Natalja Menold, Technische Universität Dresden

Praktische Signifikanz von unterschiedlichen Methoden der Messinvarianzanalyse im Vergleich

Gemeinsame Tagung der Sektion Methoden der empirischen Sozialforschung der DGS und ASI, 21.-22. September, Dresden

Gefördert durch die DFG (ME 353810-1)

Forschungsteam: Jasmin Kadel, Hagen von Hermanni, Hend Achmed, Lena Eggert, Rhaja Horst

Inhalt

Einführung

Forschungsfrage

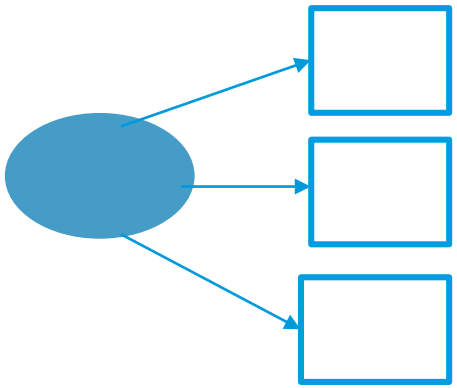
Methoden

Ergebnisse

Diskussion

Exakte Messinvarianz über Multi-Group CFA (MGCFA, Meredith, 1993)

Arabisch

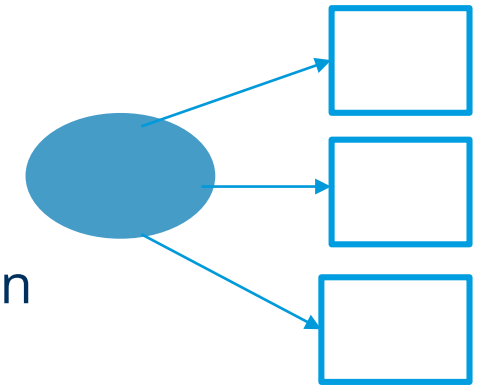


Struktur vergleichbar?
Konfigurale Invarianz

Ladungen gleich?
Metrische Invarianz:
Vergleiche der Korrelationen möglich

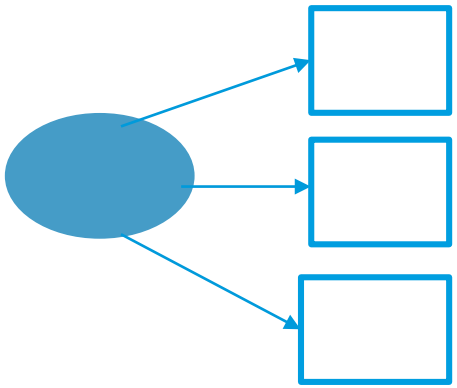
Ordinatenabschnitte vergleichbar?
Skalare Invarianz:
Vergleiche der Mittelwerte möglich

Deutsch



Bayesian Approximate Messinvarianz (BAMI, Muthén & Asparouchov, 2012)

Arabisch



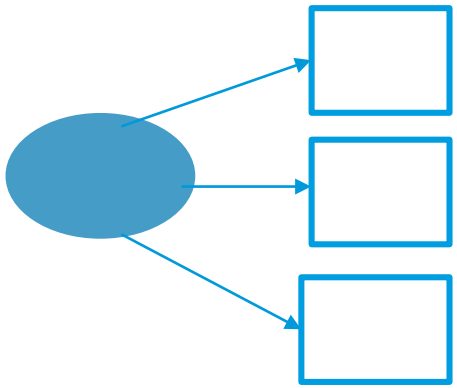
Unterschiede ~ 0
Verteilung mit dem Mittelwert 0 und einer „Prior“ Varianz

Prior: Alle Parameter erhalten eine gemeinsame Verteilung

Posterior Verteilung: Prior Knowledge + Observed Data

Der kleinste Prior = .01
Je größer Prior, desto näher ist das Modell an der konfiguralen Invarianz

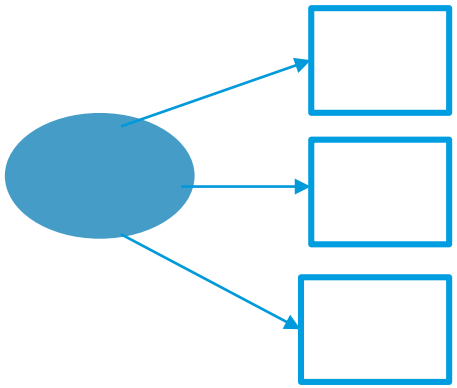
Deutsch



Rationale für die Setzung der Priors erforderlich

MG Alignment (Muthén & Asparouchov, 2014)

Arabisch



Nutzt die EFA Logik:
Rotation für die Einfachstruktur

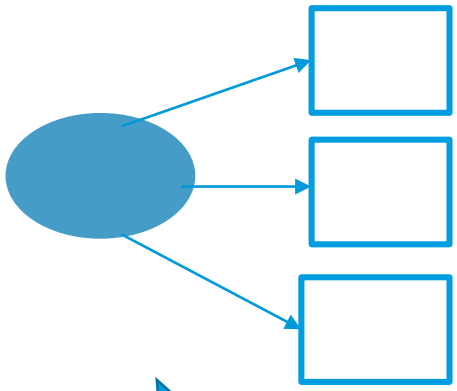
Unterschiede zwischen den Gruppen als Unterschiede zwischen latenten Variablen

Non-Invarianz: Unterschiede, die nicht durch latente Mittelwertunterschiede erklärt werden können

Konfigural: nicht rotiert
Alignment: möglichst viele invariante Parameter

Schätz alle Parameter gruppenspezifisch, minimiert die Unterschiede

Deutsch



Automatisierung der partial MI; Wie viel Invarianz ist verträglich?

Praktische Signifikanz

Messinvarianzanalysen (MI):

- Geringe Beachtung in komparativen Studien (Boer et al., 2018)
- Oft keine skalare (exakte) MI, insbesondere zwischen Sprachen/Ländern/Enthnien (Leitgöb et al., 2022)
- Non-Invarianz ist bei Vergleichen zu vernachlässigen (e.g., Robitzsch & Lüdtke, 2023)

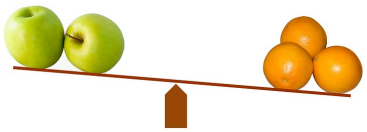
MI als Möglichkeit zur Untersuchung von Messfehlern und Bias

Cross-Cultural Comparability Bias

Comparability Bias: Unterschiedliche Werte für manifeste Variablen bei demselben Wert auf dem latenten Konstrukt (z. B. van de Vijver 2018)

1) **construct bias**: Messung unterschiedlicher Konstrukte in den Vergleichsgruppen
2) **method bias**: Unterschiedliche Methoden und Methodeneffekte
3) **item bias**: Unterschiedliche Bedeutung der Items

Studien: Korrespondenz zwischen Bias und Non-Invarianz



Method Bias:

- Non-Invarianz für interviewer- and Selbstadministrierung (Klausch et al. 2013), bei Unterschieden in Ratingskalen (Menold & Tausch 2016)

Item Bias (Benítez et al. 2022; Maitinger 2017)

- metrische und skalare Non-Invarianz beim unterschiedlichen Verständnis der Items

Fragestellung

Hypothesen zu Problemen aus kognitiven Pretests / Verletzung der Regeln der Fragebogenkonstruktion:

Wie unterscheiden sich die Methoden der MI-Analysen bei der Aufdeckung von Konstrukt- oder Itembias in Bezug auf

- **Model Fit**?
- Finden von **nicht-invarianten Indikatoren** (Korrespondenz zu Problemen wie doppelte Stimuli, Komplexität, Verständnisprobleme)?
- Schätzung der **latenten Mittelwerte**?

Studiendesign

Arabisch oder Dari sprechende Geflüchtete / Deutsche Aufnahmegesellschaft; Gesundheitsskalen aus dem SOEP

Kognitive Interviews (CI) bei GESIS (Hadler et al. 2021): Jeweils 6 Geflüchtete aus Syrien, Irak, Afghanistan mit den Unterschieden im Alter, Geschlecht, Bildung

Probleme auf der Konstrukt- und Indikatorenebene; Überarbeitung der Instrumente (ENSURE-Version)

Randomisiertes Experiment in 2021/2022: SOEP-ENSURE

- Registerstichprobe (Einwohnermeldeamt) für Dresden, Leipzig und Chemnitz, Online Umfrage mit postalischer Rekrutierung (RR: 8% bis 20%; n=231 Deutsch, n = 172 Arabisch)
- Facebook-Rekrutierung: Arabisch (N = 700) und Dari (N = 520)

Fragebögen im SOEP

Instrument	Konzept	Measurement Invarianz / Item Bias
SF-12 (Short Form Health Survey-12) 12 Items; 2 Faktoren	Physical and mental health	Factorial structure varies (Treanor & Donelli, 2015); Item Bias for ethnic comparisons in the U.S. (Desouky et al. 2013, Fleishman & Lawrence 2003);
RHS-15 (Refugee Health Screener-15) 15 Items, ein Faktor	Post traumatic disorder, anxiety, trauma	Difficulty in understanding of some items (no bias analysis) (Hollifield et al. 2016)

Non-Invarianz für Arabisch und Dari im SOEP (Tibubos & Kroeger, 2020)

Ergebnisse der Kongn. Pretests und Revisionen

Instru-ment	Probleme (Hadler et al. 2021)	Revisionen: ENSURE
SF-12	- Bezug zu psychischer, physischer und genereller Gesundheit nicht eindeutig - Unterschiedliche Zeitbezüge werden übersehen (generell / letzter Monat) - „Probleme bei Tätigkeiten im Alltag“ als “berufliche Probleme” - Gefühlsadjektive: Äquivalenz zwischen den Sprachen verletzt	- Eindeutiger Bezug für psychisch / physisch / allgemein - Alternative Formulierungen für Gefühle (Itemauswahl) - Doppente Stimulie aufteilen (Itemauswahl) - Wiederholungen vermeiden, klares Layout - Skalenorientierung von wenig/nie bis viel/immer
RHS	- Probleme bei Gefühlsadjektiven (sich taub fühlen, innerlich unruhig, schreckhaft - Wiedersprüche zwischen DS (traurig = deprimiert?) - Auch hier manchmal unklare Trennung physisch – psychisch - Unterstellung für Trauma	- Eindeutiger Bezug - Nicht zutreffend - Alternative für Adjektive - Vermeiden der DS - Klareres Layout

Ergebnisse SF-12: Model Fit Zufallsstichprobe
2 Faktoren (physical and mental health)

	model	*Exact MLR CMIN, RMSEA, CFI (metric, scalar: Δ)*	*Exact WLSMV BIC*	*BAMI CI für pp (best fit)*	*Alignment BIC*
SOEP	configural	320.77*** (106) / .12 / .81	7604.181	NA	
		173.02*** (104) /.07 / .94		.04	NA
	metric	12.68 (10) / -.002 / -.002	7618.57	1) [-4.038, 64.451] 2) [26.765, 93.735]	7997.23
	scalar	109.98*** (10) /.029/.073	7658.75		
ENSURE	configural	283.61*** (106) / .11 /.895	8109.67	NA	
		210.79*** (104) /.087 / .937		.01	NA
	metric	10.24 (10) / -.004 /-.001	8091.56	1) [9.171, 73.708] 2) [48.900,115.276]	8745.14
	scalar	28.69** (10) /.003 /.010	8102.68		

- Exakt bildet die Veränderung ab
- BAMI: Lehnt die MI ab

SF-12: Anzahl non-invarianter Indikatoren Zufallsstichprobe

	model	*Exact MLR*	*Exact WLSMV*	*BAMI .01/.02 /.04*	*Alignment*
SOEP	metric	0	NA	0/0/0	0
	scalar	6 (DS, Bezug, Gefühle)	NA	2/2/2 (Gefühle)	1 (Gefühle)
ENSURE	metric	0	NA	0/0/1	0
	scalar	1 (Tätigkeit)	NA	1/1/0 (Tätigkeit)	1(Tätigkeit)

Exakt
- bildet die Veränderung ab
- Indikatoren mit Problemen: non-invariante Interzepte

SF-12 Unterschiede in latenten Mittelwerten Zufallsstichprobe

Arabisch Baseline
Bessere psychische Gesundheit für Deutsch mit dem SOEP-Instrument

	model	*Exact MLR*	*Exact WLSMV*	*BAMI*	*Alignment*
SOEP	Physisch	0.02	0.03	0.02	0.01
	Psychisch	0.37*	0.36*	0.29*	0.42***
ENSURE	Physisch	0.09	0.02	0.11	0.09
	Psychisch	0.07	0.07	0.08	0.03

- Übereinstimmung zwischen den Methoden
- Kausaler Effekt: Instrument als Ursache
- Keine Anpassung für Alignment

Ergebnisse SF-12: Model Fit in der Facebook-Stichprobe
2 Faktoren (physical and mental health)

	model	*Exact MLR CMIN, RMSEA, CFI (metric, scalar: Δ)*	*Exact WLSMV BIC*	*BAMI CI für pp (best fit)*	*Alignment BIC*
SOEP	configural	493.71*** (106) / .12 / .08	7604.181	NA	
		294.49*** (104) /.07/.94		.04	NA
	metric	24.81**(10)/ .001 /.004	7618.573	1) [70.749, 133.106] 2) [104.367, 170.064]	31018.579
	scalar	49.49**(10) /.003/.012	7658.748		
ENSURE	configural	341.400*** (86) / .0.75 /.930	28239.627	NA	NA
	metric	16.56 (9)/.002 /.003	28239.332	.04 1) [216.597, 274.790	30662.077
	scalar	77.20*** (9)/.004 /.017	28253.9		

- Exakt bildet die Veränderung ab
- BAMI: Lehnt die MI ab

SF-12: Anzahl non-invariante Indikatoren Facebook

	model	*Exact MLR*	*Exact WLSMV*	*BAMI .01/.02 /.04*	*Alignment*
SOEP	metric	2 (DS, Bezug)	NA	1/1/1	1
	scalar	4 (DS, Tätigkeit)	NA	3/5/7 (Tätigkeit, Bezug, DS)	1 (Tätigkeit)
EN-SURE	metric	0	NA	0/0/1	0
	scalar	3 (?)	NA	3/2/3 (?)	3 (?)

Exakt und BAMI
- bilden die Veränderung ab
- Indikatoren mit Problemen: non-invariante Interzepte
- BAMI strenger als exakt

SF-12 Unterschiede in latenten Mittelwerten Facebook

Bessere physische und psychische Gesundheit für Arabisch als für Dari mit dem SOEP-Instrument

	model	*Exact MLR*	*Exact WLSMV*	*BAMI*	*Alignment*
SOEP	Physisch	-0.32***	-1.36***	-0.30*	-0.31***
	Psychisch	-0.25***	-0.78***	-0.26	-0.36***
ENSURE	Physisch	0.08	0.07	0.11	0.06
	Psychisch	0.05	0.02	0.06	-0.02

- Übereinstimmung zwischen den Methoden
- Kausaler Effekt: Instrument als Ursache
- Keine Anpassung für Alignment

Ergebnisse RHS: Model Fit Zufallsstichprobe 1 Faktor

	model	*Exact* *MLR* *CMIN, RMSEA, CFI* *(metric, scalar: Δ)*	*Exact* *WLSMV* *BIC*	*BAMI* *CI für pp (best fit)*	*Alignment* *BIC*
SOEP	configural	480.90*** (180)/ .11 / .79	9391.616		
		368.38*** (176)/.09 / .84		NA	NA
	metric		9387.679	.04	
		47.20*** (14)/ .004 / .025		1) [135.988, 218.839] 2) [173.756, 267.543]	
	scalar	187.02*** (14)/.028/.129	9688.241		
ENSURE	configural	330.21*** (208) / .07 /.90	8478.959		
		297.47***(207)/.058 / .929		NA	NA
	metric			.04	
		28.35* (15)/ .001 /.010	8477.065	1) [16.145, 117.338]	
	scalar	143.34*** (15)/.026 /.094	8554.223		

- Exakt bildet die Veränderung ab
- BAMI: Bessere Passung bei ENSURE

RHS: Indikatoren und Mittelwerte Zufallsstichprobe

- Metrisch ESNURE: Ähnliche Ergebnisse
- Alignment und BAMI: Bessere skalare MI für ENSURE

Indikatoren mit Problemen: non-invariante Interzepte

Mittelwerte: keine Unterschiede

	model	*Exact MLR*	*Exact WLSMV*	*BAMI .01/.02/.04*	*Alignment*
SOEP	metric	2	NA	5/5/5	1
	scalar	10	NA	7/7/8	6
ENSURE	metric	2	NA	1/1/2	0
	scalar	9	NA	3/5/5	2

Mehr posttraumatische Symptome für Gefluchtete

model	*Exact MLR*	*Exact WLSMV*	*BAMI*	*Alignment*
SOEP	-0.37***	-0.43***	-0.56*	-0.28*
ENSURE	-0.36***	-0.35*	-0.57*	-0.35^{+}

Ergebnisse RHS: Model Fit: Facebookstichprobe 1 Faktor

	model	*Exact MLR CMIN, RMSEA, CFI (metric, scalar: Δ)*	*Exact WLSMV BIC*	*BAMI CI für pp (best fit)*	*Alignment BIC*
SOEP	configural	765.71*** (180)/ .09 / .90	9391.616		
		497.41*** (177)/.07 / .94		NA	NA
	metric	51.56*** (14) / .001/ .006	9387.679	.02 1) [485.158, 565.268] 2) [155.313, 231.816]	
	scalar	146.76*** (14) /.008/.021	9688.241		
ENSURE	configural	557.22*** (208) / .06 /.93	37654.216		
		447.14***(207) /.05 / .95		NA	
	metric	13.53 (15) / .001 /.000	37620.295		
	scalar	184.00*** (15) /.012 /.039	37684.899		

- Exakt bildet die Veränderung ab
- BAMI: Lehnt die MI ab

RHS: Indikatoren und Mittelwerte Facebook

	model	*Exact MLR*	*Exact WLSMV*	*BAMI .01/.02 /.04*	*Alignment*
SOEP	metric	4	NA	4/4/4	0
	scalar	5	NA	6/5/5	3
ENSURE	metric	0	NA	0/0/0	0
	scalar	5	NA	4/3/3	3

model	*Exact MLR*	*Exact WLSMV*	*BAMI*	*Alignment*
SOEP	0.09			
ENSURE	0.15			

- Ergebnisse wie bei der Zufallsstichprobe

Ergebnisse zum Methodenvergleich

Model Fit

- Exakte MI (MLR): Bessere MI für die nach den KP revidierten Instrumente; Beschränkung auf Konfigural beim RHS
- BAMI: Ablehnung der MI

Indikatoren

- Mehr non-invariate Indikatoren erwartungskonform für exakt und BAMI
- Alignment am wenigsten sensitiv

Mittelwertvergleiche

- Übereinstimmung zwischen den Methoden
- Alignment: keine Justierung; kann die Unterschiede über- oder unterschätzen

Diskussion

Exakt und BAMI können Probleme in der Vergleichbarkeit identifizieren

Exakt und BAMI: Ergebnisse sind ähnlich, **BAMI ist jedoch strenger** als Exakt

In Bezug auf den Vergleich der Mittelwerte: **Methoden liefern ähnliche Ergebnisse, unabhängig von der Erfüllung der MI** - > Interpretation des Forschers; Theoriegeleitete Hypothesen wichtig

Probleme mit der Konfiguralen MI: Nur Exakt anwendbar, andere Methoden setzen sie voraus

Ausblick

RHS: Skalare MI für ENSURE-Instrument durch die Auswahl der Indikatoren erreichbar

Reliabilität möglicherweise aussagekräftiger für den Bias beim Mittelwertvergleich

Korrelationen zu Drittvariablen für die Überprüfung der Annahmen bzgl. der metrischen MI

Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:
natalja.menold@tu-dresden.de

Cognitive Pretests as Method to improve Cross-Cultural Comparability

Previous research has compared pretesting methods in their ability to

- —predict validity / reliability of survey data (Maitland & Presser 2016; Yan et al.,2012)
- —detect errors with draft questions (Presser & Blair 1994)
- —predict errors with survey questions (Forsyth et al 2004; Maitland & Presser 2017)
- —improved MI and reliability for a stereotypes measure, but more stable results by web probing and cross-cultural pretest review (Menold et al. 2022)

Little evaluation research:
- How helpful /effective with respect to measurement invariance?
- No studies for adaptations for refugees in a country