

**Die Kurzsкала BESS**  
**zur direkten Veränderungsmessung**  
**des gesundheitlichen Befindens**  
**Psychometrische Überprüfungen für den Bereich der**  
**Medizinischen Rehabilitation**

Schmidt, J.<sup>1</sup>, Nübling, R.<sup>1</sup>, Kriz, D.<sup>1</sup>, Kaiser, U.<sup>1</sup> & Wittmann, W.W.<sup>2</sup>

<sup>1</sup> Gesellschaft für Qualität im Gesundheitswesen GfQG, Karlsruhe

<sup>2</sup> Emeritus Lehrstuhl Diagnostik und Evaluation Universität Mannheim

GfQG-Schriften - Nr. 1

Karlsruhe, 24.01.2018

## Kontakt

### **Dr. Jürgen Schmidt**

Diplom-Psychologe  
Gesellschaft für Qualität im Gesundheitswesen GfQG  
Wendtstr. 1, 76185 Karlsruhe  
Tel: +49 (721) 89337040  
eMail: schmidt@gfqq.de

### **Dr. David Kriz**

Diplom-Psychologe  
Gesellschaft für Qualität im Gesundheitswesen GfQG  
Wendtstr. 1, 76185 Karlsruhe  
Tel: +49 (721) 89337040  
eMail: kriz@gfqq.de

### **Prof. Dr. Werner W. Wittmann**

Emeritus Diagnostik und Evaluation Universität Mannheim  
c/o Gesellschaft für Qualität im Gesundheitswesen GfQG  
Wendtstr. 1, 76185 Karlsruhe  
Tel. +49 (721) 89337040

### **Dr. Rüdiger Nübling**

Diplom-Psychologe  
Gesellschaft für Qualität im Gesundheitswesen GfQG  
Wendtstr. 1, 76185 Karlsruhe  
Tel: +49 (721) 89337040  
eMail: nuebling@gfqq.de

### **Dr. Udo Kaiser**

Diplom-Psychologe, Diplom-Betriebswirt  
Gesellschaft für Qualität im Gesundheitswesen GfQG  
Wendtstr. 1, 76185 Karlsruhe  
Tel. +49 (721) 89337040  
eMail: kaiser@gfqq.de

## Inhalt

Summary.....	4
1 Hintergrund.....	6
1.1 Einleitung.....	6
1.2 Die Skala BESS.....	7
1.3 Varianten der Veränderungsmessung.....	8
2 Fragestellungen.....	11
3 Methodik.....	11
3.1 Datengrundlage.....	11
3.2 Stichproben.....	13
3.3 Assessmentinstrumente.....	14
3.4 Analyseverfahren.....	15
4 Ergebnisse.....	15
4.1 Item- und Skalenanalysen.....	15
4.2 Faktorielle Struktur der BESS-Items - Explorative Faktorenanalysen.....	17
4.3 Zeitliche Stabilität / Änderungssensitivität.....	17
4.4 Facetten der Validität (Gültigkeit).....	19
4.4.1 Zusammenhänge mit zeitgleich erhobenen Outcome-bezogenen Merkmalen.....	19
4.4.2 Zusammenhänge mit Veränderungsinformationen aus der direkten Veränderungsmessung (DVM).....	20
4.4.3 Zusammenhänge mit Veränderungsinformationen aus der indirekten Veränderungsmessung (IVM).....	22
4.4.4 Zusammenhänge mit zeitgleichen Status-Messungen.....	23
4.4.5 Prädiktive Validität.....	25
4.4.5.1 Vorhersagbarkeit der BESS-Skalenwerte aus Ausgangswerten.....	25
4.4.5.2 Vorhersagevalidität der BESS-Skalenwerte.....	25
4.4.6 Diskriminante Validität - Gruppenunterschiede.....	27
4.5 Referenzwerte.....	30
4.6 Gebesserte und nichtgebesserte Patienten (kritische Werte).....	31
5 Zusammenfassung und Bewertung.....	34
6 Literatur.....	37
7 Anhang – Glossar Fragebögen.....	40

## Summary

Vorgestellt wird eine einfache Ergebnisskala, die auf dem Prinzip der direkten Veränderungsmessung basiert. Die Skala BESS umfasst 5 Items zur Veränderung des gesundheitlichen Befindens und kann Werte zwischen 5 und 25 annehmen. BESS wurde seit 1990 im Rehabilitationsbereich in vielen Datenerhebungen eingesetzt. Je höher der Skalenwert ausfällt, desto größer ist im Erleben des Patienten die Verbesserung seines gesundheitlichen Befindens bzw. seines gesundheitlichen Zustands.

**Methode:** Auf der Basis einer Reanalyse von insgesamt 12 Studien bzw. Datenerhebungen mit etwa 160.000 Reha-Patienten werden die Ergebnisse der umfangreichen psychometrischen Überprüfung der Skala vorgestellt. Diese umfasst Item- und Skalenanalysen, Faktorenanalysen und korrelationsstatistische Analysen.

**Ergebnisse:** Die dargestellten Befunde beziehen sich auf die interne Konsistenz der Skala, auf die faktorielle Struktur und auf verschiedene Aspekte der Kriteriumsvalidität. Weiterhin werden Gruppenunterschiede analysiert, Referenzwerte und kritische Werte zur Ermittlung von Besserungsquoten vorgestellt.

**Faktorielle Struktur:** Alle durchgeführten Faktorenanalysen sprechen für die Eindimensionalität der Skala. Hauptkomponentenanalysen zeigen übereinstimmend einen salienten Hauptfaktor, der etwa 65 – 80 % der Gesamtvarianz erklären kann.

**Interne Konsistenz / Homogenität:** Die ermittelten Cronbach-  $\alpha$  - Koeffizienten liegen im Bereich zwischen 0,86 und 0,94. Die durchschnittlichen Inter-Item-Korrelationen liegen zwischen  $r = 0,56$  und  $r = 0,76$ .

**Stabilität / Änderungssensitivität:** Zwischen Messungen bei Entlassung und in Nacherhebungen zeigen moderate Korrelationen im Bereich zwischen  $r = 0,32$  und  $r = 0,45$ . Im Regelfall liegen signifikante Mittelwertsunterschiede mit Effektstärken bis zu  $ES = 1,09$  vor.

**Konkurrente Validität:** BESS korreliert bei Entlassung im Bereich  $r = 0,34$  bis  $0,58$  mit ZUF-8 (Patientenzufriedenheit; Schmidt et al., 1989). Weiterhin zeigen sich deutliche Zusammenhänge mit der Einschätzung der Zielerreichung ( $r = 0,52 - 0,66$ ), mit globalen Nutzen- und Ergebnisbeurteilungen ( $r = 0,39$  bis  $0,57$ ) und konsistent mit gleichzeitigen Statusmessungen der psychischen und körperlichen Befindlichkeit (Einzelitems und psychometrische Skalen, z. B. B-L, D-S, SCL-90-R;  $r = \pm 0,24$  bis  $\pm 0,79$ , Vorzeichen polungsbedingt).

**Zusammenhänge mit anderen direkten Veränderungsmessungen:** BESS - Werte von Patienten korrelieren substantiell mit singulären patientenseitigen Veränderungseinstufungen des Gesundheitszustands ( $r = 0,76$  bis  $0,88$ ). Ebenfalls zeigen sich hohe Korrelationen mit zeitgleich erfassten multiplen Ergebniskriterien bzw. Veränderungsskalen ( $r = 0,68 - r = 0,96$ ).

Die Zusammenhänge zwischen direkten Selbst- und Fremdbeurteilungen der Veränderung sind in der Regel eher moderat ( $r = 0,27$  bis  $0,50$ ).

Zusammenhänge mit indirekten Veränderungsmessungen: Auch hier zeigen sich plausible Korrelationen mit korrespondierenden Skalen-Differenzwerten (z. B. bei den Skalen GB-10, MFHW, „Psychischen Wohlbefinden“ aus dem HEALTH-49 oder B-L) in der Größenordnung von etwa  $r = \pm 0,20$  bis  $\pm 0,67$  (Vorzeichen polungsbedingt).

Vorhersagbarkeit aus Messungen bei Aufnahme: Konsistente Korrelationen  $\geq r = [0,40]$  zeigen sich mit den Skalen „Beziehungszufriedenheit“ (HAQ;  $r = 0,51$ ) und „Frühberentungsrisiko“ (SIBAR; bis  $r = -0,45$ ), sowie mit den Single-Item-Ratings „Arbeitsfähigkeit“ ( $r = -0,42$ ) und „Gesundheitszustand“ ( $r = 0,40$ ). Weitere Zusammenhänge sind moderat bis geringfügig.

Prognostische Validität von BESS bei Entlassung: Es zeigen sich moderate Korrelationen mit katamnestischen Zielerreichungseinschätzungen, Nutzenbewertungen und Patientenzufriedenheit in der Größenordnung  $r = 0,35$  bis  $r = 0,56$ .

Vergleichswerte: Für klinische Stichproben mit fast 140000 stationären Rehabilitanden liegen Vergleichswerte (M, SD) für die vier Indikationen Psychosomatik, Orthopädie, Kardiologie und Onkologie und Rehabilitation gesamt vor.

Unterschiede zwischen Indikationen: In zwei großen Patientenbefragungen mit  $N = 50342$  und  $N = 81999$  Rehabilitanden zeigen sich zum Zeitpunkt der Entlassung signifikante Unterschiede zwischen den vier Indikationen. Konsistent weisen Kardiologie-Patienten die höchsten durchschnittlichen BESS-Skalenwerte und Psychosomatik-Patienten die niedrigsten BESS-Mittelwerte auf. Je nach angelegtem Signifikanzniveau liegen die bei Entlassung beobachtbaren Besserungsquoten der Rehabilitanden zwischen 71,9 % (0,1 %-Niveau) und 80,0 % (5 %-Niveau).

### **Gesamtfazit**

Die Re-Analyse der insgesamt 12 Studien bzw. Datenerhebungen belegt aus Sicht der Autoren die zufriedenstellende interne Konsistenz, Änderungssensitivität und Validität der Skala BESS. Die eindimensionale (direkte) Veränderungsskala ist wegen ihrer Ökonomie und aufgrund ihrer umfangreichen Überprüfung insbesondere sehr gut als globales Ergebnismaß für Routine-Patientenbefragungen im Rahmen der Qualitätssicherung geeignet.

# 1 Hintergrund

## 1.1 Einleitung

Für eine praktikable Ergebnismessung im Rahmen der Routine-Qualitätssicherung von Reha-Maßnahmen sind möglichst kurze und ökonomische Erhebungsinstrumente gefragt, diese sollten allerdings empirisch überprüft, reliabel und valide sein. Vor- und Nachteile von Verfahren der indirekten, quasi-indirekten und direkten Veränderungsmessung und deren Zusammenhänge wurden vielfach beschrieben (Stieglitz & Baumann, 1994; Nübling & Schmidt, 2000; Schmidt et al., 2003; Bleßmann, 2004; Nübling et al., 2004; Krampen, 2010a, 2010b; Pohontsch, 2012; Mittag et al., 2013). Ein Beispiel für eine Kurzsкала auf Grundlage der direkten Veränderungsmessung ist die Skala BESS, bei der 5 gesundheitsbezogene Veränderungseinstufungen zu einem Veränderungsindex aggregiert werden. Die Kurzsкала kam erstmals 1986 in der „Zauberberg-II-Studie“ in der Psychosomatischen Klinik Schömberg zum Einsatz (Schmidt et al. 1992, 1994a, 1994b; Nübling, 1992), in der Folgezeit wurde sie in weiteren Evaluationsstudien mit Mehr-Punkt-Erhebungen und im Rahmen von umfangreichen Patientenbefragungen in Rehabilitationskliniken verwendet.

Bei der Ergebnisevaluation können im Wesentlichen drei Strategien oder Modelle der Veränderungsmessung unterschieden werden, die man als indirekte Veränderungsmessung (IVM), quasi-indirekte Veränderungsmessung (QIVM) und als direkte Veränderungsmessung (DVM) kann. Die indirekte Veränderungsmessung kann als Standardstrategie in der Ergebnisevaluation betrachtet werden, die beiden anderen Strategien stellen weitere Möglichkeiten dar, deren Relevanz durchaus kontrovers diskutiert wird. Mit dem Vergleich dieser unterschiedlichen Messstrategien haben sich in den letzten 20 Jahren – mit unterschiedlichen Akzenten eine ganze Reihe von Autoren beschäftigt (z. B. Kastner und Basler 1997, Kohlmann und Raspe 1998, Bleßmann 2004, Schmidt et al. 2003). Während bei Mehr-Punkt-Erhebungen grundsätzlich alle drei Varianten der Veränderungsmessung anwendbar sind, scheidet bei sog. Ein-Punkt-Erhebungen die klassische indirekte Veränderungsmessung aus. Ein-Punkt-Erhebungen sind aber insbesondere im Rahmen der Routine-Qualitätssicherung die Methode der Wahl, z. B. in der Rehabilitandenbefragung der DRV (etwa 3 Monate nach Ende der Reha) oder auch bei systematische Patientenbefragungen am Ende der Reha, wie sie in vielen Kliniken durchgeführt werden. Wer also überlegt, Outcomes im Sinne gesundheitlicher Veränderungen mittels Ein-Punkt-Erhebungen zu erfassen, kann sich für die direkte oder die quasi-indirekte Methode oder für die Kombination der beiden Varianten entscheiden.

## 1.2 Die Skala BESS

Die Skala BESS basiert auf dem Prinzip der direkten Veränderungsmessung (siehe Abschnitt 1.3). BESS (Schmidt u. Nübling, 1986; Schmidt et al., 1994; Nübling u. Schmidt, 2000) ist eine direkte Veränderungsskala, die wegen ihrer Ökonomie mittlerweile in einigen Programmevaluationsstudien und anderen Datenerhebungen im Reha-Bereich zur globalen Ergebnismessung aus Patientensicht (mit-)eingesetzt wurde. Die Patienten werden in der Instruktion gebeten, den jetzigen Zustand mit dem Zustand vor der Reha (Behandlung) kritisch zu vergleichen und einzuschätzen, inwieweit sich 5 Aspekte des Gesundheitszustands (körperliches Befinden, seelisches Befinden, Allgemeinbefinden, Leistungsfähigkeit, Beschwerden / Beeinträchtigungen) in diesem Zeitraum verändert haben. Wie Tabelle 1 zeigt kann die wahrgenommene Veränderung bzw. Nicht-Veränderung der 5 Aspekte jeweils anhand einer 5-stufigen Skala (von 1 = deutlich besser bis 5 = deutlich schlechter) eingestuft werden.

**Tabelle 1: Items der Kurzskala BESS**

Wenn Sie Ihren <u>jetzigen</u> Zustand mit Ihrem Zustand <u>vor der Reha</u> vergleichen: Welche Veränderungen haben sich ergeben?	deutlich besser ++	etwas besser +	unverändert 0	etwas schlechter -	deutlich schlechter --
1. Mein körperliches Befinden (körperliche Verfassung) ist...	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>	<input type="checkbox"/> <sub>4</sub>	<input type="checkbox"/> <sub>5</sub>
2. Mein seelisches Befinden (psychische Verfassung) ist...	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>	<input type="checkbox"/> <sub>4</sub>	<input type="checkbox"/> <sub>5</sub>
3. Mein Allgemeinbefinden ist...	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>	<input type="checkbox"/> <sub>4</sub>	<input type="checkbox"/> <sub>5</sub>
4. Meine Leistungsfähigkeit ist...	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>	<input type="checkbox"/> <sub>4</sub>	<input type="checkbox"/> <sub>5</sub>
5. Meine Beschwerden / Beeinträchtigungen sind...	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>	<input type="checkbox"/> <sub>4</sub>	<input type="checkbox"/> <sub>5</sub>

Der BESS-Gesamtwert ergibt sich nach Umpolung durch Summation der Itemwerte. Der mögliche Skalenrange reicht von 5 – 25 Skalenpunkten, wobei der Wert 25 den maximal möglichen Gesundungsfortschritt (= 5 x 5) und der Wert 5 eine maximal mögliche Verschlechterung (= 5 x 1) anzeigt. 15 Skalenpunkte indizieren eine durchschnittliche Null-Veränderung (= 5 x 3). Die Standardinstruktion kann unverändert bei verschiedenen Messzeitpunkten verwendet werden, wobei sich die Zeitabschnitte für den intendierten Vergleich entsprechend verändern. Insofern sind auch Mehrfachmessungen mit der Skala BESS möglich. Wird die Skala z. B. am Ende der Behandlung vorgegeben, dann ist mit „jetzigem“ Zustand der Abschluss der Behandlung gemeint, bei einer Vorgabe z. B. bei einer 12-Monats-Katamnese ist mit „jetzig“ der Zustand 1 Jahr nach der Behandlung angesprochen - entsprechend mit dem Messzeitpunkt variieren also die impliziten Vergleichszeiträume (Zustand vor der Behandlung vs. Zustand am Ende der Behandlung bzw. Zustand vor der Behandlung vs. Zustand 1 Jahr nach der Behandlung).

### 1.3 Varianten der Veränderungsmessung

Fragen zur Veränderungsmessung spielen in der Ergebnisevaluation eine zentrale Rolle. Hier lassen sich drei Zugangsweisen unterscheiden, die allesamt mit unterschiedlichen Vor- und Nachteilen verbunden sind (Stieglitz & Baumann, 1994; Kohlmann & Raspe, 1998; Nübling & Schmidt, 2000b; Schmidt et al., 2001; Schmidt et al., 2002; Wittmann & Schmidt, 2002; Schmidt et al., 2003; Nübling et al., 2004; Bleßmann, 2004; Krampen, 2010a, 2010b):

- die indirekte Veränderungserfassung (IVM)
- die quasi-indirekte Veränderungserfassung (QIVM) mit retrospektivem Prä-Test
- die direkte Veränderungserfassung (DVM)

Als „Königsweg“ der Veränderungsmessung wird häufig die sog. indirekte Veränderungsmessung (IVM) bezeichnet (vgl. z. B. Meffert et al, 2012). Die IVM stellt in Form eines Prä/Post-Designs den klassischen Ansatz der Erfassung von Veränderungsinformationen dar. Bei diesem Verfahren wird die Ausprägung eines interessierenden Merkmals jeweils vor und nach der Behandlung, also über wiederholte Statusmessungen erfasst (vgl. Abbildung 1). Die durch Subtraktion bestimmbaren Messwertdifferenzen stellen somit indirekt gewonnene Veränderungsmaße dar. Ein zentraler Vorteil dieser Methode besteht darin, dass Status-Informationen über zwei Messzeitpunkte, also auch über den Ausgangszustand vorliegen. Allerdings erfordert die IVM deshalb mindestens zwei Messzeitpunkte, was einen hohen organisatorischen Aufwand bei der Datenerhebung mit sich bringt. Dieser Aspekt ist insbesondere im Rahmen von Routinebefragungen zur Qualitätssicherung problematisch. Darüber hinaus bestehen methodische Probleme wie Regression zur Mitte, die Frage nach der Reliabilität von Differenzmaßen sowie mögliche Veränderungen im Bezugssystem des Patienten, welche die Vergleichbarkeit wiederholter Statusmessungen generell in Frage stellt. Dies impliziert die Gefahr der Interpretation von statistischen Artefakten bei der Berechnung von Korrelationsmaßen zwischen Differenzmaßen und anderen Ergebnisindikatoren.

**Abbildung 1: Messrational der indirekten Veränderungsmessung**

Indirekte Veränderungsmessung (iVM) Differenzbildung zwischen Prä und Post-Status IVM = Post – Prä					
Prä-Messung (vor der Behandlung) Mein Gesundheitszustand ist ...			Post-Messung (nach der Behandlung) Mein Gesundheitszustand ist ...		
1	2	3	1	2	3
gut	mittelmässig	schlecht	gut	mittelmässig	schlecht

Die sog. quasi-indirekte Veränderungsmessung (QIVM) ist eine Quasi-Variante der indirekten Veränderungsmessung, die nur einen Messzeitpunkt erfordert, bei dem neben dem Post-Status die notwendige Prä-Messung retrospektiv, also im Nachhinein aus der Erinnerung, erhoben wird (z.B. am Ende der Intervention oder in einer Nacherhebung). Auch hier werden die Veränderungsinformationen durch Differenzbildung zwischen beiden Statusangaben (also zwischen der Post-Messung und der retrospektiven Prä-Messung) gewonnen (vgl. Abbildung 2). Vorteile sind in der ökonomischen Anwendbarkeit dieser Methode zu sehen, da nur eine einmalige Messung nach Ende der Behandlung benötigt wird und dennoch Informationen über den Ausgangszustand des Patienten gewonnen werden. Allerdings tauchen hier die gleichen methodischen Probleme (z.B. Regression zur Mitte) bei der Verwendung von Differenzwerten auf wie bei der klassischen IVM. Darüber hinaus kann die Validität von retrospektiven Angaben kritisch gesehen werden, da die Erinnerung durch das aktuelle Befinden und andere Faktoren verzerrt sein kann.

### Abbildung 2: Messrational der quasi-indirekten Veränderungsmessung

<b>Quasi-Indirekte Veränderungsmessung (QIVM)</b> Differenzbildung zwischen erinnertem Prä-Status und aktuell erfaßtem Post-Status $QIVM = Post - Prä_{RETRO}$						
Retrospektive Prä-Messung (nach der Behandlung) Mein Gesundheitszustand war ...				Post-Messung (nach der Behandlung) Mein Gesundheitszustand ist ...		
1	2	3		1	2	3
gut	mittelmässig	schlecht		gut	mittelmässig	schlecht

In der Literatur wird berichtet, dass retrospektive Angaben ein pessimistischeres Bild vom Ausgangszustand vermitteln als tatsächliche Prä-Messungen (Stieglitz, 1990; Schmidt et al., 2002), was zu höheren Effektgrößen führt, die als „Überschätzung“ der Behandlungserfolge interpretiert werden können. Andererseits kann man aber auch davon ausgehen, dass ein Jahr nach der Behandlung eine vermehrte Problemeinsicht des Patienten vorhanden ist, was zu einer realistischeren Bewertung des tatsächlichen Ausmaßes der damaligen Beeinträchtigung führt. Generell ist bei der Anwendung des Verfahrens ein Bias durch Gedächtniseffekte zu erwarten (Beeinflussung der retrospektiven Messung durch das aktuelle Befinden). Unabhängig davon hat sich gezeigt, dass retrospektive und reale Einschätzungen von Befindlichkeiten gut korrelieren und Veränderungswerte, die auf der QIVM basieren, sogar höher mit Fremdeinschätzungen korrelieren als Veränderungswerte, die anhand der IVM gewonnen wurden (Stieglitz, 1990).

Auch die sog. direkte Veränderungsmessung (DVM) erfordert nur einen Messzeitpunkt nach der Intervention (vgl. Abbildung 3). Bei diesem Zugang wird die subjektiv erlebte Veränderung im Sinne einer Vergleichsaussage (besser - unverändert - schlechter) direkt eingestuft.

**Abbildung 3: Messrational der direkten Veränderungsmessung**

Direkte Veränderungsmessung (dVM)		
Post-Messung (nach der Behandlung)		
Mein Gesundheitszustand ist im Vergleich zu vorher ...		
+1	0	-1
Besser	unverändert	schlechter

Vorteile liegen wie bei der QIVM in der besonderen Ökonomie des Verfahrens, welches sich damit für Routinebefragungen und Untersuchungen eignet, bei denen ein Prä-Test zu aufwendig oder überhaupt nicht mehr zu realisieren ist, wie etwa bei Post-Hoc-Studien. Die DVM steht von ihrem Ansatz her dem Prozesscharakter des menschlichen Erlebens nahe und entspricht der kognitiven Eigenschaft, Urteile durch Vergleichsrelationen (besser-schlechter) zu bilden. Das Hauptproblem der DVM besteht darin, dass bei ihrer ausschließlichen Anwendung keine Information über den Ausgangszustand des Patienten mehr verfügbar ist. Auch die Berechnung von Effektgrößen ist bei der DVM mit besonderen Problemen verbunden. Darüber hinaus stellt sich die Frage, ob der Ansatz tatsächlich „Veränderung“ und nicht etwa einfach das momentane Befinden erfasst (Kastner & Basler, 1997). Kohlmann und Raspe (1998) fanden lediglich mäßige Übereinstimmungen zwischen direkten und indirekten Veränderungsmaßen und kommen zu dem Schluss, dass die DVM eine eigenständige Veränderungsdimension abbildet. Auch Krampen (2010b) stellt fest, dass IVM und DVM nicht äquivalent sind, da sie zum Teil zu unterschiedlichen Ergebnissen führen und daher als komplementäre Methoden kombiniert eingesetzt werden sollten, um wahre Veränderungen bei Patienten möglichst valide abschätzen zu können.

Mittag et al. (2013) konstatieren in ihren Empfehlungen zur Veränderungsmessung in der medizinischen Rehabilitation, dass die Vorteile und Nachteile der verschiedenen Methoden der Veränderungsmessung je nach Fragestellung und Untersuchungsdesign abgewogen werden müssen. Die Autoren unterstreichen, dass eine grundsätzliche Überlegenheit einer der beschriebenen Methoden zumindest derzeit nicht behauptet werden kann. In diesem Lichte kann die häufig sehr kritische Haltung gegenüber der DVM – so spricht z. B. Fydrich (2006) von einer „kleinen Lösung“ für den Fall, dass keine Vortestdaten erhoben wurden – durchaus hinterfragt werden.

Die in diesem Bericht vorgestellte BESS-Skala stellt ein Veränderungsmaß dar, das auf dem Prinzip der DVM basiert. Fünf direkte Veränderungseinstufungen werden hierzu zu einer Skala aggregiert. Zur Abschätzung der Konstruktvalidität werden die BESS-Skalenwerte in der nachfolgend vorgestellten Reanalyse u.a. mit anderen Veränderungsinformationen aus der DVM und der IVM korreliert. Bei gegebener Konstruktvalidität sollte BESS mit Veränderungsmaßen auf der Basis beider Varianten der Veränderungsmessung korrelieren, wobei theoretisch zu erwarten ist, dass die Korrelationen mit anderen DVM-Maßen höher ausfallen.

## 2 Fragestellungen

Auf der Grundlage von Daten aus verschiedener Studien bzw. Erhebungen werden die psychometrischen Eigenschaften der Skala BESS einer umfassenden Reanalyse unterzogen. Untersucht werden hierbei

- die Mittelwerte, Standardabweichungen und die Trennschärfen der fünf Items,
- die Skalenmittelwerte und die interne Konsistenz der Skala (Cronbach  $\alpha$ ),
- die Faktorenstruktur der Skala,
- die zeitliche Stabilität bzw. Änderungssensitivität der Skala,
- die korrelativen Zusammenhänge mit zeitgleich erhobenen Outcome-bezogenen Merkmalen (konkurrente Validität I),
- die korrelativen Zusammenhänge mit Veränderungsinformationen aus direkten Veränderungsmessungen (konvergente Validität I),
- die korrelativen Zusammenhänge mit Veränderungsinformationen aus indirekten Veränderungsmessungen (konvergente Validität II),
- die korrelativen Zusammenhänge mit zeitgleichen Status-Messungen,
- die Vorhersagbarkeit von BESS aus Ausgangswerten,
- die Vorhersagevalidität von BESS und
- die diskriminante Validität von BESS (Unterschiede zwischen verschiedenen Patientengruppen).

Darüber hinaus sollen Referenzwerte für die Indikationen Orthopädie, Kardiologie, Onkologie und Psychosomatik sowie für Reha-Patienten allgemein dargestellt und kritische Werte für die BESS-Skalenwerte ermittelt werden.

## 3 Methodik

### 3.1 Datengrundlage

Die nachfolgenden Befunde zur Skala BESS stammen aus insgesamt 12 Studien bzw. Datenerhebungen, in denen die Skala miterfasst wurde. Eine Übersicht über die verwendete Datengrundlage gibt Tabelle 2. Die Daten wurden im Zeitraum 1987 bis 2016 bei über

160.000 Patienten erhoben. Die Tabelle liefert darüber hinaus Informationen über die Anzahl der Messzeitpunkte (1 – 5), die Stichprobengrößen (N = 138 bis N = 92043), die Hauptindikationsbereiche, die Diagnosen und über die Zeitpunkte, an denen die BESS-Skala erhoben wurde (1 – 4).

**Tabelle 2: Übersicht 1 über die Datenerhebungen, in denen die Skala BESS mit-erfasst wurde**

	Studie / Datenerhebung	Typ	MZP Studie	N	Hauptindikation	Diagnosen	MZP BESS
1	Zauberberg-II-Studie	EVS	3	565	Psychosomatik	gemischt	E, K1J
2	R-K-Studie	EVS	4	560	Psychosomatik	gemischt	E, K1J, K5J
3	H-K-Studie	EVS	3	317	Psychosomatik	gemischt	E, K1J
4	GR-Studie	EVS	3	292	Psychosomatik	gemischt	E, K1J
5	CED-Studie	EVS	2	175	Psychosomatik	CED	E
6	ISOR-Studie	EVS	2	248	Orthopädie	Schmerzsyndrom	K1J
7	DOS	EVS	5	892	Pneumologie Dermatologie	gemischt	E, K6M, K1J, K2J
8	Katamnese-Studie Lindenallee	EVS	3	138	Psychosomatik	gemischt	K1J
9	RAP (2 Kliniken)	R-A	2	9352	Psychosomatik	gemischt	E
10	Patientenbefragung A (1 Klinik)	PF	1	1746	Psychosomatik	gemischt	E
11	Patientenbefragung B (11 Kliniken)	PF	1	55748	Orthopädie Kardiologie Onkologie Psychosomatik	gemischt	E
12	Patienten-Befragung C (29 Kliniken)	PF	1	92043	Orthopädie Kardiologie Onkologie Psychosomatik	gemischt	E

Studien und Datenerhebungen:

- Zauberberg-II-Studie (Schmidt et al., 1994a; Schmidt et al., 1994b)
- R-K-Studie = Reinerzauer-Katamnese-Studie (Nübling et al., 1995)
- H-K-Studie = Bad Herrenalber Katamnese-Studie (Nübling et al., 2000a)
- G-R-Studie = Studie zu Gesundheitsverhalten und Risikofaktoren (Schmidt et al., 1996)
- CED-Studie (Maatz & Schmidt, 1998)
- ISOR-Studie (Schmidt et al., 2012)
- DOS = Davoser Outcome-Studie (Schmidt et al., 2014)
- Katamnese-Studie Lindenallee (Nübling et al., 2013)
- RAP = Routine-Assessment Psychosomatik (Nübling et al., 2015)
- Patientenbefragung A und B (Steffanowski et al., 2006)
- Patientenbefragung C (Kriz, Nübling & Schmidt, 2014)

Typ / Art der Datenerhebung: EVS = prospektive Evaluationsstudie, R-A = Routine-Assessment, PF = Patientenbefragung (Ein-Punkt-Messung vor Entlassung); MZP Studie: Anzahl der Messzeitpunkte in der jeweiligen Studie; N = Größe der (Ausgangs-)Stichprobe; MZP BESS: Zeitpunkt der BESS-Erhebung: E = Entlassung, K6M = Katamnese 6 Monate nach E, K1J = Katamnese 1 Jahr nach E, K2J = Katamnese 2 Jahre nach E, K5J = Katamnese 5 Jahre nach E

Die reanalysierten Datensätze beinhalten:

- 8 umfassende, prospektive Evaluationsstudien aus der psychosomatischen (6 Studien), orthopädischen (1 Studie) und pneumologischen /dermatologischen Rehabilitation (1 Studie). Die Evaluationsstudien hatten 2 – 5 Messzeitpunkte und fanden zwischen 1987 und 2014 statt (vgl. Studien 1 – 8 in Tabelle 2);

- ein umfassendes Routine-Assessment aus dem Bereich der psychosomatischen Rehabilitation (2 Messzeitpunkte; erhoben 2012 bis 2016 in zwei Kliniken; vgl. Datenerhebung 9 in Tabelle 2) und
- drei kontinuierliche Patientenbefragungen vor Entlassung, die im Zeitraum 1997 bis 2013 in 40 Rehakliniken mit unterschiedlichen Hauptindikationen stattfanden (vgl. Datenerhebungen 10 – 12 in Tabelle 2). Patientenbefragung B beinhaltet Daten aus dem Zeitraum 1997 – 2004 (erhoben in 11 Kliniken), Patientenbefragung C Daten aus dem Zeitraum 2008 – 2013 (erhoben in 29 Kliniken).

Alle Daten stammen aus der stationären medizinischen Rehabilitation.

### 3.2 Stichproben

Eine Übersicht über Stichprobenmerkmale der einbezogenen Datensätze gibt Tabelle 3. Das Durchschnittsalter der Patienten variiert zwischen 38,4 (Datensatz 5) und 57,6 Jahren (Datensatz 12). Der Frauenanteil schwankt zwischen 48,2 % (Datensatz 6) und 73,0 % (Datensatz 8). Die durchschnittliche Behandlungsdauer liegt zwischen 26,5 Tagen (Datensatz 12) und 58,1 Tagen (Datensatz 4). Auch bezüglich der Zusammensetzung der Patientenstichproben nach Kostenträger der Reha zeigen sich Unterschiede: während der Anteil der Patienten von DRV-Bund (früher BfA) zwischen 6,3 % (Datensatz 9) und 75,4 % (Datensatz 8) variiert, liegt der Anteil der Patienten regionaler DRVen (früher LVAen) zwischen 8,6 % (Datensatz 7) und 88,8 % (Datensatz 9) und der Anteil von Kassenpatienten zwischen 0,9 % (Datensatz 9) und 30,9 (Datensatz 11).

**Tabelle 3: Übersicht 2 über Merkmale der 12 Studien / Datenerhebungen**

	Studie / Datenerhebung	Ø Alter (SD)	Frauen In %	Ø Dauer (SD)	DRV Bund %	reg. DRVen %	GKV %
1	Zauberberg-II-Studie	42,2 (10,2)	51,9	52,3 (19,8)	41,6	33,6	15,2
2	R-K-Studie	41,4 (10,2)	57,0	47,6 (20,6)	33,8	55,5	7,9
3	H-K-Studie	39,9 (11,6)	59,6	54,8 (21,4)	38,8	16,0	25,2
4	GR-Studie	44,7 (10,4)	60,3	58,1 (21,7)	60,1	18,4	20,5
5	CED-Studie	38,4 (11,1)	60,0	51,0 (22,6)	47,4	20,6	22,9
6	ISOR-Studie	52,6 (7,7)	48,2	34,5 (2,2)	--	--	--
7	DOS	49,1 (13,0)	59,7	27,4 (6,5)	55,9	8,6	16,4
8	Katamnese Lindenallee	51,2 (7,8)	73,0	39,1 (5,6)	75,4	21,0	3,6
9	RAP	47,9 (9,1)	51,5	35,4 (17,9)	6,3	88,3	0,9
10	Patientenbefragung A	46,3 (11,3)	66,1	45,4 (20,8)	50,9	17,7	18,2
11	Patientenbefragung B	55,8 (14,9)	58,1	30,0 (14,3)	35,0	24,3	30,9
12	Patientenbefragung C	57,6 (12,7)	60,8	26,5 (13,8)	48,0	29,1	4,4

### 3.3 Assessmentinstrumente

In den reanalysierten Studien wurde eine Vielzahl unterschiedlicher Fragebogenverfahren verwendet. Viele dieser Instrumente wurden wiederholt vorgegeben (Mehrfachmessungen), einige Fragebögen lediglich zu einem Messzeitpunkt. Eine alphabetische Übersicht über die Instrumente, die nachfolgend zur Validierung von BESS verwendet wurden, liefert Tabelle 4. Eine kurze inhaltliche Beschreibung inkl. Zitation findet man im Anhang (vgl. 7 Glossar Fragebögen).

**Tabelle 4: Übersicht über Instrumente, die nachfolgend zu Validierung von BESS herangezogen wurden (alphabetisch geordnet)**

Instrument	Anzahl Items	Bezeichnung	Autoren	Verwendet in Studie
BDI II	21	Beck-Depressions-Inventar, Revision	Hautzinger, Keller & Kühner (2009)	9
B-L	24	Beschwerde-Liste	von Zerssen (1975)	3, 4
D-S	16	Depressivitäts-Skala	von Zerssen (1975)	4
EMEK-27	27	Multiples Ergebniskriterium mit 27 Komponenten	Schmidt (1987; 1995)	1, 2, 3
GB-10	10	Gesundheitliches Befinden	Schmidt & Nübling R (2010a)	6, 7, 8, 9
GBB	57	Gießener Beschwerdebogen	Brähler & Scheer (1983)	5
HADS-D	14	Hospital Anxiety and Depression Scale - deutsche Version	Herrmann-Lingen, Buss & Snaith (1995)	6, 7
HAQ-D	11	Helping Alliance Questionnaire	Bassler, Potratz & Krauthauser (1995)	3, 9
HEALTH-49	49	Hamburger Module	Rabung, Harfst, Koch & Schulz (2007)	6, 8, 9
ICF AT 50-Psych	50	Selbstbeurteilung von Aktivitäten und Teilhabe bei psychischen Störungen	Nosper (2007)	9
ISR	29	ICD-10-Symptom-Rating	Tritt et al. (2008)	9
MFHW	7	Marburger Fragebogen zum habituellen Wohlbefinden	Basler, Herda & Scharfenstein (2003)	6
SCL-90-R	90	Symptom-Checkliste	Franke (2002)	3
SF-12	12	Kurzform des SF- 36 Health Survey	Morfeld, Kirchberger & Bullinger (2011)	6, 7
SIBAR	11	Frühberentungsrisiko - Screening-Instrument Beruf und Arbeit in der Rehabilitation (SIBAR)	Bürger W, Deck R, Raspe H, Koch U (2007), Bürger & Deck (2009)	7, 9
TSD	30	Test zur Erfassung der Schwere einer Depression	Obermair, Stoll & Rickels (1983)	5
VE	17	Verstimmungsstörungen (Subskala der Kieler Änderungssensitiven Symptomliste KASSL)	Zielke (1979)	1
VM-9	9	Fragebogen zur direkten Veränderungs-messung	Schmidt & Nübling (1988)	1, 2, 3, 4
VM-10	10	Fragebogen zur direkten Veränderungs-messung	Schmidt & Nübling (2010b)	7, 9
ZUF-8	8	Fragebogen zur Patientenzufriedenheit	Schmidt, Lamprecht & Wittmann (1989)	

### 3.4 Analyseverfahren

Zum Zwecke dieser Reanalyse wurden alle BESS – Messungen in den 12 Datensätzen erneut item- und faktorenanalytisch überprüft. Alle Datenanalysen erfolgten mit SPSS Version 23. Für die Skalen- und Itemanalysen wurde die Prozedur RELIABILITY verwendet, wobei auch die interne Konsistenz über Cronbach  $\alpha$  geschätzt wurde (basierend auf der durchschnittlichen Korrelation zwischen den Items). Weiterhin wurden die Korrelationsmuster der Skalenwerte analysiert. Zur Ermittlung der faktoriellen Struktur der BESS-Items wurde die Prozedur FACTOR verwendet und explorative Hauptkomponentenanalysen (PC, principal components) durchgeführt. Zur Ermittlung verschiedener Aspekte der Validität wurden Korrelationen nach Pearson bestimmt und Gruppenunterschiede mittels *t*-Tests, einfachen Varianzanalysen oder dem parameterfreien Kruskal-Wallis-Test analysiert. Mittels *t*-Tests für unabhängige Stichproben wurde überprüft, ob die Mittelwerte zweier unabhängiger Stichproben verschieden sind. Mittels einfaktorieller Varianzanalysen wurde getestet, ob sich die Mittelwerte mehrerer unabhängiger Gruppen (oder Stichproben) unterscheiden, die durch eine kategoriale unabhängige Variable definiert werden. Mittels Kruskal-Wallis-Tests für unabhängige Stichproben wurde getestet, ob sich die zentralen Tendenzen mehrerer unabhängiger Stichproben unterscheiden. Der Kruskal-Wallis-Test wurde verwendet, wenn die Voraussetzungen für eine Varianzanalyse nicht erfüllt waren. Mit dem Kolmogorov-Smirnov-Anpassungstest (KS-Test in SPSS) wurde weiterhin die Annahme der Normalverteilung überprüft.

## 4 Ergebnisse

### 4.1 Item- und Skalenanalysen

Die wichtigsten Befunde der durchgeführten Item- und Skalenanalysen sind in Tabelle 5 zusammengefasst. Jeweils aufgelistet sind der Messzeitpunkt der BESS-Erhebung, die Stichprobengröße, der durchschnittliche Item-Mittelwert, die durchschnittliche Inter-Item-Korrelation der 5 BESS-Items, die Trennschärfen (min, max), der Skalen-Mittelwert, die Standardabweichung der Skala und der Wert Cronbach  $\alpha$ .

10 von 20 Item- und Skalenanalysen basieren auf BESS-Messungen bei Entlassung, bei 7 liegen Messungen 1 Jahr nach Entlassung zugrunde und 3 betreffen weitere Messzeitpunkte 6 Monate, 2 Jahre und 5 Jahre nach Entlassung. Während die durchschnittlichen Item-Mittelwerte zwischen 3,16 (Zeile 16, Datensatz 6) und 4,22 (Zeile 6, Datensatz 7) variieren, liegen die beobachteten BESS-Skalenwerte entsprechend auch zwischen 15,81 (Zeile 16,

Datensatz 6) und 21,11 (Zeile 6, Datensatz 7). Im Durchschnitt können zum Entlassungszeitpunkt höhere Skalenwerte als in den Katamnesen beobachtet werden, d.h. mit wachsendem zeitlichen Abstand zum Ende der stationären Reha-Maßnahmen schildern die Patienten weniger positive Veränderungen als bei Entlassung.

**Tabelle 5: Ergebnisse der Item- und Skalenanalysen der Skala BESS (5 Items) in den 12 Datenerhebungen**

Nummer	MZP	Studie	N	Ø Item-Mittelwert	Ø Inter-Item-Korrelation	Trennschärpen	M Skala	SD Skala	Cronbach $\alpha$
1	E	1	405	3,90	0,64	0,68 – 0,85	19,51	3,69	<b>0,89</b>
2	E	2	251	4,01	0,60	0,57 – 0,81	20,08	3,31	<b>0,88</b>
3	E	3	258	4,07	0,56	0,65 – 0,73	20,35	3,61	<b>0,86</b>
4	E	4	253	3,79	0,64	0,69 – 0,83	18,93	3,91	<b>0,89</b>
5	E	5	162	3,79	0,57	0,60 – 0,81	18,93	3,78	<b>0,87</b>
6	E	7	759	4,22	0,58	0,59 – 0,80	21,11	3,09	<b>0,87</b>
7	E	9	9352	3,79	0,76	0,80 – 0,87	18,98	4,38	<b>0,94</b>
8	E	10	1522	3,81	0,68	0,72 – 0,85	19,07	3,83	<b>0,91</b>
9	E	11	51386	4,08	0,64	0,64 – 0,81	20,42	3,63	<b>0,90</b>
10	E	12	80647	4,06	0,62	0,68 – 0,81	20,29	3,49	<b>0,89</b>
11	K6M	7	694	3,77	0,66	0,65 – 0,86	18,86	4,04	<b>0,91</b>
12	K1J	1	359	3,63	0,73	0,78 – 0,88	18,15	4,43	<b>0,92</b>
13	K1J	2	386	3,66	0,76	0,80 – 0,90	18,30	4,86	<b>0,94</b>
14	K1J	3	184	4,03	0,64	0,68 – 0,80	20,17	4,26	<b>0,89</b>
15	K1J	4	206	3,61	0,70	0,76 – 0,86	18,05	4,23	<b>0,92</b>
16	K1J	6	104	3,16	0,69	0,68 – 0,85	15,81	4,34	<b>0,92</b>
17	K1J	7	632	3,71	0,62	0,54 – 0,84	18,57	3,93	<b>0,89</b>
18	K1J	8	133	3,66	0,70	0,71 – 0,88	18,28	4,23	<b>0,94</b>
19	K2J	7	564	3,62	0,61	0,48 – 0,85	18,10	4,02	<b>0,89</b>
20	K5J	2	293	3,75	0,76	0,80 – 0,90	18,76	5,34	<b>0,93</b>

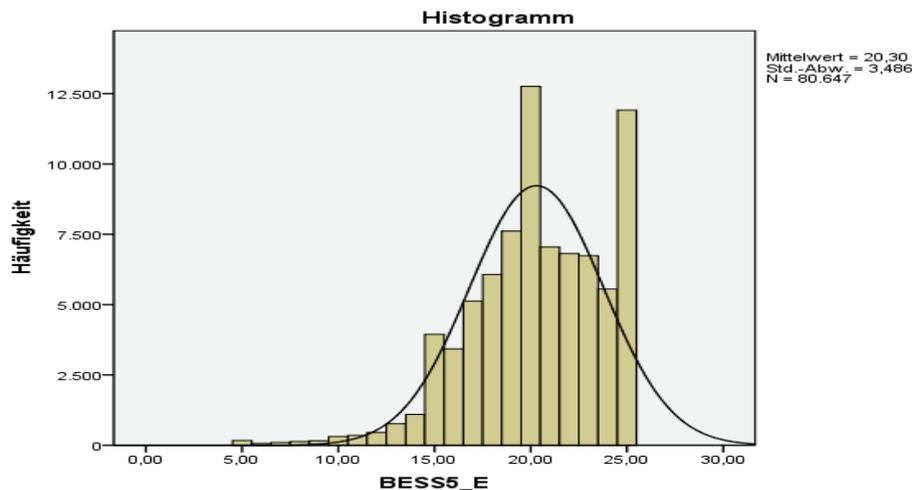
MZP=Zeitpunkt der Erhebung: E = Entlassung, K6M = Katamnese 6 Monate nach E, K1J = Katamnese 1 Jahr nach E, K2J = Katamnese 2 Jahre nach E, K5J = Katamnese 5 Jahre nach E; Trennschärpen = korrigierte Item-Skala-Korrelationen; M=Mittelwert; SD=Streuung

Die Standardabweichungen liegen zwischen SD = 3,09 (Zeile 6, Datensatz 7) und SD = 5,34 (Zeile 20, Datensatz 2). In der Gesamtschau zeigen die 5 BESS-Items eine deutliche Homogenität, wobei die durchschnittlichen Inter-Item-Korrelationen im Bereich zwischen 0,56 und 0,76, die interne Konsistenz der Skala (Cronbach  $\alpha$ ) - zwischen  $\alpha = 0,86$  und  $\alpha = 0,94$  liegen. Die ermittelten korrigierten Item-Skala-Korrelationen (Trennschärpen) liegen in den 20 Analysen zwischen 0,48 und 0,90. In keiner Analyse gibt es Items mit Trennschärpen-Werten unterhalb 0,30 und nur in einem einzigen Fall wird die 0,50-Bariere unterschritten.

Eine typische Verteilung der BESS-Skalenwerte zeigt Abbildung 4 in Form eines Histogramms. Dargestellt sind die BESS-Skalenwerte in der Patientenbefragung C (Daten aus 28

Rehakliniken; Messung kurz vor Entlassung). Die Verteilung ist zweigipflig und linksschief bzw. rechtssteil mit einem Schiefe-Wert von - 0,712 und einem Exzess (Kurtosis) von 0,833.

**Abbildung 4: Verteilung der Skala BESS – Datensatz Nr. 12 (Patientenbefragung in Rehabilitationskliniken vor Entlassung) N = 80.647**



Der Kolmogorov-Smirnov-Test ist signifikant, d.h. die BESS-Skalenwerte sind nicht normalverteilt (Prüfwert = 0,96, df = 80647, p = 0,000 mit Signifikanzkorrektur nach Lilliefors)

## 4.2 Faktorielle Struktur der BESS-Items - Explorative Faktorenanalysen

Die wichtigsten Befunde der durchgeführten Hauptkomponentenanalysen sind in Tabelle 6 zusammengefasst. Tabelle 6 zeigt, dass in allen 20 Analysen jeweils nur 1 Faktor mit einem Eigenwert > 1 beobachtet werden kann; sie legen insgesamt klar eine 1-Faktoren-Lösung nahe, die etwa 65 – 81 % der Gesamtvarianz erklären kann. Die 5 BESS-Items haben durchgängig hohe Ladungen auf diesem Generalfaktor, die mehrheitlich etwa im Bereich zwischen 0,78 und 0,90 liegen.

## 4.3 Zeitliche Stabilität / Änderungssensitivität

Tabelle 7 zeigt wie die bei Entlassung gemessenen BESS-Skalenwerte (MZP 1) mit zeitlich späteren BESS-Messungen (MZP 2) korrelieren. In fünf Fällen konnte der Zusammenhang zwischen BESS-Skalenwerten bei Entlassung und BESS-Skalenwerten in der 1-Jahres-Katamnese überprüft werden, in den anderen Fällen wurde der Zusammenhang mit BESS-Werten in der 6-Monats- bzw. 2-Jahres- bzw. 5-Jahres-Katamnese ermittelt. Tabelle 7 zeigt, dass moderate signifikante Korrelationen im Bereich  $r = 0,32$  bis  $r = 0,45$  beobachtet werden

können. Der höchste Zusammenhang ergibt sich zur BESS-Messung 6 Monate nach Entlassung, der vergleichsweise geringste Zusammenhang zeigt sich zur BESS-Messung nach 5 Jahren.

**Tabelle 6: Ergebnisse der Hauptkomponentenanalysen der Skala BESS (5 Items) in den 12 Datenerhebungen (SPSS FACTOR)**

Nummer	MZP	Studie	N	EW >1	Erklärte Varianz in %	Faktorenloadungen (range)
1	E	1	405	1	71,7	0,79 - 0,91
2	E	2	251	1	68,5	0,71 - 0,89
3	E	3	258	1	64,9	0,79 - 0,84
4	E	4	253	1	71,3	0,80 - 0,90
5	E	5	162	1	66,0	0,74 - 0,89
6	E	7	759	1	66,6	0,73 - 0,88
7	E	9	9352	1	80,6	0,87 - 0,92
8	E	10	1522	1	74,3	0,82 - 0,91
9	E	11	51386	1	71,1	0,76 - 0,89
10	E	12	80647	1	70,1	0,78 - 0,89
11	K6M	7	694	1	73,4	0,76 - 0,92
12	K1J	1	359	1	78,2	0,86 - 0,93
13	K1J	2	386	1	80,9	0,87 - 0,94
14	K1J	3	184	1	71,2	0,79 - 0,88
15	K1J	4	206	1	76,1	0,84 - 0,92
16	K1J	6	104	1	75,5	0,79 - 0,91
17	K1J	7	632	1	70,2	0,67 - 0,91
18	K1J	8	133	1	76,4	0,87 - 0,92
19	K2J	7	564	1	69,6	0,61 - 0,92
20	K5J	2	293	1	80,8	0,87 - 0,94

MZP=Zeitpunkt der Erhebung: E = Entlassung, K6M = Katamnese 6 Monate nach E, K1J = Katamnese 1 Jahr nach E, K2J = Katamnese 2 Jahre nach E, K5J = Katamnese 5 Jahre nach E; EW=Eigenwert

**Tabelle 7: Zeitliche Stabilität der BESS-Skalenwerte**

Studie	MZP 1	MZP 2	BESS M(SD) 1	BESS M (SD) 2	r <sub>12</sub>	N	t	p	ES
7	E	K6M	21,25 (2,97)	18,87 (4,01)	<b>0,45</b>	695	16,19	.000	0,80
1	E	K1J	19,54 (3,58)	18,29 (4,35)	<b>0,36</b>	276	4,59	.000	0,35
2	E	K1J	20,29 (3,33)	18,24 (5,08)	<b>0,44</b>	184	5,96	.000	0,62
3	E	K1J	20,61 (3,54)	20,52 (4,26)	<b>0,35</b>	163	0,29	.766	0,03
4	E	K1J	19,04 (3,90)	18,07 (4,24)	<b>0,41</b>	192	3,03	.003	0,24
7	E	K1J	21,23 (2,95)	18,56 (3,89)	<b>0,39</b>	596	16,88	.000	0,90
7	E	K2J	21,28 (2,95)	18,05 (4,00)	<b>0,36</b>	530	18,42	.000	1,09
2	E	K5J	20,08 (3,51)	18,89 (5,22)	<b>0,32</b>	129	2,55	.012	0,34

M=Mittelwert, SD=Standardabweichung; MZP=Messzeitpunkte: E = Entlassung / K1J = 1-Jahres-Katamnese / K5J = 5-Jahres-Katamnese; r<sub>12</sub> = Produkt-Moment-Korrelation nach Pearson; t = Wert der Prüfgröße t ( t-Test für abhängige Stichproben), p=Unterschied (zweis.); ES = (Mittelwert<sub>BESS1</sub> - Mittelwert<sub>BESS2</sub>) / SD<sub>BESS1</sub> (Effektgröße des Unterschieds)

Zwischen BESS-Messungen bei Entlassung und katamnestischen BESS-Messungen können in 7 von 8 Vergleichen signifikante Mittelwertsunterschiede beobachtet werden. Die Effektgrößen des Unterschieds liegen zwischen ES = 0,24 und 0,61. In den Nacherhebungen zeigen sich in der Regel weniger positive Änderungen als bei der Messung am Ende der Reha. Dies steht im Einklang mit vielen Befunden, die eine Abnahme positiver Veränderungen mit wachsendem Abstand zum Ende der stationären Rehabilitation aufzeigen und sprechen für die Änderungssensitivität der Skala.

## 4.4 Facetten der Validität (Gültigkeit)

Zur Erhellung der Validität der BESS-Skala wurden insbesondere verschiedene Aspekte der Kriteriumsvalidität bzw. Konstruktvalidität untersucht: die konkurrente Validität (= Zusammenhang mit gleichzeitigen Messungen anderer Merkmale), die konvergente Validität (= Zusammenhang mit anderen Veränderungsinformationen) und die prädiktive Validität (= Zusammenhang mit zeitlich später erhobenen Kriterien). Weiterhin wurde überprüft, wie stark sich unterschiedliche Gruppen bezüglich BESS unterscheiden (diskriminative Validität) und in welchem Maße BESS-Werte bei Entlassung mit Messungen bei Reha-Beginn korrelieren (Vorhersagbarkeit).

### 4.4.1 Zusammenhänge mit zeitgleich erhobenen Outcome-bezogenen Merkmalen

Tabelle 8 macht deutlich, dass zwischen BESS-Skalenwerten und gleichzeitig erhobenen Selbst- und Fremdbeurteilungen, die sich auf das Ergebnis beziehen, moderate bis deutliche und hochsignifikante Zusammenhänge bestehen ( $r = 0,30$  bis fast  $r = 0,70$ ). Gleichzeitig bzw. zeitgleich erhoben meint hier, dass die zum Zeitpunkt X gemessenen BESS-Werte mit anderen zum Zeitpunkt X gemessenen Outcome-Variablen korreliert wurden.

Die BESS-Skalenwerte korrelieren deutlich mit anderen patientenseitigen Beurteilungen, die das Ergebnis betreffen, z.B. mit der Patientenzufriedenheit (ZUF-8; bis  $r = 0,58$ ), mit der Nutzenbewertung der Behandlung (bis  $r = 0,68$ ), mit der Einschätzung der Zielerreichung (bis  $r = 0,66$ ) und mit der Erfolgs- bzw. Gesamtbewertung der Reha (bis  $r = 0,63$ ).

Ein deutlicher Zusammenhang zeigt sich auch zwischen BESS und der HAQ-Skala „Erfolgszufriedenheit“ (bis  $r = 0,69$ ). Ebenfalls hochsignifikante Zusammenhänge gibt es zwischen BESS und Fremdbeurteilungen der Zielerreichung ( $r = 0,54$ ) und des Ergebnisses ( $r = 0,51$ ) durch Ärzte.

**Tabelle 8: BESS – Skala - Zusammenhänge mit zeitgleich erhobenen Outcome-bezogenen Daten**

Kennwert	Typ	MZP	Studie	r	N	p
<b>Selbstbeurteilungen (Patienten)</b>						
Nutzenbewertung der Behandlung (1-4, 4 = großer Nutzen)	EI	E	9	<b>0,68</b>	3452	0,000 ***
	EI	K1J	1	<b>0,39</b>	351	0,000 ***
	EI	K1J	2	<b>0,55</b>	381	0,000 ***
	EI	K1J	4	<b>0,39</b>	206	0,000 ***
	EI	K1J	8	<b>0,55</b>	132	0,000 ***
Einschätzung der Zielerreichung (1-5, 5 = vollständig erreicht)	EI	E	11	<b>0,66</b>	39750	0,000 ***
	EI	E	7	<b>0,59</b>	766	0,000 ***
	EI	K1J	7	<b>0,52</b>	644	0,000 ***
Erfolgsbewertung der Reha (1-5, 5 = ausgezeichnet)	EI	E	12	<b>0,57</b>	80344	0,000 ***
Gesamtbewertung der Reha (0-10, 10 = ausgezeichnet)	EI	E	7	<b>0,63</b>	3452	0,000 ***
	EI	K1J	8	<b>0,38</b>	133	0,000 ***
Gesamtbewertung der Reha (1-5, 5 = ausgezeichnet)	EI	E	12	<b>0,51</b>	88775	0,000 ***
HAQ - Beziehungszufriedenheit	S	E	9	<b>0,46</b>	3436	0,000 ***
	S	E	3	<b>0,31</b>	255	0,000 ***
HAQ - Erfolgswzufriedenheit	S	E	9	<b>0,69</b>	3436	0,000 ***
	S	E	3	<b>0,61</b>	258	0,000 ***
HAQ - Gesamtwert	S	E	9	<b>0,59</b>	3436	0,000 ***
	S	E	3	<b>0,45</b>	253	0,000 ***
ZUF-8 (Zufriedenheit mit der Behandlung)	S	E	9	<b>0,65</b>	3452	0,000 ***
	S	E	11	<b>0,58</b>	51275	0,000 ***
	S	E	12	<b>0,50</b>	82682	0,000 ***
	S	E	1	<b>0,43</b>	385	0,000 ***
	S	E	3	<b>0,34</b>	253	0,000 ***
	S	E	4	<b>0,47</b>	250	0,000 ***
	S	E	5	<b>0,44</b>	153	0,000 ***
	S	E	7	<b>0,55</b>	767	0,000 ***
	S	E	10	<b>0,56</b>	1323	0,000 ***
	S	K1J	6	<b>0,47</b>	107	0,000 ***
Zufriedenheit mit Behandlungsergebnis (1-5, 5 = sehr zufrieden)	EI	E	9	<b>0,64</b>	3452	0,000 ***
	EI	K1J	8	<b>0,50</b>	131	0,000 ***
<b>Fremdbeurteilungen (Ärzte/Psychotherapeuten)</b>						
Einschätzung der Zielerreichung (1-5, 5 = vollständig erreicht)	EI	E	9	<b>0,54</b>	1829	0,000 ***
Ergebnisbewertung (1-5, 5 = sehr gelohnt)	EI	E	9	<b>0,51</b>	1835	0,000 ***

Legende: Typ: EI = Einzelitem (single-item-rating), S = Skala; MZP = Zeitpunkt der Erhebung von BESS: E = Entlassung, K6M = Katamnese 6 Monate nach E, K1J = Katamnese 1 Jahr nach E, K2J = Katamnese 2 Jahre nach E, K5J = Katamnese 5 Jahre nach E; ; r = Produkt-Moment-Korrelation nach Pearson; N = Stichprobengröße; p = Signifikanz (2-seitig).

#### 4.4.2 Zusammenhänge mit Veränderungsinformationen aus der direkten Veränderungsmessung (DVM)

Als Veränderungsinformation aus der DVM wurden singuläre Veränderungseinstufungen, Veränderungsskalen auf DVM-Basis und ein multiples Ergebniskriterium auf DVM-Basis herangezogen. Tabelle 9 zeigt, dass es zwischen BESS und anderen direkten Veränderungsinformationen (Einzelitems und Skalen) teilweise hochsignifikante und substantielle Zusammenhänge gibt, die aus Sicht der Autoren für die Konstruktvalidität der Skala sprechen. So korrelieren die BESS-Skalenwerte z. B. hoch mit den Single-Item-Ratings „Veränderung des

Gesundheitszustand“ ( $r = 0,76$  bis  $r = 0,88$ ) und „Veränderung der Belastbarkeit“ ( $r = 0,84$  bis  $r = 0,85$ ).

**Tabelle 9: Skala BESS - Zusammenhänge mit Veränderungsinformationen aus der direkten Veränderungsmessung**

Kennwert	Typ	MZP	Studie	r	N	p
<b>Selbstbeurteilungen (Patienten)</b>						
Veränderung Gesundheitszustand (1-5, 5 = deutlich besser)	EI	E	9	<b>0,88</b>	3452	0,000 ***
	EI	E	12	<b>0,78</b>	87248	0,000 ***
	EI	K1J	1	<b>0,87</b>	358	0,000 ***
	EI	K1J	8	<b>0,88</b>	137	0,000 ***
	EI	K	3	<b>0,76</b>	178	0,000 ***
Veränderung Belastbarkeit (1-5, 5 = deutlich besser)	EI	E	9	<b>0,85</b>	3452	0,000 ***
	EI	K1J	8	<b>0,84</b>	138	0,000 ***
Veränderung Entspannungsfähigkeit (1-5, 5 = deutlich besser)	EI	E	9	<b>0,73</b>	3452	0,000 ***
	EI	K1J	8	<b>0,73</b>	138	0,000 ***
Veränderung Ausgeglichenheit (1-5, 5 = deutlich besser)	EI	E	9	<b>0,78</b>	3452	0,000 ***
	EI	K1J	8	<b>0,75</b>	138	0,000 ***
VM-9 (Veränderungsskala, 9 Items)	S	E	3	<b>0,68</b>	250	0,000 ***
	S	K1J	1	<b>0,74</b>	335	0,000 ***
	S	K1J	2	<b>0,79</b>	337	0,000 ***
	S	K1J	3	<b>0,78</b>	174	0,000 ***
	S	K1J	4	<b>0,76</b>	198	0,000 ***
VM-10 (Veränderungsskala, 10 Items)	S	E	9	<b>0,96</b>	3452	0,000 ***
	S	E	7	<b>0,93</b>	766	0,000 ***
	S	K1J	7	<b>0,95</b>	647	0,000 ***
EMEK-27 (Multiples Ergebniskriterium)	S	K1J	1	<b>0,82</b>	359	0,000 ***
	S	K1J	2	<b>0,84</b>	386	0,000 ***
	S	K1J	3	<b>0,83</b>	185	0,000 ***
<b>Fremdbeurteilungen (Ärzte/Psychotherapeuten)</b>						
Veränderung körperliches Befinden (1-5, 5 = deutlich gebessert)	EI	E	9	<b>0,48</b>	1810	0,000 ***
	EI	E	5	<b>0,41</b>	159	0,000 ***
	EI	E	1	<b>0,31</b>	201	0,000 ***
	EI	E	2	<b>0,28</b>	251	0,000 ***
Veränderung seelisches Befinden (1-5, 5 = deutlich gebessert)	EI	E	9	<b>0,50</b>	1831	0,000 ***
	EI	E	1	<b>0,30</b>	200	0,000 ***
	EI	E	2	<b>0,29</b>	251	0,000 ***

Typ: EI = Einzelitem (single-item-rating), S = Skala; MZP = Zeitpunkt der Erhebung von BESS: E = Entlassung, K6M = Katamnese 6 Monate nach E, K1J = Katamnese 1 Jahr nach E, K2J = Katamnese 2 Jahre nach E, K5J = Katamnese 5 Jahre nach E; ; r = Produkt-Moment-Korrelation nach Pearson; N = Stichprobengröße; p = Signifikanz (2-seitig).

Zum Teil noch deutlichere Zusammenhänge gibt es auf der Selbstbeurteilungsebene zwischen BESS-Skalenwerten und den Veränderungsskalen VM-9 und VM-10 ( $r = 0,68$  bis  $r = 0,96$ ) und dem multiplen Ergebniskriterium EMEK-27 ( $r = 0,82$  bis  $r = 0,84$ ). Eher im moderaten Bereich liegen die Korrelationen der BESS-Skalenwerte mit Single-Item-Ratings durch Ärzte. Hier liegen die Zusammenhänge mit den Veränderungseinschätzungen des körperlichen und seelischen Befindens zwischen  $r = 0,30$  und  $r = 0,50$ . In der Gesamtschau liegen etwa 75 % der 29 Korrelationskoeffizienten im Bereich  $r > 0,50$  und nur 2 Korrelationen im Bereich  $r < 0,30$ .

#### 4.4.3 Zusammenhänge mit Veränderungsinformationen aus der indirekten Veränderungsmessung (IVM)

Tabelle 10 zeigt, dass es zwischen BESS-Skalenwerten und zeitlich korrespondierenden Differenzwerten vielfältige hochsignifikante Zusammenhänge gibt, die ebenfalls für die Konstruktvalidität der BESS-Skala sprechen.

**Tabelle 10: BESS-Skala - Zusammenhänge mit Veränderungsinformationen aus der indirekten Veränderungsmessung (Skalen-Differenzwerte)**

Variable	Differenzwert zwischen	Studie	r	N	p
Gesundheitliches Befinden (GB-10)	A - E	9	<b>-0,58</b>	3452	0,000 ***
	A - E	7	<b>-0,48</b>	753	0,000 ***
	A - K1J	6	<b>-0,67</b>	111	0,000 ***
	A - K1J	7	<b>-0,54</b>	632	0,000 ***
	A - K1J	8	<b>-0,67</b>	136	0,000 ***
ISR - Gesamtscore	A - E	9	<b>0,43</b>	3437	0,000 ***
ICF – AT 50 - Beeinträchtigungsindex	A - E	9	<b>0,43</b>	3441	0,000 ***
Beschwerden (B-L)	A - E	3	<b>0,41</b>	255	0,000 ***
	A - K1J	3	<b>0,43</b>	178	0,000 ***
	A - E	4	<b>0,44</b>	249	0,000 ***
	A - K1J	4	<b>0,47</b>	204	0,000 ***
Beschwerden (GBB – Gesamtwert)	A - E	5	<b>0,42</b>	95	0,000 ***
Somatoforme Beschwerden (HEALTH-49)	A - E	9	<b>0,39</b>	3445	0,000 ***
	A - K1J	8	<b>0,31</b>	136	0,000 ***
Depressives Syndrom (ISR)	A - E	9	<b>0,47</b>	3437	0,000 ***
Depressivität (HEALTH-49)	A - E	9	<b>0,40</b>	3445	0,000 ***
Depression (HADS-D)	A - E	7	<b>0,30</b>	749	0,000 ***
	A - K1J	6	<b>0,60</b>	110	0,000 ***
	A - K1J	8	<b>0,45</b>	136	0,000 ***
Depression (BDI II)	A - E	9	<b>0,38</b>	3352	0,000 ***
Depression (D-S)	A - E	4	<b>0,37</b>	250	0,000 ***
	A - K1J	4	<b>0,32</b>	204	0,000 ***
Depression (TSD)	A - E	5	<b>0,38</b>	97	0,000 ***
Angstsyndrom (ISR)	A - E	9	<b>0,30</b>	3437	0,000 ***
Phobische Angst (HEALTH-49)	A - E	9	<b>0,16</b>	3445	0,000 ***
	A - K1J	8	<b>0,18</b>	136	0,038 *
Angst (HADS)	A - E	7	<b>0,29</b>	749	0,000 ***
	A - K1J	6	<b>0,47</b>	110	0,000 ***
Psychisches Wohlbefinden (HEALTH-49)	A - E	9	<b>0,54</b>	3445	0,000 ***
	A - K1J	8	<b>0,62</b>	135	0,000 ***
Wohlbefinden (MFHW)	A - K1J	6	<b>-0,45</b>	100	0,000 ***
Verstimmungsstörungen (VE)	A - K1J	1	<b>0,48</b>	356	0,000 ***
Selbstwirksamkeit (HEALTH-49)	A - E	9	<b>0,54</b>	3445	0,000 ***
	A - K1J	8	<b>0,50</b>	136	0,000 ***
Körperliche Gesundheit (SF-12)	A - E	7	<b>-0,27</b>	662	0,000 ***
	A - K1J	7	<b>-0,33</b>	564	0,000 ***
	A - K1J	6	<b>-0,22</b>	63	0,090 ns
Psychische Gesundheit (SF-12)	A - E	7	<b>-0,24</b>	662	0,000 ***
	A - K1J	7	<b>-0,31</b>	564	0,000 **
	A - K1J	6	<b>-0,39</b>	63	0,002 **
SCL-90-R-GSI	A - E	3	<b>0,36</b>	241	0,000 ***
	A - K1J	3	<b>0,36</b>	176	0,000 ***

MZP = Zeitpunkt der Erhebung von BESS: E = Entlassung, K6M = Katamnese 6 Monate nach E, K1J = Katamnese 1 Jahr nach E, K2J = Katamnese 2 Jahre nach E, K5J = Katamnese 5 Jahre nach E; r = Produkt-Moment-Korrelation nach Pearson; N = Stichprobengröße; p = Signifikanz (2-seitig).

Als Veränderungsinformation aus der IVM wurden Skalen-Differenzwerte aus wiederholten Statusmessungen herangezogen. Zur Ermittlung des Zusammenhangs zwischen direkter

und indirekter Veränderungsmessung wurden z. B. die bei E (Entlassung) gemessenen BESS-Skalenwerte mit den korrespondierenden A-E-Differenzwerten (= Unterschied zwischen dem Status bei Aufnahme und dem Status bei Entlassung) und die bei der 1-Jahres-Katamnese gemessenen BESS-Werte mit den entsprechenden A-K1J-Differenzwerten (Unterschied zwischen dem Status bei Aufnahme und dem Status bei der 1-Jahres-Katamnese) korreliert.

Verglichen mit den direkten Veränderungsinformationen fallen die Korrelationen mit den indirekten Veränderungsinformationen im Durchschnitt geringer aus, die Korrelationskoeffizienten in Tabelle 10 variieren zwischen  $r = \pm 0,16$  und  $r = \pm 0,67$ . Von den 42 Korrelationskoeffizienten liegen zwei Drittel im Bereich zwischen  $r = \pm 0,31$  –  $r = \pm 0,50$  und knapp 20 % im Bereich über bzw. unter  $r = \pm 0,50$ . Die vergleichsweise höchsten Zusammenhänge zeigen sich zwischen BESS und den GB-10-Differenzwerten (bis zu  $r = - 0,67$ ): je mehr positive gesundheitliche Veränderungen patientenseitig berichtet werden (BESS), desto negativer sind die korrespondierenden GB-10-Differenzwerte dieser Patienten (die negativen Vorzeichen der Korrelationskoeffizienten sind polungsbedingt, da negative Differenzwerte eine Besserung des Gesundheitlichen Befindens anzeigen). Die deutlichen Korrelationen sind auch Folge einer gewissen inhaltlichen Symmetrie zwischen BESS und GB-10: 4 der 5 BESS-Beurteilungsaspekte (Körperliches Befinden, seelisches Befinden, Allgemeinbefinden und Leistungsfähigkeit) sind auch in den 10 Items der Statusskala GB-10 enthalten.

#### 4.4.4 Zusammenhänge mit zeitgleichen Status-Messungen

Inhaltlich plausible Zusammenhänge zeigen sich auch zwischen BESS und zeitgleichen Status-Messungen (vgl. Tabelle 11). Zeitgleich bedeutet hier wiederum, dass die zum Zeitpunkt X gemessenen BESS-Skalenwerte mit ebenfalls zum Zeitpunkt X gemessenen Statusmerkmalen in Beziehung gesetzt wurden. Die in Tabelle 11 zusammengefassten 56 Korrelationen beziehen sich auf drei Single-Item-Statusratings (Einzelitems zu Gesundheitszustand, körperlichem und seelischem krank Empfinden) und 30 psychometrische Skalen, deren Zusammenhang mit BESS zum Teil für die Messzeitpunkte E und K1J ermittelt wurde.

Die ermittelten Korrelationen sind hochsignifikant und plausibel. Sie variieren zwischen  $r = \pm 0,21$  und  $r = \pm 0,79$  und entsprechen durchgängig der Erwartungsrichtung (die Vorzeichen sind polungsbedingt). Knapp 60 % der Koeffizienten liegen im Bereich über bzw. unter  $r = \pm 0,50$ , diese Zusammenhänge sind somit recht deutlich.

Tabelle 11: BESS-Skala - Zusammenhänge mit gleichzeitigen Statusmessungen

Kennwert	Typ	MZP	Studie	r	N	p
Gesundheitliches Befinden (GB-10)	S	E	7	<b>0,55</b>	762	0,000 ***
	S	E	9	<b>0,79</b>	3452	0,000 ***
	S	K1J	6	<b>0,68</b>	111	0,000 ***
	S	K1J	7	<b>0,55</b>	762	0,000 ***
	S	K1J	8	<b>0,79</b>	137	0,000 ***
Krank empfinden – körperlich (1-5, 5 = überhaupt nicht)	EI	E	3	<b>0,40</b>	258	0,000 ***
	EI	E	5	<b>0,47</b>	162	0,000 ***
Krank empfinden – seelisch (1-5, 5 = überhaupt nicht)	EI	E	3	<b>0,48</b>	257	0,000 ***
	EI	E	5	<b>0,31</b>	162	0,000 ***
Gesundheitszustand (1-6, 6 = sehr gut)	EI	E	9	<b>0,77</b>	3452	0,000 ***
	EI	E	5	<b>0,75</b>	253	0,000 ***
	EI	E	7	<b>0,52</b>	762	0,000 ***
	EI	K1J	7	<b>0,54</b>	647	0,000 ***
Beschwerden (B-L)	S	E	3	<b>-0,49</b>	255	0,000 ***
	S	K1J	3	<b>-0,53</b>	178	0,000 ***
	S	E	4	<b>-0,45</b>	249	0,000 ***
	S	K1J	4	<b>-0,50</b>	204	0,000 ***
Beschwerdedruck (GBB)	S	E	5	<b>-0,50</b>	104	0,000 ***
Somatisierung (SCL-90-R)	S	E	3	<b>-0,43</b>	253	0,000 ***
Somatoformes Syndrom (ISR)	S	E	9	<b>-0,64</b>	3439	0,000 ***
Somatoforme Beschwerden (HEALTH-49)	S	E	9	<b>-0,66</b>	3445	0,000 ***
	S	K1J	8	<b>-0,54</b>	136	0,000 ***
Depressivität (SCL-90-R)	S	E	3	<b>-0,50</b>	253	0,000 ***
Depression (HADS-D)	S	E	7	<b>-0,29</b>	761	0,000 ***
	S	K1J	6	<b>-0,65</b>	110	0,000 ***
Depressivität (HEALTH-49)	S	E	9	<b>-0,64</b>	3445	0,000 ***
	S	K1J	8	<b>-0,68</b>	138	0,000 ***
Depressives Syndrom (ISR)	S	E	9	<b>-0,67</b>	3439	0,000 ***
Depressivität (D-S)	S	E	4	<b>-0,47</b>	250	0,000 ***
	S	K	4	<b>-0,53</b>	204	0,000 ***
Depression (BDI II)	S	E	9	<b>-0,64</b>	3356	0,000 ***
Depression (TSD)	S	E	5	<b>-0,45</b>	104	0,000 ***
Ängstlichkeit (SCL-90-R)	S	E	3	<b>-0,46</b>	253	0,000 ***
Angst (HADS-D)	S	E	7	<b>-0,24</b>	761	0,000 ***
	S	K1J	6	<b>-0,51</b>	110	0,000 ***
Phobische Angst (SCL-90-R)	S	E	3	<b>-0,39</b>	253	0,000 ***
Phobische Ängste (HEALTH-49)	S	E	9	<b>-0,50</b>	3445	0,000 ***
	S	K1J	8	<b>-0,50</b>	136	0,000 ***
	S	K1J	6	<b>-0,61</b>	110	0,000 ***
HADS-D-Gesamtwert	S	E	7	<b>-0,28</b>	761	0,000 ***
Verstimmungsstörungen (VE)	S	K1J	1	<b>-0,47</b>	358	0,000 ***
Psychisches Wohlbefinden (HEALTH-49)	S	E	9	<b>-0,71</b>	3445	0,000 ***
	S	K1J	8	<b>-0,67</b>	135	0,000 ***
Wohlbefinden (MFHW)	S	K1J	6	<b>0,65</b>	104	0,000 ***
Selbstwirksamkeit (HEALTH-49)	S	E	9	<b>-0,63</b>	3445	0,000 ***
	S	K1J	8	<b>-0,69</b>	136	0,000 ***
Psychotizismus (SCL-90-R)	S	E	3	<b>-0,50</b>	253	0,000 ***
Globalwert GSI (SCL-90-R)	S	E	3	<b>-0,52</b>	250	0,000 ***
Körperliche Gesundheit (SF-12)	S	E	7	<b>0,28</b>	705	0,000 ***
	S	K1J	6	<b>0,41</b>	81	0,000 ***
Psychische Gesundheit (SF-12)	S	E	7	<b>0,21</b>	705	0,000 ***
	S	K1J	6	<b>0,49</b>	81	0,000 ***
ISR - Zusatzskala	S	E	9	<b>-0,60</b>	3439	0,000 ***
ISR - Gesamtscore	S	E	9	<b>-0,61</b>	3439	0,000 ***
Anforderungen erfüllen (ICF AT 50)	S	E	9	<b>-0,63</b>	3442	0,000 ***
Beeinträchtigung (ICF AT 50)	S	E	9	<b>-0,62</b>	3442	0,000 ***

Typ: EI = Einzelitem (single-item-rating), S = Skala; Zeitpunkt der Erhebung: E = Entlassung, K6M = Katamnese 6 Monate nach E, K1J = Katamnese 1 Jahr nach E, K2J = Katamnese 2 Jahre nach E, K5J = Katamnese 5 Jahre nach E; r = Produkt-Moment-Korrelation nach Pearson; N = Stichprobengröße; p = Signifikanz (2-seitig).

Die relativ höchsten Zusammenhänge können zwischen BESS und dem zeitgleich erfassten gesundheitlichen Befinden (GB-10, bis  $r = 0,79$ ) und dem Single-Item-Rating Gesundheitszustand (bis  $r = 0,77$ ) beobachtet werden: Patienten, die viele positive Veränderungen berichten, schildern zeitgleich ein besseres gesundheitliches Befinden und einen besseren Gesundheitszustand und umgekehrt.

Deutliche Zusammenhänge zeigen sich auch zwischen BESS und Wohlbefinden (bis  $r = \pm 0,71$ ), Depressivität (bis  $r = -0,68$ ), Beschwerden (bis  $r = -0,66$ ), Selbstwirksamkeit (bis  $r = -0,69$ ) und den geschilderten Beeinträchtigungen (bis  $r = -0,60$ ).

#### 4.4.5 Prädiktive Validität

##### 4.4.5.1 Vorhersagbarkeit der BESS-Skalenwerte aus Ausgangswerten

Die Zusammenhänge zwischen Statusmessungen bei Aufnahme (A) und den späteren BESS-Skalenwerten sind in Tabelle 12 zusammengefasst. Die 41 Korrelationen betreffen zwei Single-Item-Statusratings (Gesundheitszustand, Arbeitsfähigkeit) und 32 psychometrische Skalen. Die bei Aufnahme gemessenen Skalen bilden wiederum unterschiedliche Facetten der körperlichen und psychischen Gesundheit ab.

Die ermittelten Korrelationen sind im Durchschnitt deutlich geringer als diejenigen mit zeitgleichen Statusmessungen und variieren zwischen  $r = \pm 0,08$  und  $r = \pm 0,51$ . Alle Koeffizienten entsprechen der Erwartungsrichtung und sind nur teilweise signifikant. Knapp 60 % der Koeffizienten liegen im Bereich unterhalb von  $r = \pm 0,30$ , diese Zusammenhänge sind somit relativ schwach. Den relativ höchsten Vorhersagewert haben die initialen Messungen der HAQ-Skala „Erfolgszufriedenheit“ ( $r = 0,51$ : je höher, desto höher BESS), des SIBAR-Index ( $r = -0,45$ : je weniger sozialmedizinisch auffällig, desto höher BESS), des gesundheitlichen Befindens ( $r = 0,38$ : je besser, desto höher BESS) und der Single-Item-Ratings „Gesundheitszustand“ ( $r = 0,40$ : je besser, desto höher BESS) und „Arbeitsfähigkeit“ ( $r = -0,42$ : je besser, desto höher BESS).

##### 4.4.5.2 Vorhersagevalidität der BESS-Skalenwerte

Zur Überprüfung des Vorhersagewerts der BESS-Skalenwerte bei Entlassung für zeitlich später erhobene Bewertungen wurden vier Parameter aus den Nacherhebungen herangezogen:

- die Einschätzung der Zielerreichung,
- die Nutzenbewertung,
- die Patientenzufriedenheit (ZUF-8) und

- das gesundheitliche Befinden (GB10).

Die ermittelten Zusammenhänge sind in Tabelle 13 zusammengefasst.

**Tabelle 12: Vorhersagbarkeit der Skala BESS (Messung bei E bzw. K1J) aus Statusmessungen bei Aufnahme (A)**

Merkmale, gemessen bei Aufnahme	Typ	Studie	MZP BESS	r	N	p
Gesundheitliches Befinden (GB-10)	S	9	E	<b>0,38</b>	3452	0,000 ***
Gesundheitszustand (1-6, 6 = sehr gut)	EI	9	E	<b>0,40</b>	3452	0,000 ***
Arbeitsfähigkeit (1-4, 1 = überhaupt nicht arbeitsfähig)	EI	9	E	<b>-0,42</b>	3452	0,000 ***
Beschwerden (B-L)	S	3	E	<b>-0,13</b>	258	0,036 *
	S	4	E	<b>-0,10</b>	253	0,087
Beschwerdedruck (GBB)	S	5	E	<b>-0,12</b>	147	0,151
Somatisierung (SCL-90-R)	S	3	E	<b>-0,12</b>	248	0,057
Somatoformes Syndrom (ISR)	S	9	E	<b>-0,26</b>	3450	0,000 ***
Somatoforme Beschwerden (HEALTH-49)	S	9	E	<b>-0,34</b>	3452	0,000 ***
		8	K1J	<b>-0,28</b>	138	0,001 **
Depressivität (SCL-90-R)	S	3	E	<b>-0,08</b>	248	0,188
Depression (HADS-D)	S	6	K1J	<b>-0,27</b>	111	0,004 **
Depressivität (HEALTH-49)	S	9	E	<b>-0,30</b>	3452	0,000 ***
	S	8	K1J	<b>-0,30</b>	138	0,000 ***
Depressives Syndrom (ISR)	S	9	E	<b>-0,31</b>	3450	0,000 ***
Depressivität (D-S)	S	4	E	<b>-0,12</b>	253	0,052
Depression (BDI II)	S	9	E	<b>-0,34</b>	3447	0,000 ***
Depression (TSD)	S	5	E	<b>-0,12</b>	158	0,116
Ängstlichkeit (SCL-90-R)	S	3	E	<b>-0,09</b>	248	0,166
Angst (HADS-D)	S	6	K1J	<b>-0,16</b>	111	0,105
Phobische Angst (SCL-90-R)	S	3	E	<b>-0,09</b>	247	0,143
Phobische Ängste (HEALTH-49)	S	9	E	<b>-0,33</b>	3452	0,000 ***
	S	8	K1J	<b>-0,34</b>	138	0,000 ***
Verstimmungsstörungen (VE)	S	1	E	<b>-0,08</b>	403	0,121
Psychisches Wohlbefinden (HEALTH-49)	S	9	E	<b>-0,25</b>	3452	0,000 ***
	S	8	K1J	<b>-0,04</b>	138	0,615
Wohlbefinden (MFHW)	S	6	K1J	<b>0,35</b>	105	0,000 ***
Selbstwirksamkeit (HEALTH-49)	S	9	E	<b>-0,27</b>	3452	0,000 ***
	S	8	K1J	<b>-0,11</b>	138	0,222
Psychotizismus (SCL-90-R)	S	3	E	<b>-0,10</b>	247	0,108
Globalwert GSI (SCL-90-R)	S	3	E	<b>-0,11</b>	258	0,075
Körperliche Gesundheit (SF-12)	S	6	K1J	<b>0,27</b>	83	0,012 *
Psychische Gesundheit (SF-12)	S	6	J1J	<b>0,23</b>	83	0,035 *
ISR - Zusatzskala	S	9	E	<b>-0,36</b>	3450	0,000 ***
ISR - Gesamtscore	S	9	E	<b>-0,35</b>	3450	0,000 ***
Anforderungen erfüllen (ICF AT 50)	S	9	E	<b>-0,38</b>	3450	0,000 ***
Beeinträchtigungsindex (ICF AT 50)	S	9	E	<b>-0,38</b>	3450	0,000 ***
SIBAR-Index	S	9	E	<b>-0,45</b>	1746	0,000 ***
	S	7	E	<b>-0,11</b>	459	0,018 *
HAQ - Beziehungszufriedenheit	S	9	E	<b>0,30</b>	2442	0,000 ***
HAQ - Erfolgswzufriedenheit	S	9	E	<b>0,51</b>	2442	0,000 ***

Typ: EI = Einzelitem (single-item-rating), S = Skala; Zeitpunkt der Erhebung: E = Entlassung, K6M = Katamnese 6 Monate nach E, K1J = Katamnese 1 Jahr nach E, K2J = Katamnese 2 Jahre nach E, K5J = Katamnese 5 Jahre nach E; r = Produkt-Moment-Korrelation nach Pearson; N = Stichprobengröße; p = Signifikanz (2-seitig).

**Tabelle 13: Prognostische Validität der BESS-Skalenwerte bei Entlassung (E) bezüglich zeitlich später erhobener Outcome-bezogener Merkmale**

Datenquelle	Kriteriumsvariable	MZP	r <sub>12</sub>	N	p
7	Einschätzung der Zielerreichung (1-5, 5 = vollständig erreicht)	K6M	<b>0,50</b>	652	0,000 ***
7		K1J	<b>0,41</b>	599	0,000 ***
7		K2J	<b>0,41</b>	531	0,000 ***
7	Nutzenbewertung der Behandlung (1-4, 4 = großer Nutzen)	K6M	<b>0,54</b>	652	0,000 ***
2		K1J	<b>0,35</b>	276	0,000 ***
3		K1J	<b>0,46</b>	185	0,000 ***
4		K1J	<b>0,56</b>	192	0,000 ***
7		K1J	<b>0,42</b>	596	0,000 ***
7		K2J	<b>0,45</b>	534	0,000 ***
7	Zufriedenheit mit der Behandlung (ZUF-8)	K6M	<b>0,52</b>	652	0,000 ***
7		K1J	<b>0,43</b>	600	0,000 ***
7		K2J	<b>0,42</b>	537	0,000 ***
7	Gesundheitliches Befinden (GB-10)	K6M	<b>0,25</b>	655	0,000 ***
7		K1J	<b>0,22</b>	604	0,000 ***
7		K2J	<b>0,16</b>	532	0,000 ***

MZP = Messzeitpunkt der Erhebung der Kriteriumsvariable: K6M = Katamnese 6 Monate nach Entlassung / K1J = 1-Jahres-Katamnese / K2J = 2-Jahres-Katamnese / K5J = 5-Jahres-Katamnese; r = Produkt-Moment-Korrelation nach Pearson; N = Stichprobengröße; p = Signifikanz (2-seitig).

Tabelle 13 zeigt, dass BESS hochsignifikant und in moderater Höhe mit der zeitlich späteren Einschätzung der Zielerreichung der Reha (bis  $r = 0,50$ : je höher BESS, desto besser die spätere Zielerreichungseinschätzung), der zeitlich späteren Nutzenbeurteilung der Reha (bis  $r = 0,56$ : je höher BESS, desto besser die spätere Nutzenbewertung) und der späteren Patientenzufriedenheit (ZUF-8) korreliert (bis  $r = 0,52$ : je höher BESS, desto höher die spätere Zufriedenheit). Die Beziehung zum späteren gesundheitlichen Befinden (GB10) ist zwar ebenfalls hochsignifikant, in der Größenordnung jedoch schwächer (bis  $r = 0,25$ : je höher BESS, desto besser das spätere gesundheitliche Befinden).

#### 4.4.6 Diskriminante Validität - Gruppenunterschiede

Es wurde weiterhin überprüft, ob die Skala BESS bei Vorliegen unterschiedlicher Ausprägungen bestimmter Merkmale statistisch bedeutsame Differenzierungen zeigt, die theoretisch und/oder klinisch in der Erwartungsrichtung liegen. Hierzu haben wir die Variablen

- „beruflicher Status vor der Reha“ (Erwerbsstatus; 1-9),
- „Einschätzung der Arbeitsfähigkeit zu Beginn der Reha (1-4)“,
- „Sozialmedizinischer Risikoindex (SIBAR; 1-3),
- „globale Schweregradeinschätzung zu Beginn der Reha“ (Fremdbeurteilung; 1-4) und
- „Rentenbegehren“ (Fremdbeurteilung; 0-2)

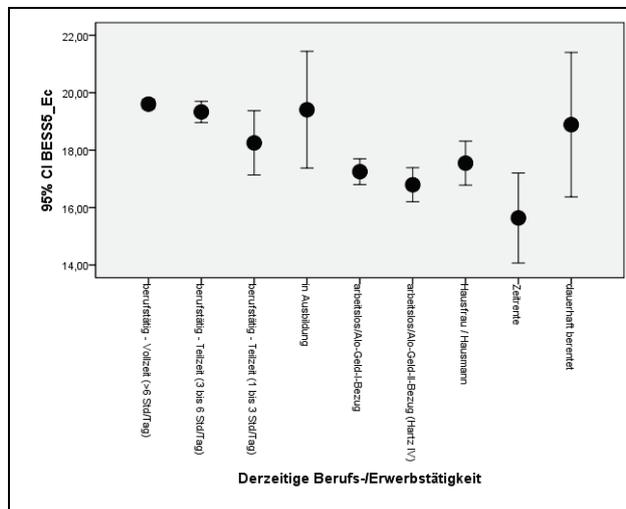
aus Studie 9 herangezogen und die BESS-Mittelwerte bei Entlassung der jeweiligen Untergruppen miteinander verglichen. Abbildung 5 zeigt, dass sich die BESS-Mittelwerte verschiedener Patientengruppen, die man aufgrund ihres Erwerbsstatus vor der Reha bilden

kann, signifikant unterscheiden ( $F = 27,206$ ,  $p < .001$ ; Kruskal-Wallis-Test:  $\chi^2 = 192,23$ ,  $df = 8$ ,  $p < .001$ ).

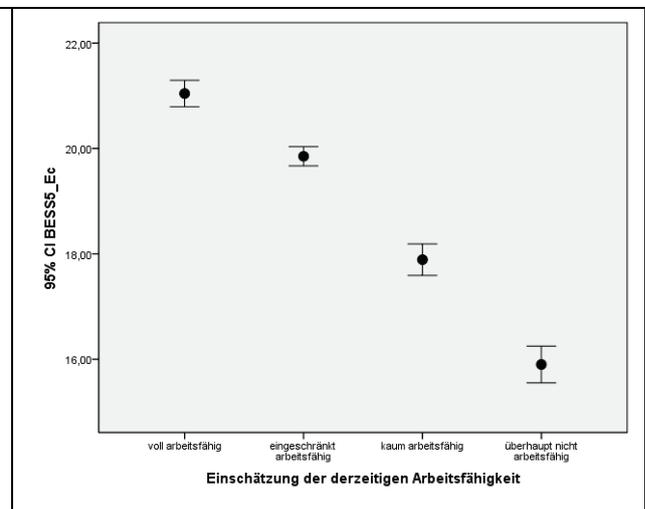
Die im Durchschnitt höchsten BESS-Skalenwerte haben bei Entlassung die Vollzeit-Berufstätigen ( $M = 19,6$ ), die vergleichsweise geringsten weisen Zeitrentner ( $M = 15,6$ ), Hartz IV- Bezieher ( $M = 16,8$ ) und Arbeitslose mit Alo-Geld I-Bezug ( $M = 17,2$ ) auf.

Abbildung 6 verdeutlicht, dass sich die BESS-Mittelwerte von vier Patientengruppen, die man aufgrund der Selbsteinschätzung der Arbeitsfähigkeit bilden kann, signifikant unterscheiden ( $F = 250,48$ ,  $p < .001$ ; Kruskal-Wallis-Test:  $\chi^2 = 597,60$ ,  $df = 3$ ,  $p < .001$ ).

**Abbildung 5: Erwerbsstatus bei Aufnahme - BESS-Mittelwerte der unterschiedlichen Patientengruppen (N = 3452)**



**Abbildung 6: Einschätzung der Arbeitsfähigkeit bei Aufnahme - BESS- Mittelwerte der unterschiedlichen Patientengruppen (N = 3452)**



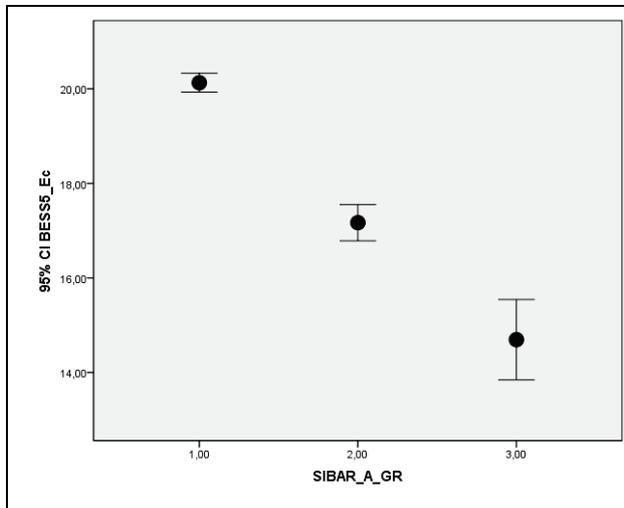
Die durchschnittlich höchsten BESS-Skalenwerte bei Entlassung können bei denjenigen Patienten beobachtet werden, die sich zu Beginn der Reha als „voll arbeitsfähig“ bezeichneten ( $M = 21,0$ ), die vergleichsweise geringsten zeigen hingegen diejenigen Patienten, die sich als „überhaupt nicht arbeitsfähig“ einstufen ( $M = 15,0$ ).

Abbildung 7 zeigt, dass sich die BESS-Mittelwerte von drei Patientengruppen, die man aufgrund ihrer SIBAR-Skalenwerte bei Aufnahme bilden kann, signifikant unterscheiden ( $F = 165,017$ ,  $p < .001$ ; Kruskal-Wallis-Test:  $\chi^2 = 266,16$ ,  $df = 2$ ,  $p < .001$ ). Gruppe 1 sind Patienten mit SIBAR-Werten von 0 - 7 (in Studie 9: 71,2 %), Gruppe 2 sind Patienten mit SIBAR-Werte von 8 – 15 (23,7 %) und Gruppe 3 Patienten mit SIBAR-Werten  $> 15$  (5,1 %). Ein deutlich erhöhtes Frühberentungsrisiko und ein berufsbezogener Behandlungsbedarf besteht laut Bürger et al. (2007), wenn der Punktwert bei mindestens 8 Punkten liegt. Die durchschnittlich höchsten BESS-Skalenwerte bei Entlassung haben Patienten der Gruppe 1 ( $M = 20,1$ ), also die sozialmedizinisch eher wenig auffälligen Patienten. Die niedrigsten

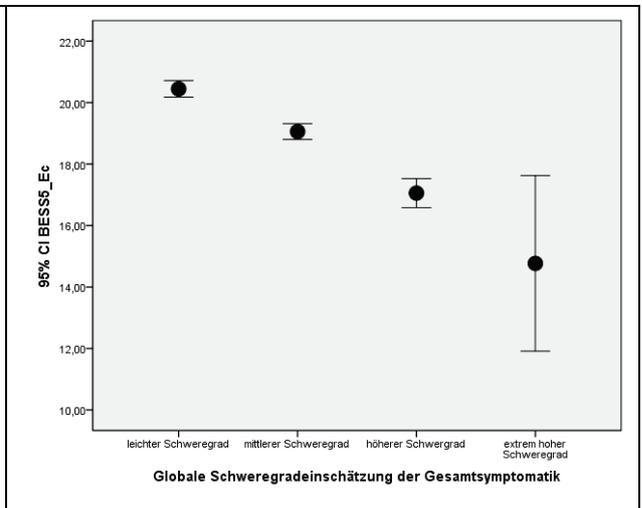
BESS-Werte können bei Patienten der Gruppe 3 (M = 14,7) beobachtet werden, also bei denjenigen mit dem höchsten sozialmedizinischen Risiko.

Abbildung 8 zeigt, dass sich die BESS-Mittelwerte verschiedener Patientengruppen, die man aufgrund von Arztbeurteilungen des Schweregrads der Gesamtsymptomatik zu Beginn der Reha bilden kann, signifikant unterscheiden (F = 60,088, p < .001; Kruskal-Wallis-Test: Chi<sup>2</sup> = 158,59, df = 3, p < .001). Die durchschnittlich höchsten BESS-Werte bei Entlassung zeigen Patienten mit leichtem Schweregrad (M = 20.4), die niedrigsten BESS-Werte können bei Patienten mit extrem hohen Schweregrad beobachtet werden (M = 14,7).

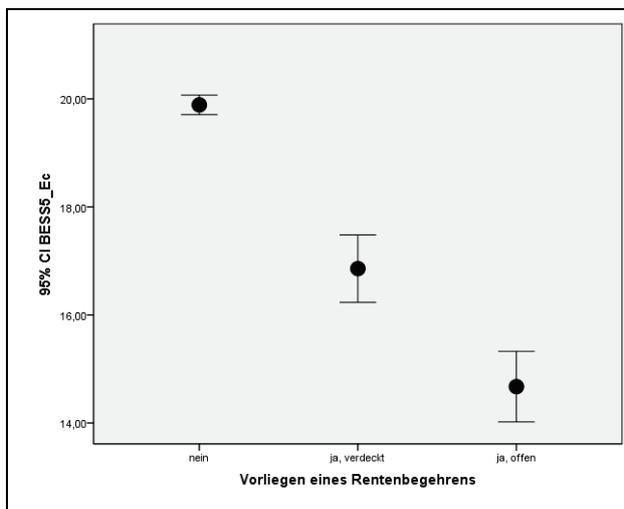
**Abbildung 7: Sozialmedizinischer Risikoindex (SIBAR) bei Aufnahme - BESS-Mittelwerte unterschiedlicher Patientengruppen (N = 1702)**



**Abbildung 8: Schweregradeinschätzung zu Beginn der Reha (Fremdbeurteilung) – BESS-Mittelwerte unterschiedlicher Patientengruppen (N = 1702)**



**Abbildung 9: Vorliegen eines Rentenbegehrens (Fremdbeurteilung) - BESS-Mittelwerte unterschiedlicher Patientengruppen (N = 1945)**



In Abbildung 9 werden bezüglich des Vorliegens eines Rentenbegehrens (= Arztbeurteilungen zu Beginn der Reha) drei Patientengruppen unterschieden: Patienten ohne Rentenbegehren, Patienten mit einem verdeckten Rentenbegehren und Patienten mit offenem Rentenbegehren. Abbildung 9 zeigt, dass sich die BESS-Mittelwerte der Gruppen bei Entlassung signifikant unterscheiden ( $F = 177,206$ ,  $p < .001$ ; Kruskal-Wallis-Test:  $\text{Chi}^2 = 257,06$ ,  $df = 2$ ,  $p < .001$ ).

## 4.5 Referenzwerte

Vergleichswerte in Form von Mittelwerten und Standardabweichungen der Skala BESS für die Hauptindikationen Psychosomatik, Orthopädie, Kardiologie sowie für Reha-Patienten insgesamt findet man in Tabelle 14. Die Vergleichswerte wurden insbesondere auf Grundlage der großen Stichproben aus den kontinuierlichen Patientenbefragungen B und C (Datensätze 11 und 12) ermittelt, für die Psychosomatik wurden zusätzlich 5 weitere Datenerhebungen herangezogen.

**Tabelle 14: Referenzwerte für die Skala BESS – stationäre Rehabilitation - Skalenwerte bei Entlassung**

	Studie	Indikation	M	SD	N
1	11	<b>Psychosomatik</b>	<b>19,69</b>	3,77	16323
2	12		<b>19,76</b>	3,92	15185
3	9		<b>18,98</b>	4,38	9352
4	1-4		<b>19,69</b>	3,68	1167
5	11	<b>Orthopädie</b>	<b>20,33</b>	3,71	19448
6	12		<b>20,19</b>	3,57	28968
7	11	<b>Kardiologie</b>	<b>21,77</b>	2,91	7211
8	12		<b>20,71</b>	3,47	7311
9	11	<b>Onkologie</b>	<b>20,79</b>	3,17	3603
10	12		<b>20,45</b>	3,15	30535
11	11	<b>Reha-Patienten</b>	<b>20,42</b>	3,63	51386
12	12		<b>20,25</b>	3,49	87619

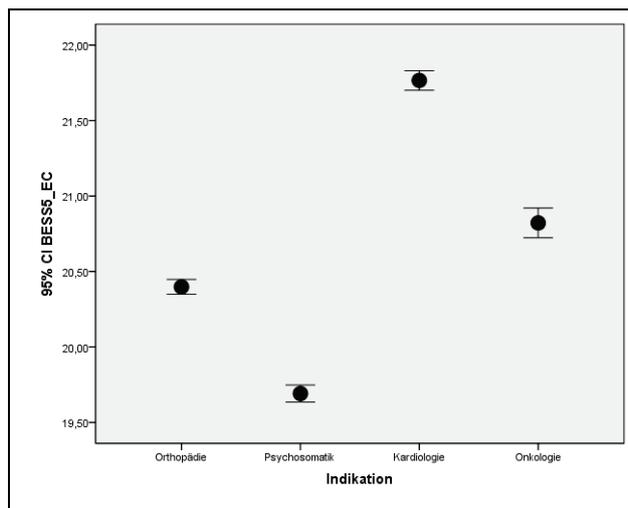
M = Mittelwert; SD = Standardabweichung

In der Patientenbefragung B (vgl. Abbildung 10) können signifikante BESS-Mittelwertsunterschiede zwischen den vier Indikationen Orthopädie, Psychosomatik, Kardiologie und Onkologie beobachtet werden ( $F = 619,569$ ,  $p < 0,001$ ). Die höchsten durchschnittlichen BESS-Skalenwerte gibt es in der Kardiologie ( $M = 21,7$ ), die niedrigsten in der Psychosomatik ( $M = 19,7$ ).

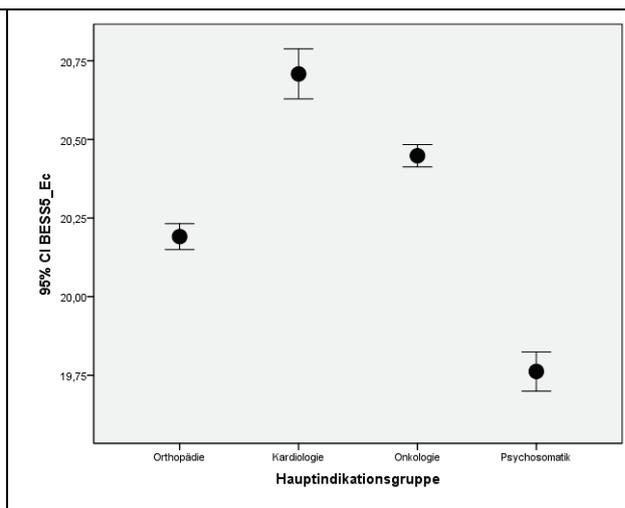
Ein analoges Ergebnis zeigt sich auch in Patientenbefragung C (vgl. Abbildung 111). Auch in dieser Befragung gibt es zwischen den vier Indikationen signifikante BESS-Mittelwerts-

unterschiede ( $F = 177,116$ ,  $p < 0,001$ ). Die im Durchschnitt niedrigsten BESS-Skalenwerte zeigen sich wiederum für die Psychosomatik vor ( $M = 19,8$ ). Auch hier liegt der höchste BESS-Mittelwert für die Kardiologie vor ( $M = 20,7$ ), wobei dieser Mittelwert allerdings um etwa 1 Skalenpunkt unter dem in der Patientenbefragung B liegt.

**Abbildung 10: BESS-Mittelwerte der Indikationen Psychosomatik, Orthopädie, Kardiologie, Onkologie in der Patientenbefragung B (Datensatz 11) N=50342**



**Abbildung 11: BESS-Mittelwerte der Indikationen Psychosomatik, Orthopädie, Kardiologie, Onkologie in der Patientenbefragung C (Datensatz 12) N = 81999**



## 4.6 Gebesserte und nichtgebesserte Patienten (kritische Werte)

Bei Prä/Post-Vergleichen kann man anhand der Skalenstreuung der Prä-Messung und der Reliabilität eine Mindestdifferenz berechnen, welche erreicht werden muss, damit die beobachtete Veränderung bei einem einzelnen Patienten mit 95%-tiger Wahrscheinlichkeit nicht mehr durch Messfehler erklärbar ist (Schmitz & Davies-Osterkamp, 1997; Kordy, 1997; Jacobson & Truax, 1991). Hat man es mit direkten Veränderungseinstufungen im Sinne von „verbessert – unverändert – verschlechtert“ zu tun, so lässt sich ein Konfidenzintervall um den theoretisch zu erwartenden Wert der „Nichtveränderung“ berechnen (vgl. Schmidt et al., 2003; Zielke & Kopf-Mehnert, 2001a). Die Grenzen des Konfidenzintervalls (CI) berechnen sich dann nach Gleichung 1.

$$(1) CI = X_0 \pm z SD \sqrt{(1 - r_{tt})}$$

wobei

$X_0$  = Wert der theoretischen Nicht-Veränderung,  $z$  = kritischer z-Wert,  $SD$  = Standardabweichung und  $r_{tt}$  = Reliabilität.

Bei direkten Veränderungsmessungen (DVM) werden ja keine Prä/Post-Differenzen gebildet, weshalb eine andere Berechnungsformel als bei der indirekten Veränderungsmessung zu

wählen ist. Als Streuungsmaß wird die Standardabweichung der DVM verwendet, als Reliabilitätsmaß lässt sich die interne Konsistenz der DVM-Skala einsetzen. Bei einer tolerierten Irrtumswahrscheinlichkeit von 5% ist auch hier für  $z$  der Wert 1,96 einzusetzen. Patienten innerhalb des Konfidenzintervalls lassen sich nach dieser Vorschrift als „unverändert“ klassifizieren. Patienten mit Messwerten außerhalb der Intervallgrenzen gelten als gebessert. bzw. „verschlechtert“.

Wenngleich die Verwendung des Kriteriums der „kritischen“ Differenzen zur Unterscheidung zwischen gebesserten und nicht gebesserten Rehabilitanden unter dem Aspekt einer „fairen“ Bewertung sowie Sensitivitäts- und Spezifitätsgesichtspunkten wegen seiner extrem hohen Anforderungen an die Reliabilität der jeweils verwendeten Skala durchaus diskutabel ist, hat wir es auch bei der BESS-Skala verwendet, um differenzielle Besserungsquoten aufzuzeigen.

Zur Bestimmung der kritischen Differenz wurden BESS-Kennwerte der Patientenbefragung C herangezogen:  $X_0 = 15$ ,  $SD = 3,48$  und  $\alpha = 0,90$ . Eine Nullveränderung wird bei der Skala BESS angenommen, wenn alle Veränderungsaussagen mit einer „3“ markiert werden. Dies entspricht einem Skalenwert von  $5 \times 3 = 15$  Punkten. Bei Anwendung von Formel 1 ergeben sich folgende Konfidenzintervalle:

- auf dem 5 % - Niveau:  $CI = 15 \pm 2,16 \rightarrow [12,84 - 17,16]$
- auf dem 1 % -Niveau:  $CI = 15 \pm 2,84 \rightarrow [12,16 - 17,84]$
- auf dem 0,1 % -Niveau:  $CI = 15 \pm 3,62 \rightarrow [11,38 - 18,62]$ .

Aufgrund dieser Cut-off-Werte lassen sich 7 Veränderungsgruppen unterscheiden, wobei Gruppe 1 Verschlechterungen auf dem 0,1%-Niveau, Gruppe 4 Nullveränderungen und Gruppe 7 Besserungen auf dem 0,1%-Niveau beinhalten (vgl. Tabelle 14).

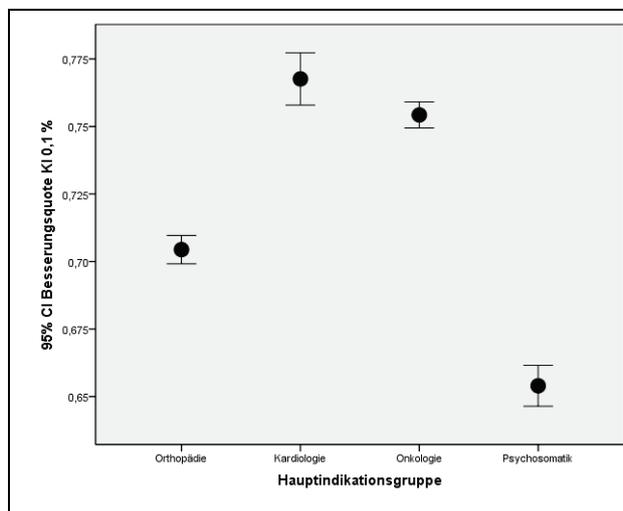
**Tabelle 14: Gebesserte, unveränderte und verschlechterte Reha-Patienten in der Patientenbefragung C (Skala BESS, Messung bei Entlassung)**

Veränderungsgruppen	N	%	Kumulierte %
1,00 verschlechtert auf dem 0,1 %-Nivea	1467	1,6	1,6
2,00 verschlechtert auf dem 1 %-Niveau	527	0,6	2,2
3,00 verschlechtert auf dem 5%-Niveau	57	0,1	2,2
4,00 unverändert	16341	17,8	20,0
5,00 gebessert auf dem 5%-Niveau	522	0,6	20,5
6,00 gebessert auf dem 1%-Niveau	6925	7,5	28,1
7,00 gebessert auf dem 0,1%-Niveau	66204	71,9	100,0
Gesamt	92043	100,0	

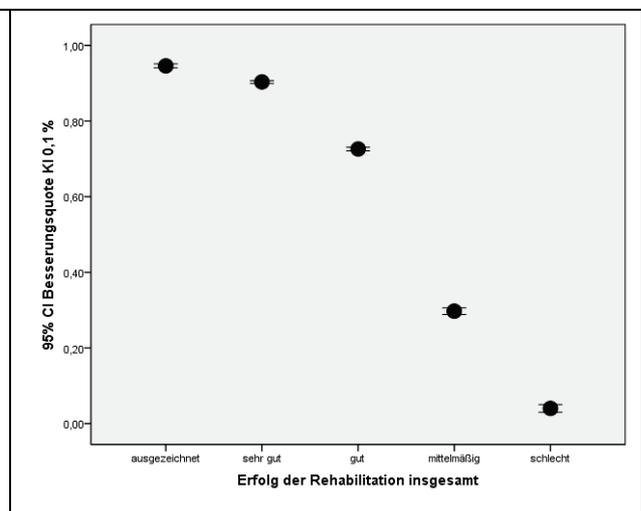
Tabelle 14 zeigt, dass zum Zeitpunkt der Entlassung bei insgesamt 80,0 % der Rehabilitanden (= Addition der Gruppen 5–7) eine signifikante Besserung des gesundheitlichen Befindens festgestellt werden kann. Nach Tabelle 14 können 2,2 % der Rehabilitanden als signifi-

kant verschlechtert (= Addition der Gruppen 1–3) und (mindestens) 17,8 % als „unverändert“ klassifiziert werden. Würde man hinsichtlich Besserung und Verschlechterung das härtere 0,1%-Niveau-Kriterium heranziehen, dann liegt die Besserungsrate der Reha-Patienten bei 71,9 %, 26,5 % sind unverändert und 1,6 % signifikant verschlechtert. Je nach angelegtem Signifikanzniveau liegen die beobachtbaren Besserungsquoten somit bei 80,0 % (5 %-Niveau), 79,5 % (1 %-Niveau) und 71,9 % (0,1 %-Niveau). Die oben berichteten BESS-Mittelwertsunterschiede bei Entlassung zwischen den untersuchten Indikationen spiegeln sich in entsprechend unterschiedlichen Besserungsraten.

**Abbildung 12: Differenzielle Besserungsraten (BQ\_4-Mittelwerte) bei den 4 Indikationen Orthopädie, Kardiologie, Onkologie, Psychosomatik (Skala BESS) –0,1 % Signifikanzniveau – Patientenbefragung C (N=81999)**



**Abbildung 13: Differenzielle Besserungsraten (BQ\_4-Mittelwerte) in Abhängigkeit zur Erfolgsbeurteilung – 0,1 % Signifikanzniveau – Patientenbefragung C (N=80344)**



BQ\_4 ist die Besserungsquote bei Anwendung des 0,1 %-Signifikanzniveaus. BQ\_4 kann Werte zwischen 0 (= 0% Gebesserte) und 1,0 (= 100 % Gebesserte) annehmen. Dargestellt sind die jeweiligen BQ\_4-Mittelwerte.

Abbildung 12 zeigt, dass bei Anwendung des strengen 0,1%-Niveaus die differenziellen Besserungsquoten zwischen 65,4 % (Psychosomatik) und 76,8 % (Kardiologie) variieren. Dazwischen liegen die Besserungsquoten für die Orthopädie (70,4 %) und für die Onkologie (75,4 %). Ähnlich kann gezeigt werden, dass sich der Patientenbefragung C (Studie 12) gefundene deutliche Zusammenhang zwischen BESS und Erfolgsbewertung ( $r = 0,57$ ; vgl. Tabelle 8), in entsprechend unterschiedlichen Besserungsquoten widerspiegelt.

Abbildung 13 verdeutlicht, dass Reha-Patienten, die den Erfolg der Reha bei Entlassung als „ausgezeichnet“ oder „sehr gut“ bewerten Besserungsquoten von 94,6 % bzw. 90,3 % aufweisen. Bei der Bewertung „gut“ kann eine Besserungsquote von 72,6 %, bei „mittelmäßig“ eine Besserungsquote von 29,7 % und bei „schlecht“ eine Besserungsquote von 0,4 % beobachtet werden.

## 5 Zusammenfassung und Bewertung

Auf der Basis einer Reanalyse von insgesamt 12 Studien bzw. Datenerhebungen werden Ergebnisse umfangreicher psychometrischer Überprüfungen der BESS-Skala<sup>1</sup> (Selbstbeurteilungsform) vorgestellt. Die 12 Studien bzw. Datenerhebungen beinhalten 1 – 4 Erhebungen der Skala BESS (10 Mal bei Entlassung und 10 Mal in zeitlich unterschiedlichen Nacherhebungen) bei etwa 160.000 Reha-Patienten.

Die Item- und Skalenanalyse zeigen durchweg linksschiefe (rechtssteile) Verteilungen der Skalenwerte, d.h. höhere Skalenwerte sind häufiger als niedrige. Zum Zeitpunkt der Entlassung können im Durchschnitt höhere Skalenwerte als in den Katamnesen beobachtet werden. Durchgängig kann eine hohe interne Konsistenz der Skala ( $\alpha = 0,86$  -  $\alpha = 0,94$ ) ermittelt werden, die sehr zufriedenstellend ist. Insgesamt legen alle Hauptkomponentenanalysen eine 1-Faktoren-Lösung nahe, die etwa 65 – 81 % der Gesamtvarianz erklären kann. Alle 5 BESS-Items haben durchgängig hohe Ladungen auf diesem Generalfaktor, die mehrheitlich etwa im Bereich zwischen 0,78 und 0,90 liegen.

Zwischen BESS-Messungen bei Entlassung und in Nachbefragungen zeigen sich moderate signifikante Korrelationen im Bereich zwischen  $r = 0,32$  und  $r = 0,45$ . In 7 von 8 Vergleichen gibt es signifikante Mittelwertsunterschiede. Die Effektstärken des Unterschieds variieren zwischen  $ES = 0,24$  und  $ES = 0,61$  und belegen eine akzeptable Änderungssensitivität der Skala.

Erwartungskonform korreliert das Ergebnismaß BESS stringent mit anderen Outcome-Maßen. So korrelieren die BESS-Skalenwerte deutlich mit anderen ergebnisbezogenen Selbstbeurteilungen der Patienten, z.B. mit der Patientenzufriedenheit (ZUF-8; bis  $r = 0,58$ ), mit Nutzenbewertungen (bis  $r = 0,68$ ), mit Zielerreichungseinschätzungen (bis  $r = 0,66$ ) und mit der Erfolgs- bzw. Gesamtbewertung der Reha (bis  $r = 0,63$ ). Ein deutlicher Zusammenhang kann auch zwischen BESS und der HAQ-Skala „Erfolgszufriedenheit“ beobachtet werden (bis  $r = 0,69$ ). Hochsignifikante Zusammenhänge zeigen sich auch zwischen BESS und Arztbeurteilungen der Zielerreichung ( $r = 0,54$ ) und des Ergebnisses ( $r = 0,51$ ).

Weiterhin liegen zum Teil hochsignifikante und substantielle Zusammenhänge zwischen BESS und anderen direkten Veränderungsinformationen (Einzelitems und Skalen nach dem DVM-Prinzip) vor. BESS korreliert z. B. hoch mit den Single-Item-Ratings „Veränderung des Gesundheitszustand“ ( $r = 0,76$  bis  $r = 0,88$ ) und „Veränderung der Belastbarkeit“ ( $r = 0,84$  bis  $r = 0,85$ ). Sehr hohe Korrelationen zeigen sich auch zwischen BESS und den Veränderungsskalen VM-9 und VM-10 ( $r = 0,68$  bis  $r = 0,96$ ) sowie mit dem multiplen Ergebniskriterium

---

<sup>1</sup> Untersucht wurde die BESS-Skala in der Version für Patienten (Selbstbeurteilung). Daneben gibt es auch eine Fremdbeurteilungsform von BESS, die nicht Gegenstand dieser Reanalyse war.

EMEK-27 ( $r = 0,82$  bis  $r = 0,84$ ). Vergleichsweise eher im moderaten Bereich liegen die Korrelationen von BESS mit singulären Veränderungseinstufungen durch Ärzte.

Auch mit Veränderungsinformationen aus der indirekten Veränderungsmessung (IVM), operationalisiert durch zeitlich kohärente Differenzwerte (z. B. BESS, gemessen bei Entlassung vs. Differenzwerte zwischen Aufnahme und Entlassung), zeigen sich vielfältige hochsignifikante Korrelationen, die jedoch durchschnittlich etwas geringer ausfallen als diejenigen mit direkten Veränderungsmaßen (von 42 Korrelationskoeffizienten liegen zwei 2/3 im Bereich zwischen  $r = \pm 0,31$  –  $r = \pm 0,50$ ). Die vergleichsweise höchsten Korrelationen zeigen sich zwischen BESS und GB-10-Differenzwerten (bis zu  $r = - 0,67$ ).

Hohe plausible Zusammenhänge weist BESS mit zeitgleichen Status-Messungen auf: Die höchsten Korrelationen findet man zwischen BESS und dem zeitgleich erfassten gesundheitlichen Befinden (Skala GB-10, bis  $r = 0,79$ ) und der Einschätzung des „Gesundheitszustands“ (bis  $r = 0,77$ ; Single-Item-Rating).

Deutlich geringer korreliert BESS hingegen mit Statusmessungen zu Beginn der stationären Rehabilitation. Knapp 60 % der ermittelten Koeffizienten liegen hier im Bereich unterhalb von  $r = \pm 0,30$ . Den höchsten Vorhersagewert haben die initialen Messwerte in der HAQ-Skala „Erfolgszufriedenheit“ ( $r = 0,51$ : je höher zu Beginn, desto höher BESS), der SIBAR-Wert ( $r = - 0,45$ : je weniger sozialmedizinisch auffällig zu Beginn, desto höher BESS) und das gesundheitliche Befinden ( $r = 0,38$ : je besser zu Beginn, desto höher BESS).

BESS verfügt weiterhin über eine moderate Vorhersagevalidität (BESS korreliert z. B. mit zeitlich späteren Zielerreichungs- und Nutzenbewertungen im Bereich von  $r = 0,50$ ) und es zeigen sich plausible Gruppenunterschiede in der Ausprägung der BESS-Skala (diskriminative Validität).

Zusammenfassend zeigen die vorgestellten Befunde zur BESS-Skala (5 Items, Skalenspanne: 5 – 25) somit,

- dass es sich faktorenanalytisch um eine eindimensionale Skala handelt,
- dass die Skala eine hohe interne Konsistenz besitzt,
- BESS substantiell mit umfangreicheren zeitgleich erhobenen selbstberichteten direkten Veränderungsmaßen korreliert,
- BESS ebenfalls substantiell – allerdings moderater - mit korrespondierenden indirekten patientenseitigen Veränderungsmessungen (Differenzwerten) korreliert,
- BESS deutlich mit zeitgleichen Ergebnisbewertungen der Patienten (z. B. Nutzen, Zielerreichung) und der Zufriedenheit mit der Behandlung korreliert.
- BESS substantiell und kohärent mit zeitgleichen Statusmessungen (z. B. Gesundheitszustand, Wohlbefinden) korreliert,

- BESS größtenteils eher moderat bis geringfügig mit Statusmessungen korreliert, die zu Beginn der stationären Rehabilitation erhoben werden. Die Daten deuten jedoch darauf hin, dass höhere BESS-Werte eher mit einem besseren Gesundheitszustand zu Beginn der Reha einhergehen bzw. dass initial stärker beeinträchtigte Patienten eher niedrigere BESS-Werte aufweisen.

In der Gesamtschau erfüllt die eindimensionale (direkte) Veränderungsskala BESS die Hauptgütekriterien Objektivität, Reliabilität bzw. interne Konsistenz und Validität in zufriedenstellender Weise. BESS ist weiterhin auch änderungssensitiv, da sich im zeitlichen Verlauf deutliche Unterschiede zeigen. Als Veränderungsmaß („je höher der Skalenwert ausfällt, desto größer ist im Erleben des Patienten die Verbesserung seines gesundheitlichen Befindens“) zeigt BESS eine zufriedenstellende Konstruktvalidität. Die Skala korreliert zum Teil sehr deutlich mit anderen direkten und indirekten Veränderungsmessungen. Hierbei fällt insbesondere ins Auge, dass diese sehr kurze Skala sehr hohe Korrelationen mit deutlich umfangreicheren Veränderungsskalen zeigt. Mit lediglich 5 Items ist BESS sehr ökonomisch und insbesondere dann als Ergebnismaß sehr nützlich, wenn für eine Datenerhebung nur ein Messzeitpunkt realisierbar ist. Auch liegen umfangreiche Vergleichswerte aus Patientenbefragungen mit fast 140000 Rehabilitanden unterschiedlicher Indikationen vor. Gerade wegen ihrer Ökonomie und ihrer umfangreichen Überprüfung eignet sich die BESS-Skala insbesondere sehr gut als globales Ergebnismaß Rahmen der Routine-Qualitätssicherung (z. B. Patientenbefragungen). Im Rahmen von evaluativen Mehr-Punkte-Erhebungen kann sie eine wertvolle und gut untersuchte Ergänzung zur indirekten Veränderungsmessung darstellen.

## 6 Literatur

- Baumann U, Sodemann U, Tobien H (1980): Direkte versus indirekte Veränderungsdiagnostik. Zeitschrift für Differentielle und Diagnostische Psychologie, 1, 201-216.
- Bereiter C (1962): Using tests to measure change. Personnel and Guidance Journal, 41, 6-11.
- Bereiter C (1963): Some persisting dilemmas in the measurement of change. In CW Harris (Ed.), Problems in measurement change (pp. 3-20). Madison, WI: University of Wisconsin Press.
- Bleißmann A (2004): Krankheitsverläufe bei chronischen Erkrankungen: Welche Methode der Veränderungsmessung eignet sich zu ihrer Beschreibung und Prognose? Dissertation. Universität Bielefeld.
- Bleißmann A, Kohlmann T, Raspe H (2002): Indirekte versus direkte Veränderungsmessung und ihre prognostische Bedeutung. 11. Rehabilitationswissenschaftliches Kolloquium 04.-06.03.2002 München. Tagungsband, 273-2751.
- Bürgy R, Nübling R, Meyerberg J, Oppl M, Kieser J, Schmidt J, Wittmann WW (2000): Stationäre psychosomatische Rehabilitation im Rahmen eines schulenübergreifenden Behandlungskonzepts: Ergebnisse der 1-Jahres-Katamnese der Bad Herrenalber Katamnese studie. In VDR (Hrsg), Individualität und Reha-Prozeß. DRV-Schriften, Band 20 (S. 374). Frankfurt, VDR.
- Fydrych T (2006): Diagnostik und Intervention in der Klinischen Psychologie. In M Amelang, L Schmidt-Atzert (Hrsg.), Psychologische Diagnostik und Intervention (4. Auflage, S. 495-532). Heidelberg: Springer.
- Jacobson NS, Truax P (1991): Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of consulting and Clinical psychology, 59, 12-19.
- Kastner S, Basler HD (1997): Messen Veränderungsfragebögen wirklich Veränderungen? Untersuchungen zur Erfolgsbeurteilung in der psychologischen Schmerztherapie. Der Schmerz, 11, 254-262.
- Kohlmann T, Raspe H (1998): Zur Messung patientennaher Erfolgskriterien in der medizinischen Rehabilitation: Wie gut stimmen „indirekte“ und „direkte“ Methoden der Veränderungsmessung überein? Rehabilitation, 37, 30-37.
- Kordy H 1997: Das Konzept der klinischen Signifikanz in der Psychotherapieforschung. In J Bengel, B Strauß (Hrsg.). Forschungsmethoden in der Medizinischen Psychologie 14 (S. 129-145). Göttingen: Hogrefe.
- Krampen G (2010a): Direkte Veränderungsmessung der Effekte von Entspannungstherapie/-training und Psychotherapie: Ihre Vor- und Nachteile sowie psychometrische Hilfsmittel. Entspannungsverfahren, 27, 48-64.
- Krampen G (2010b): Experimentelle Konstruktion eines Kurzfragebogens zur direkten Veränderungsmessung psychotherapeutischer Effekte im Befinden. Diagnostica, 56, Heft 4, 212-221.
- Krampen G, Delius A v (1981): Zur direkten Messung subjektiv erlebter gesundheitlicher Veränderungen. Medizinische Psychologie, 7, 166-174.
- Kriz D, Nübling R, Schmidt J (2014): Psychometrische Analyse einer kontinuierlichen Patientenbefragung in der medizinischen Rehabilitation. Band 103 (S. 185-187). Berlin: Deutsche Rentenversicherung Bund.
- Maatz E, Schmidt J (1998): Psychosomatische Rehabilitation von Patienten mit chronisch-entzündlichen Darmerkrankungen - Erste Ergebnisse der Gengenbacher CED-Studie. In: Verband Deutscher Rentenversicherungsträger (Hrsg): Interdisziplinarität und Vernetzung. DRV-Schriften Band 11. Frankfurt, VDR, 462-463.
- Meffert C, Kohlmann T, Raspe t et al. (2012): Verwendung von Verlaufstypen zur Veränderungsmessung in der medizinischen Rehabilitation: Ein Beitrag zum Vergleich verschiedener Methoden der Ergebnismessung. Rehabilitation, 51 (3): 153-161.
- Mittag O, Kohlmann T, Meyer T, Meyer-Moock S et al (2013): Empirisch gestützte Empfehlungen für die patientenberichtete Veränderungsmessung in der medizinischen Rehabilitation. Rehabilitation, 52, 119-125.
- Nübling R, Schmidt, J, Kriz D, Kobelt A, Bassler M (2015): Routine-Assessment in der psychosomatischen Rehabilitation – Behandlungsergebnisse auf der Grundlage eines EDV-gestützten Routine-Assessment-Systems. DRV-Schriften, Band 107 (S. 336-338). Berlin: Deutsche Rentenversicherung Bund.
- Nübling R, Schmidt J, Reichelt W, Kriz D (2013): Behandlungsergebnisse ein Jahr nach stationärer psychosomatischer Rehabilitation. Prävention und Rehabilitation, 25, 150-165.
- Nübling R, Steffanowski A, Körner M, Rundel M, Kohl CFR, Löschmann C (2007): Kontinuierliche Patientenbefragung als Instrument für das interne Qualitätsmanagement in Einrichtungen der Gesundheitsversorgung. Gesundheitsökonomie & Qualitätsmanagement, 12, 44-50.

- Nübling R, Steffanowski A, Wittmann WW, Schmidt J (2004): Strategien der Ergebnismessung am Beispiel der psychosomatischen Rehabilitation. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 65, 35-44.
- Nübling R, Bürgy R, Meyerberg J, Oppl M, Kieser J, Schmidt J, Wittmann WW. (2000a): Stationäre psychosomatische Rehabilitation im Rahmen eines schulübergreifenden Behandlungskonzepts. Erste Ergebnisse der 1 Jahres Katamnese der Bad Herrenalber Katamnesestudie. In M. Bassler (Hrsg), Leitlinien zur stationären Psychotherapie (S. 274-300). Gießen: Psychosozial-Verlag.
- Nübling R, Schmidt J (2000b): Methodische Grundlagen der Ergebnisevaluation. In J Bengel & U Koch (Hrsg.), *Grundlagen der Rehabilitationswissenschaft* (S. 324-346). Berlin: Springer.
- Nübling R, Puttendörfer J, Wittmann WW, Schmidt J, Wittich A. (1995): Evaluation psychosomatischer Heilverfahren. Ergebnisse einer Katamnese-Studie. *Die Rehabilitation*, 34, 65-118.
- Pohontsch N (2012): Gründe für die Diskordanz zwischen indirekter und direkter Messung der Veränderung des subjektiven Gesundheitszustands von orthopädischen und onkologischen Rehabilitanden – Eine qualitative Analyse. Inauguraldissertation. Universität Lübeck.
- Schmidt J, Kaiser U, Kriz D, Nübling R (2016): Psychische Komorbidität bei stationären Patienten mit Atemwegs- und Lungenkrankheiten – Korrelate und Verlauf. *Prävention und Rehabilitation*, 28 (3); 85-101.
- Schmidt J, Kaiser U, Kriz D, Nübling R (2014): Die Davoser Outcome-Studie (DOS) – Ergebnisse stationärer pneumologischer und dermatologischer Heilbehandlungen im Spiegel wiederholter Nacherhebungen. *Prävention und Rehabilitation*, 26 (4); 119-140.
- Schmidt J, Zucker A, Wöllstein K, Kriz D, Keck M, Nübling R (2012): Interdisziplinäre Schmerztherapie in der Orthopädischen Rehabilitation (ISOR) – eine Studie zur Inanspruchnahme und zu längerfristigen Ergebnissen eines multimodalen Behandlungsangebots für Patienten mit Schmerzsyndromen. *Prävention und Rehabilitation*, 24 (4); 132-150.
- Schmidt J, Nübling R (2010a): Der Fragebogen GB-10 zur Statusmessung des gesundheitlichen Befindens. Unveröffentlichtes Manuskript. Karlsruhe: Gesellschaft für Qualität im Gesundheitswesen (GfQG).
- Schmidt J, Nübling R (2010b): Der Fragebogen VM-10 zur direkten Veränderungsmessung des gesundheitlichen Befindens. Unveröffentlichtes Manuskript. Karlsruhe: Gesellschaft für Qualität im Gesundheitswesen (GfQG).
- Schmidt J, Nübling R, Steffanowski A, Kriz D, Wittmann WW (2008): Die Kurzsкала BESS zur direkten Veränderungsmessung - Zusammenfassende Befunde aus mehreren Studien. *DRV-Schriften*, Band 77 (S. 101-103). Berlin: Deutsche Rentenversicherung Bund.
- Schmidt J, Steffanowski A, Nübling R, Lichtenberg S, Wittmann WW (2003): Ergebnisqualität stationärer psychosomatischer Rehabilitation: Vergleich unterschiedlicher Evaluationsstrategien. Regensburg: Roderer.
- Schmidt J, Nübling R, Steffanowski A, Wittmann WW (2002): Evaluation der Effektivität psychosomatischer Rehabilitation. Wie gut stimmen echte und retrospektive Vorher-Nachher-Vergleiche überein. Ergebnisse der EQUA-Studie. 11. Rehabilitationswissenschaftliches Kolloquium 04.-06.03.2002 München. Tagungsband, 271-273.
- Schmidt J, Nübling R, Wittmann WW (2001): Ergebnisqualität psychosomatischer Rehabilitation. Zusammenhänge zwischen direkten und indirekten Veränderungsmessungen. 10. Rehabilitationswissenschaftliches Kolloquium 12.-14.03.2001 Halle. Tagungsband, 386-388.
- Schmidt J, Kieser J, Amann K, Nübling R, Johnen R, Lamprecht F, Wittmann WW (1996): Kurz- und längerfristige Veränderungen gesundheitsrelevanter Merkmalsbereiche (Risikofaktoren, Verhaltensweisen, Einstellungen) durch psychosomatische Rehabilitationsbehandlungen. In VDR (Hrsg), *Evaluation in der Rehabilitation*. DRV-Schriften Band 6 (S. 160-162). Frankfurt: VDR.
- Schmidt J, Nübling R, Wittmann WW (1995): Praktikable Möglichkeiten der Messung von Ergebnisqualität im Rahmen der Qualitätssicherung. Erfahrungen mit multiplen Ergebniskriterien im Bereich der psychosomatischen Rehabilitation. In Verband Deutscher Rentenversicherungsträger VDR (Hrsg.), *Zusammenarbeit von Forschung und Praxis* (S. 71-73). DRV-Schriften Band 5. Frankfurt/Main, VDR.
- Schmidt J, Nübling R, Lamprecht F, Wittmann WW (1994a): Patientenzufriedenheit am Ende psychosomatischer Reha-Behandlungen. Zusammenhänge mit Behandlungs- und Ergebnisvariablen und prognostische Bedeutung. In F Lamprecht, R Johnen (Hrsg): *Salutogenese – ein neues Konzept in der Psychosomatik?* Frankfurt: VAS (S. 271-283).
- Schmidt J, Lamprecht F, Nübling R, Wittmann WW (1994b): Veränderungsbeurteilungen von Patienten und von Haus- und Fachärzten nach psychosomatischer Rehabilitation – Ein katamnestischer Vergleich. *PPmP Psychother. Med. Psychol.*, 44, 108-114.

- Schmidt J, Lamprecht F (1992): Psychosomatische Rehabilitation. Ergebnisse von Verlaufsstudien. In Bundesarbeitsgemeinschaft für Rehabilitation (BAR; Hrsg.), Rehabilitation – Zukunft 2000. Essen. A. Sutter-Messe-Verlag (S. 261-267).
- Schmidt J, Nübling R (1988): Die Skala VM-9. Unveröffentlichtes Paper. Psychosomatische Klinik Schömburg.
- Schmidt J, Bernhard P, Wittmann WW, Lamprecht F (1987): Die Unterscheidung zwischen singulären und multiplen Ergebniskriterien – Ein Beitrag zur Kriterienproblematik in der Evaluation. In F Lamprecht (Hrsg.), Spezialisierung und Integration in Psychosomatik und Psychotherapie (S. 293-299). Berlin Heidelberg: Springer.
- Schmidt J, Nübling R (1986): Die Skala BESS. Unveröffentlichtes Paper. Psychosomatische Klinik Schömburg.
- Schmidt J (1985): Fragebogen zur direkten Veränderungsmessung. Unveröffentlichtes Paper. Psychosomatische Klinik Schömburg.
- Schmitz N, Davies-Osterkamp S (1997): Klinische und Statistische Signifikanz – diskutiert am Beispiel der Symptom Check Liste (SCL-90-R). Diagnostica, 43 (1), 80-96.
- Steffanowski A (2008): Evidenz psychosomatischer Rehabilitation im Spiegel multipler Ergebniskriterien. Inauguraldissertation. Universität Mannheim.
- Steffanowski A, Nübling R, Schmidt J, Löschmann C (2006): Patientenbefragungen in den Medizinischen Rehabilitation – Computergestütztes Routinemonitoring der Struktur-, Prozess- und Ergebnisqualität. PraxKlin-Verhaltensmed Rehabilitation, 71, 35-46.
- Stieglitz RD, Baumann U (2001): Veränderungsmessung. In RD Stieglitz, U Baumann, J Freyberger (Hrsg.), Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie (2. Auflage, S. 21-38). Stuttgart: Thieme.
- Stieglitz RD, Baumann U (1994): Veränderungsmessung. In RD Stieglitz, U Baumann (Hrsg.), Psychodiagnostik psychischer Störungen (S. 21-36). Stuttgart: Thieme.
- Wittmann WW, Schmidt J (2002): Varianten der Veränderungsmessung auf dem Prüfstand: Probleme der Konsistenz/Inkonsistenz und Validität von direkten, indirekten und quasi-indirekten Assessmentstrategien. 11. Rehabilitationswissenschaftliches Kolloquium 04.-06.03.2002 München. Tagungsband, 270-271.
- Zielke M, Kopf-Mehnert C. (2001a): Der VEV-R-2001: Entwicklungs- und testtheoretische Reanalyse der revidierten Form des Veränderungsfragebogens des Erlebens und Verhaltens (VEV). Praxis Klinische Verhaltensmedizin und Rehabilitation, 53, 7-19.
- Zielke M, Kopf-Mehnert C. (2001b): 22 Jahre wissenschaftliche und klinische Erfahrungen mit dem Veränderungsfragebogen des Erlebens und Verhaltens VEV. Praxis Klinische Verhaltensmedizin und Rehabilitation, 53, 3-6.
- Zielke M, Kopf-Mehnert C. (1978): Veränderungsfragebogen des Erlebens und Verhaltens (VEV). Weinheim: Beltz.

# 7 Anhang – Glossar Fragebögen

## Glossar der zur Validierung verwendeten Fragebogenverfahren

Zur Validierung der BESS-Skala wurden überprüfte psychometrische Skalen herangezogen, die in den verschiedenen Studien / Datenerhebungen 1 – 9 teilweise mehrfach verwendet wurden. Die Skalen bilden unterschiedliche Facetten der körperlichen und psychischen Gesundheit dar (z.B. Gesundheitliches Befinden, Beschwerden, Depressivität, Angst, Wohlbefinden), die zum Teil über verschiedene Verfahren erfasst wurden (Depressivität: Skalen aus SCL-90-R, HADS-D, ISR, HEALTH-49, D-S, TSD, BDI II). Einbezogen wurden weiterhin Skalen aus einem ICF-orientierten Fragebogen (ICF AT 50-Psych), eine sozialmedizinische Risikoskala (SIBAR), drei Veränderungsskalen auf der Basis direkter Veränderungseinstufungen (EMEK-27, VM-9, VM-10) und eine Skala zur Patientenzufriedenheit (ZUF-8).

Wesentliche Merkmale der Fragebögen bzw. Skalen (alphabetisch geordnet) sind nachfolgend kurz beschrieben.

### BDI-II - Beck-Depressions-Inventar

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

Selbstbeurteilungsverfahren (21 Items). Das Beck-Depressions-Inventar (BDI; Original: Beck, Ward, Mendelson, Mock & Erbaugh, 1961) ist ein psychologisches Testverfahren, das die Schwere depressiver Symptomatik im klinischen Bereich erfasst. Die Skala kann Werte zwischen 0 und 63 annehmen. BDI-II ist eine revidierte Version der Ursprungsversion des BDI aus dem Jahr 2006. Das Instrument wird heutzutage nicht nur in der Praxis, sondern auch in der Forschung als diagnostisches Mittel eingesetzt und international verwendet. Laut Manual zeigen Werte im Bereich 0 – 8 keine Depression und Werte im Bereich 29 – 63 eine schwere Depression an.

Bedeutung des Skalenwertes: hohe Werte = ausgeprägte depressive Symptomatik / niedrige Werte = keine bzw. geringfügige depressive Symptomatik.

### B-L - Die Beschwerdenliste

Zerssen D von (1976); Zerssen, D. v. (2002). **B-L, B-L' Beschwerden-Liste. In E Brähler, J Schumacher & B Strauß (Eds.), Diagnostische Verfahren in der Psychotherapie (pp. 56-59). Göttingen: Hogrefe.**

Selbstbeurteilungsverfahren (24 Items). Die Beschwerden - Liste (B-L) ist ein Fragebogentest, mit dem subjektive, körperliche Beschwerden quantitativ erfasst werden können. Der Test besteht aus 24 Fragen zu allgemeinen Beschwerden, die häufig von Patienten in Arztpraxen angegeben werden, wie z.B. Schwächegefühl, Übelkeit, Sodbrennen, Rückenschmerzen, Schlaflosigkeit, Zittern etc. Die Beschwerden werden vom Patienten nach ihrer momentanen subjektiven Stärke als „gar nicht = 0“, „kaum = 1“, „mäßig = 2“ oder „stark = 3“ eingestuft. Zur Auswertung wird ein Summenscore errechnet.

Bedeutung des Skalenwertes: hohe Werte = viele Beschwerden / niedrige Werte = keine bzw. wenige Beschwerden.

## D-S- Depressivitätsskala

Zerssen D von (2002). D-S, D-S' Depressivitätsskala. In E. Brähler, J. Schumacher, & B. Strauß (Eds.), *Diagnostische Verfahren in der Psychotherapie* (pp. 76-78). Göttingen: Hogrefe.

Selbstbeurteilungsverfahren (16 Items). Der Fragebogen beinhaltet Kurzaussagen, deren Zutreffen für sich der Patient über vier verschiedene gegebene Antwortmöglichkeiten entscheidet: 0 - trifft gar nicht zu, 1 - trifft etwas zu, 3 - trifft überwiegend zu, 4 - trifft ausgesprochen zu. Zur Auswertung werden die Rohwerte der einzelnen Items zum Summen-Score aufsummiert, wobei Item 1 nach der Formel „y = Punktwert-3“ recodiert wird. Die Summen-Scores liegen zwischen 0 (Minimum) und 48 (Maximum). Die Summen-Scores geben Anhaltspunkte für das Vorliegen und gegebenenfalls für das Ausmaß einer psychopathologischen Symptomatik im Sinne einer depressiven, ängstlichen oder auch reizbaren (dysphorischen) Verstimmung, ohne jedoch bei hohem D-Score eine eindeutige Syndromdiagnose, etwa die einer endogenen Depression, zu ermöglichen. Erhöhte Scores können auf eine depressive Symptomatik aufmerksam machen, den rein subjektiven Anteil objektivieren und in gewissem Maße quantifizieren.

**Bedeutung des Skalenwertes:** hohe Werte = deutliche depressive Symptomatik / niedrige Werte = keine bzw. geringe depressive Symptomatik.

## EMEK-27 – Multiples Ergebniskriterium

Schmidt J, Bernhard P, Wittmann WW, Lamprecht F (1987): Die Unterscheidung zwischen singulären und multiplen Ergebniskriterien – Ein Beitrag zur Kriterienproblematik in der Evaluation. In F Lamprecht (Hrsg.), *Spezialisierung und Integration in Psychosomatik und Psychotherapie* (S. 293-299). Berlin Heidelberg: Springer.

Schmidt J, Nübling R, Wittmann WW (1994): Effekte und Wirkungen psychoanalytischer Therapie. Die Bedeutung von Aggregationsprinzipien, von Dosisdimensionen und multiplen Ergebniskriterien. Vortrag auf der 1. Tagung der Fachgruppe „Methoden“ in der Deutschen Gesellschaft für Psychologie. 27.-29.09.1994 in Kiel.

Schmidt, J, Nübling R, Wittmann WW (1995). Praktikable Möglichkeiten der Messung von Ergebnisqualität im Rahmen der Qualitätssicherung. Erfahrungen mit multiplen Ergebniskriterien im Bereich der psychosomatischen Rehabilitation. In *Verband Deutscher Rentenversicherungsträger VDR* (Hrsg.), *Zusammenarbeit von Forschung und Praxis* (S. 71-73). DRV-Schriften Band 5. Frankfurt/Main, VDR.

Steffanowski A (2008): *Evidenz psychosomatischer Rehabilitation im Spiegel multipler Ergebniskriterien. Inauguraldissertation. Universität Mannheim.*

Selbstbeurteilungsverfahren (27 Items). Multiples Ergebniskriterium zum Zeitpunkt der Katamnese. Basis sind Veränderungseinstufungen und Ratings von Patienten, die nach einem Bewertungsalgorithmus zu einer Skala aufsummiert werden. Anzahl der Komponenten (Items): 27. Verrechnung: jede Komponente wird mit 1 oder 0 bewertet, wobei eine 1 für ein positiv bewertbares Ergebnis und eine 0 für ein neutrales bzw. negativ bewertbares Ergebnis vergeben wird. Theoretischer Range: 0 – 27.

**Bedeutung des Skalenwertes:** hohe Werte = viele positiv bewertbare Ergebnisaspekte, gute Ergebnisqualität / niedrige Werte = keine bzw. wenig positiv bewertbare Ergebnisaspekte, schlechte Ergebnisqualität.

Aufgrund der positiven Erfahrungen in der Zauberberg-I-Studie wurde in den Folgejahren der EMEK-Ansatz in fünf weiteren Programmevaluationsstudien zur stationären psychosomatischen Rehabilitation angewandt, wobei jeweils 27 Einzelaspekte zur Skala EMEK\_27 aggregiert wurden.

Vorteile multipler Ergebniskriterien wie z. B. EMEK-27:

- Inhaltlich heterogene Veränderungsinformationen können unabhängig von der jeweils verwendeten Skalierung zu einem Gesamtindex zusammengefasst werden. Dies entspricht der Forderung, eine möglichst breite Palette von bio-psychozialen Ergebnisaspekten bei der Evaluation zu berücksichtigen.
- Der evaluative Prozess des Messens und Bewertens (Kordy & Scheibler, 1984a) wird bereits auf Einzel-Item-Ebene vollzogen und definiert für jeden singulären Ergebnisaspekt klare Bewertungskriterien im Sinne von „vorhanden“ versus „nicht vorhanden“. Dadurch wird eine präzise und gleichberechtigte Abbildung aller erfassten Stakeholder-Interessen in der EVA-Box im Datenmodell von Wittmann (1990) ermöglicht.
- Die Mittelwerte der 27 neu gebildeten dichotomen Items lassen sich direkt im Sinne von Prozentwerten interpretieren. Der Summenwert der Gesamtskala kann ähnlich interpretiert werden. Erreicht z.B. ein Patient 20 Punkte auf der Skala EMEK\_27, so entspricht dies einer „Erfolgsquote“ von  $20 / 27 = 74,1\%$ . Diese Darstellungsform erleichtert erheblich den Transfer der Ergebnisse in die Praxis. Damit ist dieser Ansatz für alle Anwendungsfelder ideal, wo es auf eine anschauliche Vermittlung der Ergebnisse ankommt, etwa bei der Erstellung von Routineauswertungen, Präsentationen und Qualitätsberichten oder im Dialog mit klinischen Praktikern und politischen Entscheidungsträgern.

## GB-10 – Gesundheitliches Befinden

Schmidt J & Nübling R (2010) / Schmidt J et al. (in Vorbereitung)

Selbstbeurteilungsverfahren (10 Items). GB-10 ist eine aus 10 Items bestehende Statusskala des gesundheitlichen Befindens (Range: 10 – 60), die sich in einigen Evaluationsstudien als reliabel ( $\alpha = 0,90$  bis  $\alpha = 0,97$ ) und änderungssensitiv erwiesen hat. Die 10 Bewertungsaspekte umfassen Gesundheitszustand, körperliches Befinden, seelisches Befinden, Allgemeinbefinden, Leistungsfähigkeit, Belastbarkeit, Entspannungsfähigkeit, Ausgeglichenheit, Selbstvertrauen und Fähigkeit, mit alltäglichen Belastungen umzugehen. Die 10 Items werden jeweils auf einer 6-stufigen Skala eingestuft, wobei 1 = sehr schlecht bis 6 = sehr gut.

Bedeutung des Skalenwertes: hohe Werte = gutes gesundheitliches Befinden / niedrige Werte = schlechtes gesundheitliches Befinden.

## GBB - Gießener Beschwerdebogen

Brähler E und Scheer JW (1995; 1. Aufl. 1983).

Selbstbeurteilungsverfahren (57 Items). Klinischer Test. Fragebogen zur Erfassung von Einzelbeschwerden. Der GBB umfasst 57 Items aus den Bereichen Allgemeinbefinden, Vegetativum, Schmerzen und Emotionalität. Subskalen sind: Erschöpfung, Magenbeschwerden, Gliederschmerzen, Herzbeschwerden und Beschwerdedruck. Durchführungszeit ca. 10 bis 15 Min. Der GBB kann eine Fremdbildform eingesetzt werden.

Bedeutung der Skalenwerte (5 Skalen): hohe Werte = viele Beschwerden / hoher Beschwerdedruck / niedrige Werte = wenig Beschwerden / geringer Beschwerdedruck.

## HADS-D - Hospital Anxiety and Depression Scale - Deutsche Version

Herrmann Ch, Buss U, Snaith R P (1995). HADS-D - Hospital Anxiety and Depression Scale - Deutsche Version: Ein Fragebogen zur Erfassung von Angst und Depressivität in der somatischen Medizin. Bern: Huber

Selbstbeurteilungsverfahren (14 Items). Die HADS bzw. ihre deutsche Version HADS-D ist ein kurzes, rasch zu bearbeitendes und gut akzeptiertes Selbstbeurteilungsverfahren mit je sieben alternierend dargebotenen Angst- und Depressionsitems. Mittlerweile liegen nahezu tausend internationale Publikationen vor, die sowohl die Validität und Reliabilität als auch die Praktikabilität und klinische Relevanz der HADS(-D) in den verschiedensten Settings belegen. Auch existieren aus mehreren großen Patienten- und Bevölkerungsstichproben mehrerer Länder Referenz- bzw. Normwerte. Es handelt sich bei der HADS damit mittlerweile in vielen insbesondere europäischen Ländern um eines der Standardverfahren zum Screening sowie zur Verlaufsdokumentation von Angst und Depressivität, wobei der Einsatz schwerpunktmäßig im Bereich der somatischen Medizin erfolgt.

Bedeutung der Skalenwerte (2 Skalen): hohe Werte = hohe Angst- bzw. Depressionssymptomatik / niedrige Werte = geringe Angst- bzw. Depressionssymptomatik.

## HAQ – Helping Alliance Questionnaire

Bassler M, Pongratz P, Krauthauser H (1995); Originalversion: Alexander LB, Luborsky L (1986).

Selbstbeurteilungsverfahren (11 Items). Erfasst Merkmale der Therapiebeziehung und kann insbesondere zur Therapieevaluation (Prozess und Outcome; Psychotherapieforschung) eingesetzt werden. Der HAQ kann sowohl zur Erhebung der Therapeuten- als auch der Patientenperspektive eingesetzt werden. Der Fragebogen umfasst 11 Items (z. B. «Ich habe das Gefühl, dass der Therapeut mich versteht.», «Ich glaube, dass mir die Behandlung hilft») mit 6-stufigen Antwortformat («1» – «sehr unzutreffend» bis «6» – «sehr zutreffend»). Es können ein Gesamtwert und zwei Subskalenwerte „Beziehungszufriedenheit“ „Erfolgsszufriedenheit“ werden. Die zweifaktorielle Struktur wird faktorenanalytisch gestützt. Insbesondere die Erfolgsskala ist ein guter Prädiktor für den Therapieerfolg bzw. Kriterien des Therapieerfolgs (z. B. Angst, Depressivität).

Bedeutung der Skalenwerte (2 Skalen + Gesamtwert): hohe Werte = hohe Erfolgs- bzw. Beziehungszufriedenheit bzw. positive Einschätzung der Beziehung / niedrige Werte = geringe Erfolgs- bzw. Beziehungszufriedenheit bzw. keine positive Einschätzung der Beziehung.

## HEALTH-49 – Hamburger Module

Rabung S, Harfst T, Koch U, Wittchen HU, Schulz H (2007)

Selbstbeurteilungsverfahren (49 Items). Die "Hamburger Module" (HEALTH) sind ein Selbstbeurteilungsinstrument mit insgesamt 9 Skalen, das zur multidimensionalen Erfassung allgemeiner Aspekte psychosozialer Gesundheit entwickelt wurde. Nach eingehender psychometrischer Überprüfung der 79 Items umfassenden Ursprungsversion des HEALTH-Fragebogens auf der Grundlage umfangreicher gesunder und klinischer Stichproben wurde der HEALTH auf 49 Items verkürzt.

Der HEALTH-49 erfasst mit sechs inhaltlich eigenständigen Modulen die folgenden Bereiche:

- Modul A: Psychische und somatoforme Beschwerden (Skala SOM, Skala DEP, Skala PHO, 18 Items)
- Modul B: Psychisches Wohlbefinden (Skala WOHL, 5 Items)

- Modul C: Interaktionelle Schwierigkeiten (Skala INT, 7 Items)
- Modul D: Selbstwirksamkeit (Skala SELB, 5 Items)
- Modul E: Aktivität und Partizipation (Skala A&P, 6 Items)
- Modul F: Soziale Unterstützung/Soziale Belastung Skala SOZU und Skala SOZB, 8 Items)

**Bedeutung der Skalenwerte:** hohe Werte = hohe Belastung, viel Symptome (SOM, DEP, PHO, INT) bzw. hohe negative Ausprägung (WHOL, INT, SELB, A&Ü, SOZU, SOZB) / niedrige Werte = geringe Belastung, wenige Symptome (SOM, DEP, PHO, INT) bzw. positive Ausprägung (WHOL, INT, SELB, A&Ü, SOZU, SOZB).

## ICF-AT 50 Psych

Nosper M (2007)

Selbstbeurteilungsverfahren (50 Items). Mit dem ICF-AT 50 Psych erschien erstmals ein handhabbar erscheinendes Rating-Instrument für die Erfassung potentieller Einschränkungen in den Fähigkeiten von Patienten. Mit Hilfe eines Fragebogens werden hierbei Items abgefragt, die wiederum Einschränkungen in 6 Domänen der ICF abbilden (sollen). In einem einfachen Fragebogen werden anhand von 50 Items Selbsteinschätzungen zur Beeinträchtigung der Fähigkeiten von Aktivitäten und Teilhabe in verschiedenen relevanten Bereichen erfasst bzw. abgefragt. Alle Aussage (z. B. „Ich kann mich sportlich betätigen“, „Ich kann Freundschaften eingehen und pflegen“, „Ich kann die Anforderungen des Arbeitsplatzes erfüllen“) können 5-stufig beantwortet werden, wobei 0 = kein Problem und 4 = volles Problem bedeuten.

Die 6 identifizierten Faktoren des ICF AT-50 Psych haben folgende Bezeichnungen:

- Faktor 1: Anforderungen erfüllen
- Faktor 2: Soziale Beziehungen und Aktivitäten
- Faktor 3: Verbale Kompetenz
- Faktor 4: Fitness und Wohlbefinden
- Faktor 5: Nähe in Beziehungen
- Faktor 6: Soziale Rücksichtnahme

**Bedeutung der Skalenwerte (6 Skalen):** hohe Werte = hohe Problembelastung, starke Beeinträchtigung (negative Richtung) / niedrige Werte = geringe Problembelastung, geringe Beeinträchtigung (positive Richtung).

## ISR – ICD-10-Symptom-Rating

Tritt K et al., 2008

Selbstbeurteilungsverfahren (29 Items). Der Selbstbeurteilungsfragebogen „ICD-10-Symptom-Rating (ISR)“ stellt ein neues Instrument zur psychometrischen Symptomerfassung häufiger psychischer Erkrankungen dar, welches eine verlässliche Diagnosevergabe anhand der ICD-10-Diagnosekriterien ermöglichen soll. Das ISR wurde von einer interdisziplinären Gruppe von Wissenschaftlern in und für die Praxis entwickelt. Die Items des Fragebogens können zu 5 Syndromskalen (Depression: 4 Items, Angst: 4 Items, Zwang: 3 Items, Somatisierung: 3 Items, Essstörung: 3 Items), einem Gesamtscore und einer Zusatzskala (12 Items) verrechnet werden.

**Bedeutung der Skalenwerte (5 Syndromskalen, Gesamtscore, 1 Zusatzskala):** hohe Werte = hohe Symptombelastung (negative Richtung) / niedrige Werte = geringe Symptombelastung (positive Richtung)

## MFHW - Marburger Fragebogen zum habituellen Wohlbefinden

Basler HD, Herda C, Scharfenstein A (2003)

Selbstbeurteilungsverfahren (7 Items). Die eindimensionale Skala dient der Messung positiven Wohlbefindens und ist im Kontext der Gesundheitspsychologie entwickelt worden. Die Zustimmung zu den als Aussagen zum positiven Wohlbefindensbereich formulierten Items ist auf einer sechsstufigen Likertskala auszudrücken. Durch verschiedene Analysen wurde eine eindimensionale faktorielle Struktur bestätigt.

Bedeutung des Skalenwertes: hohe Werte = gutes Wohlbefinden (positiv) / niedrige Werte = schlechter Wohlbefinden (negativ).

## SCL-90-R - Symptom-Checkliste

Franke GH (2002)

Selbstbeurteilungsverfahren (90 Items). Der SCL-90 misst, wie sehr ein Patient an seinen körperlichen und psychischen Symptomen leidet. Die Checkliste wurde von dem Psychologen Leonard Derogatis entwickelt (SCL-90-R = revidierte Version 1994). Der SCL-90 erfasst in seinen 90 Items sowohl körperliche als auch psychische Symptome. Dazu kann der Patient seine Beschwerden innerhalb der letzten 7 Tage anhand einer 5-stufigen Likertskala bewerten („überhaupt nicht“ (0) bis „sehr stark“ (4)). Der SCL-90 enthält folgende 9 Skalen und 3 globale Kennwerte:

- Skala 1: Somatisierung (SOMA)
- Skala 2: Zwanghaftigkeit (AGGR)
- Skala 3: Unsicherheit im Sozialkontakt (UNSI)
- Skala 4: Depressivität (DEPR)
- Skala 5: Ängstlichkeit (ANGS)
- Skala 6: Aggressivität (AGGR)
- Skala 7: Phobische Angst (PHOB)
- Skala 8: Paranoides Denken (PARA)
- Skala 9: Psychotizismus (PSYC)
- GSI: Global Severity Index. Durchschnittliche psychische Belastung auf 90 Items
- PSDI: Positive Symptom Distress Index. Bezieht sich auf die Schwere der Beeinträchtigung bei den Items, bei denen eine psychische Belastung vorliegt
- PST: Positive Symptom Total. Anzahl aller Items, bei denen psychische Beschwerden vorliegen

Bedeutung der Skalenwerte (Skalen 1-9): hohe Werte = hohe Symptomausprägung (negative Richtung) / niedrige Werte = geringe Symptomausprägung (positive Richtung).

## SF-12 – Kurzform des Fragebogens zum Gesundheitszustand (SF-36)

Morfeld M, Kirchberger I, Bullinger M (2011). 2., ergänzte und überarbeitete Auflage

Selbstbeurteilungsverfahren (12 Items). Der SF-12 ist eine Kurzform der Short Form SF-36, ein krankheitsübergreifendes standardisiertes Messinstrument zur Erfassung der subjektiven Gesundheit bzw. der gesundheitsbezogenen Lebensqualität von Probanden. Der Fragebogen umfasst 12 Items zur Erfassung der körperlichen und psychischen Gesundheit, die zu zwei Skalenwerten (Körperliche Summenskala und Psychische Summenskala)

verrechnet werden können. Beide Skalen haben einen Wertebereich von 0 - 100, wobei niedrige Werte ein hohes und hohe Werte ein niedriges Beschwerdeniveau indizieren. Nach Kuhl, Farin & Follert (2003, S.237) liegen zahlreiche Belege dafür vor, dass der SF-12 in verschiedenen Indikationsgruppen in der Lage ist, die zwei Summenskalen des SF-36 zu annähernd 90% zu reproduzieren. Ein wenig beachteter Nachteil im Algorithmus zur Bildung der Skalenindizes liegt darin, dass die Bildung der Skalenwerte nur dann empfohlen wird, wenn alle Items von einer Person beantwortet wurden. Die relativ komplizierte Auswertung der Skalenwerte erfolgte mittels einer SPSS-Syntax des RFV Würzburg.

**Bedeutung der Skalenwerte (2 Skalen):** hohe Werte = guter somatischer bzw. psychischer Status (positive Richtung) / niedrige Werte = schlechter somatischer bzw. psychischer Status (negative Richtung).

## SIBAR – Frühberentungsrisiko

**Bürger W, Deck R, Raspe H, Koch U (2007): SIBAR - "Screening-Instrument Beruf und Arbeit in der Rehabilitation". Entwicklung und Implementierungsmöglichkeiten eines generischen Screening-Instrumentes zur Identifikation von beruflichen Problemlagen und des Bedarfes an berufsorientierten und beruflichen Rehabilitationseinrichtungen. Projektbericht.**

**Bürger W, Deck R. (2009): SIBAR - ein kurzes Screening-Instrument zur Messung des Bedarfs an berufsbezogenen Behandlungsangeboten in der medizinischen Rehabilitation. Die Rehabilitation 48(4): 211–221**

Selbstbeurteilungsverfahren (11 Items). Der generische Screening-Fragebogen SIBAR (Screening-Instrument für Beruf und Arbeit in der Rehabilitation) besteht aus drei Skalen unterschiedlicher Bedarfsbereiche:

- Sozialmedizinische Risikofaktoren (Frühberentungsrisiko)
- Berufliches Belastungserleben
- Subjektiver berufsbezogener Behandlungsbedarf

Die Endversion von SIBAR besteht aus einem kurzen Screening-Bogen, der 11 Items umfasst. Eine längere Version umfasst neben den wichtigsten demographischen Daten zusätzlich eine differenziertere Erfassung der beruflichen Belastungen und des subjektiven Bedarfs an unterschiedlichen berufsbezogenen Behandlungsangeboten.

**Bedeutung des Skalenwertes (Frühberentungsrisiko):** hohe Werte = viele sozialmedizinische Risikofaktoren vorhanden (negative Richtung) / niedrige Werte = wenig sozialmedizinische Risikofaktoren vorhanden (positive Richtung)

## TSD - Test zur Erfassung der Schwere einer Depression

**Obermair W, Stoll KF und Rickels K (1983)**

Selbstbeurteilungsverfahren (30 Items). Klinischer Test zur Depressivität. Der TSD ermöglicht eine quantifizierte Abschätzung depressiver Zustände und dient vor allem zur Verlaufskontrolle und Effizienzprüfung von Behandlungen. Die Selbstbeurteilung wird anhand von vier Subscores (Schuldgefühle, Schlafstörungen, agitierte Depression und ängstlich gehemmte Depression) mit 30 Items sowie des Gesamtpunktwertes interpretiert.

**Bedeutung des Gesamtwertes:** hohe Werte = hoher Schweregrad der Depression (negative Richtung) / niedrige Werte = geringer Schweregrad der Depression (positive Richtung).

## VE – Skala Verstimmungsstörungen

Zielke (1979)

VE ist eine Subskala der Kieler Änderungssensitive Symptomliste (KASSL), die vornehmlich depressiv getönte Stimmungen und Stimmungsschwankungen zum Inhalt hat.

Selbstbeurteilungsverfahren (VE = 17 Items). Die KASSL umfasst vier Skalen, denen insgesamt 50 Items zugeordnet werden, die verschiedene Beschwerden erfassen. Die Skalen erfassen die Bereiche: 1. Soziale Kontaktstörung: z.B. mangelndes Selbstvertrauen, Angst vor anderen Menschen, Minderwertigkeitsgefühle, allgemeine Kontaktarmut etc. 2. Verstimmungsstörungen: z.B. Lebensüberdruß, Überlastungsgefühle, Einsamkeit, Gefühl, den Alltagsanforderungen nicht gewachsen zu sein etc. 3. Berufsschwierigkeiten: z.B. mangelndes Interesse am Beruf, Entschlossenlosigkeit, mangelnde Ausdauer, Unsicherheit bei der Berufswahl etc. 4. Konzentrations- und Leistungsstörungen: z.B. Konzentrationsschwierigkeiten, Leistungsschwäche, Angst zu versagen, Lernschwierigkeiten etc. Alle vier Skalen gemeinsam bilden den Gesamtwert der Symptombelastung.

**Bedeutung des Skalenwertes VE:** hohe Werte = viele Verstimmungsstörungen (negative Richtung) / niedrige Werte = wenige Verstimmungsstörungen (positive Richtung).

## VM-9 – Veränderungsskala

Schmidt J, Nübling R (1988): Die Skala VM-9. Unveröffentlichtes Paper. Psychosomatische Klinik Schömburg.

Schmidt J (1985): Fragebogen zur direkten Veränderungsmessung. Unveröffentlichtes Paper. Psychosomatische Klinik Schömburg.

Nübling R (1992): Psychotherapiemotivation und Krankheitskonzept: zur Evaluation psychosomatischer Heilverfahren. Frankfurt a.M.: VAS (siehe Seiten 112 und Anhang 1b).

Schmidt J, Schmid-Ott G, Lamprecht F (1998): Prävalenz der Adipositas in einer analytisch orientierten psychosomatischen Fachklinik und Behandlungsergebnisse durch Psychosomatische Rehabilitation. Teil II: Ergebnisse von Verlaufsstudien. Psychotherapie 3, Bd. 3, 162-175 (siehe S. 164).

Selbstbeurteilungsverfahren (9 Items). VM-9 ist eine Ergebnisskala, bei der 9 reharelevante direkte bipolare Veränderungseinstufungen (Wohlfühlen, Umgang mit Problemen, Selbsthilfefähigkeit, Frustrationstoleranz, Zurechtkommen mit Arbeit, Belastbarkeit, Auskommen mit Mitmenschen, Leben können mit Einschränkungen, Ausgeglichenheit) zu einem Gesamtwert aufsummiert werden (Cronbach-alpha = .88). Die Items können 7-stufig beantwortet werden, die Skalenwerte können zwischen 9 und 63 liegen. Der theoretische Neutralpunkt (= keine Veränderung) liegt bei 36 (= 9x4). VM-9 korreliert in der Zauberberg-II-Studie hoch mit dem multiplen Ergebniskriterium EMEK-27 ( $r = 0,83$ ) und dem Kriterium BESS ( $r = 0,74$ ).

**Bedeutung des Skalenwertes:** hohe Werte = viele positive gesundheitsbezogene Veränderungen, gute Ergebnisqualität (positive Richtung) / niedrige Werte = wenige positive gesundheitsbezogene Veränderungen oder Verschlechterungen, schlechte Ergebnisqualität (negative Richtung).

## VM-10 – Veränderungsskala

Nübling R, Schmidt J, Reichelt W, Kriz D (2013): Behandlungsergebnisse ein Jahr nach stationärer psychosomatischer Rehabilitation. *Prävention und Rehabilitation*, 25, 150-165.

Schmidt J, Kaiser U, Kriz D, Nübling R (2016): Psychische Komorbidität bei stationären Patienten mit Atemwegs- und Lungenkrankheiten – Korrelate und Verlauf. *Prävention und Rehabilitation*, 28 (3); 85-101.

Selbstbeurteilungsverfahren (10 Items). Die Skala VM-10 beinhaltet 10 direkte Veränderungseinschätzungen (DVM-Prinzip). Es handelt sich um die gleichen 10 Bewertungsaspekte, für die in der Skala GB-10 Statuseinschätzungen erhoben werden. Jedes Item ist 5-stufig skaliert, wobei 1 = stark verschlechtert bis 5 = stark verbessert bedeutet. Die Skalenwerte können zwischen 10 und 50 liegen. Der theoretische Neutralpunkt (= keine Veränderung) liegt bei 30 (= 10x3). Die Vorgängerversion VM-13 mit 13 Items (Schmidt et al., 2003) wies ebenfalls eine hohe interne Konsistenz auf (Entlassungszeitpunkt:  $\alpha = 0,94$ , Katamnesezeitpunkt:  $\alpha = 0,97$ ).

**Bedeutung des Skalenwertes:** hohe Werte = viele positive gesundheitsbezogene Veränderungen, gute Ergebnisqualität (positive Richtung) / niedrige Werte = wenige positive gesundheitsbezogene Veränderungen oder Verschlechterungen, schlechte Ergebnisqualität (negative Richtung).

## ZUF-8 – Fragebogen zur Patientenzufriedenheit

Schmidt, J; Lamprecht, F; Wittmann, WW (1989): Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen. *Psychother med Psychol*, 39, 248-255.

Selbstbeurteilungsverfahren (8 Items). ZUF-8 ist ein Messinstrument zur globalen, eindimensionalen Erfassung von Patientenzufriedenheit („brief global measure of client satisfaction“). Es handelt es sich um die deutschsprachige Adaptation des amerikanischen CSQ-8 von Attkisson & Zwick (1982). Über acht Items wird die generelle Zufriedenheit („general satisfaction“) mit Aspekten der Klinik bzw. der erhaltenen Behandlung erfasst. ZUF-8 ist insbesondere für ein ökonomisches Screening der Patientenzufriedenheit geeignet. Zum Zwecke differenzierterer Zufriedenheitsanalysen mit bestimmten Versorgungsaspekten sind Erweiterungen bzw. zusätzliche Erhebungsinstrumente empfehlenswert.

**Bedeutung des Skalenwertes:** hohe Werte = hohe Zufriedenheit mit der Behandlung bzw. mit dem Klinikaufenthalt (positive Richtung) / niedrige Werte = geringe Zufriedenheit mit der Behandlung bzw. mit dem Klinikaufenthalt (negative Richtung).