

Fragebögen Abkürzung	Ausgangsitempool Abgekürzter Itemtext	Itemausschluss nach theoretischer Prüfung Delphi-Prozess	Items vorhanden in Dimension		Statistische Itemselektion				resultierender Itempool
			Erlebte Belastung	Belastungs- reaktion	Unidimen- sionalität	Kategorien- funktion	DIF	slope	
Anzahl	128	-24	62	42	-21	-10	-2	-2	69
sub31	Bei meiner Arbeit gibt es viel unnötigen Leerlauf, an dem ich nicht Schuld bin.		•		x				
sub32	Manchmal denke ich, dass ich mir mit meiner Arbeit zuviel zumute.	x							
sub40_A	Manchmal wächst mir die Arbeit über den Kopf.		•						•
sub40_B	Manchmal wächst mir die Hausarbeit über den Kopf.	x							

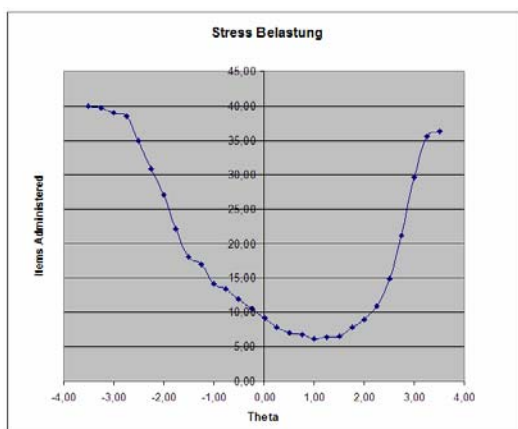
Die graue Hervorhebung zeigt an bis zu welchem Schritt, das Item in den Analysen berücksichtigt wurde. Das x markiert ab wann ein Item nicht mehr berücksichtigt wurde.

Ein • zeigt an in welcher Dimension das Item vorkam.

6 Simulationsexperimente

Im *ersten* Simulationsexperiment mit einer simulierten Stichprobe von N=2900 Testpersonen (s. Methodik) wurde die Güte der Itembank des Stress-CAT und des Itemabfolgealgorithmus untersucht. Abbildung 13 stellt die Anzahl der im Rahmen der Dimensionen darzubietenden Items dar, damit ein $SE \leq 0,32$ erreicht wird.

A



B

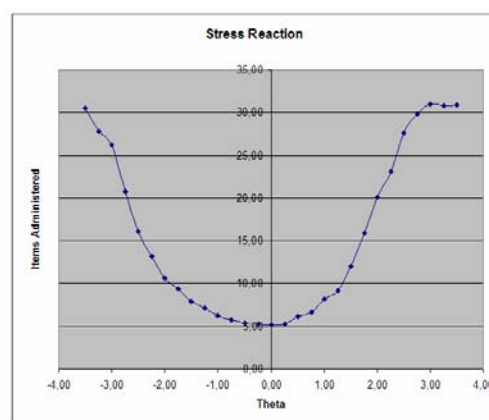


Abbildung 13: Anzahl der dazubietenden Items in Abhängigkeit von der geschätzten Merkmalsausprägung Theta (jeweils für die Dimension A: ‚erlebte Belastung‘ und B: ‚Belastungsreaktion‘) bei einer voreingestellten Messgenauigkeit von $SE \leq 0,32$.

Die Stoppfunktion des Stress-CAT lässt sich als maximal applizierte Anzahl von Items programmieren und / oder als gewünschte Messpräzision. Wir entschieden uns hier für die Definition eines Konfidenzintervalls, das mit einem Standardfehler $SE \leq 0,32$ festgesetzt wurde. Das Abbruchkriterium von $SE \leq 0,32$ entspricht einer Reliabilität von $\rho \geq 0,9$. Ist die zuvor erwartete Genauigkeit der Messung erreicht oder die maximale Anzahl von Items beantwortet, ist die Stoppfunktion erfüllt. Da die Skala Teil einer Testbatterie ist (zwei Dimensionen), schließt sich bei der praktischen Anwendung des Stress-CAT die Messung des nächsten Konstruktes an (‚Belastungsreaktion‘).

Bei mittleren Merkmalsausprägungen ($-2 \leq \theta \leq +2$) der simulierten Testpersonen betrug die durchschnittliche Anzahl der für eine präzise Messung von ‚erlebter Belastung‘ benötigten Items $10,3 \pm 1,7$ Items ($M \pm SD$). Analog dazu waren es $8,6 \pm 2,2$ Items ($M \pm SD$) für die Dimension ‚Belastungsreaktion‘. So werden den Patienten für die Dimension ‚erlebte Belastung‘, die innerhalb von zwei Standardabweichungen um den Mittelwert der Stichprobe liegen (ca. 90% der Patienten), 8-12 Items vorgelegt. Bei simulierten Testpersonen, deren Merkmalsausprägungen außerhalb der zweifachen Standardabweichung liegen, steigt die Anzahl der vorzulegenden Items auf $30,9 \pm 3,2$ ($M \pm SD$).

Die totale Testinformationsfunktion zeigt für die Dimension ‚erlebte Belastung‘ ein Maximum von einer Standardabweichung über dem Durchschnitt in Abhängigkeit der zugrunde liegenden Stichprobe (siehe Abbildung 14).

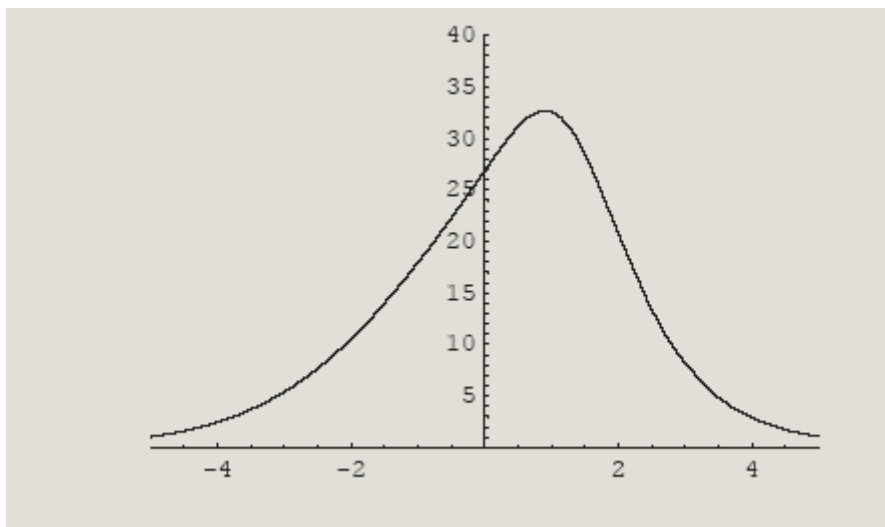


Abbildung 14: Totale Testinformation für die Dimension ‚erlebte Belastung‘ (n=38 Items) in Abhängigkeit von der geschätzten Merkmalsausprägung

Den Patienten, die innerhalb von zwei Standardabweichungen um den Mittelwert der Stichprobe liegen (ca. 90% der Patienten), werden für die Dimension ‚Belastungsreaktion‘ 6-10 Items vorgelegt. Bei simulierten Testpersonen, deren Merkmalsausprägungen außerhalb der zweifachen Standardabweichung liegen, steigt die Anzahl der vorzulegenden Items auf $25,2 \pm 6,1$ ($M \pm SD$).

Die totale Testinformationsfunktion zeigt hier ein gewünschtes Maximum über dem Durchschnitt (siehe Abbildung 15).

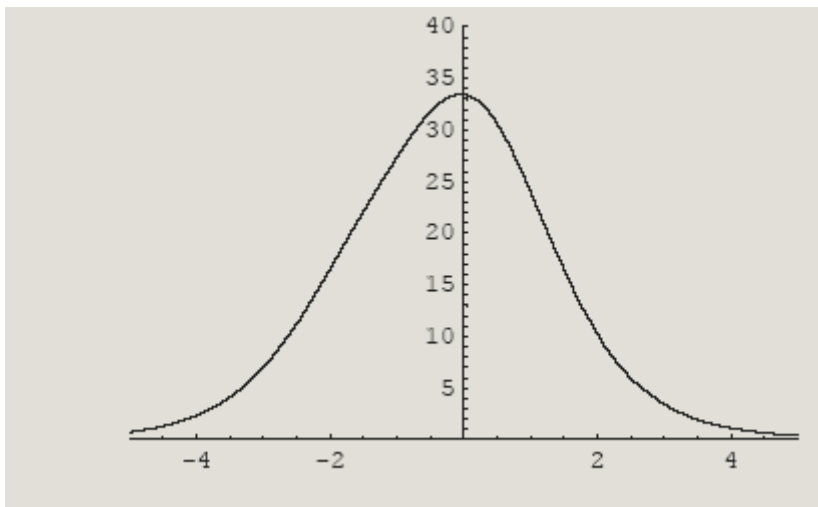


Abbildung 15: Totale Testinformation für die Dimension ‚Belastungsreaktion‘ (n=31 Items) in Abhängigkeit von der geschätzten Merkmalsausprägung

Im *zweiten* Simulationsexperiment wurden Schätzungen der Theta-Werte mittels eines simulierten Stress-CAT an realen Itemantworten der Personenstichprobe vorgenommen. Die Patienten beantworteten die Items der Itembank im Zusammenhang der jeweiligen Standard-Fragebögen. Der adaptive Prozess und die damit einhergehende Auswahl der Items wurden nachträglich simuliert. Vergleicht man die Schätzungen der Theta-Werte auf der Basis des Stoppkriteriums von $SE \leq 0,32$ mit der Schätzung der Theta-Werte, bei der alle zur Verfügung stehenden Items herangezogen wurden, zeigt sich eine sehr hohe Übereinstimmung der ermittelten Werte für die Dimension ‚erlebte Belastung‘ ($r = 0,96$, siehe Abbildung 16). Auch für die Dimension ‚Belastungsreaktion‘ zeigt sich eine sehr hohe Übereinstimmung, vergleicht man die Schätzungen der Theta-Werte auf der Basis des Stoppkriteriums von

$SE \leq 0,32$ mit der Schätzung der Theta-Werte, bei der alle zur Verfügung stehenden Items herangezogen wurden ($r = 0,97$, siehe Abbildung 16). Dies weist darauf hin, dass gegenüber der Darbietung aller zur Verfügung stehenden Items der Itembank (‚erlebte Belastung‘: $n=38$ Items; ‚Belastungsreaktion‘: $n=31$ Items) die mit dem Stress-CAT mögliche Itemreduktion zu keinem nennenswerten Informationsverlust führt.

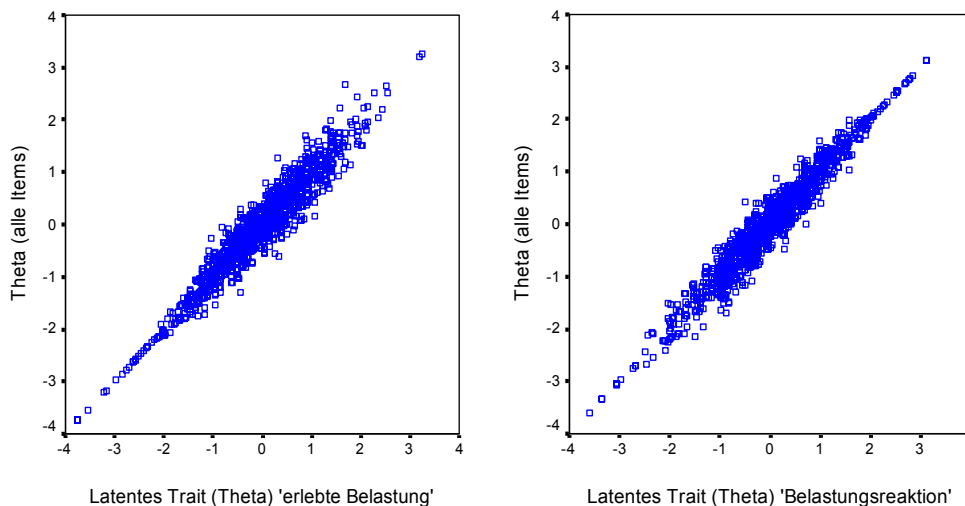


Abbildung 16: Beziehung zwischen dem simulierten Stress-CAT Scores (jeweils für ‚erlebte Belastung‘ und ‚Belastungsreaktion‘) und dem Score bei Berücksichtigung aller Items (‚erlebte Belastung‘: $n=38$ Items; ‚Belastungsreaktion‘: $n=31$ Items).

Wird das Ergebnis der Schätzung der Theta-Werte mittels des simulierten Stress-CAT für die Dimension ‚erlebte Belastung‘ mit der konventionell berechneten KTT-basierten Summenskala des ABF (reale Werte, die Patienten bearbeiteten den ABF mit allen 58 Items in üblicher Reihenfolge) in Beziehung gesetzt (siehe Abbildung 17), deutet sich eine S-förmige Punktwolke an, die zum einen eine enge Beziehung zwischen dem Alltagsbelastungsfragebogen (Traue et al., 2000) und den Theta-Schätzungen anzeigt ($r=.83$), zum anderen darauf hinweist, dass der IRT-basierte Algorithmus vor allem im unteren Merkmalsbereich eine bessere Diskriminationsfähigkeit als der KTT-basierte Summenscore bietet.

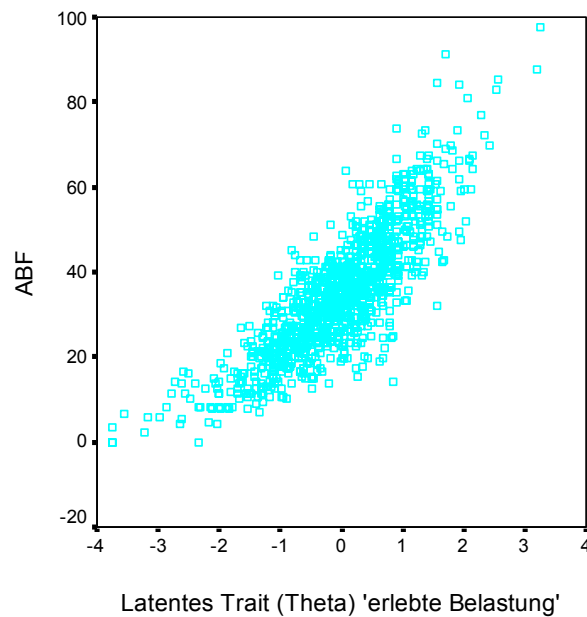


Abbildung 17: Beziehung zwischen dem simulierten Stress-CAT-Score (für die Dimension ‚erlebte Belastung‘) und dem Alltagsbelastungsfragebogen von Traue et al. (2000).

Wird das Ergebnis der Schätzung der Theta-Werte mittels des simulierten Stress-CAT für die Dimension ‚Belastungsreaktion‘ mit der konventionell berechneten KTT-basierten Summenskala des PSQ (reale Werte, die Patienten bearbeiteten den PSQ mit allen 20 Items in üblicher Reihenfolge) in Beziehung gesetzt (siehe Abbildung 18), zeigt sich eine S-förmige Punktwolke, die zum einen eine enge Beziehung zwischen dem *Perceived Stress Questionnaire* (Levenstein et al., 1993; Fliege et al., 2001) und den Theta-Schätzungen anzeigt ($r=.82$).

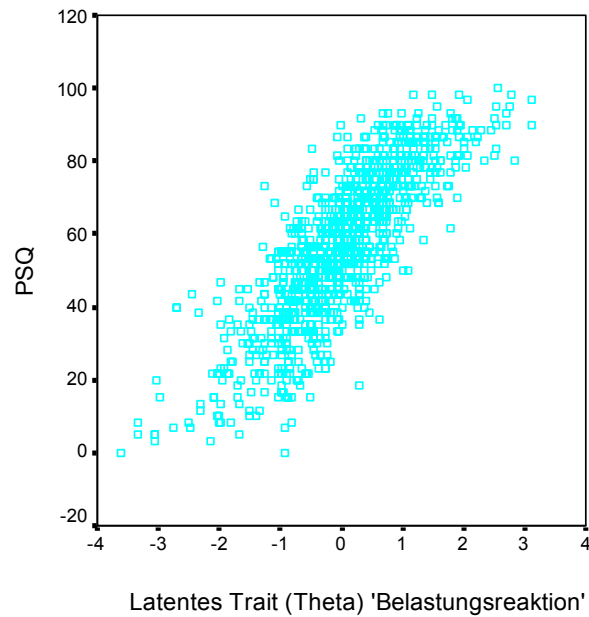


Abbildung 18: Beziehung zwischen dem simulierten Stress-CAT-Score (für die Dimension ‚Belastungsreaktion‘) und der Perceived Stress Scale (Levenstein, S. et al. 1993a); (Fliege, H. et al. 2001b).

Auch hier zeigt sich, dass der IRT-basierte Algorithmus in den extremen Merkmalsbereichen eine bessere Diskriminationsfähigkeit als der KTT-basierte Summenscore bieten scheint.