

TESTDOKUMENTATION

Dokumentation der Stressskala

Psychometrische Evaluation

First Last^a

^a *University of Instances*

Dezember 2025

Abstract

In diesem Abschnitt ist zusammengefasst, welches Konstrukt erfasst wird und für welchen Anwendungsbereich das Instrument geeignet ist. Die Zusammenfassung besteht aus maximal 120 Wörtern.

Keywords: Stress, Diagnostik, Psychometrie

Author Note

First Last  <https://orcid.org/0000-1111-1111-1111>

We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to First Last, Address String, Email: email@example.com

Überblick

Zusammenfassung

In diesem Abschnitt ist zusammengefasst, welches Konstrukt erfasst wird und für welchen Anwendungsbereich das Instrument geeignet ist. Die Zusammenfassung besteht aus maximal 120 Wörtern. Die Dokumentation ist in sechs Kapitel gegliedert. Die Darstellung orientiert sich an den formalen Richtlinien der Deutschen Gesellschaft für Psychologie (2019): Blocksatz, Schriftart Arial, die Schriftgröße 11, 1,5-facher Zeilenabstand.

Instrument

Instruktion

Hier wird die Instruktion der Skala dokumentiert.

Items

Table 1

Itemübersicht mit Polung und Subskalenzuordnung

Nr.	Item	Polung	Subskala
1	Hier ist das erste Item dargestellt.		
2	Hier das zweite Item...	+	A
3	Drittes Item...	-	B

Antwortvorgaben

Hier werden die Antwortvorgaben/Antwortkategorien spezifiziert. Es wird dokumentiert, wie viele Antwortkategorien es gibt und wie diese benannt sind. Alternativ können die Antwortkategorien bereits im vorherigen Abschnitt dargestellt werden.

Auswertungshinweise

Dieser Abschnitt beschreibt, welche numerischen Werte den Antwortkategorien zugeordnet werden. Hier wird ebenfalls beschrieben, welche Items negativ kodiert sind, welche Items zu einer Skala zusammengefasst werden, ob Subskalen getrennt ausgewertet werden, ob einfache oder gewichtet Summenwerte gebildet werden.

Anwendungsbereich

Hier wird beschrieben, für welchen Zweck die Skala eingesetzt werden kann, in welchem Modus (z.B. schriftliche oder mündliche Befragung) die Skala üblicherweise eingesetzt wird und für welche Zielgruppe die Skala entwickelt wurde.

Theoretischer Hintergrund

Hier wird der theoretische Hintergrund der Skala beschrieben. Es wird erklärt, warum die Skala relevant ist und aus welcher Theorie das zugrundeliegende Konstrukt abgeleitet wird. Es kann auf relevante Literatur verwiesen werden.

Skalenentwicklung

Itemkonstruktion und Itemselektion

Die Items wurden auf Basis theoretischer Überlegungen zu Stress und Stresserleben entwickelt. Die Itemselektion erfolgte nach folgenden Kriterien:

- Theoretische Fundierung der Items
- Inhaltliche Abdeckung verschiedener Stressfacetten
- Verständlichkeit und sprachliche Klarheit
- Vermeidung von Redundanzen

Hinweis: Weitere Details zur Itemkonstruktion werden von einem Kollegen ergänzt.

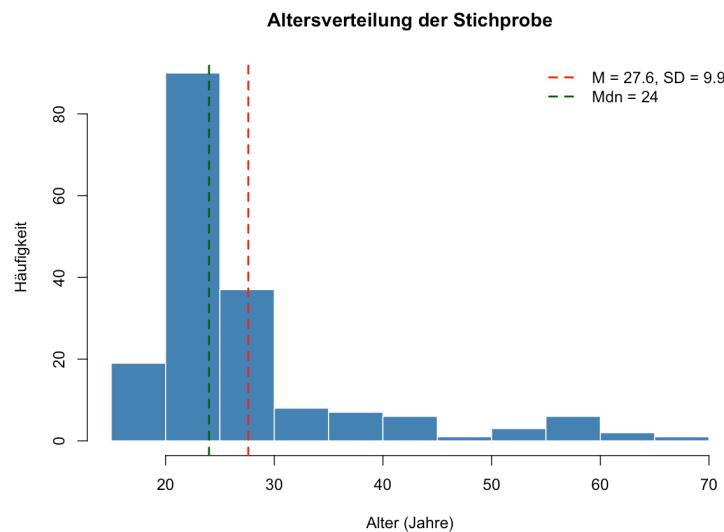
Stichproben

An der Studie nahmen insgesamt 232 Personen teil. Nach Datenbereinigung und Ausschluss von Retest-Fällen (QUESTNNR = B) verblieben 211 Teilnehmer. Von diesen wurden weitere 29 Personen ausgeschlossen, die den Aufmerksamkeitstest nicht bestanden haben (SOo2_14 ≠ 4). Die finale Stichprobe umfasste somit N = 180 Personen.

Demographische Merkmale

Das durchschnittliche Alter der Stichprobe betrug $M = 27.62$ Jahre ($SD = 9.90$ Jahre). Alle Teilnehmer waren mindestens 18 Jahre alt. Die Altersverteilung ist in Abbildung 1 dargestellt.

Figure 1
Altersverteilung der Stichprobe (N = 180)



Die Geschlechterverteilung umfasste 36 Männer (20.0%), 140 Frauen (77.8%) und 4 diverse Personen (2.2%). Hinsichtlich der Beschäftigung wurden die Freitextangaben in drei Kategorien klassifiziert: Studenten (alle Personen im Studium) umfassten 116 Personen (64.4%), Erwerbstätige (alle berufstätigen Personen, einschließlich Angestellte, Selbstständige, Beamte, etc.) umfassten 49 Personen (27.2%), Andere (Arbeitslose, Rentner, Schüler, Auszubildende) umfassten 14 Personen (7.8%), und 1 Person machte keine Angabe (0.6%).

Die Bildungsniveaus wurden auf Basis des höchsten Schulabschlusses in drei Kategorien gruppiert: Niedrig (Hauptschule oder Realschule) umfasste 9 Personen (5.0%), Mittel (Fachhochschulreife oder Hochschulreife) umfasste 77 Personen (42.8%), und Hoch (Bachelor, Master oder Staatsexamen) umfasste 94 Personen (52.2%).

Für Subgruppenanalysen wurden zusätzlich Altersgruppen gebildet: Jung (unter 30 Jahre) mit 134 Personen (74.4%), Mittel (30 bis 45 Jahre) mit 24 Personen (13.3%), und Alt (über 45 Jahre) mit 22 Personen (12.2%).

Hinweis: Sämtliche R-Skripte zur Datenbereinigung, Skalenkonstruktion und statistischen Analyse sind öffentlich zugänglich unter <https://github.com/linozen/dipra>.

Ausreißeranalyse

Zur Qualitätssicherung der Daten wurde eine umfassende Ausreißeranalyse durchgeführt. Dabei kamen drei komplementäre Methoden zum Einsatz:

1. **Z-Score-Methode** ($|z| > 3.29$): Identifizierte 7 Personen (3.9%) mit extremen Werten auf mindestens einer Variable
2. **IQR-Methode** (Interquartilsabstand): Identifizierte 48 Personen (26.7%) mit Werten außerhalb der Boxplot-Whiskers
3. **Mahalanobis-Distanz** (multivariat, $p < .001$): Keine multivariaten Ausreißer identifiziert
Konsistente Ausreißer, die von mindestens zwei Methoden identifiziert wurden, umfassten 7 Personen (3.9%). Diese Fälle wurden für Sensitivitätsanalysen gespeichert, aber nicht aus den Hauptanalysen ausgeschlossen, da keine Hinweise auf Messfehler vorlagen und die Werte valide Extremausprägungen repräsentieren können.

Dimensionalität und Itemkennwerte

Varianzanalyse

Eine umfassende Varianzanalyse wurde für alle 65 Items durchgeführt, um Items mit problematischer Streuung zu identifizieren. Die Analyse ergab:

- Items mit Deckeneffekt (>15% wählen Maximum): 14 Items (21.5%)
- Items mit Bodeneffekt (>15% wählen Minimum): 28 Items (43.1%)
- Gesamt problematische Items: 42 Items (64.6%)

Besonders betroffen waren die Skalen für Stressbelastung (lang) und Stresssymptome (lang), bei denen jeweils über 80% der Items Varianzprobleme aufwiesen. Die Varianzstatistiken nach Skalen zeigten folgende Muster:

- **Stressbelastung (kurz):** 2 von 5 Items (40%) mit Varianzproblemen
- **Stressbelastung (lang):** 10 von 11 Items (91%) mit Varianzproblemen
- **Stresssymptome (kurz):** 3 von 5 Items (60%) mit Varianzproblemen
- **Stresssymptome (lang):** 11 von 13 Items (85%) mit Varianzproblemen
- **Coping-Items:** 14 von 22 Items (64%) mit Varianzproblemen

Diese Befunde unterstreichen die Relevanz der Kurzskalen, die weniger stark von Varianzproblemen betroffen sind. Items mit geringer Varianz wurden nicht automatisch entfernt, da geringe Streuung bei bestimmten Stressindikatoren theoretisch bedeutsam sein kann (z.B. bei adaptiven Stichproben).

Faktorenstruktur

Die theoretisch postulierte 7-Faktoren-Struktur wurde mittels konfirmatorischer Faktorenanalyse (CFA) überprüft. Das Modell umfasste folgende Faktoren:

1. Stressbelastung (5 Items - Kurzskala)
2. Stresssymptome (5 Items - Kurzskala)
3. Aktives Coping (4 Items)
4. Drogen-Coping (5 Items)
5. Positives Coping (4 Items)
6. Soziales Coping (5 Items)
7. Religiöses Coping (4 Items)

Fit-Indizes des Gesamtmodells:

- $\chi^2(443) = 782.65, p < .001$
- CFI = 0.871 (problematisch, Schwellenwert: ≥ 0.90)
- TLI = 0.856 (problematisch, Schwellenwert: ≥ 0.90)
- RMSEA = 0.065 [90% KI: 0.058, 0.073] (akzeptabel, Schwellenwert: ≤ 0.08)
- SRMR = 0.079 (akzeptabel, Schwellenwert: ≤ 0.08)

Die Fit-Indizes zeigen einen akzeptablen bis verbesserungswürdigen Modellfit. Während RMSEA (0.065) und SRMR (0.079) im akzeptablen Bereich liegen, erreichen CFI (0.871) und TLI (0.856) nicht den empfohlenen Schwellenwert von 0.90. Dies deutet darauf hin, dass die theoretisch postulierte 7-Faktoren-Struktur von den Daten nur teilweise gestützt wird. Mögliche Gründe hierfür könnten Überlappungen zwischen den Faktoren oder modellbedingte Spezifikationsprobleme sein.

Faktorladungen

Die standardisierten Faktorladungen sind in Tabelle 1 dargestellt. Die meisten Items zeigen substanzelle Ladungen (> 0.40) auf ihren zugehörigen Faktoren. Zwei Items wiesen jedoch schwache Ladungen auf:

- SO23_15 (Positive Neubewertung): $\lambda = 0.33$
- NI07_02 (Religiöses Coping): $\lambda = 0.23$

Eine vollständige Visualisierung aller Faktorladungen findet sich in Abbildung 2.

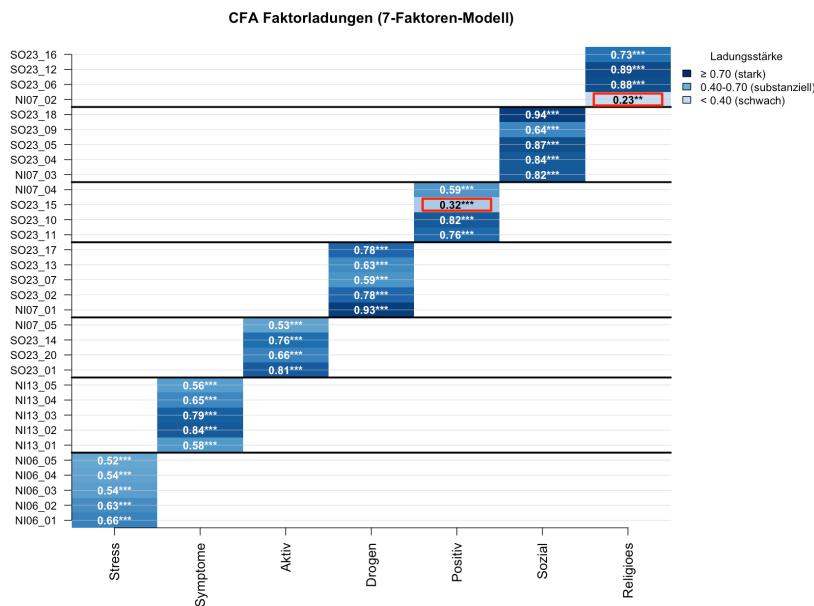
Table 2
Standardisierte Faktorladungen (λ) aller Items im 7-Faktoren-Modell

Faktor	Item	λ	p
Stress	NI06_01	0.66	<.001
	NI06_02	0.63	<.001
	NI06_03	0.54	<.001
	NI06_04	0.54	<.001
	NI06_05	0.52	<.001
Symptome	NI13_01	0.58	<.001
	NI13_02	0.84	<.001
	NI13_03	0.79	<.001
	NI13_04	0.65	<.001
	NI13_05	0.56	<.001
Aktiv	SO23_01	0.82	<.001
	SO23_20	0.66	<.001
	SO23_14	0.76	<.001
	NI07_05	0.53	<.001
Drogen	NI07_01	0.93	<.001
	SO23_02	0.78	<.001
	SO23_07	0.59	<.001
	SO23_13	0.64	<.001
	SO23_17	0.78	<.001
Positiv	SO23_11	0.76	<.001
	SO23_10	0.82	<.001
	SO23_15	0.33	.001
	NI07_04	0.59	<.001
Sozial	NI07_03	0.82	<.001
	SO23_04	0.84	<.001
	SO23_05	0.87	<.001
	SO23_09	0.64	<.001
	SO23_18	0.94	<.001
Religiös	NI07_02	0.23	.007
	SO23_06	0.88	<.001
	SO23_16	0.89	<.001

Anmerkung. Faktorladungen > 0.40 gelten als substanzIELL, > 0.70 als stark. Schwache Ladungen (< 0.40) sind kursiv dargestellt.

Figure 2

Heatmap der standardisierten Faktorladungen im 7-Faktoren-Modell



Itemstatistiken

Tabelle 2 zeigt die deskriptiven Statistiken der Items der Kurzskalen für Stressbelastung und Stresssymptome.

Table 3
Itemstatistiken der Kurzskalen (N = 180)

Item	M	SD	Trennschärfe
NI06_01	4.24	1.28	0.52
NI06_02	3.86	1.41	0.48
NI06_03	3.78	1.47	0.41
NI06_04	4.52	1.36	0.42
NI06_05	4.09	1.38	0.39
NI13_01	3.12	1.37	0.46
NI13_02	3.23	1.46	0.63
NI13_03	3.24	1.52	0.61
NI13_04	3.28	1.58	0.51
NI13_05	3.28	1.52	0.43

Anmerkung. Skala von 1 (trifft nicht zu) bis 6 (trifft zu). Trennschärfen wurden als korrigierte Item-Skala-Korrelationen berechnet.

Interfaktor-Korrelationen

Die Korrelationen zwischen den Faktoren (Tabelle 3) zeigen theoriekonform hohe Zusammenhänge zwischen Stressbelastung und Stresssymptomen ($r = 0.82$). Adaptive Coping-Strategien (positiv, sozial) korrelieren negativ mit Stress und Symptomen, während Drogen-Coping positive Zusammenhänge aufweist. Alle Interfaktor-Korrelationen lagen unter 0.85, was auf hinreichende Diskriminanz zwischen den Faktoren hindeutet.

Table 4
Interfaktor-Korrelationen im 7-Faktoren-Modell

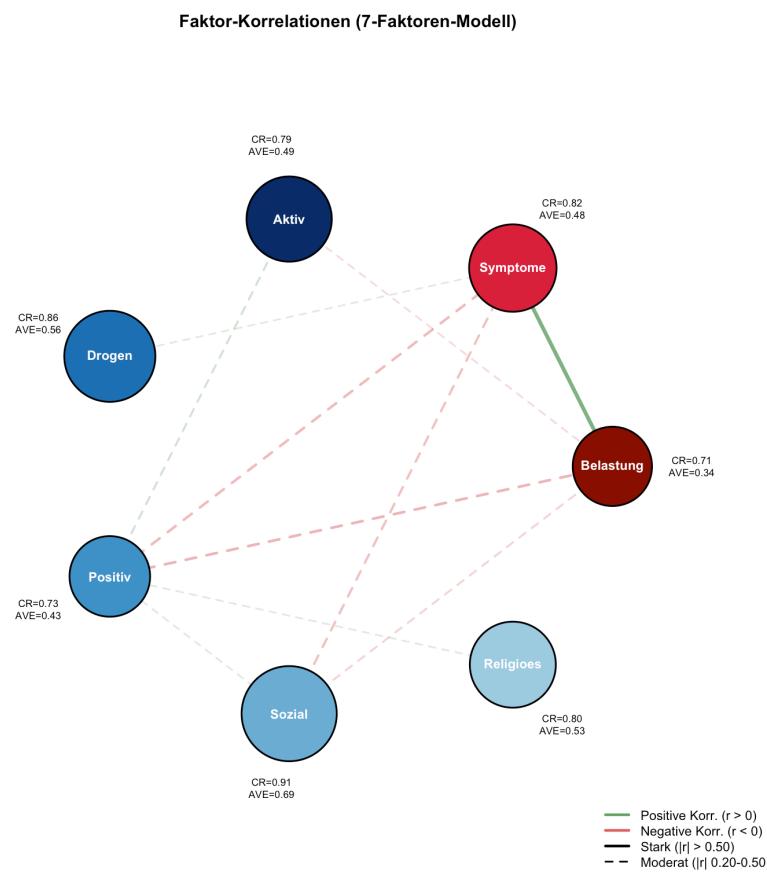
Faktor 1	Faktor 2	r	p
Stress	Symptome	0.82	<.001
Stress	Aktiv	-0.24	.022
Stress	Drogen	0.19	.033
Stress	Positiv	-0.49	<.001
Stress	Sozial	-0.29	.002
Stress	Religiös	0.02	.888
Symptome	Aktiv	-0.20	.073
Symptome	Drogen	0.21	.008
Symptome	Positiv	-0.45	<.001
Symptome	Sozial	-0.40	<.001
Symptome	Religiös	-0.08	.363
Aktiv	Drogen	-0.13	.195
Aktiv	Positiv	0.32	.038
Aktiv	Sozial	0.07	.435
Aktiv	Religiös	0.11	.185
Drogen	Positiv	-0.11	.295
Drogen	Sozial	-0.14	.057
Drogen	Religiös	-0.06	.403
Positiv	Sozial	0.21	.021
Positiv	Religiös	0.22	.007
Sozial	Religiös	-0.07	.397

Anmerkung. Alle Korrelationen < 0.85 indizieren hinreichende diskriminante Validität zwischen den Faktoren.

Eine graphische Darstellung der Faktor-Korrelationen als Netzwerk findet sich in Abbildung 3.

Figure 3

Netzwerkdarstellung der Interfaktor-Korrelationen. Stärkere Korrelationen sind durch dicke Linien repräsentiert.



Reliabilität und konvergente Validität der Faktoren

Die Composite Reliability (CR) und Average Variance Extracted (AVE) wurden für jeden Faktor berechnet (Tabelle 4). Die CR ist ein Maß für die interne Konsistenz der Faktoren, während die AVE angibt, wie viel Varianz durch die Items eines Faktors im Durchschnitt erklärt wird.

Table 5

Composite Reliability (CR) und Average Variance Extracted (AVE) der Faktoren

Faktor	CR	AVE	Bewertung
Stress	0.71	0.34	CR: gut, AVE: problematisch
Symptome	0.82	0.48	CR: gut, AVE: grenzwertig
Aktiv	0.79	0.49	CR: gut, AVE: grenzwertig
Drogen	0.86	0.56	CR: gut, AVE: gut
Positiv	0.73	0.43	CR: gut, AVE: problematisch
Sozial	0.91	0.69	CR: exzellent, AVE: gut
Religiös	0.80	0.53	CR: gut, AVE: gut

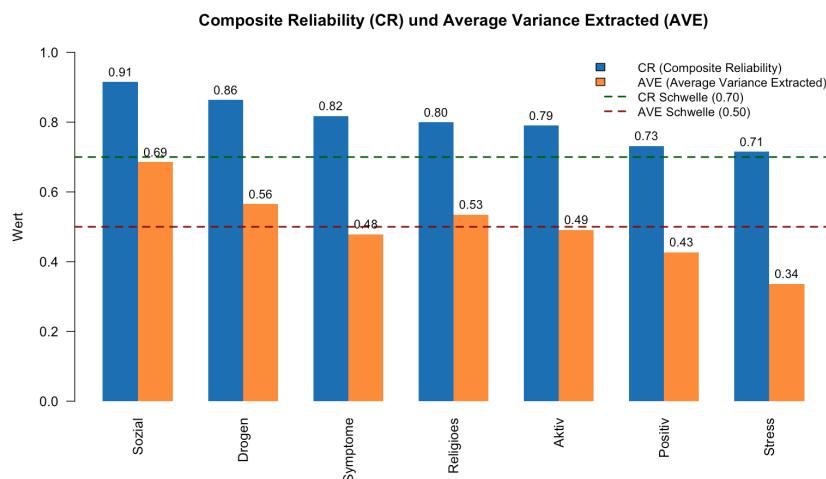
Anmerkung. Interpretationsrichtlinien: CR > 0.70 = gute Reliabilität, CR > 0.90 = exzellente Reliabilität; AVE > 0.50 = ausreichende konvergente Validität, AVE < 0.50 = problematisch.

Interpretation: Alle Faktoren erreichen eine gute bis exzellente Reliabilität (CR > 0.70), wobei der Faktor "Soziales Coping" mit CR = 0.91 herausragt. Die konvergente Validität ist jedoch bei vier Faktoren eingeschränkt (AVE < 0.50 bei Stress, Symptome, Aktiv und Positiv). Dies deutet darauf hin, dass bei diesen Skalen ein substanzialer Anteil der Varianz durch Messfehler oder andere Einflüsse erklärt wird. Besonders kritisch ist der Faktor "Stressbelastung" mit AVE = 0.34, bei dem nur 34% der Varianz durch die latente Variable erklärt werden.

Die drei Coping-Faktoren Drogen (AVE = 0.56), Sozial (AVE = 0.69) und Religiös (AVE = 0.53) zeigen sowohl gute Reliabilität als auch gute konvergente Validität.

Figure 4

Vergleich von Composite Reliability (CR) und Average Variance Extracted (AVE) für alle Faktoren. Die gestrichelten Linien markieren die Schwellenwerte (CR = 0.70, AVE = 0.50).



Gütekriterien

Objektivität

Hier kann die Durchführungs-, Auswertungs-, und Interpretationsobjektivität der Skala bewertet werden. Dieser Abschnitt entspricht dem Standard 5 (Minimierung des Prozessfehlers) des Rats für Sozial und Wirtschaftsdaten (RatSWD, 2014).

Reliabilität

Hier werden Schätzungen der Reliabilität der Skalenwerte berichtet. Wurden Messmodelle der klassischen Testtheorie verwendet, sind Cronbachs Alpha, die Split-Half- oder die Re-Test-Korrelation üblich. Alternativ oder ergänzend kann auch die Skalenreliabilität nach Raykov (1997) oder McDonals Omega (1999) berichtet werden. Wurden Messmodelle der Latent State-Trait Theorie (Steyer, Schmitt & Eid, 1999) verwendet, werden Schätzer der Reliabilität, Konsistenz, Messzeitpunkt-Spezifität (und ggf. Methodenspezifität) berichtet. Wurden IRT-Modelle verwendet, kann Anrichs Reliabilität oder die Skalenreliabilität nach Raykov, Dimitrov und Asparouhov (2010) berichtet werden. Dieser Abschnitt entspricht dem Standard 4 (Reliabilität) des Rats für Sozial und Wirtschaftsdaten (RatSWD, 2014).

Validität

In diesem Abschnitt werden Ergebnisse berichtet, die auf die Inhaltsvalidität, Konstruktvalidität oder Kriteriumsvalidität des Instruments hindeuten. Dieser Abschnitt entspricht dem Standard 2 (Validität) und dem Standard 3 (Minimierung methodenspezifischer Effekte) des Rats für Sozial und Wirtschaftsdaten (RatSWD, 2014).

Deskriptive Statistiken (Normierung)

Hier werden Mittelwert, Standardabweichung, Schiefe und Exzess der Skalenwerte beschrieben.

Nebengütekriterien

Hier werden Ergebnisse berichtet, die es erlauben Nebengütekriterien wie Testfairness, Ökonomie oder Verfälschbarkeit des Instruments zu bewerten. Dieser Abschnitt entspricht dem Standard 6 (Weitere Qualitätsmerkmale) des Rats für Sozial und Wirtschaftsdaten (RatSWD, 2014).

Literatur

Kontakt zu Autor(en)

Vorname Nachname, Institution, evtl. Adresse, E-Mailadresse.

Literaturverzeichnis

Das Literaturverzeichnis enthält alle Literaturangaben, auf die in der Dokumentation verwiesen wird. Es gelten die Zitationsrichtlinien der Deutschen Gesellschaft für Psychologie (2019):

Deutsche Gesellschaft für Psychologie (Hrsg.). (2019). *Richtlinien zur Manuskriptgestaltung* (5., überarbeitete Aufl.). Göttingen: Hogrefe.

Kersting, M. (2006). Zur Beurteilung von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 54(4), 243-253.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah: Erlbaum.

RatSWD (Hrsg.). (2014). *Qualitätsstandards zur Entwicklung, Anwendung und Bewertung von Messinstrumenten in der sozialwissenschaftlichen Umfrageforschung*. Zugriff am 01.07.2014 <http://www.ratswd.de/themen/qualitaetsstandards>

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21, 173-184.

Raykov, T., Dimitrov, D., Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling*, 17(2), 265-279.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, 13, 389-408.