable disability patients who choose to litigate. *Journal of the American Academy of Psychiatry and the Law Online, 30*, 400-404.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17*, 410-425.

Larrabee, G. J., Greiffenstein, M. F., Greve, K. W. & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Hrsg.), *Assessment of malingered neuropsychological deficits* (S. 335-371). Oxford: Oxford University Press.

Lees-Haley, P. R., English, L. T. & Glenn, W. J. (1991). A fake bad scale on the MMPI-2 for personal injury claimants. *Psychological Reports, 68*, 203-210.

Lilienfeld, S.O., Thames A. D. & Watts, A.L. (2013). Symptom validity testing: Unresolved questions, future directions. *Journal of Experimental Psychopathology, 4*, 78-87.

Locke, D. E. C., Smigielski, J. S., Powell, M. R. & Stevens, S. R. (2008). Effort issues in post-acute outpatient acquired brain injury rehabilitation seekers. *NeuroRehabilitation, 23*, 273-281.

Loewen, P. J., Dawes, C. T., Mazar, N., Johannesson, M., Koellinger, P. & Magnusson, P. K. E. (2013). The heritability of moral standards for everyday dishonesty. *Journal of Economic Behavior & Organization, 93*, 363-366.

Mazar, N. & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing, 25*, 117-126.

Mehrhoff, F., Meeßen, A., Cibis, W., S., D., Diedrich, U., Rombach, W. … Link, M. (2012). Rechtsgrundlagen der Auftraggeber von ärztlichen Gutachten. In J. Fritze, F. Mehrhoff & E. Fritze (Hrsg.), *Die ärztliche Beguatchtung*. Berlin: Springer.

Merten, T. (2005). Der Stellenwert der Symptomvalidierung in der neuropsychologischen Begutachtung. *Zeitschrift für Neuropsychologie, 16*, 29-45.

Merten, T. (2011). Beschwerdenvalidierung bei der Begutachtung kognitiver und psychischer Störungen. *Fortschritte der Neurologie, Psychiatrie, 79*, 102-116.

Merten, T. (2014). Coaching und Coaching-Forschung. In A. Thöne-Otto, H. Flor, S. Gauggel, S. Lautenbacher & H. Niemann (Hrsg.), *Beschwerdenvalidierung* (S. 61-63). Göttingen: Hogrefe.

Merten, T., Diederich, C. & Stevens, A. (2008). Vorgetäuschte Beschwerden nach Distorsionstrauma der Halswirbelsäule: eine experimentelle Simulationsstudie. *Aktuelle Neurolologie, 35*, 8-15.

Merten, T., Friedel, E., Mehren, G. & Stevens, A. (2007). Über die Validität von Persönlichkeitsprofilen in der nervenärztlichen Begutachtung. *Der Nervenarzt, 78*, 511-520.

Merten, T., Friedel, E. & Stevens, A. (2006). Eingeschränkte Kooperativität in der neurologisch-psychiatrischen Begutachtung: Schätzungen zu Auftretenswahrscheinlichkeiten an einer Begutachtungspopulation. *Versicherungsmedizin, 58*, 19-21.

Merten, T., Green, P., Henry, M., Blaskewitz, N. & Brockhaus, R. (2005). Analog validation of German-language symptom validity tests and the influence of coaching. *Archives of Clinical Neuropsychology, 20*, 719-726.

Merten, T., Henry, M. & Hilsabeck, R. (2004). Symptomvalidierungstests in der neuropsychologischen Diagnostik: eine Analogstudie. *Zeitschrift für Neuropsychologie, 15*, 81-90.

Merten, T., Stevens, A. & Blaskewitz, N. (2007). Beschwerdenvalidität und Begutachtung: eine Einführung. *Praxis der Rechtspsychologie, 17*, 7-28.

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. E. & Volbrecht, M. (1998). Validation of reliable digits for detection of malingering. *Assessment, 5*, 303-307.

Mittenberg, W., Patton, C., Canyock, E.-M. & Condit, D.-C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*, 1094-1102.

Morel, K. R. (1998). Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. *Journal of Personality Assessment, 70*, 299-314.

Musil, C. M., Warner, C. B., Yobas, P. K. & Jones, S. L. (2002). A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research, 24*, 815-829.

Mykletun, A., Overland, S., Dahl, A. A., Krokstad, S., Bjerkeset, O., Glozier, N. … Prince, M. (2006). A population-based cohort study of the effect of common mental disorders on disability pension awards. *American Journal of Psychiatry, 163*, 1412-1418.

Nelson, N.W. & Sweet, J.J. (2010). Malingering of psychiatric presentation: Divergence of cognitive effort measures and psychological test validity indicators. In J. Morgan & J. Sweet (Hrsg.), *Neuropsychology of malingering casebook*. (S.195-214) New York: Taylor & Francis.

Nelson, N.W., Sweet, J.J., Berry, D.T.R, Bryant, F.B. & Granacher, R.P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society, 13*, 440-449.

Nietzsche, F. (1998). *Friedrich Nietzsche: Werke, 1: Menschliches, Allzumenschliches und andere Schriften*. Köln: Könemann.

Noeker, M. & Petermann, F. (2011). Simulation neurologischer versus psychischer Beschwerden - Notwendigkeit unterschiedlicher Validierungsstrategien. *Psychotherapeut, 56*, 449-454.

Petermann, F. (2012). *Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV, deutsche Version)*. Frankfurt: Pearson Assessment.

Poythress, N. G., Edens, J. F. & Watkins, M. M. (2001). The relationship between psychopathic personality features and malingering symptoms of major mental illness. *Law and Human Behavior, 25*, 567-582.

Rabung, S., Harfst, T., Kawski, S., Koch, U., Wittchen, H.-U. & Schulz, H. (2009). Psychometrische Überprüfung einer verkürzten Version der „Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis“ (HEALTH-49). *Zeitschrift für Psychosomatische Medizin und Psychotherapie, 55*, 162-179.

Reinhardt, H. (2013). *Sozialgesetzbuch VI: Gesetzliche Rentenversicherung*. Baden-Baden: Nomos.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie, 28*, 286-340.

Rey, A. (1958). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Rogers, R. (2008). *Clinical assessment of malingering and deception* (3rd edition). New York: Guilford.

Rogers, R., Harrell, E. H. & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical consideration. *Clinical Psychology Review, 13*, 255-275.

Rogers, R., Robinson, E. V. & Gillard, N. D. (2014). The SIMS screen for feigned mental disorders: The development of detection-based scales. *Behavioral Sciences & the Law, 32*, 455-466.

Rogers, R., Sewell, K. W. & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms, 2nd edition (SIRS)*. Lutz, FL: Psychological Assessment Resources.

Rogers, R., Sewell, K. W., Martin, M. A. & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*, 160-177.

Rosen, G. M. (2006). DSM's cautionary guideline to rule out malingering can protect the PTSD data base. *Journal of Anxiety Disorders, 20*, 530-535.

Rosenhan, D. L. (1973). On being sane in insane places. *Science, 179*, 250-258.

Schenk, K. & Sullivan, K. A. (2010). Do warnings deter rather than produce more sophisticated malingering? *Journal of Clinical and Experimental Neuropsychology, 32*, 752-762.

Schickel, S. & Henningsen, P. (2010). Begutachtung von somatoformen Störungen. *Psychotherapeut, 55*, 409-414.

Schmand, B. & Lindeboom, J. (2005). *Amsterdam Short-Term Memory Test. Amsterdamer Kurzzeitgedächtnis.* Leiden, NL: PITS.

Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T. & Hamburger, H. L. (1998). Cognitive complaints in patients after whiplash injury: the impact of malingering. *Journal of Neurology, Neurosurgery & Psychiatry, 64*, 339-343.

Schmidt, C., Bernert, S. & Spyra, K. (2014). Zur Relevanz psychischer Komorbiditäten bei chronischem Rückenschmerz: Häufigkeitsbild, Erwerbsminderungsrenten und Reha-Nachsorge im Zeitverlauf der Reha-Kohorten 2002–2009. *Die Rehabilitation, 53*, 384-389.

Schmidt, T., Lanquillon, S. & Ullmann, U. (2011). Kontroverse zu Beschwerdenvalidierungsverfahren bei der Begutachtung psychischer Störungen. *Forensische Psychiatrie, Psychologie, Kriminologie, 5*, 177–183.

Schneider, W. (2007). Standards der sozialmedizinischen Leistungsbegutachtung in der Psychosomatischen Medizin und Psychotherapie. *Psychotherapeut, 52*, 447-462.

Schneider, W., Becker, D., Dohrenbusch, R., Freyberger, H. J., Gündel, H. & Henningsen, P. (2010). Berufliche Leistungsfähigkeit. *Psychotherapeut, 55*, 373-379.

Schneider, W., Becker, D., Vogel, M., Freyberger, H. J., Gündel, H., Henningsen, P. … Dohrenbusch, R. (2011). Begutachtung der Leistungsfähigkeit bei Personen mit psychischen und psychosomatischen Störungen - ein diagnostisches Modell. *Versicherungsmedizin, 63*, 68-75.

Shalvi, S., Dana, J., Handgraaf, M. J. J. & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes, 115*, 181-190.

Sharland, M. J. & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*, 213-223.

Slick, D. J., Sherman, E. M. S. & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*, 545-561.

Smith, G. P. & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law, 25*, 183-189.

Sperling, W., Biermann, T., Löwe, A., Kornhuber, J. & Reulbach, U. (2010). Frühberentung aufgrund psychiatrischer Erkrankungen. *Fortschritte der Neurologie, Psychiatrie, 78*, 213-218.

Statistisches Bundesamt (2013). *Migration in Deutschland 2012*. Wiesbaden: Statistisches Bundesamt.

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 Research, 157*, 191-200.

Strauss, E., Hultsch, D. F., Hunter, M., Slick, D. J., Patry, B. & Levy-Bencheton, J. (1999). Using intraindividual variability to detect malingering in cognitive performance. *The Clinical Neuropsychologist, 13*, 420-432.

Suhr, J., Gunstad, J., Greub, B. & Barrash, J. (2004). Exaggeration index for an expanded version of the Auditory Verbal Learning Test: Robustness to coaching. *Journal of Clinical and Experimental Neuropsychology, 26*, 416-427.

Suhr, J. A. & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology, 15*, 415-424.

Sullivan, B. K., May, K. & Galbally, L. (2007). Symptom exaggeration by college adults in Attention-Deficit Hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology, 14*, 189-207.

Sullivan, K., Keane, B. & Deffenti, C. (2001). Malingering on the RAVLT: Part I. Deterrence strategies. *Archives of Clinical Neuropsychology, 16*, 627-641.

Sullivan, K. & Richer, C. (2002). Malingering on subjective complaint tasks: An exploration of the deterrent effects of warning. *Archives of Clinical Neuropsychology, 17*, 691-708.

Sullivan, K. A. & Elliott, C. (2012). An investigation of the validity of the MMPI-2 response bias scale using an analog simulation design. *Clinical Neuropsychologist, 26*, 160-176.

Sweet, J. J., Malina, A. & Ecklund-Johnson, E. (2006). Application of the new MMPI-2 malingered depression scale to individuals undergoing neuropsychological evaluation: Relative lack of relationship to secondary gain and failure on validity indices. *The Clinical Neuropsychologist, 20*, 541-551.

Tombaugh, T. N. (1996). *Test of Memory Malingering*. Toronto: Multi-Health Systems.

Tydecks, S., Merten, T. & Gubbay, J. (2006). The Word Memory Test and the One-in-Five-Test in an analogue study with russian speaking participants. *International Journal of Forensic Psychology, 1*, 29-37.

van Impelen, A., Merckelbach, H., Jelicic, M. & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A Systematic Review and Meta-Analysis. *The Clinical Neuropsychologist, 28*, 1336-1365.

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*, 73-85.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester: Wiley.

Walter, F., Geissler, N., Petermann, F. & Kobelt, A. (2015). Beeinflussen Instruktionen und die Motivation für eine Rente die Beschwerdedarstellung in Fragebogen? *Die Rehabilitation*, in druck.

Walter, F., Petermann, F., Dietrich, D. E. & Kobelt, A. (2013). Wie können Beschwerden im Rahmen der medizinischen Rehabilitation validiert werden? *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin, 23*, 334-340.

Walter, F., Petermann, F. & Kobelt, A. (2012). Beschwerdenvalidierung: Ein aktueller Überblick. *Die Rehabilitation, 51*, 342-348.

Walter, F., Petermann, F. & Kobelt, A. (2015). Erfassung von negativen Antwortverzerrungen – Entwicklung und Validierung des BEVA. eingereicht.

Walter, F., Petermann, F., Kobelt, A. (2015). Entwicklung eines Screeningverfahrens für die Beschwerdenvalidierung von Erkrankungen mit depressiver Symptomatik. [Abstract] In Deutsche Rentenversicherung Bund (Hrsg.), *24. Rehabilitationswissenschaftliches Kolloquium – Psychische Störungen – Herausforderungen für Prävention und Rehabilitation vom 16. bis 18. März 2015 in Augsburg.* Berlin: Deutsche Rentenversicherung Bund. In Druck.

Weltgesundheitsorganisation (WHO) (2005). *Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F). Klinische diagnostische Leitlinien.* Bern: Huber.

Widder, B. (2011). Beurteilung der Beschwerdenvalidität. In B. Widder & P. Gaidzik (Hrsg.), *Begutachtung in der Neurologie* (2. vollständig überarbeitete Auflage, S. 64-95). Stuttgart: Thieme.

Widder, B., Dertwinkel, R., Egle, U. T., Foerster, K. & Schiltenwolf, M. (2007). Leitlinie für die Begutachtung von Schmerzen. *Psychotherapeut, 52*, 334-346.

Wiener, D. N. (1948). Subtle and obvious keys for the Minnesota Multiphasic Personality Inventory. *Journal of Consulting Psychology, 12*, 164-170.

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

Wong, J. L., Lerner-Poppen, L. & Durham, J. (1998). Does warning reduce obvious malingering on memory and motor tasks in college samples? *International Journal of Rehabilitation and Health, 4*, 153-165.

Youngjohn, J. R., Less-Haley, P. R. & Binder, L. M. (1999). Comment: Warning malingerers produces more sophisticated malingering. *Archives of Clinical Neuropsychology, 14*, 511-515.

Studie I

Walter, F., Petermann, F. & Kobelt, A. (2012). Beschwerdenvalidierung: Ein aktueller Überblick. *Die Rehabilitation, 51*, 342-348.

Studie II

Walter, F., Petermann, F., Dietrich, D. E. & Kobelt, A. (2013). Wie können Beschwerden im Rahmen der medizinischen Rehabilitation validiert werden? *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin, 23*, 334-340.

Studie III

Walter, F., Geissler, N., Petermann, F. & Kobelt, A. (2015). Beeinflussen Instruktionen und die Motivation für eine Rente die Beschwerdedarstellung in Fragebogen? *Die Rehabilitation*. In Druck.

Studie IV

Walter, F., Petermann, F. & Kobelt, A. (2015). Erfassung von negativen Antwortverzerrungen – Entwicklung und Validierung des BEVA. Eingereicht.

Anhang E

Fragebogen der Expertenbefragung

Anhang F

Materialien der Studien II und IV:

Soziodemografischer Fragebogen

Szenario (Depression)

Fragebogen zur Rollenkonformität (Depression)

Fragebogen zur Rollenerfüllung (Depression)

Szenario (Schmerz)

Fragebogen zur Rollenkonformität (Schmerz)

Fragebogen zur Rollenerfüllung (Schmerz)

Therapeutenfragebogen

Anhang G

Materialien Studie III

Soziodemografischer Fragebogen

Information (inklusive Warnung)

Anhang H

Eigenleistung der Doktorandin

Darstellung des Eigenanteils der Veröffentlichungen

Die vorliegende kumulative Dissertation basiert auf vier Veröffentlichungen, die von nationalen Fachzeitschriften bereits abgedruckt, zum Druck angenommen oder eingereicht wurden. Die Manuskripte wurden jeweils durch die Doktorandin in Erstautorinnenschaft erstellt.

Nachfolgend soll der Arbeitsanteil der promotionsrelevanten Publikationen dargelegt werden. Der Hauptanteil, dies betrifft die Konzeption, Literaturrecherche, die Durchführung der Studien, die Auswertung und Interpretation der Daten, das Erstellen der Manuskripte sowie deren Revisionen, lag bei der Promovendin. In der Tabelle 16 sind die genauen Arbeitsanteile für jede einzelne Publikation genannt. Es werden die einzelnen Arbeitsschritte dargelegt und die jeweiligen Anteile (vollständig, überwiegend, teilweise) zugeordnet. *Vollständig* bedeutet in diesem Zusammenhang, dass der Schritt von der Promovendin allein durchgeführt wurde, die Promovendin jedoch immer im regelmäßigen Austausch zu allen Mitautoren stand.

Tabelle 16: Darlegung der Arbeitsanteile der Promovendin an allen promotionsrelevanten Publikationen

	Publikation I	**Publikation II**	**Publikation III**	**Publikation IV**
Konzeption	vollständig	vollständig	teilweise	vollständig
Literaturrecherche	vollständig	vollständig	überwiegend	vollständig
Durchführung und Datenerhebung	nicht relevant	überwiegend	teilweise	teilweise
Auswertung der Daten	nicht relevant	vollständig	überwiegend	vollständig
Interpretation der Ergebnisse	nicht relevant	überwiegend	überwiegend	überwiegend
Erstellen des Manuskripts	überwiegend	überwiegend	vollständig	vollständig
Revision	vollständig	vollständig	vollständig	steht noch aus

Anmerkung: **Publikation I**: Walter, F., Petermann, F. & Kobelt, A. (2012). Beschwerdenvalidierung: Ein aktueller Überblick. *Die Rehabilitation, 51*, 342-348.; **Publikation II**: Walter, F., Petermann, F., Dietrich, D. E. & Kobelt, A. (2013). Wie können Beschwerden im Rahmen der medizinischen Rehabilitation validiert werden? *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin, 23*, 334-340.; **Publikation III**: Walter, F., Geissler, N., Petermann, F. & Kobelt, A. (2015). Beeinflussen Instruktionen und die Motivation für eine Rente die Beschwerdedarstellung in Fragebogen? *Die Rehabilitation*. In Druck. **Publikation IV**: Walter, F., Petermann, F. & Kobelt, A. (2015) Erfassung von negativen Antwortverzerrungen – Entwicklung und Validierung des BEVA. (eingereicht).

Die Autorin der vorliegenden Dissertation, Dipl. Psych. Franziska Walter, und die Mitautoren der promotionsrelevanten Publikationen, Prof, Dr. Franz Petermann, PD Dr. Axel Kobelt, Prof. Dr. Detlef E. Dietrich sowie Dr. Norbert Geissler, stimmen anhand ihrer Unterschrift der Darstellung der Eigenleistungen der Promovendin zu. Zudem geben die Mitautoren ihr Einverständnis, die vorliegenden Publikationen für das Promotionsverfahren von Dipl. Psych. Franziska Walter zu verwenden.

Bremen, im Januar 2015

Prof. Dr. Franz Petermann

PD Dr. Axel Kobelt

Prof. Dr. Detlef E. Dietrich

Dr. Norbert Geissler

Dipl. Psych. Franziska Walter

Anhang I

Eidesstattliche Erklärung

Eidesstattliche Erklärung

Hiermit versichere ich, Dipl. Psych. Franziska Walter, dass ich die vorliegende kumulative Dissertation selbstständig verfasst und außer den aufgeführten keine anderen Quellen und Hilfsmittel verwendet habe. Das aus fremden Quellen wörtlich oder inhaltlich übernommene Gedankengut wurde als solches kenntlich gemacht. Die vorliegende Arbeit wurde bisher weder national noch international in gleicher oder ähnlicher Form als Prüfungsleistung vorgelegt.

_____ _____

Ort, Datum Unterschrift