

Aus der Medizinischen Klinik mit Schwerpunkt Psychosomatik der Medizinischen
Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Der Einsatz von Moderner Testtheorie zur Messung des
Stressresilienzkonstruktes

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Dipl.-Psych. Nina Obbarius

aus Ettlingen

Datum der Promotion: 05.03.2021

Inhaltsverzeichnis

A. Zusammenfassung	4
Abstract (deutsch).....	4
Abstract (englisch)	6
1. Einleitung	8
1.1 Die Entwicklung des Stresskonzeptes	9
1.2 Überlappung der Stress- und Resilienzforschung	10
1.3 Die Entwicklung des Resilienzkonzeptes	11
1.4 Probleme der Operationalisierung von Resilienz	12
1.5 Vorteile der Entwicklung von Instrumenten auf Basis der Modernen Testtheorie	13
1.6 Spezifische Ziele des Promotionsprojektes	14
2. Methoden	14
2.1 Konzeptuelle Einbettung	15
2.2 Generierung des Itempools und schrittweise Überarbeitung der Items.....	15
2.3 Datenerhebung.....	15
2.4 Psychometrische Itembankentwicklung	15
2.4.1 Schiefe.....	16
2.4.2 Eindimensionalität.....	16
2.4.3 Lokale Unabhängigkeit und heuristische Analysen	17
2.4.4 Differential Item Functioning.....	17
2.4.5 Itemparameterschätzung und Itemfit.....	17
2.4.6 Bewertung der Itembankeigenschaften und Entwicklung einer Kurzform	18
2.4.7 Konstruktvalidität.....	18
2.5 Verwendete Software	18
3. Ergebnisse	18
3.1 Konzeptuelle Einbettung	18
3.2 Generierung des Itempools und schrittweise Überarbeitung der Items.....	19
3.3 Stichprobenszusammensetzung	20
3.4 Psychometrische Itembankentwicklung	20
3.4.1 Schiefe.....	20
3.4.2 Eindimensionalität.....	20
3.4.3 Lokale Unabhängigkeit und heuristische Analysen	21
3.4.4 Differential Item Functioning.....	21
3.4.5 Itemparameterschätzung und Itemfit.....	21
3.4.6 Bewertung der Itembankeigenschaften und Entwicklung einer 10-Item-Kurzform	21
3.4.7 Konstruktvalidität.....	21

4. Limitationen	22
5. Diskussion	23
6. Schlussfolgerung	26
Literaturverzeichnis	27
B. Eidesstattliche Versicherung/Anteilserklärung	31
C. Auszug aus der Journal Summary List	32
D. Druckversion der Publikation	33
E. Lebenslauf	46
F. Publikationsliste	48
G. Danksagung	49

A. Zusammenfassung

Abstract (deutsch)

Titel

Der Einsatz von Moderner Testtheorie zur Messung des Stressresilienzkonstruktes

Einleitung

Stressassoziierte Erkrankungen nehmen stetig zu, wobei bisher noch weitestgehend unklar ist, warum manche Personen unter widrigen Umständen, wie z.B. hoher familiärer und beruflicher Belastung gesund bleiben während andere erkranken. Die Resilienzforschung widmet sich dieser Fragestellung. Resilienz wird häufig als patientenberichtetes Merkmal in Form von statischen Fragebögen erfasst. Das zugrundeliegende Konstrukt wird hierbei jedoch sehr unterschiedlich konzeptualisiert und operationalisiert, wodurch die Integration von Forschungsbefunden erschwert ist. State of the Art der Erfassung patientenberichteter Merkmale ist die Entwicklung von Itembanken auf Basis der Item-Response-Theorie (IRT). Itembanken besitzen unter anderem die Vorteile, dass stichprobenspezifische Kurzformen flexibel zusammengestellt werden können, dass auf ihrer Grundlage Computer-Adaptive Tests (CATs) entwickelt werden können, durch die Patient*innen weniger belastet werden, und dass durch die Entwicklung einer gemeinsamen Metrik auch der Vergleich zwischen verschiedenen Instrumenten ermöglicht wird. Die vorliegende Arbeit beschäftigt sich mit der Entwicklung einer Itembank zur Erfassung von Resilienz, die aufgrund der konzeptuellen Einbettung in das Transaktionale Stressmodell von Lazarus als Stressresilienz (SR) bezeichnet wird.

Methodik

Anhand von Fokusgruppen, kognitiven Interviews und Pilottests wurde iterativ ein initialer Itempool entwickelt, der Stressresilienz auf der individuellen Ebene möglichst breit erfassen sollte. Diese Items wurden in einer Stichprobe von N=521 psychosomatischen Ambulanzpatient*innen erhoben, die sich durch eine hohe Varianz an gesundheitlichen Problemen charakterisierten. Für die psychometrische Itembankentwicklung wurden Standardkriterien angewendet: Evaluation der Antwortverteilungen und Ausschluss stark schief verteilter Items, Eindimensionalität der Itembank und Faktorladungen $> 0,5$, lokale Unabhängigkeit und heuristische Analysen der Itemantwortkurven, Differential Item Functioning (DIF), Itemparameterschätzung und Evaluation des Itemfits. Zudem wurden die Itembankmerkmale evaluiert und eine generische statische

Kurzform aus den besten Items zusammengestellt. Zur Anwendung kamen hierbei explorative und konfirmatorische Faktorenanalysen, weitere state-of-the-art Itemanalysen und Item-Response-Theorie-Analysen.

Ergebnisse

Im Entwicklungsprozess wurden 131 Items generiert, von denen 64 Items ausgeschlossen wurden (54 Items aufgrund einer Faktorladung $< 0,5$, vier Items aufgrund von Residualkorrelationen $> 0,3$, zwei Items aufgrund nicht diskriminativer Item Response Kurven und vier Items aufgrund von DIF). Für die 67 verbleibenden Items wurde anhand der konfirmatorischen Faktorenanalyse und IRT-Analysen ausreichender Modellfit gefunden. Sowohl die SR-Itembank als auch die 10-Item-Kurzform wiesen hohe Korrelationen mit einem statischen Resilienz-Fragebogen (Connor-Davidson Resilience Scale) auf.

Schlussfolgerung

Die finale SR-Itembank und die 10-Item-Kurzform zeigten gute psychometrische Eigenschaften. Im Anschluss an eine weitere Validierung können sie im Rahmen von CATs zur umfassenden Messung von Stress- und Resilienzprozessen eingesetzt werden. Zudem könnte zukünftig durch die Entwicklung einer gemeinsamen standardisierten Metrik für bereits etablierte statische Resilienzfragebögen der Vergleich von Forschungsergebnissen verschiedener Studien ermöglicht werden.

Modifiziert und übersetzt aus: Obbarius N, Fischer F, Obbarius A, Nolte S, Liegl G, Rose M. A 67-item stress resilience item bank showing high content validity was developed in a psychosomatic sample. J Clin Epidemiol. 2018 Aug;100:1-12. doi: 10.1016/j.jclinepi.2018.04.004. Epub 2018 Apr 10

Abstract (englisch)

Title

The Application of Modern Test Theory for the Measurement of Stress Resilience

Introduction

Despite increasing incidence of stress-associated diseases, little is known about why some people stay healthy in the face of adversity, while others become ill. Resilience research aims to address this issue, where the construct of resilience is often assessed by patient self-report using static questionnaires. A major challenge in measuring resilience is that the underlying construct lacks consistent conceptualization and operationalization hampering the integration of results across different studies. State-of-the-art of patient-reported outcomes measurement is the development of item banks based on Item Response Theory (IRT). Item banks offer several advantages including the possibility to compile population-specific short forms, develop Computer-adaptive tests (CATs) that substantially reduce patient burden, and compare scores of different instruments based on a common metric. The present thesis was aimed at the development of an item bank assessing resilience. Due to its conceptual embedment in the Transactional Model of Lazarus it is designated as stress resilience (SR).

Methods

Qualitative item development resulted in an initial item pool covering a broad SR concept at an individual level. All items were tested in a psychosomatic outpatient population (n=521) that was characterized by a wide range of health problems. Psychometric item bank development followed established standard criteria: evaluation of response distributions and elimination of items with a strongly skewed distribution, unidimensionality of the item bank and factor loadings $> .5$, local independence and heuristic analyses of the item response curves, differential item functioning (DIF), estimation of item parameters, and evaluation of item fit. Additionally, item bank characteristics were evaluated and a generic short form was developed. Exploratory and confirmatory factor analyses, further state-of-the-art item analyses and IRT analyses were used.

Results

Of the initial 131 items, 64 items were excluded (54 factor loadings $< .5$, four residual correlations $> .3$, two nondiscriminative item response curves, and four DIF). The remaining 67 items indicated sufficient model fit in confirmatory factor analysis and IRT analyses. Both the SR item bank and

the 10-item short form were highly correlated with a static legacy tool (Connor-Davidson-Resilience Scale).

Conclusion

The final SR item bank and the 10-item short form showed good psychometric characteristics. Following further validation, they will be ready to be used within a framework of CATs for a comprehensive assessment of stress and resilience processes. Future research could link existing resilience measures to an established common SR metric which will allow for the comparison of research results across studies.

Modifiziert aus: Obbarius N, Fischer F, Obbarius A, Nolte S, Liegl G, Rose M. A 67-item stress resilience item bank showing high content validity was developed in a psychosomatic sample. J Clin Epidemiol. 2018 Aug;100:1-12. doi: 10.1016/j.jclinepi.2018.04.004. Epub 2018 Apr 10

1. Einleitung

Mit dem enormen Anstieg zivilisationsbedingter Erkrankungen in den letzten Jahrzehnten gewinnt die Frage danach, welche Faktoren für die Erhaltung der körperlichen und psychischen Gesundheit maßgeblich sind, zunehmend an Bedeutung. Stress spielt eine wesentliche Rolle bei der Entwicklung und dem Verlauf vieler schwerer Erkrankungen [1-5]. So konnte beispielsweise vielfach belegt werden, dass Stress einen Risikofaktor für die Entwicklung einer koronaren Herzerkrankung (KHK) [6, 7] darstellt und sich negativ auf den Verlauf der KHK auswirkt [8]. In der Schwangerschaft kann Stress neben einem negativen Einfluss auf den Schwangerschaftsverlauf auch negative Auswirkungen auf die Gesundheit des Kindes haben [9].

Die Weltgesundheitsorganisation (WHO) erklärte Stress zu einer der größten Gesundheitsgefahren des 21. Jahrhunderts [10]. Die Zunahme stressassoziierter Erkrankungen ist im Kontext unserer derzeitigen gesellschaftspolitischen Situation mit dem Streben nach immer mehr Leistungseffizienz, den zunehmenden Alltagsanforderungen, der Verdichtung von Arbeitsabläufen, der ständigen Erreichbarkeit und Informationsflut, sowie dem Mangel an Pausen zu sehen [11]. Häufig stellen sich die Betroffenen erst in der Primärversorgung vor, wenn es bereits zu tiefgreifenden Beeinträchtigungen der Lebensqualität und zu körperlichen oder psychischen Krankheitssymptomen gekommen ist. Immer häufiger kommt es zu Burnout, depressiven und Angsterkrankungen, zu temporärer Arbeitsunfähigkeit oder Frühberentungen [11, 12].

Individuen reagieren jedoch sehr unterschiedlich auf Stressoren [13, 14]. Personen mit einer hohen Resilienz (psychischen Widerstandsfähigkeit) sind deutlich weniger anfällig dafür, psychische Erkrankungen, wie bspw. Angst oder Depression zu entwickeln [15-17]. Zudem zeigen Personen mit hoher Resilienz in der Auseinandersetzung mit negativen Lebensereignissen, wie bspw. einer Krebsdiagnose, stabileres emotionales Wohlbefinden [18, 19]. Die uneinheitliche Konzeptualisierung und Operationalisierung von Resilienz behindert jedoch den Erkenntnisgewinn hinsichtlich der Faktoren, die Personen widerstandsfähig gegenüber Stress machen [20].

Das Ziel der vorliegenden Arbeit ist es, die Operationalisierung des Resilienzkonstruktes zu verbessern, indem an state-of-the-art Ansätze der Entwicklung patientenberichteter Merkmale angeknüpft und eine Stressresilienz-Itembank entwickelt wird. Am Anfang einer Instrumententwicklung erfolgt eine umfassende konzeptuelle Einbettung des Zielkonstruktes. Daher werden nachfolgend zunächst die Konzepte Stress und Resilienz beschrieben und ihre Überschneidung dargelegt.

1.1 Die Entwicklung des Stresskonzeptes

Das Verständnis von Stress hat sich seit der ersten Beschreibung mehrfach deutlich geändert und unterscheidet sich auch heute noch zwischen und innerhalb verschiedener Disziplinen. Bis zur Mitte des 20. Jahrhunderts wurde Stress als biologische Reaktion gesehen. Schon Claude Bernard zeigte 1859, dass Organismen versuchen, gegenüber Umwelteinflüssen ein inneres Gleichgewicht zu erhalten [21]. Cannon prägte darauf aufbauend 1932 den Begriff der Homöostase [22]. Er beschrieb die Kampf-oder-Flucht-Reaktion (fight-or-flight), bei der das sympathische Nervensystem aktiviert und der Organismus hierdurch zum Kampf oder zur Flucht befähigt wird. Das Stresskonzept von Selye [23] baute auf dem Konzept von Cannon auf. Er definierte Stress als „non-specific response of the body to any demand“ [24] und verstand Stress somit als eine unspezifische Anpassungsreaktion. Er prägte den Begriff des Allgemeinen Anpassungssyndroms, also körperlicher Prozesse, die durch physische oder psychologische Anforderungen aus der Umwelt hervorgerufen werden. Es beinhaltet drei Phasen: die Alarmreaktion, die der Kampf-oder-Flucht-Reaktion nach Cannon entspricht, die Phase des Widerstands, in der eine Anpassung an die Anforderung erfolgt, und eine Phase der Erschöpfung, wobei der Organismus unter dauerhafter Stressbelastung mit Adaptationskrankheiten und Erschöpfung bis hin zum Tod reagieren kann. Selye und Cannon beschrieben somit die körperlichen Veränderungen, die unter Stress erfolgen und zeigten die zwei Hauptachsen der Stressreaktion auf: die Hypothalamus-Hypophyse-Nebennierenrinde-Achse und die Sympathikus-Nebennierenmark-Achse. Zudem konnte Selye schon 1936 den Einfluss von Stress auf das Immunsystem nachweisen [25] und war somit Vorreiter heutiger psychoneuroimmunologisch/-endokriner Stressforschung [10].

Das Konzept der Homöostase wurde durch das Konzept der Allostase abgelöst. Unter Allostase wird das Erreichen von Stabilität durch Veränderung verstanden [26]. Es gibt keinen fixen Zustand, zu dem der Organismus versucht zurückzukehren, sondern er passt sich durch Veränderungen an die äußeren Anforderungen an, indem die Sollwerte, die physiologische Parameter der Stressantwort regulieren, angepasst werden [10].

Holmes & Rahe [27] widmeten ihr Augenmerk hingegen den Stressauslösern und erstellten eine Liste an Lebensereignissen, die in Bezug auf ihr stressauslösendes Potenzial gewertet wurden. Zahlreiche Studien beschäftigten sich daraufhin mit dem Zusammenhang von (negativen sowie positiven) Lebensereignissen und dem Auftreten von Erkrankungen [28].

Im Rahmen der kognitiven Wende wurde der Mensch nicht mehr als „black box“ verstanden, der auf Reize nach lerntheoretischen Prinzipien mit Reaktionen antwortet, sondern individuelle

Kognitionen und Bewertungsprozesse wurden zunehmend zum Forschungsgegenstand [29]. Eines der international bekanntesten psychologischen Stressmodelle ist das transaktionale Stressmodell von Lazarus und Folkman [13], das die Interaktion von Person und Umwelt und die Bedeutung der individuellen Bewertungsprozesse für die Stressantwort betont. In der Konfrontation mit Stressoren bewertet die Person die Relevanz der Stressoren (primäre Bewertung) sowie seine Ressourcen, um mit den Stressoren zurecht zu kommen (sekundäre Bewertung). Auf Basis dieser Bewertungsprozesse setzt die Person bestimmte Copingstrategien ein. Die Copingstrategien beeinflussen wiederum die kurzfristige Stressreaktion und langfristig das psychische Wohlergehen, die körperliche Gesundheit oder Krankheit und die soziale Funktionsfähigkeit. In den Folgejahren erweiterten Lazarus und seine Koautoren das transaktionale Stressmodell zu einem kognitiv-emotional-motivationalem Modell und analysierten den Zusammenhang von Bewertungsprozessen mit bestimmten Emotionen [30, 31].

Die Konzeptualisierungen von Stress unterscheiden sich folglich in Hinsicht auf ihren Fokus auf Stressauslöser (die wiederum in akute und chronische Stressoren, sowie alltägliche und kritische oder traumatische Lebensereignisse unterschieden werden können), Stressreaktionen oder individuelle Bewertungsprozesse [32]. Das Gemeinsame der verschiedenen Ansätze ist jedoch „das Interesse an den Prozessen, die einsetzen, wenn Umwelteinflüsse die Anpassungsfähigkeit der Person übersteigen und somit zu biologischen und psychologischen Veränderungen führen, die eine Erkrankung der Person nach sich ziehen können“ [32] (Übersetzung durch die Autorin).

1.2 Überlappung der Stress- und Resilienzforschung

Die Entwicklung der Stressforschung erfolgte parallel zum Wandel von defizitorientierter Forschung, die sich mit Vulnerabilität und Risikofaktoren beschäftigte, zu ressourcenorientierten Ansätzen, die ihren Fokus auf protektive Faktoren und die Erhaltung von Gesundheit legen [33, 34]. Die Resilienzforschung widmet sich der Fragestellung, warum manche Personen unter widrigen Umständen gesund bleiben, während andere unter diesen Umständen erkranken und weist somit starke Bezüge zum Salutogenese-Konzept von Antonovsky auf [35].

Resilienz kann als „dynamischer Prozess, der positive Anpassung im Kontext signifikanter Widrigkeiten umfasst“ [36] definiert werden. Die zwei Schlüsselkonzepte Widrigkeit und positive Anpassung bilden die Grundlage für verschiedene Definitionen von Resilienz [20]. Die Sichtweise von Resilienz als dynamischer Prozess, der auf diesen Schlüsselkonzepten basiert, weist Ähnlichkeiten zum transaktionalen Stressmodell von Lazarus [13] auf. Dieses betont die Interaktion von Person und Umwelt in der Konfrontation mit Stressoren (also Widrigkeiten), die

entweder zu einer erfolgreichen Anpassung an Stressoren (positive Anpassung) oder zu negativen Stressfolgen, wie bspw. Erkrankung führen. In der vorliegenden Arbeit wird das Resilienzkonstrukt daher in das transaktionale Stressmodell von Lazarus [13] eingebettet und im Rahmen der Neuentwicklung der Itembank als Stressresilienz (SR) bezeichnet, um die inhaltliche Nähe zu verdeutlichen. Einen grundlegenden konzeptuellen Unterschied zwischen SR und Resilienz gibt es jedoch nicht.

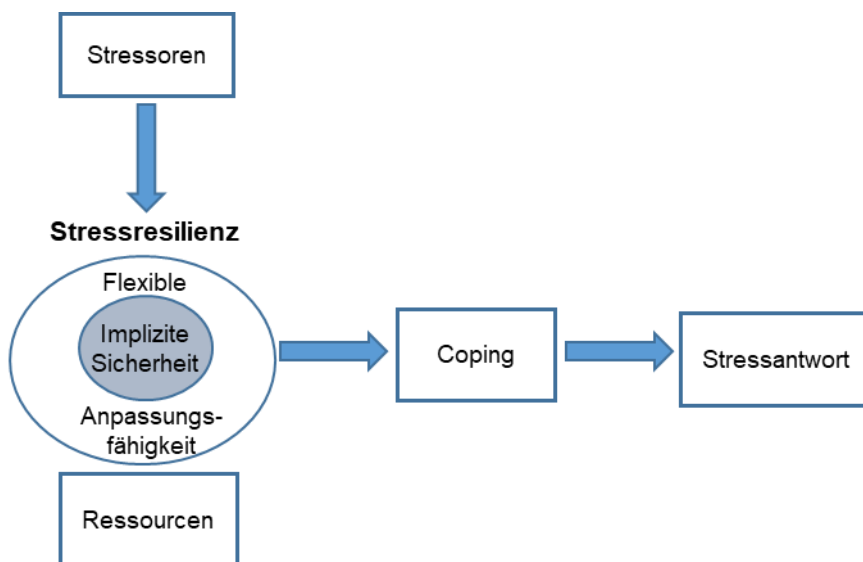


Abbildung 1. Konzeptuelle Einbettung von Stressresilienz in das Transaktionale Stressmodell von Lazarus (Figure 2 auf Seite 5 der Publikation, Übersetzung durch die Autorin)

1.3 Die Entwicklung des Resilienzkonzeptes

Geschichtlich entwickelte sich die Resilienzforschung aus der entwicklungspsychologischen und der persönlichkeitspsychologischen Forschungslinie. Die entwicklungspsychologische Resilienzforschung widmete sich der Beobachtung von Kindern, die unter widrigen Umständen wider Erwarten keine Auffälligkeiten zeigten [37-39]. So wurden als eine der wegweisenden Pionierarbeiten von Resilienz in den 50er Jahren Kinder auf Kauai in einer umfassenden Längsschnittstudie mit mehreren Messzeitpunkten im Laufe ihres Lebens untersucht. Ein Drittel der Kinder, die unter widrigen Umständen (massive psychosoziale Belastungsfaktoren, wie bspw. Alkoholabhängigkeit der Eltern) aufwuchsen, entwickelte sich zu kompetenten sozialen Erwachsenen mit einer hohen Funktionsfähigkeit [38, 40]. Die Forschungsgruppe um Garmezy [41] beobachtete in ihrem Minnesota Risk Research Project die Kinder von an Schizophrenie erkrankter Eltern und kam ebenfalls zu dem Schluss, dass einige Kinder trotz der widrigen Umstände zu Erwachsenen mit gutem Funktionsniveau heranwuchsen. Im Fokus stand die Suche nach protektiven Faktoren. Als gemeinsame Ergebnisse verschiedener Studien zu Schutzfaktoren

gegenüber Widrigkeiten konnten unter anderem Selbstwirksamkeit, positive Emotionen, Extraversion, Selbstwert, positiver Affekt und Spiritualität identifiziert werden [20]. Es lassen sich internale protektive Faktoren (wie z.B. hohe Selbstwirksamkeitserwartung, Problemlösefähigkeit) und externale protektive Faktoren (wie z.B. soziale Unterstützung, enge emotionale Beziehung zu einer Bezugsperson) unterscheiden. Block und Block begründeten 1950 mit ihrem Konstrukt der Ego-Resilienz, das Resilienz als „trait“, also als überdauerndes Persönlichkeitsmerkmal konzipiert, die persönlichkeitspsychologische Resilienzforschung [42]. Als Persönlichkeitsmerkmal ist Resilienz nur schwer von anderen Konzepten, wie z.B. „hardiness“ [43] und „Sense of Coherence“ [44] abzugrenzen.

Während sich die Resilienzforschung zunächst der Suche nach protektiven Faktoren widmete, wurde das Interesse an den zugrundeliegenden Prozessen, mit denen die Widrigkeiten überwunden werden können, immer größer [36]. In der Zusammenschau der verschiedenen Studien zu Resilienz lässt sich konstatieren, dass Resilienz teils als Persönlichkeitsmerkmal („trait“), als Prozess oder als Ergebnis konzeptualisiert wurde [20].

1.4 Probleme der Operationalisierung von Resilienz

Unterschiedliche Definitionen und Konzeptualisierungen von Resilienz führten somit zu einer Vielzahl verschiedener Fragebögen zur Erfassung des zugrundeliegenden Konstruktes, deren Vergleichbarkeit schwierig ist. Verschiedene Reviews weisen auf diese Inkonsistenz hin und geben unterschiedliche Handlungsempfehlungen darüber, welche Fragebögen genutzt werden sollten [45-47]. Windle, Bennett und Noyes [46] evaluierten 19 Resilienzinstrumente, von denen vier Überarbeitungen der Originalinstrumente waren. Die Autoren bewerteten die Connor-Davidson Resilience Scale [48] zusammen mit der Brief Resilience Scale [49] und der Resilience Scale for Adults [50] am besten. Sie gaben jedoch zu bedenken, dass die Qualität dieser Fragebögen im Hinblick auf alle Qualitätskriterien ihres Bewertungsschemas nur als mäßig betrachtet werden könnte. Die qualitative Entwicklung der Items komme häufig zu kurz oder werde nicht umfassend berichtet, so dass Itemauswahl und -formulierungen nur schwer nachzuvollziehen seien. Die Zielpopulation sei häufig nur unzureichend in den Fragebogenentwicklungsprozess einbezogen worden. So basiere die Itementwicklung oftmals auf rein theoretischen Überlegungen. Zusammenfassend kamen sie zu dem Schluss, dass es keinen Goldstandard zur Messung von Resilienz gibt [46].

Weiterhin ist unklar, ob Resilienz durch einen globalen Faktor beschrieben werden kann oder ob mehrere Subfaktoren existieren, die konsistent über verschiedene Stichproben hinweg bestehen.

Die ursprüngliche 5-Faktoren Lösung („Persönliche Kompetenz“, „Vertrauen in die eigenen Instinkte, Toleranz negativer Affekte, stärkende Effekte von Stress“ „Positive Akzeptanz von Veränderung und sichere Beziehungen“, „Kontrolle“, „Spiritualität“) für die Connor-Davidson Resilience Scale [48] konnte beispielsweise nicht konsistent repliziert werden [51].

Fast alle Fragebögen, die zur Erfassung von Resilienz konzipiert wurden, basieren auf Methoden der Klassischen Testtheorie [42, 48-50, 52, 53]. Diese Fragebögen besitzen jeweils eine eigene Metrik. Das heißt, dass eine Person mit einem gegebenen Resilienzniveau bei der Beantwortung unterschiedlicher Instrumente jeweils unterschiedliche Werte zugeschrieben bekommt. Selbst wenn alle Resilienzfragebögen tatsächlich das gleiche latente (nicht direkt messbare) Konstrukt erfassen würden, wäre ein Vergleich aufgrund der jeweils eigenen Fragebogenmetrik schwierig. Außerdem können Fragebögen, die auf der Klassischen Testtheorie basieren, nur mit einer hohen Anzahl an Items eine hohe Messpräzision über einen breiten Bereich des Kontinuums der latenten Variable erreichen. Dies führt zu einer starken Belastung der Proband*innen und erschwert den Einsatz entsprechender Erhebungsinstrumente im klinischen Alltag [54].

1.5 Vorteile der Entwicklung von Instrumenten auf Basis der Modernen Testtheorie

Aufgrund der genannten Nachteile erfolgt derzeit ein Paradigmenwechsel von instrumentenbasierter Messung zu konstruktbasierter Messung. Zur standardisierten Erfassung patientenberichteter Merkmale werden zunehmend Methoden der Item-Response-Theorie (IRT) eingesetzt [55]. Hierdurch kann eine beliebige Anzahl an Items, die dasselbe latente Konstrukt messen, auf einer gemeinsamen Skala kalibriert werden. So entsteht eine konstruktsspezifische, aber instrumentunabhängige Itembank [54, 56]. Das statistische Modell, das anhand von IRT geschätzt wird, liefert Wahrscheinlichkeitsfunktionen, welche die Beziehung zwischen jedem Item und dem latenten Konstrukt beschreiben [57, 58]. Das hat den Vorteil, dass das Antwortmuster für jede Untergruppe an Items verwendet werden kann, um die Ausprägung der latenten Variable für den/die Proband*in auf einer standardisierten Metrik zu berechnen [59]. Somit können Untergruppen an Items ausgewählt werden und zu stichprobenspezifischen Fragebogenkurzformen zusammengestellt werden [60]. Darüber hinaus können Items anderer Fragebögen, die das gleiche latente Konstrukt messen, ebenfalls auf diese gemeinsame standardisierte Metrik („common metric“) gebracht werden [61]. Das ermöglicht den Vergleich von Forschungsergebnissen verschiedener Studien und wurde bereits für einige Konstrukte wie Depression [56], Angst [62] und physische Funktionsfähigkeit [63] durchgeführt. Für diese Konstrukte wurde eine Webanwendung entwickelt, die den direkten Vergleich der Werte

verschiedener etablierter Fragebögen ermöglicht [61]. Ein weiterer Vorteil besteht darin, dass auf Basis von Itembanken Computer-Adaptive Tests (CATs) erstellt werden können [60]. CATs passen die Auswahl der Items automatisch an das Antwortverhalten der Proband*innen an, wodurch die Belastung für den/die Proband*in verringert wird, da er/sie häufig nur eine geringere Anzahl für ihn/sie relevanter Fragen beantworten muss als bei statischen Instrumenten [64]. Darüber hinaus sind Itembanken dynamische Instrumente, die eine kontinuierliche Verbesserung durch Hinzunahme neuer Items erlauben [54]. Bisher gibt es lediglich ein Resilienzinstrument, das auf Basis von IRT entwickelt wurde [65]. Victorsen et al. [65] entwickelten ihre Itembank jedoch spezifisch zur Erfassung von Resilienz bei Patient*innen mit Rückenmarksverletzungen, während sich die vorliegende Arbeit mit der Erfassung generischer Stressresilienz befasst, die möglichst über verschiedene Stichproben hinweg anwendbar sein sollte.

1.6 Spezifische Ziele des Promotionsprojektes

Um die genannten Kritikpunkte an den statischen Resilienzinstrumenten basierend auf der Klassischen Testtheorie zu adressieren, beschäftigt sich die vorliegende Arbeit mit der Entwicklung einer Stressresilienz-Itembank auf Basis der Modernen Testtheorie. Durch Einbindung einer relevanten Zielpopulation in den Itemgenerierungsprozess und die theoretische Einbettung des Resilienzkonstruktes in das Transaktionale Stressmodell von Lazarus [13] soll die Inhaltsvalidität des Instruments gewährleistet und das Verständnis von SR verbessert werden.

Die Beschreibung der Entwicklung einer inhaltsvaliden Stressresilienz-Itembank und einer SR-Kurzform wurde in dem Paper „A 67-item stress resilience item bank showing high content validity was developed in a psychosomatic sample“ im Journal of Clinical Epidemiology veröffentlicht (*Volltext auf Seite 33-45*).

2. Methoden

Die Itembankentwicklung folgte gut etablierten Standardprozeduren, die bereits umfassend in der Literatur beschrieben sind [54, 56, 60, 66] und beinhaltete vier Schritte, die im Folgenden näher spezifiziert werden: (1) Konzeptuelle Einbettung (2) Entwicklung eines Itempools und iterative Verbesserung (3) Datenerhebung (4) Psychometrische Itembankentwicklung und Entwicklung einer Kurzform. *Die einzelnen Schritte des Itembankentwicklungsprozesses sind auf Seite 3 der Publikation im Abschnitt „Methods“ in Figure 1 veranschaulicht und auf Seite 3-5 ausführlich beschrieben.*

2.1 Konzeptuelle Einbettung

Zur Spezifizierung der konzeptuellen Einbettung von SR wurden Fokusgruppen durchgeführt. Diese bestanden teils aus klinischen Expert*innen (N=7) teils aus stationären Patient*innen, die in der Medizinischen Klinik m.S. Psychosomatik der Charité behandelt wurden (N=16). Um die Diskussion in den Fokusgruppen anzuregen, wurden Fokusfragen gestellt, die zunächst breit und dann immer spezifischer formuliert wurden.

2.2 Generierung des Itempools und schrittweise Überarbeitung der Items

Auf Basis der erarbeiteten Konzeptspezifikation wurde ein Itempool generiert. Die Items wurden als Aussagen formuliert, die möglichst alle Aspekte des zuvor definierten Stressresilienzkonzeptes abdeckten. Diese wurden in 10 iterativen Fokusgruppensitzungen mit klinischen Expert*innen (N=7) unter Einbezug der Ergebnisse von kognitiven Interviews mit psychosomatischen Patient*innen in Einzel- (N=7) und Gruppensettings (N=5) überarbeitet. *Nähere Erläuterungen zur qualitativen Itementwicklung finden sich auf Seite 4 der Publikation im Abschnitt „Item pool generation and iterative refinement“.*

2.3 Datenerhebung

Der initiale Itempool wurde anhand eines Pilottests an einer Ad-hoc-Stichprobe stationärer psychosomatischer Patient*innen (N=45) auf seine Verständlichkeit hin überprüft, indem die Patient*innen bei Verständnisschwierigkeiten eine eigene Antwortkategorie ankreuzen konnten. Zudem wurden die Antwortverteilungen evaluiert. Im Zeitraum von Juni 2014 bis Juni 2015 wurden die Items im Rahmen der Routinebefragung in einer Stichprobe von N=521 Ambulanzpatient*innen der Medizinischen Klinik m.S. Psychosomatik der Charité elektronisch erhoben. Zusätzlich wurden soziodemografische Fragen, das ICD-Symptom Rating ISR, das klinische Symptome erfasst, sowie ein statischer Resilienzfragebogen (Connor-Davidson Resilience Scale) beantwortet.

2.4 Psychometrische Itembankentwicklung

Der Itempool wurde iterativ anhand etablierter psychometrischer Kriterien reduziert. Jedes eliminierte Item wurde auf Verlust von Inhaltsvalidität hin überprüft. Die verbleibenden Items wurden mit dem IRT-basierten graded response model kalibriert [67]. Dieses ist für polytome (mehr als zwei Antwortkategorien) Items, die ein eindimensionales Konstrukt messen und ranggeordnete Antwortkategorien aufweisen geeignet [60]. Die Schätzung eines IRT-Modells führt zu logistischen Funktionen für jedes Item [60]. Diese modellieren die Beziehung zwischen

der Ausprägung der Person auf der latenten Variablen und der Wahrscheinlichkeit, dass diese Person eine bestimmte Antwortkategorie wählt. Diese Funktionen können anhand von Item Characteristic Curves (ICC) veranschaulicht werden. Die Itemparameter, die durch das Modell geschätzt werden und die ICC charakterisieren, sind ein Steigungsparameter (slope parameter) und mehrere Schwellenwerte (thresholds). Der Steigungsparameter zeigt an, wie stark ein Item mit der latenten Variable zusammenhängt. Die Schwellenwerte geben die Schwierigkeit eines Items an, indem der Bereich auf der latenten Variable angegeben wird, an dem das Item am informativsten ist. Auf Grundlage dieser Parameter kann die Ausprägung der latenten Variable für Antwortmuster auf jede Untergruppe an Items geschätzt werden [59].

2.4.1 Schiefe

Die Verteilung der Antwortoptionen jedes Items wurde untersucht, um die Items mit einer Schiefe über einem absoluten Wert von vier zu eliminieren. Items mit sehr schiefen Antwortverteilungen besitzen wenig informativen Wert, da viele Personen dieselbe Antwortkategorie wählen.

2.4.2 Eindimensionalität

Zur Überprüfung der Dimensionalität wurde zunächst eine explorative Faktorenanalyse (EFA) durchgeführt, um die zugrundeliegende korrelative Struktur und die Höhe der Eigenwerte zu evaluieren [66]. Alle Items mit Ladungen über 0,5 auf dem ersten Faktor wurden beibehalten, da davon ausgegangen wird, dass diese einen substantiellen Beitrag zum latenten Konstrukt liefern [56, 68]. Cronbach's α und die korrigierten Item-Total-Korrelationen wurden für diese Items bestimmt [69]. Zur Evaluation der Eindimensionalität der verbleibenden Items wurde eine einfaktorielle konfirmatorische Faktorenanalyse (CFA) durchgeführt. Für die Modellparameterschätzung wurde der weighted least squares, mean- and variance-adjusted (WLSMV) Schätzer verwendet. Zur Bestimmung der Modellgüte wurden Bentlers Comparative-Fit-Index (CFI), Tucker-Lewis-Index (TLI, jeweils Cutoff-Wert $> 0,95$) und Root Mean Square Error of Approximation (RMSEA, Cutoff-Wert $< 0,08$) herangezogen [70]. Da die Orientierung an traditionellen CFA-Fit-Indizes für Itembankentwicklungen als zu restriktiv kritisiert wurde [71], erfolgte zusätzlich eine Bifaktormodellberechnung mit einem Generalfaktor und drei explorativen Gruppenfaktoren. Eine erklärte gemeinsame Varianz (explained common variance, ECV) von über 0,6 gilt als Indikator für einen starken gemeinsamen Faktor und ausreichende Eindimensionalität der Daten [72].

2.4.3 Lokale Unabhängigkeit und heuristische Analysen

Lokale Unabhängigkeit bedeutet, dass die gesamte gemeinsame Varianz zwischen den Items durch das latente Konstrukt erklärt wird, und ist eine wichtige Annahme der IRT [60]. Residualkorrelationen zwischen zwei Items mit einem Wert größer als 0,3 wurden als Indikator für lokale Abhängigkeit betrachtet. Von diesen Itempaaren wurde das Item entfernt, das die weniger günstigen Itemantwortkurven zeigte oder mehr als eine Residualkorrelation über 0,3 aufwies. *Nähere Erläuterungen zu den heuristischen Analysen der Itemantwortkurven finden sich auf Seite 4 und 5 der Publikation im Abschnitt „Local independence and heuristic analyses“.* *Exemplarische grafische Beispiele für günstige und weniger günstige Itemantwortkurven sind in Figure 3 auf Seite 7 veranschaulicht.*

2.4.4 Differential Item Functioning

Differential Item Functioning (DIF) liegt vor, wenn sich die Wahrscheinlichkeit eine bestimmte Antwortkategorie zu wählen, bei gleicher Ausprägung der latenten Variable durch die Zugehörigkeit zu einer bestimmten Gruppe unterscheidet [60]. Um eine solche systematische Verzerrung durch Gruppenunterschiede auszuschließen, wurde ein ordinales logistisches Regressionsmodell zur Erfassung von DIF für Alter, Geschlecht und Beziehungsstatus gerechnet. Nagelkerkes Determinationskoeffizient R^2 [73] wurde berechnet, um die Items auszuschließen, die einen $\Delta R^2 > 0,02$ und somit auffallendes DIF zeigten. Um systematische Fehler auf Basis der Krankheitskategorie zu kontrollieren, wurden zusätzlich Post-hoc-Analysen gerechnet, in denen Patient*innen mit einer bestimmten Diagnose den Patient*innen ohne diese Diagnose gegenübergestellt wurden. Zu den Diagnosekategorien gehörten Depressionen (N=271), Angststörungen (N=82), somatoforme oder Schmerzstörungen (N=192), Essstörungen einschließlich Adipositas (N=116) und körperliche Erkrankungen (N=220).

2.4.5 Itemparameterschätzung und Itemfit

Die Itemparameterschätzung der verbleibenden Items erfolgte mit einem graded response model [67]. Die IRT (Theta, θ) -Werte stellen Schätzungen der Ausprägung des latenten Traits dar. Für die verschiedenen Ausprägungen von θ wurden die vorhergesagten mit den beobachteten Häufigkeiten der Antwortkategorien des Items verglichen und so die S_{-X^2} Fit Statistik (eine Pearson X^2 Statistik) berechnet. Da ein p-Wert $< 0,05$ auf eine schlechte Anpassungsgüte, also eine hohe Diskrepanz zwischen den durch das Modell vorhergesagten und beobachteten Itemantworten, hinweist, wurden Items mit einem $p.S_{-X^2} < 0,05$ ausgeschlossen.

2.4.6 Bewertung der Itembankeigenschaften und Entwicklung einer Kurzform

Iteminformations-Funktionen zeigen den Beitrag jedes Items zur Präzision der ganzen Itembank für jede Ausprägung von θ [54]. Sie können auf der Basis von IRT-Modellparametern berechnet werden. Die Testinformation für jede beliebige Anzahl an Items oder die gesamte Itembank auf einem bestimmten θ -Niveau kann durch die Summe der Iteminformation (ΣI) für diese Auswahl an Items oder die gesamte Itembank auf diesem Ausprägungsgrad von θ berechnet werden. Der erwartete Standardfehler (SE) wird durch die Formel $SE(\theta) = \sqrt{1/\Sigma I(\theta)}$ berechnet. Je höher die Iteminformation ist, desto höher ist die Präzision des Items oder Tests auf diesem θ -Niveau. Die Reliabilität kann mit folgender Formel bestimmt werden: $rel(\theta) = \text{ABS}(1 - SE(\theta)^2)$ [57]. Alle Iteminformationskurven wurden analysiert. IRT-basierte Itembanken ermöglichen die Entwicklung stichprobenspezifischer Kurzformen durch die Auswahl der präzisesten Items auf einem bestimmten θ -Niveau [60], was mit einer geringeren Belastung der Proband*innen einhergeht, da sie nur wenige für sie relevante Items beantworten müssen. Für verschiedene Untergruppen von Items wurden Testinformation, Standardfehler und Reliabilität bestimmt. Diese Kriterien, sowie Überlegungen zur Inhaltsvalidität bestimmten schließlich die Auswahl der Items für die SR-Kurzform.

2.4.7 Konstruktvalidität

Zur vorläufigen Bestimmung der konvergenten Validität wurden die Pearson-Korrelationen des θ -Werts der finalen Itembank und Kurzform mit dem Summenwert der Connor-Davidson Resilience Scale (CD-RISC) [48] berechnet.

2.5 Verwendete Software

Die explorativen statistischen Analysen wurden mit IBM SPSS® Statistics 22 durchgeführt. Für die weiteren Analysen wurde R.3.3.2 [74] mit den folgenden Softwarepaketen verwendet: *lavaan* für die CFA [75], *psych* für die Rechnung von Bifaktor-Modellen [76], *lordif* für die DIF-Analysen [77], *KernsmoothIRT* und *mirt* für IRT-Modell und Parameterschätzung [78, 79], sowie *ggplot2* für die Grafiken [80].

3. Ergebnisse

3.1 Konzeptuelle Einbettung

In iterativen Fokusgruppendifkussionen wurde Resilienz in das transaktionale Stressmodell von Lazarus [13] eingebettet und als Stressresilienz (SR) bezeichnet. SR wurde als breites Konstrukt definiert, um möglichst alle Facetten der psychischen Widerstandsfähigkeit gegenüber Stress zu

erfassen. Es wurde angenommen, dass SR auf der persönlichen Ebene erfasst wird und über die Zeit und durch Interventionen veränderbar ist. Auf Basis von Literaturrecherche und Fokusgruppendifkussionen wurde folgende Definition von SR erarbeitet:

„Stressresilienz ist die Fähigkeit psychische Gesundheit unter widrigen Umständen zu erhalten und umfasst implizite Sicherheit und flexible Anpassungsfähigkeit. (...) basierend auf Ressourcen und vorangegangenen Erfahrungen spiegelt sie eine Einstellung gegenüber der Welt wider, die beeinflusst, wie die Person Stress bewertet, wie sie internale und externale Ressourcen evaluiert, wie sie Stress bewältigt und wie sie infolge auf Stress reagiert.“ (wörtliches Zitat aus der Publikation auf Seite 5 im Abschnitt „Concept specification“, Übersetzung durch die Autorin)

Wie bereits in der Einleitung ausgeführt, bewertet gemäß des transaktionalen Stressmodells von Lazarus [13] jede Person die Stressoren und die internalen und externalen Ressourcen, die zur Bewältigung der Stressoren zur Verfügung stehen, unterschiedlich. Wir gehen davon aus, dass die individuelle SR die Bewertungs- und Copingprozesse maßgeblich beeinflusst. SR wurde als Kombination aus impliziter Sicherheit und den Entwicklungsbedingungen für implizite Sicherheit, nämlich positiven Beziehungserfahrungen in der Kindheit, sowie flexibler Anpassungsfähigkeit konzeptualisiert. *Die Einbettung von Stressresilienz in das Stresskonzept von Lazarus ist auf Seite 5 der Publikation in Figure 2 im Abschnitt „Concept specification“ veranschaulicht und wird in diesem Abschnitt näher erläutert.*

3.2 Generierung des Itempools und schrittweise Überarbeitung der Items

Auf Basis der Konzeptspezifikation wurden 156 Items entwickelt. In kognitiven Interviews und 10 iterativen Fokusgruppendifkussionen erfolgte eine Reduktion auf 131 Items. Hauptergebnis der kognitiven Interviews waren Verständnisschwierigkeiten der Items mit doppelten Verneinungen. Die Formulierung wurde auf Basis der Ergebnisse angepasst und alle Änderungen der Items in einer Item-Tracking-Matrix festgehalten. Der Pilottest zeigte keine konsistenten Verständnisschwierigkeiten und die Antwortverteilungen waren unauffällig. Die Items bestanden aus Aussagen, die auf einer 4-Punkt-Likertskala mit den Antwortoptionen „trifft gar nicht zu“, „trifft eher nicht zu“, „trifft eher zu“ und „trifft völlig zu“ beantwortet werden sollten. Der zeitliche Bezugsrahmen umfasste die letzten sieben Tage mit Ausnahme der Items, die sich auf Kindheitserfahrungen bezogen. Um eine Akquieszenzverzerrung zu vermeiden, wurden auch invers formulierte Items eingeschlossen [81, 82]. Höhere Werte weisen auf höhere SR hin.

3.3 Stichprobenzusammensetzung

Der 131 Items umfassende Itempool wurde von 521 ambulanten Patient*innen der Medizinischen Klinik m.S. Psychosomatik, zu ungefähr zwei Dritteln bestehend aus Frauen mit einem mittleren Alter von 42 Jahren, elektronisch auf Smartphones beantwortet. *Die genaue Stichprobenbeschreibung ist in Table 1 auf Seite 6 der Publikation zu finden.*

3.4 Psychometrische Itembankentwicklung

Die Ergebnisse der psychometrischen Itembankentwicklung, die hier verkürzt wiedergegeben werden, sind ausführlich auf den Seiten 6 bis 8 der Publikation beschrieben. Die gekürzten Items, die im Anschluss an die explorative Faktorenanalyse beibehalten wurden, die Faktorladungen und sukzessiven Ausschlusskriterien der CFA und IRT-Analysen sind in Table A1 im Online Only Appendix des Papers auf der Webseite des Journals zu finden (<https://linkinghub.elsevier.com/retrieve/pii/S0895435617305619>).

3.4.1 Schiefe

Keines der Items musste anhand des Kriteriums einer Schiefe über einem absoluten Wert von vier ausgeschlossen werden.

3.4.2 Eindimensionalität

Die Annahme der Eindimensionalität wurde durch das zugrundeliegende Ladungsmuster in der EFA, das Verhältnis des ersten zum zweiten Eigenwert, das größer als vier war, und eine Varianzaufklärung von mehr als 20% durch den ersten Faktor bestätigt [66]. Von den 131 Items wurden 77 Items beibehalten, da sie in der EFA auf dem ersten unrotierten Faktor $> 0,5$ luden. Cronbach's α für die 77 Items betrug 0,976 und alle Items zeigten Item-Total-Korrelationen über 0,4, was auf einen starken Zusammenhang zwischen den Items und dem latenten Konstrukt schließen ließ. Eine einfaktorielle CFA mit den beibehaltenen 77 Items wies auf eine Abweichung von einem eindimensionalen Modell hin (Chi-square = 7933 (df = 2849), $p < .001$). RMSEA zeigte hingegen einen ausreichenden Modellfit (RMSEA=.059 [.057-.060]), während CFI and TLI leicht unter 0,9 lagen (CFI=0,89, TLI=0,88). Die erklärte gemeinsame Varianz (ECV) im Bifaktormodell mit 3 explorativen Gruppenfaktoren lag bei 0,68, was als Indikator für ausreichende Eindimensionalität der Daten gewertet werden konnte [72].

3.4.3 Lokale Unabhängigkeit und heuristische Analysen

Fünf Itempaare wiesen Residualkorrelationen über 0,3 auf. Zwei Items aus diesen Itempaaren wurden aufgrund mehrfachen Residualkorrelationen über 0,3 eliminiert. Die Analysen der Item-Antwortkurven ergab den Ausschluss weiterer vier Items und die Zusammenlegung von Antwortoptionen aufgrund von ungeordneten Antwortkategorien für insgesamt 13 Items.

3.4.4 Differential Item Functioning

Aufgrund von DIF hinsichtlich Alter und Geschlecht wurden insgesamt vier Items aus der Itembank entfernt. Drei Items zeigten DIF für Essstörungen einschließlich Adipositas. Die Items „Ich weiß genau, was mir guttut und was mir nicht guttut.“, „Ich mag mich so wie ich bin.“ und „Ich fühle mich wohl in meinem Körper.“ wurden von Personen mit Essstörungen und Adipositas bei gleicher SR-Ausprägung systematisch anders beantwortet als von Personen ohne diese Diagnosen. *Weitere Ausführungen zu DIF finden sich auf Seite 7 der Publikation im Abschnitt „Differential item functioning“.*

3.4.5 Itemparameterschätzung und Itemfit

Für die beibehaltenen 67 Items wurden Modellparameter geschätzt. Alle Items zeigten einen guten Fit ($p.S_X^2 > 0,05$). Die finale Itembank wies eine hohe Messpräzision ($SE \leq 0,32$) in einem Thetabereich von -3,4 bis 3,6 auf.

3.4.6 Bewertung der Itembankeigenschaften und Entwicklung einer 10-Item-Kurzform

Für jedes Item der finalen Itembank wurden die Iteminformationsfunktionen analysiert. Die maximale Iteminformation lag zwischen 0,29 und 1,76. Das Maximum der Testinformation für die gesamte Itembank betrug 44,82 bei einem Thetawert von -0,3 ($SE = 0,15$, dies entspricht in etwa einer klassischen Reliabilität von 0,98). Zehn Items, von denen drei invers kodiert sind, wurden auf Basis der Analysen von Testinformation, SE, Reliabilität und Inhaltsvalidität für eine Kurzform ausgewählt. Die Pearson-Korrelation zwischen dem Thetawert der gesamten Itembank und der 10-Item-Kurzform betrug 0,95 ($p < 0,001$). *Die Eigenschaften der 10-Item-Kurzform sind auf Seite 8 der Publikation in Table 2 dargestellt.*

3.4.7 Konstruktvalidität

Die Pearson-Korrelation von 0,83 zwischen dem Thetawert der gesamten Itembank und der CD-RISC, sowie von 0,79 zwischen dem Thetawert der Kurzform und der CD-RISC unterstützte die konvergente Validität des neuen Instruments ($p < 0,001$).

4. Limitationen

Die vorliegende Arbeit weist einige Limitationen auf. Die Itembankentwicklung erfolgte in einem einzigen Setting an einer begrenzten Stichprobe psychosomatischer Patient*innen. Weitere Validierungsstudien sind notwendig, um eine Generalisierbarkeit der SR-Itembank auf andere Stichproben und Kontexte zu überprüfen. Darüber hinaus waren die Stichprobengrößen der unterschiedlichen Krankheitskategorien für die Überprüfung von DIF relativ klein, weshalb eine größere Stichprobe nötig sein wird, um DIF zwischen den verschiedenen Diagnosekategorien umfassender zu evaluieren und die Validität der Modellparameter für unterschiedliche Krankheitskategorien zu zeigen.

Des Weiteren erreichten die Fitindices zur Überprüfung des eindimensionalen Modells nicht alle festgelegten Cut-off-Werte. Die Annahme der Eindimensionalität für die IRT-Itembankentwicklung ist häufig nur schwer auf komplexe psychische Konstrukte anzuwenden. Daher wurden die traditionellen Fit-Indizes als zu restriktiv diskutiert [71] und explorative Bifaktor-Analysen mit einer erklärten gemeinsamen Varianz über 60% als hinreichend für die Eindimensionalitätsannahme vorgeschlagen [72]. Dennoch sollte die Art des Messmodells in weiteren Studien mit unterschiedlichen Stichproben überprüft werden.

Der Validierungsfragebogen (CD-RISC) weist gegenüber der SR-Itembank einen anderen zeitlichen Bezugsrahmen (letzter Monat vs. letzte sieben Tage) auf. Dies könnte den Zusammenhang kleiner erscheinen lassen, als er bei gleichem Bezugsrahmen wäre. Die Validierung neu entwickelter Instrumente mittels anderer bereits etablierter Fragebögen, die dasselbe zugrundeliegende Konstrukt messen, birgt generell Schwierigkeiten. Ist die Korrelation zu hoch, ist die Notwendigkeit der Neuentwicklung in Frage zu stellen, ist sie jedoch zu niedrig, wirft das Fragen bezüglich der Validität auf. Längsschnittstudien zur Untersuchung der prädiktiven Validität der SR-Itembank, wie z.B. der Wirkung von SR auf Unterschiede in der psychischen Gesundheit vor und nach einem bedeutenden Stressor, stehen noch aus.

Darüber hinaus sind Deckeneffekte der SR-Itembank nicht auszuschließen, da sie anhand einer Stichprobe psychosomatischer Patient*innen entwickelt wurde. Eine Überprüfung der Validität in einer hoch stressresilienten Stichprobe wäre somit interessant, obwohl wir davon ausgehen, dass durch den Einbezug klinischer Expert*innen in den Itementwicklungsprozess auch der höhere Bereich der SR durch die Items abgedeckt wird.

Auch wenn die vorliegende Arbeit erste Hinweise auf gute psychometrische Eigenschaften der SR-Itembank liefert, sollte die Robustheit des Modells in zukünftigen Studien weiter untersucht

werden. Mittels Multitrait-Multimethod-Analysen sowie Langzeitstudien sollen konvergente und divergente Validität zukünftig evaluiert werden. Da die Erhebung in einem Querschnitt-Design erfolgte, können keine Aussagen über die angenommene konzeptuelle Einbettung von SR in das transaktionale Stressmodell von Lazarus getroffen werden. Hierfür sollten Strukturgleichungsmodelle mit SR und unterschiedlichen patientenberichteten Merkmalen, wie internalen und externalen Ressourcen, wahrgenommenem Stress, Coping und Stressreaktion, sowie psychischer Gesundheit gerechnet werden. Auch Aussagen zur Veränderungssensitivität der SR-Itembank können auf Basis der Ergebnisse nicht gemacht werden und erfordern Längsschnittstudien.

5. Diskussion

In der vorliegenden Arbeit wurden eine inhaltsvalide SR-Itembank und eine SR-10-Item-Kurzform auf Basis Moderner Testtheorie und unter Einbezug einer relevanten Zielpopulation entwickelt. Hierbei wurde ein besonderes Augenmerk auf die iterative qualitative Entwicklung der Items und umfassende statistische Analysen gepaart mit Überlegungen zur Inhaltsvalidität gelegt. Aufgrund der inhaltlichen Überschneidungen des transaktionalen Stressmodells von Lazarus [13] und der Sichtweise von Resilienz als dynamischen Prozess [36], erschien uns eine konzeptuelle Einbettung in das Modell von Lazarus und eine Bezeichnung der neu entwickelten Itembank als Stressresilienz-Itembank sinnvoll. Die SR-Itembank und die SR-Kurzform zeigten in einer psychosomatischen Stichprobe gute psychometrische Kriterien und hohe Konstruktvalidität. Um Aussagen darüber machen zu können, ob die Items auch in anderen Stichproben angewendet werden können, sind weitere Validierungsstudien an anderen Stichproben, wie z.B. Krebspatient*innen oder Flüchtlingen notwendig.

Die faktorenanalytischen Analysen setzen ein reflektives Konstrukt voraus, das heißt, es wird angenommen, dass ein zugrundeliegendes latentes Konstrukt die Beantwortung der Items beeinflusst. Somit besteht die Annahme, dass die Items hochgradig korrelieren und austauschbar sind [69, 83]. Für das SR-Konstrukt könnte jedoch genauso angenommen werden, dass ein formatives Messmodell gültig ist, bei dem die einzelnen Items nicht notwendigerweise miteinander korreliert sein müssen und verschiedene Aspekte der SR abdecken. Das Ziel dieser Arbeit bestand jedoch nicht in der Entwicklung eines Instruments, das unterschiedliche unkorrelierte Aspekte der positiven Anpassung an Stress erfasst, wie bspw. verschiedene Copingstrategien im Umgang mit unterschiedlichen Stressoren, sondern ein distinktes zugrundeliegendes Kernkonstrukt der SR zu operationalisieren.

Das SR-Konstrukt wurde zunächst sehr breit konzeptualisiert und der Brainstormingprozess in den Fokusgruppen nicht limitiert, um sicher zu stellen, dass alle Aspekte der SR erfasst werden, wobei das Konstrukt nachfolgend durch die weitere psychometrische Itembankentwicklung gepaart mit Überlegungen zur Inhaltsvalidität verengt wurde. So wurden beispielsweise alle Items, die Kindheitserfahrungen erfassten, aufgrund von Ladungen unter 0,5 auf dem extrahierten Faktor ausgeschlossen. Ein Beispielitem hierfür ist „In meiner Kindheit wurde ich in den Arm genommen, wenn ich es gebraucht habe.“. Die geringen Faktorladungen könnten auf den unterschiedlichen zeitlichen Bezugsrahmen zurückzuführen sein. Wahrscheinlicher ist jedoch, dass diese Kindheitserfahrungen Antezedenzen von SR sind und somit ein distinktes Konstrukt der Beziehungserfahrungen erfassen. In verschiedenen Studien konnte gezeigt werden, dass eine enge emotionale Beziehung zu einer nahen Bezugsperson ein wichtiger protektiver Faktor im Umgang mit Widrigkeiten ist [36, 38, 39]. Allerdings ist das genaue Zusammenwirken der verschiedenen identifizierten protektiven Faktoren noch unklar [36, 39]. Bisher konnte noch kein umfassendes empirisch belegtes Erklärungsmodell für die dynamischen Resilienzprozesse gefunden werden [84]. Wir gehen davon aus, dass positive Beziehungserfahrungen Vorbedingungen für die Ausprägung von Resilienz auf der individuellen Ebene sind. Die meisten Items, die extrinsische Faktoren erfassten, wie z.B. „Ich habe mindestens einen Freund, auf den ich mich verlassen kann.“, wurden anhand der psychometrischen Kriterien aus der Itembank ausgeschlossen, während die intrinsischeren Items, wie z.B. „Ich kann meine Beziehungen so gestalten, dass ich mich wohl fühle.“ beibehalten wurden. Auch Items, die sich auf einen gesunden Lebensstil bezogen, wie z.B. „Ich schlafe ausreichend.“, wurden eliminiert. Es ist anzunehmen, dass ein gesunder Lebensstil eher eine Folge von SR ist und zudem eher einem formativen Messmodell zugeordnet werden kann, als einem reflektiven. Letztendlich ist jedoch nur in Langzeitstudien zu klären, welche Faktoren Antezedenzen von Resilienz, Facetten der Resilienz oder Konsequenzen von Resilienz sind. Resilienz sollte gemäß unserer zuvor bestimmten konzeptuellen Einbettung Bewertungsprozesse beeinflussen, aber ein distinktes Konstrukt sein. Übereinstimmend wurden die Items, die positive Stressbewertung erfragten, ausgeschlossen. Die 67 Items der finalen Itembank und die 10 Items der Fragebogenkurzform erfassen das zugrundeliegende Kernkonstrukt der SR am besten und spiegeln eine Kombination aus innerer Sicherheit und flexibler Anpassungsfähigkeit wider. Die hohe Korrelation mit dem CD-RISC weist zudem auf eine hohe Konstruktvalidität hin.

Immer häufiger wird auf die Notwendigkeit von Multilevel-Analysen hingewiesen, da Resilienzprozesse auf verschiedenen Ebenen wirken, wie z.B. der individuellen Ebene, der familiären Ebene und breiteren Kontextebenen [85]. Mit der Verfügbarkeit neuer Technologien gewinnen auch neurobiologische und translationale Ansätze zunehmend an Bedeutung [86]. Im Sommer 2014 wurde z.B. in Mainz das Deutsche Resilienz Zentrum gegründet, das Resilienz interdisziplinär, transdiagnostisch und multisystemisch untersucht [87]. Zusätzlich werden derzeit Alternativen zur Erfassung von Resilienz, wie bspw. der Einsatz von Situationsbewertungstests („Situational judgement tests“) [47] oder die Berechnung eines R-Scores [86, 88] diskutiert, bei dem der Quotient der Differenz psychischer Dysfunktionen und der Differenz der akkumulierten Stressbelastung zwischen zwei Testzeitpunkten vor und nach Stressexposition in Längsschnittstudien berechnet wird.

Die Uneinigkeit darüber, wie Resilienz und Stress operationalisiert werden sollten, stellt kein ursprüngliches Problem der Stress- und Resilienzkonstrukte dar, sondern betrifft auch die Operationalisierung anderer latenter (d.h. nicht direkt messbarer) psychologischer Konstrukte [47]. Da für die ganzheitliche Erfassung von Krankheit und Gesundheit psychische Aspekte maßgeblich sind, ist die wissenschaftliche Weiterentwicklung der Messung patientenberichteter Merkmale von großer Bedeutung. Gerade weil zur Erfassung psychischer Aspekte keine ausreichenden manifesten (direkt messbaren) Biomarker zur Verfügung stehen, sollte die Operationalisierung der Konstrukte kontinuierlich verbessert und vergleichbar gemacht werden.

Auf Basis der neu entwickelten SR-Itembank soll nach weiterer Validierung in unabhängigen Stichproben ein Computer-Adaptiver Test (SR-CAT) entwickelt werden. Durch die geringere Belastung der Proband*innen in der Beantwortung von CATs könnte der Einsatz im klinischen Alltag erleichtert werden, vorausgesetzt es stehen elektronische Geräte zur Datenerhebung, wie bspw. Smartphones zur Verfügung.

Zur umfassenden Analyse von Stress- und Stressresilienzprozessen ist langfristig eine Zusammenstellung verschiedener CATs zur Messung von Stressoren, Ressourcen, SR, Coping, sowie kurzfristigen (z.B. Anspannung, Freude) und langfristigen Stressreaktionen (z.B. psychische Gesundheit, soziale Funktionsfähigkeit, Lebensqualität) geplant. Zudem sollen die Items in andere Sprachen übersetzt werden, damit die Itembank international eingesetzt werden kann.

Ein weiterer wichtiger Vorteil der SR-Itembank liegt darin, dass sie ein dynamisches Instrument darstellt, das kontinuierlich durch die Hinzunahme weiterer Items verbessert werden kann [54].

Durch Einbezug von Items bereits etablierter statischer Resilienzinstrumente könnten diese auf einer gemeinsamen Metrik vergleichbar gemacht werden. Dies kann einen substanziellen Beitrag zur Integration von Forschungsergebnissen verschiedener Resilienzstudien und damit zum weiteren Erkenntnisgewinn bezüglich von Resilienzprozessen, die bisher noch nicht gut verstanden sind, liefern.

Das Ziel eines besseren Verständnisses des Resilienzkonstruktes und einer optimierten Operationalisierung sollte es jedoch nicht sein, die Verantwortung für Missstände in Gesellschaft und Arbeitswelt auf das Individuum zu projizieren. Dies könnte im Extremfall dazu führen, dass Belastungen immer weiter angezogen und Individuen dazu angehalten werden, an ihrer Resilienz zu arbeiten. Die häufig ausschließliche Fokussierung auf Leistungssteigerung, Kosteneffizienz und Selbstoptimierung, sowie der zunehmende Rückgriff auf Methoden des Neuroenhancements stellen gesellschaftspolitische Probleme dar, in deren Dienst sich die Resilienzforschung nicht stellen darf. Vielmehr sollte die neu entwickelte Stressresilienzitembank zur ganzheitlichen Erfassung von Stress- und Resilienzprozessen, zur Identifikation von Personen mit Unterstützungsbedarf und zur Evaluation von Resilienzförderungsmaßnahmen eingesetzt werden, während auf gesellschaftspolitischer Ebene die Schaffung adäquater sozialer Rahmenbedingungen und Arbeitskontexte, in der die Gesundheit und Lebensqualität des Individuums erhalten bleiben, im Vordergrund stehen sollten. *Eine weiterführende Diskussion der Ergebnisse findet sich im Abschnitt „Discussion“ auf den Seiten 8 bis 10 der Publikation.*

6. Schlussfolgerung

Die neu entwickelte 67-Item-Stressresilienzitembank und 10-Item-SR-Kurzform wiesen in einer psychosomatischen Stichprobe gute psychometrische Kriterien auf und ermöglichten eine valide Messung des zugrundeliegenden SR-Konstruktes. Weitere Validierungsstudien sind notwendig, um die Generalisierbarkeit auf andere Stichproben zu untersuchen. Weitere Schritte werden die Entwicklung eines SR-CATs, sowie die Entwicklung einer gemeinsamen Metrik sein, auf der zukünftig unterschiedliche Resilienzinstrumente verglichen werden können.

Literaturverzeichnis

- [1] Gradius JL. Prevalence and prognosis of stress disorders: a review of the epidemiologic literature. *Clin Epidemiol.* 2017;9:251-60.
- [2] Lupien SJ, McEwen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci.* 2009;10:434-45.
- [3] Sgoifo A, Montano N, Esler M, Vaccarino V. Stress, behavior and the heart. *Neurosci Biobehav Rev.* 2017;74(Pt B):257-9.
- [4] Hemmerle AM, Herman JP, Seroogy KB. Stress, depression and Parkinson's disease. *Exp Neurol.* 2012;233:79-86.
- [5] Glaser R, Kiecolt-Glaser JK. Stress-induced immune dysfunction: implications for health. *Nat Rev Immunol.* 2005;5:243-51.
- [6] Ladwig KH, Marten-Mittag B, Baumert J. Psychosoziale Belastungsfaktoren als Risiko für das Auftreten einer koronaren Herzerkrankung - Eine Bestandsaufnahme unter besonderer Berücksichtigung der KORA-Forschungsplattform. *Gesundheitswesen.* 2005;67:S86-SS93.
- [7] Brotman DJ, Golden SH, Wittstein IS. The cardiovascular toll of stress. *Lancet.* 2007;370:1089-100.
- [8] Aboa-Eboulé C, Brisson C, Maunsell E, Mâsse B, Bourbonnais R, Vézina M, Milot A, Thérioux P, Dagenais GR. Job strain and risk of acute recurrent coronary heart disease events. *JAMA.* 2007;298:1652-60.
- [9] Arck PC. Stress and pregnancy loss: role of immune mediators, hormones and neurotransmitters. *Am J Reprod Immunol.* 2001;46:117-23.
- [10] Fink G. Stress: Concepts, Definition and History. In: Reference Module in Neuroscience and Biobehavioral Psychology. Amsterdam: Elsevier; 2017. p.1-9.
- [11] Hapke U, Maske UE, Scheidt-Nave C, Bode L, Schlack R, Busch MA. [Chronic stress among adults in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1)]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz.* 2013;56:749-54.
- [12] Dragano N. Arbeit, Stress und krankheitsbedingte Frührenten : Zusammenhänge aus theoretischer und empirischer Sicht. Wiesbaden: VS Verlag für Sozialwissenschaften; 2007.
- [13] Lazarus RS, Folkman S. Transactional theory and research on emotions and coping. *Eur J Pers.* 1987;1:141-69.
- [14] Bonanno GA. Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events? *Am Psychol.* 2004;59:20-8.
- [15] Bitsika V, Sharpley CF, Bell R. The buffering effect of resilience upon stress, anxiety and depression in parents of a child with an autism spectrum disorder. *J Develop Phys Disabilities.* 2013;25:533-43.
- [16] Cleverley K, Kidd SA. Resilience and suicidality among homeless youth. *J Adolesc.* 2011;34:1049-54.
- [17] Geschwind N, Peeters F, Jacobs N, Delespaul P, Derom C, Thiery E, van Os J, Wichers M. Meeting risk with resilience: high daily life reward experience preserves mental health. *Acta Psychiatr Scand.* 2010;122:129-38.
- [18] Markovitz SE, Schrooten W, Arntz A, Peters ML. Resilience as a predictor for emotional response to the diagnosis and surgery in breast cancer patients. *Psychooncology.* 2015;24:1639-45.
- [19] Min JA, Yoon S, Lee CU, Chae JH, Lee C, Song KY, Kim TS. Psychological resilience contributes to low emotional distress in cancer patients. *Support Care Cancer.* 2013;21:2469-76.
- [20] Fletcher D, Sarkar M. Psychological Resilience. *Eur Psychol.* 2013;18:12-23.
- [21] Bernard C. Leçons sur les propriétés physiologiques et les altérations pathologiques des liquides de l'organisme. Paris: J.-B. Baillière et Fils; 1859.
- [22] Cannon WB. The wisdom of the body. New York: W W Norton & Co; 1932.
- [23] Selye H. The stress of life. New York: McGraw-Hill; 1956.
- [24] Selye H. Forty years of stress research: principal remaining problems and misconceptions. *CMAJ.* 1976;115:53-6.
- [25] Selye H. A syndrome produced by diverse nocuous agents. *Nature.* 1936;138:32.

- [26] Sterling P, Eyer J. Allostasis: A new paradigm to explain arousal pathology. In: Fisher S, Reason J, editors. *Handbook of life stress, cognition and health*. Oxford: John Wiley & Sons; 1988. p. 629-49.
- [27] Holmes TH, Rahe RH. The Social Readjustment Rating Scale. *J Psychosom Res*. 1967;11:213-8.
- [28] Dohrenwend BS, Dohrenwend BP. Some issues in research on stressful life events. *J Nerv Ment Dis*. 1978;166:7-15.
- [29] Dember WN. Motivation and the cognitive revolution. *Am Psychol*. 1974;29:161-8.
- [30] Smith CA, Lazarus RS. Emotion and adaptation. In: Pervin LA, editor. *Handbook of personality: Theory and research*. New York: The Guilford Press; 1990. p. 609-37.
- [31] Lazarus RS. From psychological stress to the emotions: A history of changing outlooks. *Annu Rev Psychol*. 1993;44:1-22.
- [32] Cohen S, Kessler RC, Gordon LU. Strategies for measuring stress in studies of psychiatric and physical disorders. In: Cohen S, Kessler RC, Gordon LU, editors. *Measuring stress: A guide for health and social scientists*. New York: Oxford University Press; 1995. p. 3-26.
- [33] Richardson GE. The metatheory of resilience and resiliency. *J Clin Psychol*. 2002;58:307-21.
- [34] Rutter M. Resilience as a dynamic concept. *Dev Psychopathol*. 2012;24:335-44.
- [35] Antonovsky A. The salutogenic perspective: Toward a new view of health and illness. *Advances*. 1987;4:47-55.
- [36] Luthar SS, Cicchetti D, Becker B. The construct of resilience: A critical evaluation and guidelines for future work. *Child Dev*. 2000;71:543-62.
- [37] Masten AS. Ordinary Magic: Resilience Processes in Development. *Am Psychol*. 2001;56:227-38.
- [38] Werner EE, Smith RS. *Journeys from Childhood to Midlife: Risk, Resilience, and Recovery*. Ithaca, NY: Cornell University Press; 2001.
- [39] Johnson JL, Wiechelt SA. Introduction to the special issue on resilience. *Subst Use Misuse*. 2004;39:657-70.
- [40] Werner EE, Smith RS. *Overcoming the odds: High risk children from birth to adulthood*. Ithaca, NY: Cornell University Press; 1992.
- [41] Garmezy N, Devine V. Project Competence: The Minnesota studies of children vulnerable to psychopathology. In: Watt NF, Anthony EJ, Wynne LC, Rolf JE, editors. *Children at risk for schizophrenia: A longitudinal perspective*. New York: Cambridge University Press; 1984. p. 289-303.
- [42] Block J, Kremen AM. IQ and ego-resiliency: conceptual and empirical connections and separateness. *J Pers Soc Psychol*. 1996;70:349.
- [43] Kobasa SC. *Stressful Life Events, Personality, and Health: An Inquiry Into Hardiness*. *J Pers Soc Psychol*. 1979;37:1-11.
- [44] Antonovsky A. *Unraveling the mystery of health: How people manage stress and stay well*. San Francisco: Jossey-Bass; 1987.
- [45] Ahern NR, Kiehl EM, Lou Sole M, Byers J. A review of instruments measuring resilience. *Issues Compr Pediatr Nurs*. 2006;29:103-25.
- [46] Windle G, Bennett KM, Noyes J. A methodological review of resilience measurement scales. *Health Qual Life Outcomes*. 2011;9:8.
- [47] Pangallo A, Zibarras L, Lewis R, Flaxman P. Resilience through the lens of interactionism: A systematic review. *Psychol Assess*. 2015;27:1-20.
- [48] Connor KM, Davidson JR. Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depress Anxiety*. 2003;18:76-82.
- [49] Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. *Int J Behav Med*. 2008;15:194-200.
- [50] Friborg O, Hjemdal O, Rosenvinge JH, Martinussen M. A new rating scale for adult resilience: what are the central protective resources behind healthy adjustment? *Int J Methods Psychiatr Res*. 2003;12:65-76.
- [51] Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the connor–davidson resilience scale (CD-RISC): Validation of a 10-item measure of resilience. *J Trauma Stress*. 2007;20:1019-28.
- [52] Wagnild G, Young H. Development and psychometric evaluation of the resilience scale. *J Nurs Meas*. 1993;1:165-78.

- [53] Bartone PT. Test-retest reliability of the dispositional resilience scale-15, a brief hardiness scale. *Psychol Rep.* 2007;101:943-4.
- [54] Rose M, Bjorner J, Becker J, Fries J, Ware JE. Evaluation of a Preliminary Physical Function Item Bank Supports the Expected Advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol.* 2008;61:17-33.
- [55] Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol.* 2010;63:1179-94.
- [56] Wahl I, Löwe B, Bjorner JB, Fischer F, Langs G, Voderholzer U, Aita SA, Bergemann N, Brähler E, Rose M. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. *J Clin Epidemiol.* 2014;67:73-86.
- [57] Embretson SE, Reise SP. *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum; 2000.
- [58] Reise SP, Waller NG. Item response theory and clinical measurement. *Annu Rev Clin Psychol.* 2009;5:27-48.
- [59] Cella D, Gershon R, Lai J-S, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res.* 2007;16 Suppl 1:133-41.
- [60] Bjorner JB, Chang C-H, Thissen D, Reeve BB. Developing tailored instruments: Item banking and computerized adaptive assessment. *Qual Life Res.* 2007;16:95-108.
- [61] Fischer HF, Rose M. www.common-metrics.org: a web application to estimate scores from different patient-reported outcome measures on a common scale. *BMC Med Res Methodol.* 2016;16:142.
- [62] Schalet BD, Cook KF, Choi SW, Cella D. Establishing a common metric for self-reported anxiety: linking the MASQ, PANAS, and GAD-7 to PROMIS Anxiety. *J Anxiety Disord.* 2014;28:88-96.
- [63] Schalet BD, Revicki DA, Cook KF, Krishnan E, Fries JF, Cella D. Establishing a common metric for physical function: Linking the HAQ-DI and SF-36 PF subscale to PROMIS® Physical Function. *J Gen Intern Med.* 2015;30:1517-23.
- [64] Rose M, Wahl I, Löwe B. *Computer Adaptive Tests in der Medizin.* *Psychother Psychosom Med Psychol.* 2013;63:48-54.
- [65] Victorson D, Tulskey DS, Kisala PA, Kalpakjian CZ, Weiland B, Choi SW. Measuring resilience after spinal cord injury: Development, validation and psychometric characteristics of the SCI-QOL Resilience item bank and short form. *J Spinal Cord Med.* 2015;38:366-76.
- [66] Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007;45:S22-S31.
- [67] Samejima F. The general graded response model. In: Nering ML, Ostini R, editors. *Handbook of polytomous item response theory models.* New York: Routledge; 2010. p. 77-107.
- [68] Nunnally J. *Psychometric theory.* 2nd Edit. . Hillsdale, NJ: McGraw-Hill; 1978.
- [69] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63:737-45.
- [70] Brown TA. *Confirmatory factor analysis for applied research.* 2nd ed. New York: The Guilford Press; 2015.
- [71] Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res.* 2009;18:447-60.
- [72] Reise SP, Scheines R, Widaman KF, Haviland MG. Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educ Psychol Meas.* 2013;73:5-26.
- [73] Nagelkerke NJ. A note on a general definition of the coefficient of determination. *Biometrika.* 1991;78:691-2.
- [74] R Development Core Team. *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing; 2008.

- [75] Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *J Stat Softw.* 2012;48:1-36.
- [76] Revelle W. psych: Procedures for psychological, psychometric, and personality research. Evanston, Illinois: Northwestern University; 2014.
- [77] Choi SW, Gibbons LE, Crane PK. Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *J Stat Softw.* 2011;39:1.
- [78] Chalmers RP. mirt: A multidimensional item response theory package for the R environment. *J Stat Softw.* 2012;48:1-29.
- [79] Mazza A, Punzo A, McGuire B. KernSmoothIRT: An R Package for Kernel Smoothing in Item Response Theory. *J Stat Softw.* 2014;58:1-34.
- [80] Wickham H. ggplot2. *Wiley Interdiscip Rev Comput Stat.* 2011;3:180-5.
- [81] Watson D. Correcting for acquiescent response bias in the absence of a balanced scale: An application to class consciousness. *Sociol Methods Res.* 1992;21:52-88.
- [82] Ray JJ. Reviving the problem of acquiescent response bias. *J Soc Psychol.* 1983;121:81-96.
- [83] Costa DSJ. Reflective, causal, and composite indicators of quality of life: A conceptual or an empirical distinction? *Qual Life Res.* 2015;24:2057-65.
- [84] Helmreich I, Lieb K. Schutzmechanismen gegen Burnout und Depression. *InFo Neurologie & Psychiatrie.* 2015;17:52-62.
- [85] Masten AS. Resilience Theory and Research on Children and Families: Past, Present, and Promise. *J Fam Theory Rev.* 2018;10:12-31.
- [86] Kalisch R, Müller MB, Tüscher O. A conceptual framework for the neurobiological study of resilience. *Behav Brain Sci.* 2014;38.
- [87] Lieb K. Resilienz – Paradigmenwechsel in der klinischen Forschung. *InFo Neurologie & Psychiatrie.* 2014;16:3.
- [88] Chmitorz A, Kunzler A, Helmreich I, Tüscher O, Kalisch R, Kubiak T, Wessa M, Lieb K. Intervention studies to foster resilience—A systematic review and proposal for a resilience framework in future intervention studies. *Clin Psychol Rev.* 2018;59:78-100.

B. Eidesstattliche Versicherung/Anteilserklärung

Eidesstattliche Versicherung

„Ich, Nina Obbarius, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „*Der Einsatz von Moderner Testtheorie zur Messung des Stressresilienzkonstruktes*“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen werden von mir verantwortet.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Betreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass mir die Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis bekannt ist und ich mich zur Einhaltung dieser Satzung verpflichte.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

Ausführliche Anteilserklärung an der erfolgten Publikation

Publikation 1: Obbarius N, Fischer F, Obbarius A, Nolte S, Liegl G, Rose M. A 67-item stress resilience item bank showing high content validity was developed in a psychosomatic sample., Journal of Clinical Epidemiology, August 2018

Beitrag im Einzelnen:

- Konzeptualisierung und Design der Studie
- Durchführung der Fokusgruppen mit allen Koautoren
- Eigenständige Organisation der Datenerhebung und Aufbereitung der Daten
- Datenauswertung und Interpretation der Ergebnisse unter Supervision
- Eigenständiges Erstellen aller Tabellen und Grafiken
- Eigenständiges Verfassen und Einreichen, sowie Revision der Publikation, Korrektur und Feedback durch alle Koautoren

Unterschrift, Datum und Stempel des betreuenden Hochschullehrers/der betreuenden Hochschullehrerin

Unterschrift des Doktoranden/der Doktorandin

C. Auszug aus der Journal Summary List

Journal Data Filtered By: **Selected JCR Year: 2017** Selected Editions: SCIE,SSCI
 Selected Categories: **“HEALTH CARE SCIENCES and SERVICES”**
 Selected Category Scheme: WoS
Gesamtanzahl: 94 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	BMJ Quality & Safety	4,293	7.226	0.016070
2	MILBANK QUARTERLY	3,552	6.000	0.005590
3	VALUE IN HEALTH	7,497	5.494	0.017360
4	HEALTH AFFAIRS	15,756	4.843	0.055270
5	ACADEMIC MEDICINE	14,301	4.801	0.025360
6	JOURNAL OF MEDICAL INTERNET RESEARCH	10,875	4.671	0.027410
7	JMIR mHealth and uHealth	1,418	4.541	0.004630
8	HEALTH TECHNOLOGY ASSESSMENT	5,630	4.513	0.011340
9	MEDICAL EDUCATION	9,440	4.405	0.011900
10	Implementation Science	7,206	4.345	0.017810
11	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	8,713	4.270	0.017580
12	JOURNAL OF CLINICAL EPIDEMIOLOGY	24,063	4.245	0.027230
13	JOURNAL OF GENERAL INTERNAL MEDICINE	17,822	4.005	0.028500
14	PHARMACOECONOMICS	4,255	3.989	0.007290
15	PALLIATIVE MEDICINE	4,636	3.780	0.008580
16	MEDICAL CARE	18,853	3.338	0.022590
17	JOURNAL OF HEALTH ECONOMICS	6,509	3.250	0.013920
18	JOURNAL OF PAIN AND SYMPTOM MANAGEMENT	9,734	3.249	0.013980
19	JOURNAL OF TELEMEDICINE AND TELE CARE	2,683	3.046	0.003930
20	MEDICAL DECISION MAKING	4,718	3.012	0.009230
21	INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS	4,584	2.957	0.006600
22	Journal of Patient Safety	785	2.683	0.002120
23	SUPPORTIVE CARE IN CANCER	10,484	2.676	0.024580
24	HEALTH SERVICES RESEARCH	6,994	2.667	0.014330
25	Patient-Patient Centered Outcomes Research	828	2.660	0.002620
26	INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE	4,172	2.554	0.004540

D. Druckversion der Publikation

Obbarius N, Fischer F, Obbarius A, Nolte S, Liegl G, Rose M. A 67-item stress resilience item bank showing high content validity was developed in a psychosomatic sample. *J Clin Epidemiol.* 2018 Aug;100:1-12. doi: 10.1016/j.jclinepi.2018.04.004. Epub 2018 Apr 10.

Impact Factor: 4,245 (2017)

(Rank 12/94 in HEALTH CARE SCIENCES & SERVICES)

Eigenfactor Score (2017): 0,027

ORIGINAL ARTICLE

A 67-item stress resilience item bank showing high content validity was developed in a psychosomatic sample

Nina Obbarius^{a,*}, Felix Fischer^a, Alexander Obbarius^a, Sandra Nolte^{a,b}, Gregor Liegl^a, Matthias Rose^{a,c}

^aCharité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charitplatz 1, Berlin 10117, Germany

^bPopulation Health Strategic Research Centre, School of Health and Social Development, Deakin University, 221 Burwood Highway, Burwood, VIC, Victoria 3125, Australia

^cDepartment of Quantitative Health Sciences, Outcomes Measurement Science, University of Massachusetts Medical School, 368 Plantation Street, The Albert Sherman Center, Worcester, MA 01605, USA

Accepted 4 April 2018; Published online 10 April 2018

Abstract

Objectives: To develop the first item bank to measure stress resilience (SR) in clinical populations.

Study Design and Setting: Qualitative item development resulted in an initial pool of 131 items covering a broad theoretical SR concept. These items were tested in $n = 521$ patients at a psychosomatic outpatient clinic. Exploratory and confirmatory factor analysis, as well as other state-of-the-art item analyses and item response theory were used for item evaluation and calibration of the final item bank.

Results: Of the initial item pool of 131 items, we excluded 64 items (54 factor loading < 0.5 , four residual correlations > 0.3 , two nondiscriminative item response curves, and four differential item functioning). The final set of 67 items indicated sufficient model fit in confirmatory factor analysis and item response theory analyses. In addition, a 10-item short form with high measurement precision ($SE \leq 0.32$ in a theta range between -1.8 and $+1.5$) was derived. Both the SR item bank and the SR short form were highly correlated with an existing static legacy tool (Connor-Davidson Resilience Scale).

Conclusion: The final SR item bank and 10-item short form showed good psychometric properties. When further validated, they will be ready to be used within a framework of computer-adaptive tests for a comprehensive assessment of the stress construct. © 2018 Elsevier Inc. All rights reserved.

Keywords: Stress; Resilience; Item response theory; Item bank development; Focus groups; Patient-reported outcomes

1. Introduction

There is ample research suggesting that stress plays a major role in the development and maintenance of several medical conditions [1–4]. However, some people thrive despite of adverse conditions and stay healthy in the face of stressors [5–7]. Still, to date, there is only little knowledge about why some people stay healthy in the face of adversity and others turn ill. Over the past years, there has been a shift from deficiency-based research to a more

resource-oriented approach [8]. Although this turn gave rise to research on protective characteristics, it is not yet well understood which factors are crucial for the maintenance of physical and mental health. Resilience is a relatively new construct that may fill this gap. It has been shown that more resilient individuals are less likely to develop mental disorders, such as depression or anxiety [9–11]. Resilience can be seen as a “dynamic process encompassing positive adaptation within the context of significant adversity” [12]. Windle [13] suggested three necessary requirements for resilience: (a) the presence of a significant adversity, (b) the presence of resources to compensate the effects of the adversity, (c) positive adaptation or the avoidance of a negative outcome. Fletcher and Sarkar [14] limited these requirements to two core concepts that form the mutual basis of different definitions of resilience: adversity and positive adaptation.

Conflict of interest: None.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author. Tel.: +4930450653890; fax: +4930450553900.

E-mail address: nina.obbarius@charite.de (N. Obbarius).

What is new?

Key findings

- The newly developed stress resilience (SR) item bank showed good psychometric properties and construct validity.

What this adds to what was known?

- This study is the first that developed an item bank assessing generic SR with modern test theory methods and target population involvement. SR item bank is the basis for computer-adaptive tests and for linking different tools measuring the SR construct on one common metric.

What is the implication and what should change now?

- The development of item banks and the establishment of a common metric by linking items from different instruments that aim to assess the same latent construct allow for better comparability of research findings. The newly developed SR item bank can therefore substantially advance the understanding and operationalization of the resilience construct.

Initial research into resilience first focused on protective factors, although it later turned to the underlying protective mechanisms and processes that help the individual to overcome adversity [12]. Fletcher and Sarkar [14] showed that resilience has been conceptualized in various different ways, i.e., as a trait, a process, or as an outcome. Hence, there is a lack of consistent conceptualization [15]. Therefore, resilience is difficult to operationalize. Various definitions led to the development of different questionnaires all claiming to measure resilience, and a number of reviews evaluated existing resilience questionnaires. As a result, there are inconsistent recommendations about which instrument should be used [16,17]. Recently, Windle, Bennett, and Noyes [17] suggested that the quality of these questionnaires may be regarded as only mediocre considering a range of quality criteria and, they also questioned the conceptual and theoretical adequacy of a number of the scales, concluding that there is no “gold standard” to measure resilience.

A major concern in the development of resilience scales is the frequent negligence of the target population in the developmental process, which generally leads to insufficient content validity of according measures [17]. This means that oftentimes, the item generation is purely theory driven, the qualitative process is not covered in much depth, and the target population is not involved in focus groups and pilot tests for item generation but only in final data

selection [17]. Moreover, it remains unclear whether or not resilience can be described by one global factor or by several subfactors. In addition, it is questionable whether or not these subfactors consistently exist in different populations. For example, the original five-factor solution for the Connor-Davidson resilience scale (CD-RISC) reported by Connor and Davidson [18] could not be replicated consistently [19].

The development of the different resilience questionnaires that exist today is mostly based on classical test theory [18,20–24]. Even if different instruments were claimed to measure the same latent construct, comparison between scores from these instruments would be difficult, as each instrument has its own metric. To overcome these problems of comparability, item response theory (IRT) methods can be used to calibrate any number of items measuring the same latent construct on a common scale [25–27], resulting in a construct-specific but instrument-independent “item bank” [28,29]. This approach has several advantages over classical test theory, including improved comparability between different subsets of items due to the established common metric, the possibility to develop computer-adaptive tests (CATs) and tailored short forms, and—when making use of CATs—a substantial reduction of response burden for patients [30,31]. A further advantage is that IRT-based item banks are dynamic tools, meaning that they can be continuously improved [28]. Therefore, the IRT approach could notably advance the operationalization of the concept of resilience, allowing not only for comparisons between studies but also for iterative refinement of the item bank by adding supplemental items. In the future, items of existing resilience measures such as the CD-RISC [18] and the RS-11 [20,32] could be linked to the established common stress resilience (SR) metric. The establishment of such a common metric covering items of several instruments has already been done for other constructs such as depression [29], anxiety [33], or physical function [34]. Recently, our group developed a web application allowing for direct comparison of scores from different depression, anxiety, or physical function instruments [35].

To date, two studies used modern test theory methods to develop or evaluate items aimed at assessing resilience. One study used IRT methods (two-parameter model) to develop resilience items to explore resilience in spinal cord injury patients [36]. Another study used a one-parameter model (RSM) to evaluate the psychometric properties at the item level of the CD-RISC [37].

In contrast to Victorson et al. [36], in the present study, it was not intended to assess resilience in a specific patient population but to develop items that work in different populations. As a result, patients with a wide range of health conditions that presented at the psychosomatic department of the medical clinic at the Charité were involved in the item generation process. This study aims to overcome aforementioned problems in the development of resilience questionnaires particularly the lack of comparability and

the negligence of target population involvement in the instrument development process. Our objectives were to develop a new SR scale thereby involving a clinical population in the item generation process and making use of modern test theory methods for item bank calibration. In detail, this article is aimed at advancing the understanding of resilience, presenting the findings of the development of a content valid item bank, and providing a SR short form.

2. Methods

We adopted an iterative process for item bank development following the guidelines of the Food and Drug Administration [38] and well-established standard procedures previously described in the literature [28–30,39]. Four

iterative steps (see Fig. 1) were performed that are described in detail below: (a) concept specification, (b) item pool generation and refinement, (c) data collection, and (d) psychometric item bank development and development of a short form. Conceptual work included focus groups with healthcare professionals and the patient population, as well as cognitive interviewing and field testing. After generation of the initial item pool, data was collected. Psychometric item bank development included exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), as well as classical item analyses and IRT methods. Dimensionality and further psychometric properties were explored. After successive elimination of items following well-established criteria [28–30], the remaining items were used for IRT-based item parameter estimation. Based on item parameters and content validity concerns, a short form was developed.

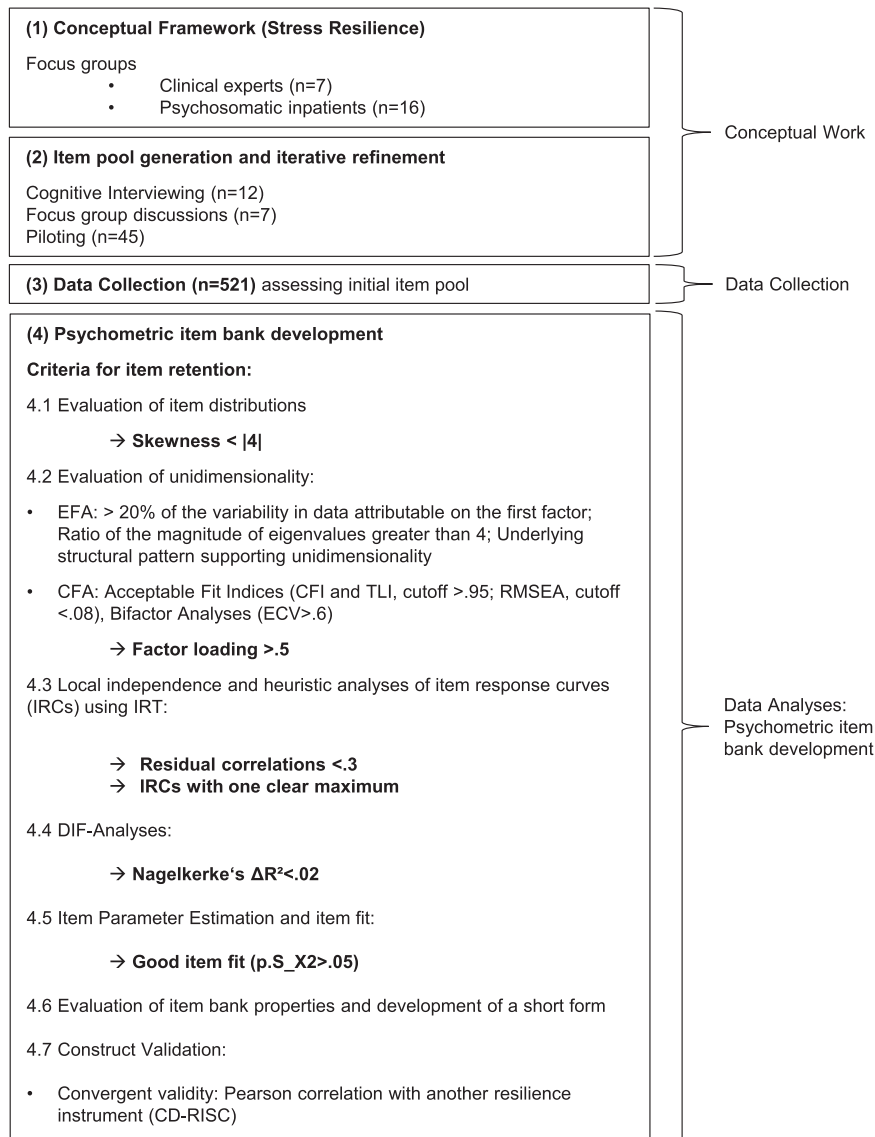


Fig. 1. Steps of item bank development.

2.1. Conceptual framework

To define a conceptual framework of SR, we conducted focus group sessions with health-care professionals and inpatients at the Department of Psychosomatic Medicine at Charité- Universitätsmedizin Berlin, Germany. The target sample entails favorable properties. Psychosomatic disorders range from mild adjustment disorders to highly chronic diseases. This allows diversity in resilience. Furthermore, patients admitting themselves for psychosomatic treatment are highly confronted with psychosocial stress factors. Focus groups with clinical experts (three physicians, three clinical psychologists, one health researcher) were held to specify the concept of SR. Several focus questions ranging from broader to more specific questions were provided to support group discussions in the psychosomatic inpatient focus groups ($n = 16$), i.e., “What does stress resilience mean to you?”.

2.2. Item pool generation and iterative refinement

Based on the aforementioned conceptual framework, an item pool was created by generating statements that capture the previously defined aspects of SR. The draft items were discussed in ten focus group sessions with clinical experts ($n = 7$) aiming at identification of gaps in concept specification, item wording, and the establishment of face validity as well as content validity. The concept specification and the initial 156 draft items were modified iteratively after each focus group session. To evaluate the understandability of the items, cognitive interviewing was performed in one-to-one settings ($n = 7$, five males, two females) and a group setting ($n = 5$, one male, four females) with the target population. During the iterative focus group sessions incorporating cognitive interviewing feedback, the entire item pool was systematically grouped by using “binning” (classification of items into [sub-] domains according to their content) and “winnowing” (elimination of items that lack face validity or are similar to better worded items), leading to an item pool of 131 items [40,41]. No items of existing scales were used due to the fact that the aim of our study was to create a new item bank assessing SR beginning with a thorough qualitative process, in which items were newly generated based on focus group input.

2.3. Data collection

The initial set of 131 items was pilot tested in an ad hoc sample of 45 psychosomatic inpatients (14 males, 31 females) who all agreed to complete a paper and pencil version of the item list. An additional response category was provided to allow highlighting questions that are not well understood. Item distributions were evaluated to identify items that showed extreme skewness.

The initial item bank was then included in the routine clinical assessment at which approximately 10% of the patients are not able to answer the questionnaires due to e.g.,

language barrier or visual impairment. Data assessment was run until approximately 500 patients were assessed.

Between June 2014 and June 2015, a total of 521 outpatients ($n = 521$) at the Department of Psychosomatic Medicine at Charité- Universitätsmedizin Berlin, Germany completed each of the 131 items electronically, using smartphones. In addition, socio-demographic data, as well as the ICD-symptom rating containing 29 items covering five syndrome scales (depression, anxiety, obsessive-compulsive, somatoform, and eating disorder) and a supplementary scale [42] were assessed, and respondents had to answer the 25-item version of the CD-RISC [18] on a 5-point rating scale referring to the last month. Higher scores indicate higher resilience.

2.4. Psychometric item bank development

Items were iteratively removed following the established criteria reported in detail below (see Fig. 1). Each removed item was reviewed for loss of content validity. The remaining items were calibrated using the IRT-graded response model.

2.4.1. Skewness

The frequency distribution of each item’s response options was evaluated, and items that displayed a skewness above an absolute value of four were eliminated.

2.4.2. Unidimensionality

Initial EFA was run to evaluate the dimensionality of the construct and identify those items that reflected the common construct best. The magnitude of the Eigenvalues and the underlying structural patterns were examined [39]. All items that loaded above 0.5 on the first unrotated factor in the EFA were retained, as they were assumed to have a sufficient contribution to the latent trait [29,43]. For the remaining items, Cronbach’s alpha as well as corrected item total correlations were calculated [44]. Subsequently, a one-factorial CFA was performed with the remaining items. We used the mean and variance adjusted weighted least squares estimator for model parameter estimation. We assessed Bentler’s comparative fit index (CFI), the Tucker-Lewis Index (TLI) (cutoff > 0.95), and the root mean square error of approximation (RMSEA) (cutoff < 0.08) as model fit criteria [45]. As traditional CFA fit indices have discussed to be misleading in item banks due to the large number of items [46], we estimated a bifactor model with one general factor representing SR and three exploratory group factors and analyzed the magnitude of explained common variance (ECV) of the general factor. An ECV above 0.6 indicates a strong general factor and sufficient unidimensionality of the data [47].

2.4.3. Local independence and heuristic analyses

To assure local independency, residual correlations were analyzed. When a pair of items showed a residual correlation above 0.3, they were considered to be indicative of

local dependence. Items showing more than one residual correlation above 0.3 and the items of the item pair that showed worse item response curves (IRCs) were eliminated. Item response curves were analyzed applying a nonparametric kernel-smoothing technique. Those items showing response curves with a clear maximum that separated well from the maximum of the other response option curves were retained. The response categories of poorly performing items were either pooled when the maximum of two response options overlapped or the items were removed due to unordered response categories.

2.4.4. Differential item functioning (DIF) analyses

To control for systematic error due to a group bias, an ordinal logistic regression model assessing DIF was applied for age, gender, and relationship status. The determination coefficient R^2 as defined by Nagelkerke [48] was calculated, and those items with $\Delta R^2 > 0.02$ indicating noticeable DIF were excluded. Additionally as post hoc analyses, we tested for DIF between patients with a certain diagnosis vs. those without this diagnosis to control for systematic error due to disease category. Disease categories were depression ($n = 271$), anxiety ($n = 82$), somatoform or pain disorder ($n = 192$), eating disorder including obesity ($n = 116$), and somatic disorder ($n = 220$).

2.4.5. Item parameter estimation and item fit

Item parameters of the remaining items were estimated using a generalized partial credit model [49]. IRT scores (θ , theta) represent estimates of the latent trait. The S_X^2 fit statistic (a Pearson X^2 statistic) was calculated, comparing the expected and observed frequencies of the item category responses for several levels of θ . A P -value < 0.05 indicates misfit of the item. Hence, items with a $p.S_X^2 < 0.05$ were removed.

2.4.6. Evaluation of item bank properties and short-form development

IRT model parameters can be used to calculate item information functions that depict the contribution of each item to the precision of the whole item bank or test for each level of theta. The test information of a subset of items or the whole item bank at a certain level of θ can be computed by the sum of the item information (ΣI) of several or all items at this θ level. The expected standard error (SE) can be calculated by the formula $SE(\theta) = \sqrt{1/\Sigma I(\theta)}$. The higher the item information, the higher the precision of the item or test at a certain θ . The reliability can be calculated by the formula $rel(\theta) = ABS(1 - SE[\theta]^2)$. Thus, for samples with a standard deviation of 1, a SE below 0.32 approximately corresponds to a classical internal consistency coefficient (i.e., Cronbach’s α) of 0.90 or better. Item information curves for each item were analyzed. Test information, SE, and reliability were calculated for different subsets of items. Based on these criteria as well as

evaluation of content validity, the best performing items were selected for the short form.

2.4.7. Construct validity

For preliminary convergent validation, Pearson correlations between the CD-RISC sum score and the full item bank and short form theta score were assessed. The CD-RISC was chosen for convergent validation as it was ranked as one of the highest regarding total quality assessment of different resilience scales by Windle [17] and assesses resilience as a personal quality [18].

Statistical exploratory analyses were conducted using IBM SPSS® Statistics 22. For subsequent analyses, R.3.3.2 [50] was used with the following packages: lavaan for CFA [51], psych for calculating bifactor models [52], lordif for DIF analyses [53], KernSmoothIRT and mirt for IRT model and parameter estimation [54,55], and ggplot2 [56] for graphics.

3. Results

3.1. Concept specification

During the iterative focus group sessions with health-care professionals, results from focus group sessions with the target population were integrated. The conceptual framework of SR was formulated. We defined resilience as a broad construct with the aim to capture different aspects of a positive aptitude in stress handling. We wanted to assess SR at a personal level and assumed that SR may change over the life course and can be enhanced by interventions. Based on our literature research and the results of the focus groups, we agree on the following definition of SR:

Stress resilience is the ability to maintain mental health in the face of adversity and consists of implicit security and the ability of flexible adaptation. Embedded in the stress concept of Lazarus [57–59] (Fig. 2) it reflects an attitude towards the world based on resources and former experiences that influences the way people appraise the stress

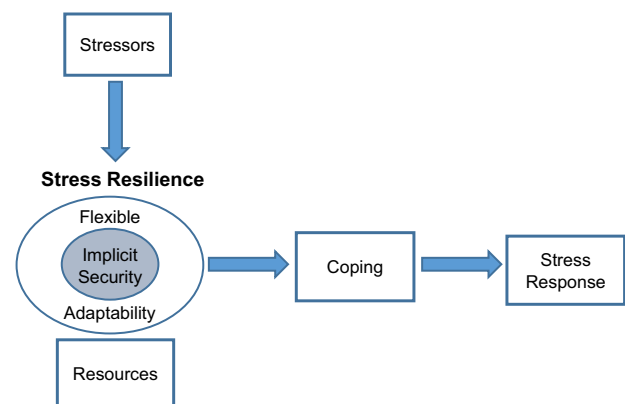


Fig. 2. Conceptual framework: stress resilience embedded in the stress concept of Lazarus.

situation, how they evaluate internal and external resources, how they cope with stress and how they subsequently respond to stress.

Lazarus suggested in his well-known transactional stress model that every individual processes stress differently depending on the appraisal of the external stressors and the internal and external resources. Reliant on the individual appraisal, different coping strategies are used leading to different stress responses. As resilience plays a major role in the guidance of appraisal and coping processes when confronted with stress, we embedded resilience in the stress concept of Lazarus. Hence, the term “stress resilience” is used in the present study.

The conceptual framework resulting from focus group discussions showed on the one hand implicit security and conditions that contributed to its development, namely important positive relationship experiences in childhood. On the other hand, there has to be a flexible adaptability to maintain homeostasis. In focus groups, we developed metaphorical images during the brain storming process. The image of a rock in the waves for internal strength and riding the wave for flexible adaptability was generated. We assume that SR is a combination of the two.

3.2. Item pool generation and iterative refinement

The initial 156 draft items that were generated based on the conceptual framework were reduced to 131 items and reformulated following cognitive interviewing with the target population and 10 iterative focus group discussions. An item tracking matrix was generated to document all the changes on the original items. Results of piloting did not lead to further changes of the item list as there were no consistent comprehension problems of items and distribution of response options was good. The resulting item pool contained 131 statements rated on a four-point Likert scale. The time referred to is the period of the last 7 days, except for some items on childhood experiences. Response options range from “not at all”, “a little bit”, “quite a bit”, to “very much”. Reverse worded items were included to control for acquiescence bias [60,61]. Higher scores indicate higher resilience. There was no missing data as an electronic data collection method was used that did not allow for skipping any of the items.

3.3. Sample characteristics

The item pool containing 131 items was assessed in 521 psychosomatic outpatients. Gender proportion was representative of psychosomatic population with two-thirds being female patients. Average age was 42 years (Table 1).

3.4. Psychometric item bank development

The abbreviated stems of the retained items following EFA, their factor loadings and successive removal criteria following CFA and IRT analyses are listed in Table A1 on

Table 1. Sample characteristics

n = 521 Psychosomatic outpatients	
Age in years (mean/SD)	42.12 (14.78)
Gender (% female)	68.8
Nationality (% German)	92.3
Employment status (% working)	52.5
Education (%)	
University entrance diploma (German “Abitur”)	48.1
Secondary school	48.3
No graduation	3.7
Marital status (%)	
Single	55.0
Married	31.0
Divorced	11.0
Widowed	3.1
Living in partnership (%)	52.9
ISR (mean [SD])	
Depression	1.93 (1.00)
Anxiety	1.63 (1.17)
Obsessive-compulsive	1.27 (1.10)
Somatization	1.30 (1.06)
Eating disorder	0.92 (1.19)
ISR sum score	1.31 (0.68)

Abbreviations: SD, standard deviation; ISR, ICD-10-symptom ranking.

Cutoff scores for mild symptoms are: 1 for depression, anxiety, and obsessive-compulsive syndrome scales, 0.75 for somatization, 0.67 for eating disorder, and 0.6 for the ISR sum score.

journal’s web site at www.elsevier.com in the online only appendix.

3.4.1. Skewness

All items showed good distribution parameters with a skewness below an absolute value of four. Hence, no item had to be excluded because of extreme skewness.

3.4.2. Unidimensionality

The underlying pattern structure in the EFA, the ratio above four of the first and the second Eigenvalue and more than 20% of the variability of the data accounted for by the first factor supported the unidimensionality assumption [39]. Seventy-seven of the 131 items were retained as they loaded higher than 0.5. On the first unrotated factor of the EFA, Cronbach’s alpha of the 77-item bank was 0.976, and all item total correlations were above 0.4 indicating strong association between the items and the latent construct. A one-factorial CFA with the retained 77 items indicated deviation from a unidimensional model (chi-square = 7,933 [df = 2,849], $P < 0.001$), but RMSEA showed a sufficient model fit (RMSEA = 0.059 [0.057–0.060]), whereas CFI and TLI were slightly below 0.9 (CFI = 0.89, TLI = 0.88). However, the use of traditional fit indices to

support unidimensionality in patient-reported outcomes has been discussed controversially [62,63]. More exploratory based methods such as the ECV in bifactorial analyses were recommended to test the unidimensionality assumption [46,47]. A bifactor model with three exploratory group factors supported the unidimensionality assumption with an ECV = 0.68.

3.4.3. Local independence and heuristic analyses

Five item pairs had a residual correlation above 0.3. We eliminated two items because of multiple violation of local independency: SRI88 (“being more misfortunate than others”) and SRI129 (“often being annoyed that life is so unfair to me”). Of the item pairs that indicated local dependence SRI38 (“knowing I matter to others”) and SRI60 (“ability to structure my everyday life”) were eliminated as they showed worse IRCs than their counterpart items. Analysis of the remaining IRCs indicated two more items that had to be eliminated because they displayed no clear maximum that separated well from the maximum of the other response option curves: SRI19 (“ability to express how I feel”) and SRI91 (“feeling particularly misfortunate”). In Figure 3, we show examples of an item that illustrates favorable IRCs (SRI80) and one that illustrates less favorable IRCs (SRI91), respectively. Furthermore, we had to collapse response options one and two for items SRI03, SRI17, SRI33, SRI39, SRI59, SRI70, SRI77, SRI93, SRI97, and SRI111; response options two and three for SRI78; and response options three and four for SRI50 and SRI56 because of unordered response categories.

3.4.4. Differential item functioning

Ordinal logistic regressions indicated DIFs for gender and age (<40 years vs. ≥40 years) with a $\Delta R^2 > 0.02$. Two items were removed because of DIF for gender: Items SRI52 (“ability to express my emotions”; $\Delta R^2 = 0.025$) and SRI64 (“good ability to structure in demanding situations”; $\Delta R^2 = 0.021$) each showed a uniform DIF. Two items were excluded because they displayed DIF for age:

Item SRI42 (“confidence in handling difficult situations”) showed a uniform DIF ($\Delta R^2 = 0.023$) and item SRI94 (“ability to handle unpredictable events”) was flagged for nonuniform DIF ($\Delta R^2 = 0.021$). There were no items that indicated DIF for partnership status. Three items showed uniform DIF for eating disorder including obesity: item SRI20 (“awareness of what is beneficial to me and what is not”; $\Delta R^2 = 0.024$), item SRI40 (“contentment with myself”; $\Delta R^2 = 0.030$), and item SRI45 (“feeling comfortable in my body”; $\Delta R^2 = 0.073$). No items indicated DIF for depression, anxiety, somatoform or pain disorder, and somatic disorder. As the DIF analyses for disease categories were run post hoc, the three items flagged for DIF were kept in the item bank but excluded from CAT and short-form development.

3.4.5. Item parameter estimation and item fit

Model parameters were estimated for the remaining 67 items including 13 collapsed items. The S-X² model fit statistics were calculated. As all items showed a good fit ($P > 0.05$), no further items had to be excluded. Consequently, the final item bank contained 67 items. Twenty of the remaining items are negatively worded and therefore reversely scored. The full item bank showed good measurement precision ($SE \leq 0.32$) in a theta range from -3.4 to 3.6.

3.4.6. Item bank properties and development of a 10-item short form

Item information functions were analyzed for each of the 67 remaining items. Maximum item information ranged from 0.29 to 1.76. Item SRI41 captured maximum item information of 1.76 at a θ of -0.5. Maximum test information of the whole item bank was 44.82 at a θ of -0.3. ($SE = 0.15$, approximate classical reliability of 0.98). Test information, SE, reliability, and content validity were compared between different subsets of items with the best item information curves. Based on these analyses, 10 core items were identified to form a 10-item SR short form (Table 2). Three of the 10 items are reversely scored. The

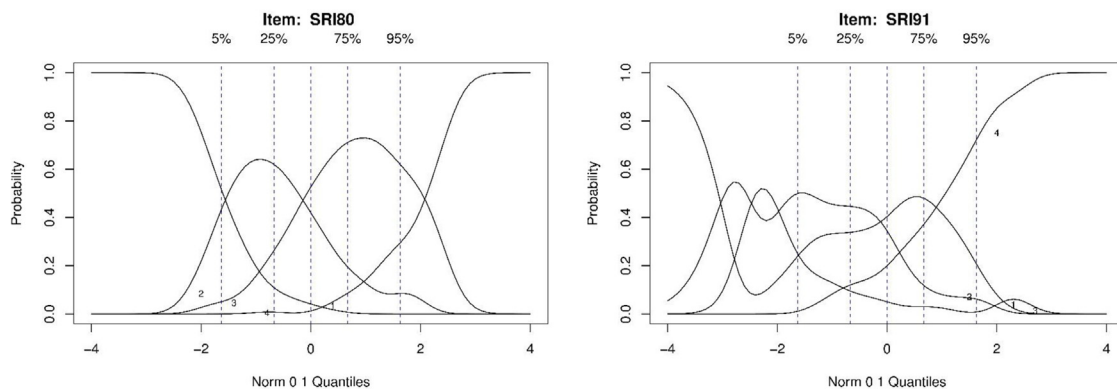


Fig. 3. Item SRI80 and item SRI91 illustrating favorable (left side) and less favorable (right side) item response curves.

Table 2. Short form properties

Item ID	Item stem	EFA F.load. ^a	CFA F.load.	Skewness	Slope ^b	Step 1	Step 2	Step 3	Fit P-value	I _{max} at θ^c
SRI02	Inner unshakeable security	0.67	0.74	0.26	2.00	-0.89	0.64	2.40	0.16	1.07 (0.6)
SRI11	Multiple things throw me off course ^d	0.68	0.73	-0.02	1.96	-1.48	0.05	1.75	0.35	1.03 (0)
SRI29	Shaping relationships to feel good	0.64	0.66	-0.28	1.59	-2.01	-0.52	1.06	0.25	0.72 (-0.6)
SRI41	Believing in myself	0.73	0.80	-0.32	2.58	-1.73	-0.49	1.02	0.15	1.76 (-0.5)
SRI57	Difficulty in handling my emotions ^d	0.66	0.70	-0.11	1.83	-1.61	-0.19	1.24	0.40	0.94 (-0.2)
SRI80	Facing everyday life's challenges	0.70	0.77	-0.18	2.21	-1.51	-0.08	1.82	0.18	1.29 (-0.1)
SRI84	Putting into practice what is beneficial to me	0.71	0.77	-0.11	2.27	-2.04	-0.32	1.51	0.49	1.33 (-0.3)
SRI89	Even when I cannot control everything, I get along	0.65	0.71	0.03	1.92	-1.66	0.04	1.79	0.21	0.98 (0)
SRI114	Knowing how to deal with burdensome situations	0.65	0.73	-0.30	2.07	-1.76	-0.06	2.37	0.10	1.11 (-0.1)
SRI132	Doubting myself permanently ^d	0.69	0.75	0.14	2.07	-1.01	0.18	1.41	0.47	1.21 (0.2)

Abbreviations: EFA, exploratory factor analysis; F.load, factor loading; CFA, confirmatory factor analysis.

^a Factor loading on the first unrotated factor in EFA.

^b Slope: item discrimination, step 1–3: response category intersections.

^c Maximum of the item information function (left number in each cell) at a particular theta (in brackets).

^d For IRT analyses, all items were inverted (higher scores = higher resilience).

mean sum score of the short form was $M = 25.16$ ($SD = 5.98$). To standardize the short form sum scores, they were calibrated on a T-metric with a mean of 50 reflecting the mean of the psychosomatic population and an SD of 10 (Table 3). The short form showed a high measurement precision ($SE \leq 0.32$) in a theta range between -1.8 and $+1.5$ (see Fig. 4). The Pearson correlation between the theta score of the full item bank and the short form was $r = 0.95$ ($P < 0.001$).

3.4.7. Construct validity

The SR item bank and the short form theta scores showed a Pearson correlation of 0.83 and 0.79 with the CD-RISC, respectively, ($P < 0.001$), confirming convergent validity of the new instrument.

4. Discussion

We developed a new content valid SR item bank and a 10-item SR short form based on modern test theory methods and involvement of the target population. The item bank development followed iterative qualitative steps and comprehensive statistical analyses to assure the identification of the most essential items assessing SR. The final item bank including 67 items and the 10-item short form showed good psychometric properties as well as high construct validity. After further validation in independent samples, on the basis of the SR item bank, SR CATs could be employed, and population-specific subsets of items could be administered resulting in substantial reduction of response burden and higher measurement precision.

The factor analysis methods that were used for data analysis assume a reflective measurement model,

presuming an underlying latent construct that influences the way the items are answered. Hence, the items are assumed to be effective indicators that are highly correlated and interchangeable [44,64]. However, one might argue that a concept as SR might be viewed as a formative construct, defined by items, which are not necessarily correlated but representing different aspects of SR. Our aim was not to generate an instrument that captures a mixture of different aspects that are uncorrelated, e.g., different coping strategies or different reactions to different stress exposures, but to obtain an instrument that is able to assess a distinct underlying core construct.

As SR is a broad construct, operationalization seems quite complicated, and the differentiation to other related concepts is difficult. Therefore, we paid attention to comprehensive conceptual work and carefully developed the items in several iterative steps. We involved psychosomatic patients as well as health-care professionals to ensure content validity of the newly developed items. In addition, items were qualitatively reviewed by cognitive interviewing and field testing. In the beginning of our conceptual work, we decided to not restrict the brain storming process in the focus groups. This was to ensure that all different facets of SR would be tapped. We followed well-established inclusion and exclusion criteria of items. Sixty-four items were removed from the initial item pool of 131 items after careful review for content validity loss because they did not conform to predefined psychometric criteria. This process led to a more narrowly defined construct of SR. Assuming a reflective measurement model, the aim was to construct an item bank that captures the underlying core concept of SR. Items that were excluded based on psychometric criteria covered among other things the childhood experiences that can probably be seen as preconditions for the

Table 3. SF sum scores, thetas, T-scores, and standard errors

SF sum score	SF theta	SE theta	SF T-score	SE T-score
10	-2.72	0.46	22.76	4.64
11	-2.38	0.39	26.24	3.93
12	-2.12	0.36	28.80	3.58
13	-1.90	0.34	30.95	3.37
14	-1.71	0.32	32.87	3.24
15	-1.54	0.32	34.64	3.17
16	-1.37	0.31	36.31	3.13
17	-1.21	0.31	37.90	3.10
18	-1.06	0.31	39.44	3.08
19	-0.91	0.31	40.94	3.07
20	-0.76	0.31	42.41	3.05
21	-0.61	0.30	43.85	3.04
22	-0.47	0.30	45.29	3.03
23	-0.33	0.30	46.72	3.03
24	-0.18	0.30	48.15	3.04
25	-0.04	0.30	49.59	3.05
26	0.10	0.31	51.04	3.07
27	0.25	0.31	52.51	3.09
28	0.40	0.31	54.00	3.11
29	0.55	0.31	55.52	3.14
30	0.71	0.32	57.07	3.15
31	0.87	0.32	58.67	3.17
32	1.03	0.32	60.32	3.18
33	1.20	0.32	62.01	3.18
34	1.38	0.32	63.76	3.20
35	1.56	0.32	65.58	3.23
36	1.75	0.33	67.50	3.30
37	1.96	0.34	69.57	3.42
38	2.19	0.36	71.88	3.62
39	2.46	0.40	74.62	3.97
40	2.82	0.46	78.24	4.64

Abbreviations: SF, short form; SE, standard error.

development of SR. The excluded items, as e.g., “In my childhood, I was hugged, when I needed it.” loaded less than 0.5 on the extracted SR factor. This could be a methodological artifact, as the items asking for childhood experiences all refer to a different time frame, or, more probably, these items capture a different construct, which we would label “attachment experiences”. The importance of close relations to supportive adults for resilience has been shown by various studies [12,65,66]. Werner and Smith [66] described shared protective factors of children, who managed to adapt well to risk factors, such as poverty, and concluded that among other factors, these children developed at least one stable close relationship to a role model or had a healthy attachment to a caregiver. However, it has been criticized that the trajectories and dynamic interplay between the identified protective factors in resilience

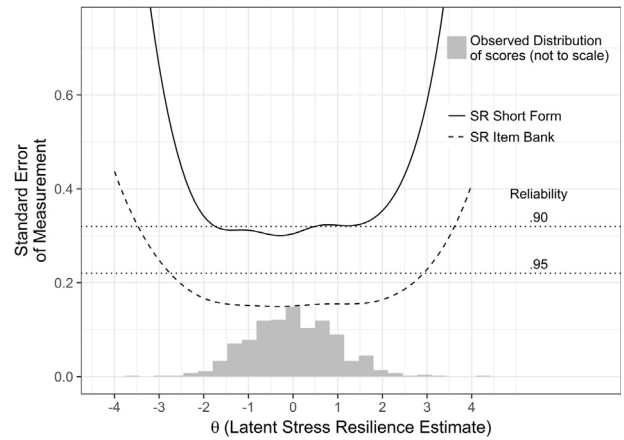


Fig. 4. Standard error of measurement, reliability and theta for the 10-item short form, and the full item bank as well as observed theta distribution (not to scale).

research are still not well understood [12,65]. We hypothesize that attachment experiences are preconditions for the development of SR but capture a distinct construct. The generalizability of the final item bank to children has to be explored in further studies, as the target population involved in the SR item bank development consisted of adults, only. In line with our attempt to focus on a latent core construct of SR, most items that captured extrinsic factors, such as “having at least one good friend to rely on” were eliminated, whereas more intrinsic items, such as “shaping relationships to feel good” were retained. Items capturing a healthy life style, such as “sleeping sufficiently” were eliminated based on the predefined psychometric criteria. A healthy life style could be hypothesized to be an effect and not a facet of SR. Furthermore, healthy life styles would fit more to a formative measurement model, than to a reflective measurement model. Yet, a clear distinction between antecedents, facets, and effects of SR can only be made by running longitudinal studies. Our previously defined framework states that resilience influences the appraisal processes but is different to them. Items that were not neatly in line with this framework by asking for positive stress appraisal were eliminated as well. To control for acquiescence bias, we initially generated negatively worded items. Twenty-three of the excluded items were reverse scored generally asking for harsh handling with oneself. The 67 items of the final item bank and the 10-item short form reflect the previously defined combination of inner security and flexible adaption. They seem to operationalize the underlying SR construct the best. In addition, construct validity was confirmed by the high correlation with the CD-RISC.

The item bank development took place in a single site, with a modestly sized sample. Therefore, further validation is needed to evaluate the generalizability to different samples and contexts. Moreover, as the different disease categories showed relatively low sample sizes, a larger

sample will be needed to evaluate DIF more elaborately across different disorder categories. Only then, the validity of the model parameters for different disorder categories can be fully shown. It seems reasonable that the item “feeling comfortable in my body” is answered systematically in another way in patients with the same SR level showing eating disorders and obesity as patients without these disorders. Yet, this has to be evaluated in a larger sample of eating disorder and obesity patients.

In addition, as the fit indices of the unidimensional model did not meet all established cutoffs, the measurement model should be confirmed in different populations with respect to its nature. The question if the final item bank is valid can only be answered in future by running longitudinal studies assessing the predictive validity in different samples.

The validation questionnaire (CD-RISC) has a different time frame (last month vs. last 7 days) and another response format (five-point-scale vs. four-point scale) compared with the SR item bank. The different time frame could bias the correlation, whereas recent research suggests that the response format does not affect the latent construct [67]. Correlation coefficients were highly indicating that they operationalize the same construct. Validation of a newly developed instrument with other already established instruments of the same underlying construct has to be seen critically. Very high correlations make the relevance of the newly developed instrument questionable, whereas too low correlations raise concerns about the construct validity of the newly developed instrument. It would be desirable to design longitudinal studies to evaluate the predictive validity and clinical relevance of SR e.g., by exploring the effect of resilience on other outcome measures such as the change in mental health scores before and after a significant adversity to avoid this trade-off. Furthermore, based on our results, we cannot make a statement about the sensitivity for change of the SR items. Hence, the SR item bank has to be further validated in longitudinal studies, e.g., with an analysis of the effect of psychotherapy on SR. As item development was performed including health-care professional input, we think that items also cover the higher range of SR. Yet, the target population did not include highly stress resilient people because by definition, psychosomatic patients all have developed some kind of disorder in the face of adversity showing difficulty in stress adaptation. Therefore, a ceiling effect cannot be ruled out. It would be interesting to run studies that evaluate whether or not the SR items also work with people that are very stress resilient.

Although the study provides preliminary support for the psychometrical soundness of the SR item bank, future studies should examine the model robustness. Exploration of convergent and divergent validity with a multitrait-multimethod approach seems to be useful to further establish construct and criterion validity. At the moment, data collection for the SR item bank is in progress allowing for further comprehensive IRT analyses and supplementary

validation. Structural equation modeling with SR scores and different patient-reported outcomes assessing internal and external resources, perceived stress, coping and stress reaction as well as mental health will have to be performed to evaluate the conceptual embedment of SR in the stress concept of Lazarus.

As a dynamic tool, the SR item bank can be continuously improved by adding new items and therefore significantly advance the construct operationalization. A great advantage of item banks is the possibility to compare scores that are measured with different subsets of items. In addition, CATs can be administered. This enables higher measurement precision by applying only the most informative items based on the respondent’s answers and decrease of respondent burden compared to the existing fixed-length questionnaires. The design and implementation of an SR-CAT are part of the next steps. The long-term aim is to develop a common metric for SR as it has been already done for other constructs such as depression [29], anxiety [33], and physical function [34] and to link different resilience measures to this metric. This allows for a better comparability across studies [68]. Hence, consolidation of research results across countries and research groups could progress. This could push knowledge about this construct that is yet not well defined a great step ahead. Comparing results of different populations and relationships to other constructs could help us better understand what resilience is and how it works. Based on a better understanding and operationalization of the resilience construct, future work could focus on the design and evaluation of resilience enforcing interventions. To date, items are only available in German. Aiming at international validation and implementation of the SR item bank, translation of items into further languages is planned.

5. Conclusions

The newly developed SR item bank allows for valid measurement of the underlying SR construct in a psychosomatic sample. To ensure content validity and utility of the newly developed items, study methods included iterative focus group discussions, target population feedback, and field testing. Following EFA, CFA, and IRT-analyses, 67 items were retained for the final item bank and a 10-item short form was generated. Initial analyses found good psychometric properties and confirmed the validity of the SR item bank. Yet, further validation in independent samples is needed to evaluate the generalizability of the SR item bank. Full German items can be provided by the authors on request.

Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.jclinepi.2018.04.004>.

References

- [1] Sgoifo A, Montano N, Esler M, Vaccarino V. Stress, behavior and the heart. *Neurosci Biobehav Rev* 2017;74(Pt B):257–9.
- [2] Hemmerle AM, Herman JP, Seroogy KB. Stress, depression and Parkinson's disease. *Exp Neurol* 2012;233:79–86.
- [3] Burns MN, Nawacki E, Kwasny MJ, Pelletier D, Mohr DC. Do positive or negative stressful events predict the development of new brain lesions in people with multiple sclerosis? *Psychol Med* 2014;44:349–59.
- [4] Glaser R, Kiecolt-Glaser JK. Stress-induced immune dysfunction: implications for health. *Nat Rev Immunol* 2005;5:243–51.
- [5] Luthar SS. Resilience and vulnerability: Adaptation in the context of childhood adversities. Cambridge: Cambridge University Press; 2003.
- [6] Masten AS. Ordinary magic: resilience processes in development. *Am Psychol* 2001;56:227.
- [7] Bonanno GA. Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events? *Am Psychol* 2004;59:20.
- [8] Richardson GE. The metatheory of resilience and resiliency. *J Clin Psychol* 2002;58:307–21.
- [9] Bitsika V, Sharpley CF, Bell R. The buffering effect of resilience upon stress, anxiety and depression in parents of a child with an autism spectrum disorder. *J Develop Phys Disabilities* 2013;25: 533–43.
- [10] Sharpley CF, Bitsika V, Wootten A, Christie DR. Does resilience 'buffer' against depression in prostate cancer patients? A multi-site replication study. *Eur J Cancer Care* 2014;23:545–52.
- [11] Fredrickson BL, Tugade MM, Waugh CE, Larkin GR. What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *J Pers Soc Psychol* 2003;84:365.
- [12] Luthar SS, Cicchetti D, Becker B. The construct of resilience: a critical evaluation and guidelines for future work. *Child Dev* 2000;71: 543–62.
- [13] Windle G. What is resilience? A review and concept analysis. *Rev Clin Gerontol* 2011;21:152–69.
- [14] Fletcher D, Sarkar M. Psychological resilience. *Eur Psychol* 2013; 18:12–23.
- [15] Davydov DM, Stewart R, Ritchie K, Chaudieu I. Resilience and mental health. *Clin Psychol Rev* 2010;30:479–95.
- [16] Ahern NR, Kiehl EM, Lou Sole M, Byers J. A review of instruments measuring resilience. *Issues Compr Pediatr Nurs* 2006;29:103–25.
- [17] Windle G, Bennett KM, Noyes J. A methodological review of resilience measurement scales. *Health Qual Life Outcomes* 2011;9:8.
- [18] Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson resilience scale (CD-RISC). *Depress Anxiety* 2003;18:76–82.
- [19] Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the Connor-Davidson resilience scale (CD-RISC): validation of a 10-item measure of resilience. *J Traumac Stress* 2007;20:1019–28.
- [20] Wagnild G, Young H. Development and psychometric evaluation of the resilience scale. *J Nurs Meas* 1993;1:165–78.
- [21] Bartone PT. Test-retest reliability of the dispositional resilience scale-15, a brief hardiness scale. *Psychol Rep* 2007;101:943–4.
- [22] Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. *Int J Behav Med* 2008;15:194–200.
- [23] Block J, Kremen AM. IQ and ego-resiliency: conceptual and empirical connections and separateness. *J Pers Soc Psychol* 1996;70:349.
- [24] Friborg O, Hjemdal O, Rosenvinge JH, Martinussen M. A new rating scale for adult resilience: what are the central protective resources behind healthy adjustment? *Int J Methods Psychiatr Res* 2003;12:65–76.
- [25] Embretson SE, Reise SP. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum; 2000.
- [26] Reise SP, Waller NG. Item response theory and clinical measurement. *Annu Rev Clin Psychol* 2009;5:27–48.
- [27] Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* 2010;63:1179–94.
- [28] Rose M, Bjorner JB, Becker J, Fries J, Ware J. Evaluation of a preliminary physical function item bank supported the expected advantages of the patient-reported outcomes measurement information system (PROMIS). *J Clin Epidemiol* 2008;61:17–33.
- [29] Wahl I, Löwe B, Bjorner JB, Fischer F, Langs G, Voderholzer U, et al. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. *J Clin Epidemiol* 2014;67:73–86.
- [30] Bjorner JB, Kosinski M, Ware JE. Computerized adaptive testing and item banking. In: Fayers P, Hays R, Eds. Assessing quality of life in clinical trials. 2nd ed. Oxford: Oxford University Press; 2005:95–112.
- [31] Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS physical function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol* 2014;67:516–26.
- [32] Schumacher J, Leppert K, Gunzelmann T, Strauß B, Brähler E. Die Resilienzskala—ein Fragebogen zur Erfassung der psychischen Widerstandsfähigkeit als Personmerkmal. *Z Klin Psychol Psychiatr Psychother* 2005;53:16–39.
- [33] Schalet BD, Cook KF, Choi SW, Cella D. Establishing a common metric for self-reported anxiety: linking the MASQ, PANAS, and GAD-7 to PROMIS Anxiety. *J Anxiety Disord* 2014;28:88–96.
- [34] Schalet BD, Revicki DA, Cook KF, Krishnan E, Fries JF, Cella D. Establishing a common metric for physical function: linking the HAQ-DI and SF-36 PF subscale to PROMIS® Physical Function. *J Gen Intern Med* 2015;30:1517–23.
- [35] Fischer HF, Rose M. www.common-metrics.org: a web application to estimate scores from different patient-reported outcome measures on a common scale. *BMC Med Res Methodol* 2016;16:142.
- [36] Victorson D, Tulskey DS, Kisala PA, Kalpakjian CZ, Weiland B, Choi SW. Measuring resilience after spinal cord injury: development, validation and psychometric characteristics of the SCI-QOL resilience item bank and short form. *J spinal cord Med* 2015;38:366–76.
- [37] González VBA, Sierra MTC, Martínez BA, Martínez-Molina A, Ponce FP. An in-depth psychometric analysis of the Connor-Davidson resilience scale: calibration with Rasch-Andrich model. *Health Qual Life Outcomes* 2015;13:154.
- [38] Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006;4:79.
- [39] Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). *Med Care* 2007;45: S22–31.
- [40] Ravens-Sieberer U, Devine J, Bevans K, Riley AW, Moon J, Salsman JM, et al. Subjective well-being measures for children were developed within the PROMIS project: presentation of first results. *J Clin Epidemiol* 2014;67:207–18.
- [41] Tamm N, Devine J, Rose M. Item generation and construction of questionnaires In: Guillemin F, Leplege A, Brianccon S, Spitz E, Coste J, editors. Perceived health and adaptation in chronic disease. Boca Raton: CRC Press, 2017.
- [42] Tritt K, von Heymann F, Zaudig M, Zacharias I, Söllner W, Loew T. Entwicklung des Fragebogens »ICD-10-Symptom-Rating« (ISR). *Z Psychosom Med Psychother* 2008;54:409–18.
- [43] Nunnally J. Psychometric theory. 2nd Ed. Hillsdale, NJ: McGraw-Hill; 1978.

- [44] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- [45] Brown TA. *Confirmatory factor analysis for applied research*. 2nd ed. New York: The Guilford Press; 2015.
- [46] Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res* 2009;18:447–60.
- [47] Reise SP, Scheines R, Widaman KF, Haviland MG. Multidimensionality and structural coefficient bias in structural equation modeling: a bifactor perspective. *Educ Psychol Meas* 2013;73:5–26.
- [48] Nagelkerke NJ. A note on a general definition of the coefficient of determination. *Biometrika* 1991;78:691–2.
- [49] Muraki E. A generalized partial credit model - application of an EM algorithm. *Appl Psychol Meas* 1992;16:159–76.
- [50] R Development Core Team. *R: a language and environment for statistical computing*. Vienna: R foundation for statistical computing 2012.
- [51] Rosseel Y. lavaan: an R package for structural equation modeling. *J Stat Softw* 2012;48:1–36.
- [52] Revelle W. *psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University; 2014:165.
- [53] Choi SW, Gibbons LE, Crane PK. lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *J Stat Softw* 2011;39:1.
- [54] Chalmers RP. mirt: A multidimensional item response theory package for the R environment. *J Stat Softw* 2012;48:1–29.
- [55] Mazza A, Punzo A, McGuiire B. KernSmoothIRT: an R package for kernel smoothing in item response theory. *J Stat Softw* 2014;58:1–34.
- [56] Wickham H. ggplot2. *Wiley interdisciplinary reviews: computational statistics* 2011;:180–5.
- [57] Lazarus RS, Folkman S. Transactional theory and research on emotions and coping. *Eur J Pers* 1987;1:141–69.
- [58] Lazarus RS. *Psychological stress and the coping process*. New York: McGraw-Hill; 1966.
- [59] Lazarus R, Folkman S. *Stress, appraisal and coping*. New York: Springer Publishing Company; 1984.
- [60] Watson D. Correcting for acquiescent response bias in the absence of a balanced scale: an application to class consciousness. *Sociol Methods Res* 1992;21:52–88.
- [61] Ray JJ. Reviving the problem of acquiescent response bias. *J Soc Psychol* 1983;121:81–96.
- [62] McIntosh CN. Improving the evaluation of model fit in confirmatory factor analysis: a commentary on Gundy, CM, Fayers, PM, Groenvold, M., Petersen, M. Aa., Scott, NW, Sprangers, MAJ, Velikov, G., Aaronson, NK (2011). Comparing higher-order models for the EORTC QLQ-C30. doi:10.1007/s11136-011-0082-6. *Qual Life Res* 2012;21:1619–21.
- [63] Fayers PM, Aaronson NK. “It ain’t over till the fat lady sings”: a response to Cameron N. McIntosh, improving the evaluation of model fit in confirmatory factor analysis. *Qual Life Res* 2012;21:1623–4.
- [64] Costa DS. Reflective, causal, and composite indicators of quality of life: a conceptual or an empirical distinction? *Qual Life Res* 2015; 24:2057–65.
- [65] Johnson JL, Wiechelt SA. Introduction to the special issue on resilience. *Subst Use Misuse* 2004;39:657–70.
- [66] Werner EE, Smith RS. *Journeys from childhood to midlife: risk, resilience, and recovery*. Ithaca, NY: Cornell University Press; 2001.
- [67] Liegl G, Gandek B, Fischer HF, Bjorner JB, Ware JE, Rose M, et al. Varying the item format improved the range of measurement in patient-reported outcome measures assessing physical function. *Arthritis Res Ther* 2017;19:66.
- [68] Liegl G, Wahl I, Berghöfer A, Nolte S, Pieh C, Rose M, et al. Using Patient Health Questionnaire-9 item parameters of a common metric resulted in similar depression scores compared to independent item response theory model reestimation. *J Clin Epidemiol* 2016;71: 25–34.

E. Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

F. Publikationsliste

PEER-REVIEWED PUBLIKATIONEN

Obbarius N, Fischer F, Obbarius A, Nolte S, Liegl G, Rose M. A 67-item stress resilience item bank showing high content validity was developed in a psychosomatic sample. *J Clin Epidemiol*. 2018 Aug;100:1-12. doi: 10.1016/j.jclinepi.2018.04.004. Epub 2018 Apr 10. (Impact Factor: 4,245, Selected JCR Year: 2017)

Obbarius A, **Obbarius N**, Fischer F, Liegl G, Rose M. [Evaluation of Factor Structure and Construct Quality of the 12-Item Short Version of the OPD Structure Questionnaire (OPD-SQS) in Psychosomatic Patients.] *Psychother Psychosom Med Psychol*. 2019 Jan;69(1):38-48. doi: 10.1055/s-0043-125394. Epub 2018 Feb 15. German. (Impact Factor: 1,128, Selected JCR Year: 2017)

Obbarius A, Fischer K, Fischer F, Liegl G, **Obbarius N**, Nolte S, Rose M. [Empirical assessment of subjective health characteristics using the example of health-related quality of life.] *Psychother Psychosom Med Psychol*. 2018 Dec;68(12):534-547. doi: 10.1055/a-0764-4691. Epub 2018 Dec 4. German. (Impact Factor: 1,128, Selected JCR Year: 2017)

BUCHARTIKEL

Tamm N, Devine J, Rose M. Item generation and construction of questionnaires. In: Guillemin F, Leplege A, Briancon S, Spitz E, Coste J, editors. *Perceived health and adaptation in chronic disease*. New York: Routledge; 2017:109-131.

KONGRESSBEITRÄGE (als Erstautorin)

Obbarius N, Obbarius A, Fischer F, Liegl G, Rose M (2018) Empirische Überprüfung des transaktionalen Stresskonzepts an psychosomatischen Patienten. *Zeitschrift für Psychosomatische Medizin und Psychotherapie* 64(1): 39–116.

Kongress: Deutscher Kongress für Psychosomatische Medizin und Psychotherapie, Berlin, Germany; 03/2018

Obbarius N, Fischer F, Obbarius A, Nolte S, Liegl G, Fischer K, Rose M (2017) Initial validation of a new content valid stress resilience item bank. *Quality of Life Research* 26(1 Supplement): 1-142.

Conference: 24th Annual Conference of the International Society for Quality of Life Research, Philadelphia, Pennsylvania, USA; 10/2017 (Poster)

Tamm N, Maas I, Obbarius A, Fischer F, Liegl G, Sautier L, Mehnert A, Nolte S, Rose M (2016) Is the Stress Computer Adaptive Test (S-CAT) a valid predictor of coping skills and quality of Life in cancer patients? *Quality of Life Research* 25(1 Supplement):1-196.

Conference: 23rd Annual Conference of the International Society for Quality of Life Research, Copenhagen, Denmark; 10/2016 (Poster)

Tamm N, Obbarius A., Nolte S., Fischer F., Engster KM., Rose M (2015) What keeps us healthy in the face of adversity? - Development of a new Stress Resilience Questionnaire (StResQ). *Quality of Life Research* 24(1 Supplement):1-191.

Conference: 22nd Annual Conference of the International Society for Quality of Life Research, Vancouver, British Columbia, Canada; 10/2015 (Vortrag)

G. Danksagung

An dieser Stelle möchte ich all jenen Menschen ausdrücklich danken, ohne die diese Dissertation nicht oder nur wesentlich erschwert möglich gewesen wäre.

Besonders bedanken möchte ich mich bei meinem Doktorvater Prof. Dr. Matthias Rose für die Förderung und Unterstützung neben der klinischen Arbeit auch in der wissenschaftlichen Arbeit voranzukommen. Vielen Dank für die stets konstruktive Kritik, die zahlreichen Anregungen und die fachliche und motivierende Unterstützung.

Des Weiteren gilt mein Dank Dr. Felix Fischer, der mich insbesondere bei allen statistischen Belangen großartig unterstützt hat.

Danken möchte ich auch der gesamten Health Outcomes Research AG für die fachlichen Diskussionen, das immer sehr hilfreiche Feedback und die freundschaftliche Zusammenarbeit. Alex, Sandra, Gregor, Felix, Kathrin: ohne Euch hätte dieses Promotionsprojekt mit Sicherheit weniger Spaß gemacht.

Nicht zuletzt gilt mein Dank meinen Eltern, Marianne und Titus Tamm, die mich über die gesamte Zeit emotional unterstützt und insbesondere in der Zeit der Finalisierung der Arbeit großartig entlastet haben, meinem Mann, meinem Bruder und meiner besten Freundin - vielen herzlichen Dank für Eure Geduld, Eure Wertschätzung und Euer Vertrauen in mich.