

7 Diskussion und Ausblick

7.1 Zusammenfassung

Die vorliegende Dissertation, welche im Rahmen eines DFG-geförderten Forschungsprojekts der Medizinischen Klinik mit Schwerpunkt Psychosomatik der Charité Universitätsmedizin Berlin entstanden ist, hatte die Entwicklung eines Computergestützten Adaptiven Tests (CAT) zur Erfassung von Stresserleben (Stress-CAT) zum Ziel.

Dies geschah aus zwei Gründen. Erstens gibt es eine beachtliche Evidenz bezüglich des Zusammenhangs von Stresserleben und Pathogenese, Progredienz und/oder Chronifizierung verschiedener Erkrankungsbilder (Arck, P. 2001; Hiller, W. & Goebel, G. 1992; Baba, V. et al. 1998; Biondi, M. et al. 1997; Cohen, S. et al. 1991; McEwen, M. 1993) und zweitens erschien es besonders lohnend zu überprüfen, ob sich die seit langem bekannten testtheoretischen Vorteile der Item Response Theory (IRT), bei einem komplexen Konstrukt, wie *Stresserleben*, in der Entwicklung umsetzen lassen.

Einer der wichtigsten Vorteile der IRT liegt in der Möglichkeit, so genannte *computeradaptive Tests (CAT)* zu konstruieren: CATs ermöglichen die individuelle Anpassung der Itemdarbietung an die Merkmalsausprägung einer Person. Dies geschieht, indem Testpersonen während der CAT-Bearbeitung nur die Items zur Beantwortung dargeboten werden, welche dem individuellen Ausprägungsniveau einer Person optimal entsprechen („adaptives Testen“). Welches Item jeweils während der CAT-Bearbeitung als ‚optimal‘ gilt, hängt dabei sowohl von der individuellen Beantwortung vorangegangener Items, als auch von der vorher an einer Kalibrierungstichprobe errechneten Iteminformation der einzelnen Items ab. Dadurch, dass einer Testperson nur die jeweils passenden Items vorgelegt werden, kann eine deutliche Itemreduktion bei einem gleichzeitig konstant hohen Messpräzisionsniveau erreicht werden.

Die Reduktion in der Testlänge kann eine Entlastung für den Diagnostiker und die Testperson bedeuten. Während IRT-basierte CATs sich aufgrund dieser und weiterer testtheoretischer Vorteile in der *Leistungsdiagnostik* bereits vielfach mit gutem Erfolg bewährt haben, ist das Ziel der vorliegenden Dissertation zu untersuchen, ob auch die klinisch-psychologische Diagnostik von dieser fortgeschrittenen Methodik profitieren kann. Dazu wurde die Entwicklung eines kurzen Screening-Instruments zur Erfassung von

Stresserleben angestrebt, welches trotz einer kürzeren Testlänge eine Messung auf einem konstant hohen Messpräzisionsniveau erlauben sollte.

Da das Forschungsfeld IRT-basierter CATs in der klinisch-psychologischen Diagnostik noch relativ jung ist, fehlt bislang ein allgemeiner Forschungskonsens über eine grundlegende methodische Strategie zur Konstruktion IRT-basierter CATs. Die hier realisierte Testentwicklung des Stress-CATs folgte daher verschiedenen Empfehlungen aus Lehrbüchern (z. B. Embretson & Reise, 2000; Hambleton et al., 1991; Wainer, 1990), Übersichtsartikeln (z. B. Hattie, 1984; Nandakumar, 1994; Meijer & Nering, 1999) und einer bereits im Rahmen der Lebensqualitätsforschung erprobten Testentwicklungsstrategie einer US-amerikanischen Forschungsgruppe (Ware et al., 2000, 2003) sowie einer Vorarbeit des hier zugrunde liegenden Forschungsprojektes (Walter, O. B. et al. 2005). Sie gliedert sich in drei prinzipielle Schritte: (1) die inhaltliche Auswahl relevanter Items, (2) die sequentielle statistische Itemanalyse und –selektion dieser Items mit dem Ziel, die Items mit der besten psychometrischen Qualität zur Konstruktion einer Itembank zu nutzen, und (3) die Implementierung der Itembank in einen computergestützten adaptiven Itemabfolge-Algorithmus, der die Präsentation der Items und die Schätzung der individuellen Ausprägung des Stresserlebens (Theta-Schätzung) von Testpersonen ermöglicht.

Den Ausgangspunkt der Entwicklung des Stress-CAT bildete die theoretische Itempoolerstellung hinsichtlich der Itemrekrutierung sowie der Zuordnung von Items zu den Dimensionen ‚erlebte Belastung‘ und ‚Belastungsreaktion‘.

Für die Entwicklung der Itembank wurden Daten an N=1092 psychosomatischen Patienten erhoben, die ein Set prospektiv ausgewählter deutschsprachiger KTT-basierter Fragebögen (Fliege, H. et al. 2001a; Traue, H. et al. 2000; Schulz, P. et al. 1999; Weyer, G. et al. (1975), Skala VIII) computergestützt beantwortet haben.

Der Stress-CAT folgt in seinen Dimensionen einem relationalen Verständnis von Stress. Obwohl die subjektive Bewertung von Anforderungen das Schlüsselkonzept des relationalen Stresskonzeptes ist, existieren bisher nur wenige Instrumente zur systematischen Erfassung solcher Einschätzungen. Eine der wenigen deutschsprachigen Mehr-Item-Skalen, die zur

Itempoolerstellung herangezogen werden konnten, ist der *Perceived Stress Questionnaire* (Levenstein, S. et al. 1993a); deutsche Adaptation: (Fliege, H. et al. 2001b); erfasst werden das Ausmaß an Sorgen, erlebter Anspannung, Anforderungen und erlebter Freude. Neben einer globalen Beurteilung subjektiv wahrgenommener Belastungsreaktionen sollten bei der Entwicklung des Stress-CAT auch spezifische Anforderungssituationen abgebildet werden (siehe Abschnitt 2.2.2). Die dafür herangezogenen Alltagsbelastungsskalen nach (Traue, H. et al. 2000), ähnlich wie die Skala VIII aus den Fragebogenskalen zur Erfassung der subjektiven Belastung nach (Weyer, G. et al. 1975), finden ihre theoretische Fundierung in einer von Stone und Shifman (1994) veröffentlichten Forschungsstrategie, die sie ‚ecological momentary assessment‘ (EMA) nennen. Dieser Forschungsansatz begründet sich in der Annahme, dass der Einfluss von Alltagsärgernissen auf psychosomatische und chronische Erkrankungen nach Möglichkeit täglich erfasst werden sollte. Allerdings hat die Stressforschung gezeigt, dass es zwischen akutem Stress und klinisch relevanten Beeinträchtigungen der Gesundheit keine konsistenten Zusammenhänge gibt (Lepore, S. J. 1995). Die zeitlichen Instruktionen des *Trierer Inventars zur Erfassung von chronischem Stress* nach (Schulz, P. et al. 1999) erfragen Stresserlebnisse in Bezug auf die letzten 6 Wochen. Die Autoren folgen einer Definition von (Pratt, L. I. & Barling, J. 1988), wonach chronischer Stress einen (1) unspezifischen, schleichenden Beginn hat, (2) von langer oder auch kurzer Dauer sein kann, (3) sehr häufig ist und (4) eine geringe oder auch hohe Intensität haben kann.

In der vorliegenden Arbeit sollten die rekrutierten Items einer einheitlichen zeitlichen Instruktion folgen, in Anlehnung an Ergebnisse der Validierung der *Perceived Stress Scale* nach (Cohen, S. et al. 1983), die, wie auch der Stress-CAT intendiert (Abschnitt 2.2.2.), auf das Ausmaß der aktuell wahrgenommenen Belastungsfaktoren und das Erleben der eigenen Belastetheit fokussiert. Cohen et al. (1983) konnten zeigen, dass die prädiktive Validität der *Perceived Stress Scale* am größten war über eine Zeitperiode von 4 Wochen.

Im Rahmen eines Delphi-Entscheidungsprozesses (Hasson, F. et al. 2000) wurden aus einem Pool von 126 rekrutierten Items insgesamt 104 Items von der Forschungsgruppe ausgewählt (s. Abschnitt 4.2.1). Davon wurden 62 Items

der Dimension ‚erlebte Belastung‘, 42 Items der Dimension ‚Belastungsreaktion‘ zugeordnet. Die Beurteilerübereinstimmung lag bei $\kappa=.80$ ($T=7,43$, $p\leq.001$). 22 Items konnten keiner Dimensionen eindeutig zugeordnet werden.

Nach einer ersten theoriegeleiteten Selektion erfolgte die statistische Itemanalyse und –selektion, beginnend mit der Überprüfung der Unidimensionalität. Es liegen verschiedene Verfahren zur Überprüfung der Unidimensionalität vor, die alle mit Vor- und Nachteilen behaftet sind (vgl. Abschnitt 3.2.1.1.1). In der vorliegenden Arbeit wurde der Auffassung von Embretson und Reise (2000) Folge geleistet, die unter Hinzuziehung der vorliegenden Literatur (u.a. Hambleton et al. 1991, Hattie, 1984; Stout, 1987) zu der Einschätzung kamen, dass die bestmögliche Information hinsichtlich der Dimensionalität der Daten erhalten werden kann, wenn die gemeinsame Varianz einem Faktor zugeordnet wird, um danach die Residualkovariationen zu analysieren (zur weiterführenden Diskussion der Überprüfung der Unidimensionalität, siehe Abschnitt 7.2).

Um der Forderung nach Homogenität und lokaler stochastischer Unabhängigkeit nachzukommen, wurden aufgrund der zu hohen residualen Korrelationen 14 der 62 Items aus dem Itempool ‚erlebte Belastung‘ ausgeschlossen. Items, deren residuelle Korrelationen zu hoch waren, wurden schrittweise aus dem Modell entfernt. Die ausgeschlossenen Items erfragten vor allem arbeitsbezogene Aspekte des Stresserlebens.

Der Itempool ‚Belastungsreaktion‘ stellte sich von Anfang an homogener dar. Nach Berechnung der konfirmatorischen Faktorenanalyse mit residualer Kovarianzanalyse der Dimension ‚Belastungsreaktion‘ wurden 5 der 42 Items aus dem Pool entfernt. Die ausgeschlossenen Items erfragten überwiegend ressourcenbezogene Aspekte des Stresserlebens.

Die Untersuchung der Antwortkategorienfunktionen der Dimension ‚erlebte Belastung‘ führte zu einem Ausschluss von weiteren 6 Items. Für den Teil der Items aus dem Itempool ‚erlebte Belastung‘ (27/48 Items), die überwiegend aus den Alltagbelastungsskalen nach (Traue, H. et al. 2000) rekrutiert wurden, musste die vierte Antwortkategorie (=‚häufig‘) und die fünfte (=‚sehr häufig‘) zusammengelegt (collapsed) werden. Der Grund hierfür war, dass die Antwortkategorien nicht ausreichend zwischen unterschiedlichen Ausprägungen auf dem latenten Traitkontinuum zu diskriminieren vermochten. Die Analyse der

Antwortkategorienfunktionen des Itempools ‚Belastungsreaktion‘ zeigte in der Mehrzahl der verbleibenden Items (33/37 Items) sehr gute Kategorienfunktionen. Es kam zu einem Ausschluss von vier weiteren Items. Die Eigenschaften der verbleibenden Items wurden anhand ihres Differential-Item-Functioning und ihrer Diskriminationsfähigkeit überprüft (siehe auch Zusammenfassung der Itemselektion, Abschnitt 5.5).

Im finalen Itempool verblieben 38 Items in der Dimension ‚erlebte Belastung‘ und 31 Items in der Dimension ‚Belastungsreaktion‘, die jeweils für die Anwendung eines polytomen Zwei-Parameter-Modells (Generalized-Partial-Credit-Model) als geeignet angesehen werden konnten (zur weiterführenden Diskussion des gewählten Modells, siehe Abschnitt 7.2). Orientiert man sich an einer Reliabilität von $\rho \geq ,90$ und legt für den computeradaptiven Testalgorithmus einen Standardfehler von $\leq ,32$ fest, so zeigen Simulationsstudien, dass die durchschnittliche Anzahl der für eine präzise Messung von ‚erlebter Belastung‘ benötigten Items $10,3 \pm 1,7$ Items ($M \pm SD$) betrug. Analog dazu waren es $8,6 \pm 2,2$ Items ($M \pm SD$) für die Dimension ‚Belastungsreaktion‘. Zudem legen die Simulationsstudien nahe, dass der CAT-Algorithmus in den extremen Merkmalsbereichen eine bessere Diskriminationsfähigkeit als der entsprechende KTT-basierte Summenscore bietet.

7.2 Methodische Implikationen

Mit der vorliegenden Arbeit werden die Entwicklungsschritte für den ersten computeradaptiven Test zur Erfassung von Stresserleben vorgestellt.

Für die Entwicklung des hier vorgestellten Stress-CAT wurden bislang ausschließlich Items etablierter Fragebögen berücksichtigt. Dies erfolgte mit der Zielsetzung Itemparameter kalkulieren zu können, die sowohl eine Vergleichbarkeit mit den Ergebnissen anderer Arbeitsgruppen ermöglichen als auch erlauben, den CAT-Score in Beziehung zu etablierten KTT-basierten Fragebogenwerten zu setzen. Nachteil dieses Vorgehens ist, dass bei der Nutzung dieser Items für die reale CAT-Anwendung mit jedem Item Frage- und Antwortformat wechseln können. Dies könnte einerseits zu einer höheren Konzentrationsbeanspruchung und schnelleren Ermüdung der Patienten führen,

andererseits aber auch über eine höhere Konzentrationsleistung einem mechanischen Antwortverhalten entgegenwirken.

Der Einsatz computerassistierter Diagnostik ist heute in der psychologischen Diagnostik bereits weitgehend akzeptiert. Die Verbreitung computerassistierter Fragebogenerhebungen ist dennoch eher gering, da bei einem Wechsel von der Papier-Bleistift-Version die Äquivalenz des Verfahrens jeweils geprüft werden muss (Kubinger, K. D. 1993; Kubinger, K. D. 1999; Fowler, R. D. 1985; Gitzinger, I. 1990; Fahrenberg, J. 1994; Rose, M. et al. 1999). Dort, wo den computerassistierten Erhebungen bislang der Vorzug gegenüber den Papier-Bleistift-Versionen gegeben wurde, waren meist ökonomische Gründe ausschlaggebend (Rose, M. et al. 1999; Rose, M. et al. 2002). Die Vorteile des computeradaptiven Testens gegenüber der computerassistierten Erhebung sind dagegen konzeptioneller Art und liegen m.E. vor allem begründet in der hohen Messpräzision bei gleichzeitig reduziertem Testumfang. Die damit verbundene geringere Patientenbelastung verspricht die Erfassung mehrerer Patientenmerkmale zu einem Untersuchungszeitpunkt zu ermöglichen, die mit herkömmlichen Fragebögen aufgrund deren Umfangs nur mit Kompromissen hinsichtlich der Messgenauigkeit zu realisieren wäre.

Während an den theoretischen Vorteilen computeradaptiver Tests kaum Zweifel bestehen (Kubinger, K. D. et al. 2000; Hornke, L. F. 1999), bleiben in der praktischen Umsetzung derzeit viele Fragen offen. So sind die meisten der in der vorliegenden Arbeit eingesetzten Entwicklungsverfahren Gegenstand einer lebhaften Diskussion. Leider zeichnet sich aus noch kein allgemeiner Konsens ab, welche Schritte für die Entwicklung eines CAT als unabdingbar angesehen werden müssen. Die vorliegende Arbeit orientiert sich daher an pragmatischen Vorschlägen verschiedener Übersichtsartikel und entsprechender Lehrbücher: So liegen z.B. verschiedene Verfahren zur Überprüfung der Unidimensionalität vor, die alle mit Vor- und Nachteilen behaftet sind. Die vorliegende Arbeit folgt der Auffassung von Embretson und Reise (2000), die unter Hinzuziehung der vorliegenden Literatur (u.a. (Hattie, J. 1984) (Stout, W. 1987) zu der Einschätzung kamen, dass die bestmögliche Information hinsichtlich der Dimensionalität der Daten erhalten werden kann, wenn die gemeinsame Varianz einem Faktor zugeordnet wird, um danach die Residualkovariationen zu analysieren. Dabei scheint es offenbar nachrangig zu sein, mit welcher

Methodik der gemeinsame Faktor bestimmt wird. Hambleton et al. (1991) sehen in der Sicherstellung der Unidimensionalität einen der wesentlichsten Schritte für die Güte bzw. den Fit des Gesamtmodells an. Hambleton et al. (1991) empfehlen, systematisch verschiedene Ansätze zu kombinieren. Sie schlugen hierfür insgesamt 22 verschiedene Ansätze vor. Diese lassen sich drei Gruppen zuordnen: 1) Prüfung der zugrunde liegenden Modellannahmen, z.B. die Prüfung der Unidimensionalität, 2) Prüfung der erwarteten Modelleigenschaften, z.B. die Prüfung in verschiedenen Stichproben, 3) Prüfung der Modellvorhersagen, z.B. in Simulationsstudien. In der vorliegenden Arbeit wurden sechs der genannten Ansätze berücksichtigt: konfirmatorische Faktorenanalyse, Prüfung residualer Kovariationen, heuristische Beurteilung der Kategorienfunktionen, Prüfung der Differential-Item-Function, Analyse des Modells sowie die Prüfung der Eigenschaften des Itempools in zwei Simulationsstudien.

Im Hinblick auf die Itemparameterschätzung steht die Modellwahl zur Diskussion. Die vorliegende Arbeit wählte das Generalized Partial Credit Modell (GPCM, Muraki, 1997;) aus den möglichen IRT-Modellen aus (siehe Kapitel 3.2.1.1.4), da es eine unidimensionale zweiparametrische IRT-Modellierung polytomer Daten mit einer simultanen Analyse unterschiedlicher Antwortformate erlaubt sowie die Variation der Diskriminationsfähigkeit unterschiedlicher Antwortkategorien und unterschiedlicher Items bei der Modellierung berücksichtigt. Es gilt als wenig restriktiv. Bei Zwei-Parameter-Modellen stellt die formale Prüfung des Modell-Fits ein bislang nicht gelöstes Problem dar. Das Programm Parscale bietet hierfür zwar einen Prüfalgorithmus an, doch zeigten Simulationsstudien von Hambleton, et al. (1991), dass die Anzahl zufälliger ‚Item-Misfits‘ mit zunehmender Stichprobengröße steigt, wie dies für χ^2 -Statistiken allgemein bekannt ist. Im Rahmen ihrer Simulationsstudie mit 50 Items und einer Stichprobengröße von N=1200 Personen fanden die Autoren 11 zufällige ‚Item-Misfits‘. McDonald (1989) formulierte dies provokant so: „A failure to reject an IRT model by fit statistics is simply a sign that the sample size was too small.“

Die Diskussion um das ‚beste‘ IRT-Modell währt bereits drei Jahrzehnte. Der Standpunkt, je mehr Parameter ein Modell berücksichtigt, desto besser kann es

die empirische Realität modellieren, läuft dem ‚Prinzip der Sparsamkeit‘ (‚principle of parsimony‘, Embretson & Hershberger, 1997) zuwider. In der Tat erscheinen in manchen Anwendungsfällen komplexe (mehr-parametrische) IRT-Modelle jedoch besser zu den empirischen Daten zu passen, da sie weniger restriktive Annahmen setzen. Allerdings unterliegen sie im Falle geringer Personenstichprobengrößen in ihrer Datenanpassung IRT-Modellen mit wenigen Parametern. Dies äußert sich dann in instabilen Parameterschätzungen. Mitunter kann ein Mangel an identifizierbaren Parametern auch der Anwendung komplexerer IRT-Modelle im Wege stehen (Van der Linden, W. J. & Glas, C. A. W. 2003).

Ziel ist es, ein Modell zu wählen, welches möglichst gut zu den empirischen Daten passt, bzw. die Daten (z. B. mittels Itemselektion) oder Konstrukte (z. B. durch Re-Konzeptualisierungen) so zu verändern, dass sie zu dem Modell passen. Hierbei ist von vorneherein wichtig, sich vor Augen zu führen, dass Modelle stets Idealisierungen darstellen, die nie gänzlich der Realität entsprechen.

Prinzipiell sind zwei Konsequenzen zur gezielten Verbesserung der Fit-Statistiken denkbar: eine gezielte Itemselektion oder eine Lockerung der Restriktionen eines Modells (oder die Wahl eines weniger restriktiven Modells). Diese Strategien sind jedoch nur sinnvoll, wenn man diese Fit-Statistiken für gültig und damit handlungsleitend hält. Generell halten sich die meisten der IRT-Forscher bezüglich der Nennung spezifischer Richtlinien zum Umgang mit ungenügenden Ergebnissen in der Fit-Statistik bedeckt. Allgemein empfehlen Van der Linden und Hambleton (2003), dass der Umgang mit Misfits von folgenden Faktoren abhängig sei:

1. der Art des Misfits,
2. der Verfügbarkeit von Ersatzitems,
3. dem mit dem Neuschreiben von Items verbundenen Aufwand,
4. der Verfügbarkeit von Kalibrierungsstichproben und dem Testziel.

Da drei dieser Punkte (2.-4.) Praktikabilitätsabwägungen beinhalten, deutet sich hier an, dass oftmals praktische Einschränkungen zur (vorläufigen) Akzeptanz von Misfits, von denen vermutet wird, dass sie lediglich statistische ‚Artefakte‘ darstellen, führen.

Zu diskutieren ist weiterhin, inwieweit *Stresserleben*, bzw. *Stress* im Allgemeinen, als psychisches Merkmal für die CAT-Erfassung geeignet ist.

Bei der theoretischen Auswahl der Items war es intendiert, in dem Stress-CAT alle Facetten des Stresserlebens, die aus klinischer Sicht relevant erscheinen, in einem Score vereinen zu können, aufgeteilt nach den Dimensionen ‚erlebte Belastung‘ und ‚Belastungsreaktion‘.

Im Zuge der empirischen Prüfung zeigte sich jedoch, dass die Dimension ‚erlebte Belastung‘ diese Erwartung nicht ganz erfüllte. Schaut man sich die totale Testinformation an (vgl. Abbildung 14, Kapitel 6) bemerkt man eine Verschiebung der maximalen Testinformation von einer Standardabweichung über dem Durchschnitt. Das deutet darauf hin, dass nicht ausreichend Items für den unteren Merkmalsbereich rekrutiert bzw. generiert wurden. Mit der Neurekrutierung von Items könnte zukünftig einer der Vorteile der Item Response Theory genutzt werden, die es erlaubt, fortlaufend Items aus dem Itempool zu entfernen und neue (nach entsprechender Kalkulation der Itemparameter) hinzuzufügen, ohne dass die Vergleichbarkeit der CAT-Werte eingeschränkt ist (Reise, S. P. 2000). Im Vergleich zur Dimension ‚erlebte Belastung‘, zeigt die Dimension ‚Belastungsreaktion‘ in der totalen Testinformation ein gewünschtes Maximum über dem Durchschnitt (vgl. Abbildung 15, Kapitel 6).

Die Entwicklung des Stress-CAT hat gezeigt, dass die Entwicklung IRT-basierter Test deutlich höhere personelle und technische Ressourcen erfordert, als eine Testentwicklung auf der Grundlage der klassischen Testtheorie. Angefangen bei den erforderlichen großen Stichprobenumfängen bis zur Anschaffung verschiedenster Hardware und spezieller Software.

Um diesem höheren Entwicklungsaufwand zu entsprechen, wurde im Frühjahr 2004 vom US Department of Health and Human Services gemeinsam mit dem National Institute of Health ein umfangreiches Förderprogramm zur computeradaptiven Erfassung subjektiver Gesundheitsmerkmale aufgelegt (RFA-RM-04-011), so dass für die nahe Zukunft zumindest im angloamerikanischen Raum eine grundlegende Änderung dieser Situation zu erwarten ist.

7.3 Ausblick

Die Stresstheorie nach Lazarus (1984) impliziert eine theoretisch begründete Trennung der Konstrukte Selbstwahrnehmung, Stressreaktion und Coping. Mit Hilfe eines Selbsteinschätzungsverfahrens zum Belastungserleben, wie ihn der Stress-CAT darstellt, ist grundsätzlich nicht zu erfassen, ob das Vorhandensein von Coping-Ressourcen durch die befragte Person überprüft wird, zu welchem Ergebnis sie dabei gelangt und ob wahrgenommene Ressourcen eingesetzt werden oder nicht. Insofern ist allein aufgrund eines ‚Stress-Scores‘ grundsätzlich nicht erfassbar, ob bzw. welche Copingstrategien von der Person bereits eingesetzt wurden und ob die Facetten der Stressreaktion, die hier als Belastungserleben konzipiert werden, vor oder nach dem Einsatz einer Copingstrategie liegen.

Dabei ist auch in Rechnung zu stellen, dass einige der Items wiederum selbst mit Copingstrategien konfundiert sein könnten, ohne dass es bislang im Einzelnen bekannt ist. So könnte etwa die Formulierung „*Sie haben das Gefühl, dass zu viele Forderungen an Sie gestellt werden*“ (STRESS02, siehe auch Abschnitt 5.1) mit einer externalen Verantwortlichkeitsattribution konfundiert sein, die sich wiederum als eine Bewältigungsstrategie mit dem Ziel der eigenen emotionalen Entlastung verstehen lässt. Kommende Studien könnten diese theoretischen Differenzierungen voranbringen, etwa indem über Prozessanalysen Phasen der Stresswahrnehmung und –verarbeitung aufgedeckt werden und durch Erhebungen von Copingvariablen Konfundierungen aufgedeckt werden – auch oder gerade unter zu Hilfenahme der Weiterentwicklung des Stress-CAT auf der Basis der Item Response Theory.

Denkbar wäre auch die Entwicklung eines computeradaptiver Tests zur Erfassung der ‚Ressourceneinschätzung‘, wobei man auf gesundheitliche Protektivfaktoren fokussieren könnte. Dabei wäre an eine Einordnung im Rahmen einer salutogenetischen Perspektive zu denken (Antonovsky, A. 1979) unter Einbeziehung weiterer personaler Ressourcenkonstrukte wie ‚Hardiness‘, ‚dispositionaler Optimismus‘ und ‚Selbstwirksamkeitserwartung‘.