

# Psychologische Testtheorie

Sitzung 12

**Validität**



We are happy to share our materials openly:

The content of these Open Educational Resources by Lehrstuhl für Psychologische Methodenlehre und Diagnostik, Ludwig-Maximilians-Universität München is licensed under CC BY-SA 4.0. The CC Attribution-ShareAlike 4.0 International license means that you can reuse or transform the content of our materials for any purpose as long as you cite our original materials and share your derivatives under the same license.

1. Eine CFA hat ergeben, dass die drei Items eines neu konstruierten Tests dem  $\tau$ -kongenerischen Modell folgen. Der vorliegende R-Output zeigt, dass die Reliabilität des ersten Items bei  $rel(x_1) = 0.568$  liegt.

Latent Variables:						
	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
f1 =~						
item1	0.436	0.020	21.712	0.000	0.436	0.568
item2	0.390	0.019	20.359	0.000	0.390	0.538
item3	0.504	0.020	25.496	0.000	0.504	0.647

2. Ihnen liegen drei Items eines parallelen Modells vor, mit  $rel(x_1) = rel(x_2) = rel(x_3) = 0.50$ . Die Reliabilität des Itemmittelwerts liegt bei  $rel(\bar{x}) = 0.75$ .

Hinweis: Vereinfachten Spearman-Brown-Formel:  $rel(\bar{x}) = \frac{k}{(k-1) + \frac{1}{rel(x_i)}}$

3. Bei Im  $\tau$ -kongenerischen Modell fällt Conbachs Alpha stets höher oder gleich hoch wie McDonalds Omega aus und sollte somit nicht zur Reliabilitätsschätzung verwendet werden.

4. Ihnen liegt folgende Reliabilitätsschätzung in R für den Summenwert eines Tests vor, der einem essentiell tau-äquivalenten Modell folgt.

```
$est  
[1] 0.82503821  
  
$se  
[1] 0.01914212  
  
$ci.lower  
[1] 0.78034891  
  
$ci.upper  
[1] 0.84181121  
  
$conf.level  
[1] 0.95  
  
$type  
[1] "alpha"
```

- a) Sie können davon ausgehen, dass die Reliabilität des Tests über .80 liegt.
- b) Durch die Hinzunahme eines weiteren Items erhöht sich die Reliabilität des Tests gewiss.

Sitzung	Datum	Thema	Themenblock
1	13.10.25	Einführung	Begriffe, Modellierung von Antwortverhalten durch Zufallsvariablen & mathematische Grundlagen der Testtheorie
2	20.10.25	Wahrscheinlichkeitstheoret. Grundlagen	
3	27.10.25	Testtheoretische Modelle I	Testtheoretische Modelle
4	03.11.25	Testtheoretische Modelle II	
5	10.11.25	Testtheoretische Modelle III	
6	17.11.25	Skalierung I	Gütekriterien psychologischer Tests
7	24.11.25	Skalierung II	
8	01.12.25	Faktorenanalyse I	
9	08.12.25	Faktorenanalyse II	
10	15.12.25	Reliabilität I	
11	22.12.25	Sitzung entfällt wegen Weihnachten!	
	30.12.25	Offizielle Winterpause	
	06.01.26		
12	12.01.26	Reliabilität II	
13	20.01.26	Validität	

➔ In der heutigen Vorlesung beschäftigen wir uns mit dem Gütekriterium der Validität.

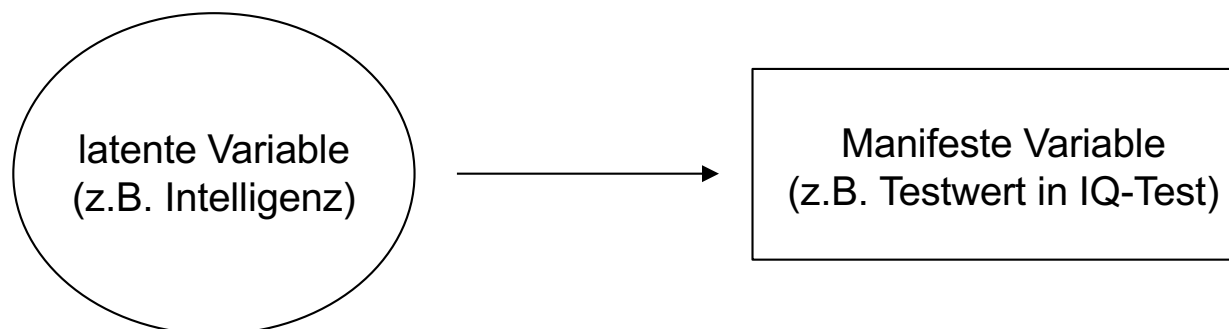
# 1. Einleitung

Die **Validität** gibt an, ob ein psychologischer Test auch wirklich das misst, was er zu messen beansprucht.

- Zentrale Frage:

**Verursacht eine Variation in der latenten Variable eine Variation im Testwert?**

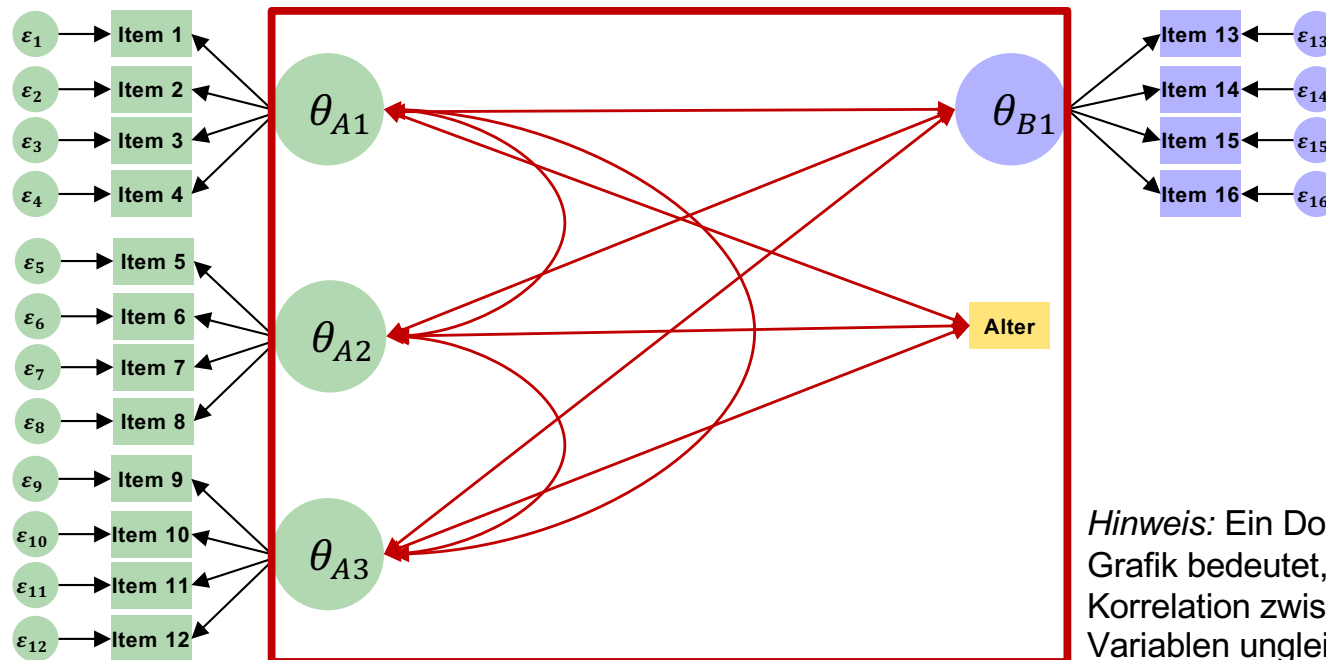
- Das (und nur das) ist Validität – es gibt nicht mehrere „Validitäten“
- Bei Validität geht es also eigentlich um Kausalität!



- Problem: Wir können die latente Variable nicht beobachten (oder gar manipulieren) und die Validität somit nicht direkt prüfen!

- Es gibt verschiedene Ansätze um die Validität dennoch zu untersuchen
- Wir brauchen eine **Arbeitsdefinition** zur Untersuchung der Validität eines Tests A:

Wir testen Hypothesen über den Zusammenhang der latenten Variable(n)  
 $\theta_{A1}, \theta_{A2}, \theta_{A3}$  des Tests A untereinander, sowie mit anderen Variablen (z.B. latente  
Variablen anderer Tests B, C, ... oder manifeste Variablen wie Alter)



*Hinweis:* Ein Doppelpfeil in der Grafik bedeutet, dass die Korrelation zwischen zwei Variablen ungleich 0 ist.

- Das theoretisch angenommene Beziehungsgeflecht zwischen den latenten Variablen innerhalb eines Tests, zwischen Tests und mit „messfehlerfreien“ Variablen wird als **nomologisches Netzwerk** bezeichnet
- Je präziser und umfassender das nomologische Netzwerk, desto besser kann die Validität eines Tests überprüft werden
- Wichtig ist, dass das nomologische Netzwerk und die damit verbundenen Hypothesen **vor der Datenerhebung** einer Validierungsstudie aufgestellt werden
- Die Hypothesen sollten **logisch begründbar** sein (in der Regel mit Bezug zu theoretischen Überlegungen oder bestehenden empirischen Befunden). Im Zweifel müssen kritische Stimmen durch gute Argumente von der Validität des Tests überzeugt werden. → Es herrscht nicht immer Einigkeit über die Validitätsprüfung eines Tests!
- **Je besser die Übereinstimmung empirischer Daten mit den zuvor aufgestellten Hypothesen, desto besser die Validität des Tests** → Ob ein Test valide ist, ist selten eine Ja/Nein Entscheidung, da oft manche Teile des nomologischen Netzwerks bestätigt werden und andere nicht!

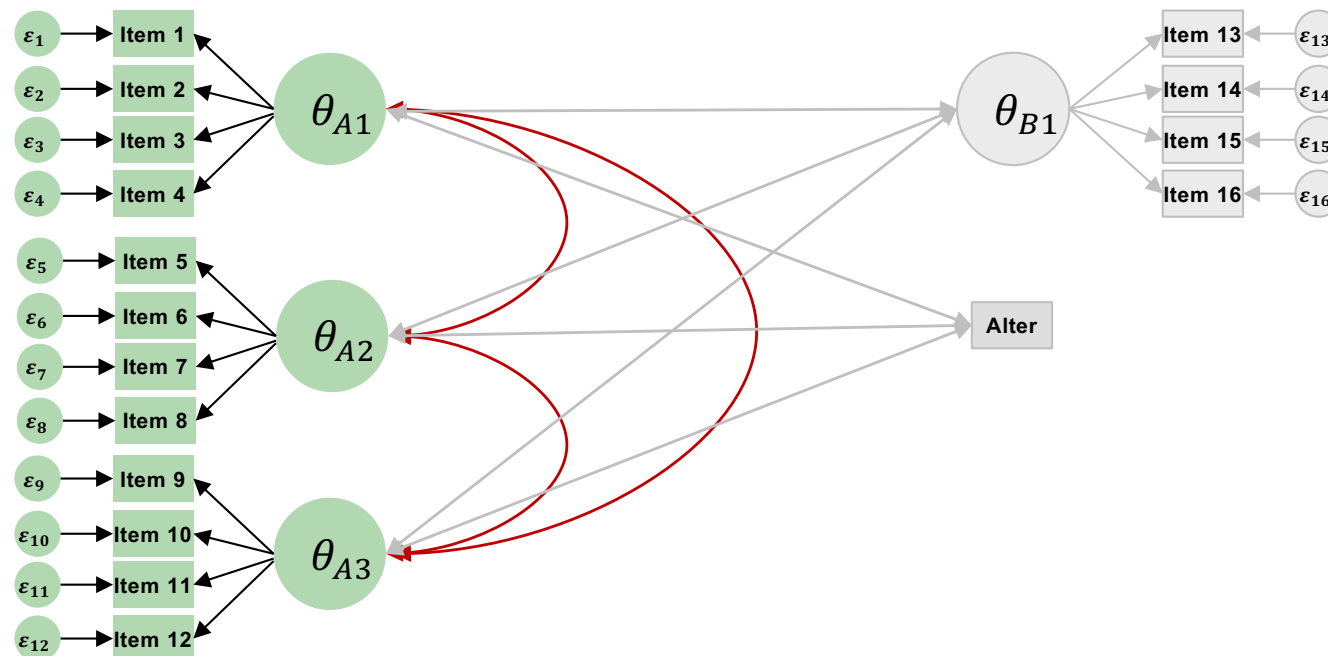


## 2. Zusammenhänge der latenten Variablen

## 2.1. Zusammenhänge zwischen den latenten Variablen eines Tests

## Zusammenhänge zwischen den latenten Variablen **eines** Tests I

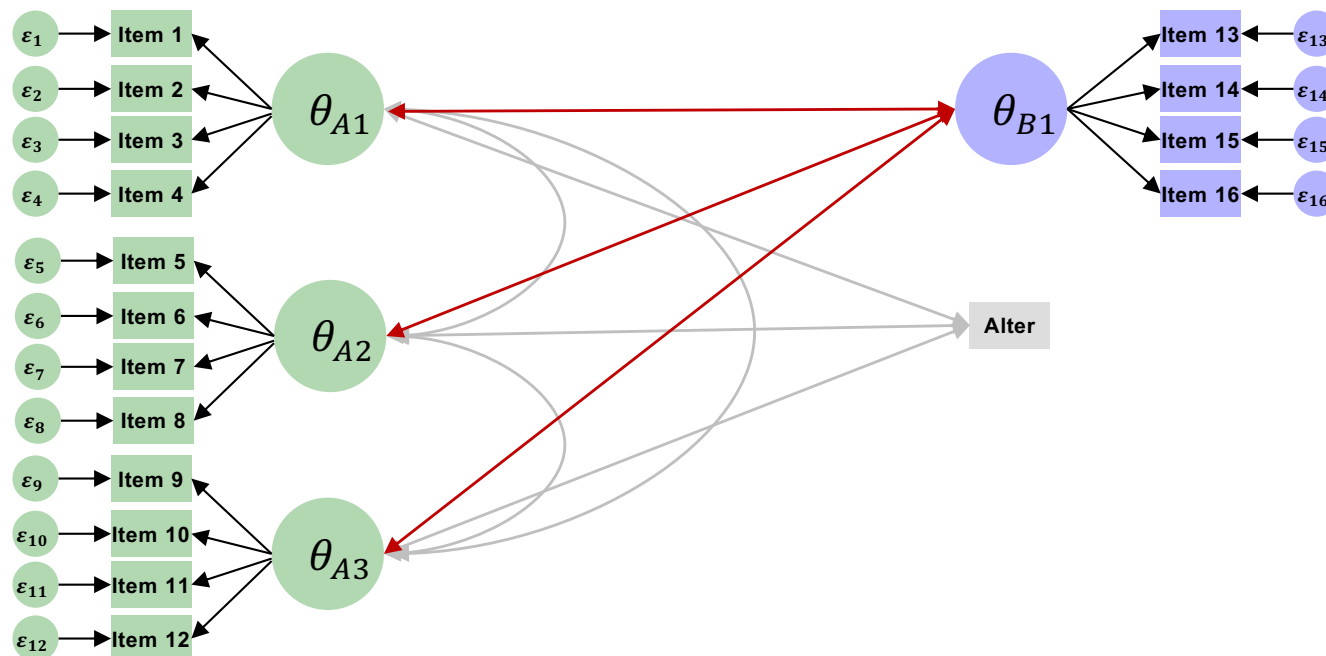
- Fragestellung: Sind die Korrelationen zwischen den latenten Variablen  $\theta_{A1}, \theta_{A2}, \theta_{A3}$  des Tests A mit der psychologischen Theorie vereinbar?
- Eng verwandt mit dem Gütekriterium der Skalierung: Faktorstruktur eines Tests ist wie erwartet



- Fragestellung: Sind die Korrelationen zwischen den latenten Variablen  $\theta_{A1}, \theta_{A2}, \theta_{A3}$  des Tests A mit der psychologischen Theorie vereinbar?
- Mögliche Hypothesen:
  - $COR(\theta_{A1}, \theta_{A2}) = 0$
  - $COR(\theta_{A1}, \theta_{A2}) > 0$
  - $COR(\theta_{A1}, \theta_{A2}) < 0.8$
  - $COR(\theta_{A1}, \theta_{A2}) > 0$  und  $COR(\theta_{A1}, \theta_{A3}) < 0$  und  $COR(\theta_{A2}, \theta_{A3}) = 0$
  - ...
- Hinweis: Je nach inhaltlicher Interpretation der latenten Variablen und deren Theorie, werden auch negative Korrelationen erwartet!

## 2.2. Zusammenhänge zwischen den latenten Variablen mehrerer Tests

- Fragestellung: Sind die Korrelationen zwischen der latenten Variablen  $\theta_{A1}$  des Tests A und den latenten Variablen  $\theta_B, \theta_C$  anderer Tests B, C mit der psychologischen Theorie vereinbar?

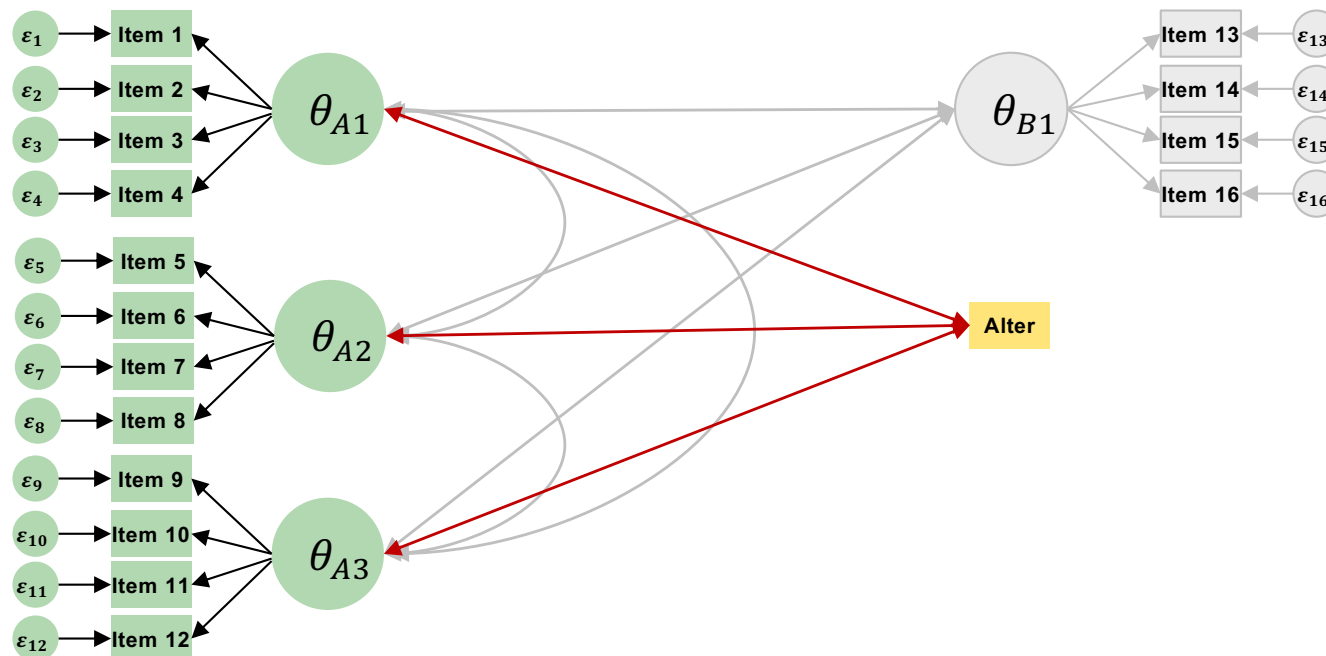


- Fragestellung: Sind die Korrelationen zwischen der latenten Variablen  $\theta_{A1}$  des Tests A und den latenten Variablen  $\theta_B, \theta_C$  anderer Tests B, C mit der psychologischen Theorie vereinbar?
- Mögliche Hypothesen:
  - $COR(\theta_{A1}, \theta_B) = 0$
  - $COR(\theta_{A1}, \theta_B) > 0$
  - $COR(\theta_{A1}, \theta_B) = 1$
  - $COR(\theta_{A1}, \theta_B) > 0.6$
  - $COR(\theta_{A1}, \theta_B) < 0.2$
  - $COR(\theta_{A1}, \theta_B) > COR(\theta_{A2}, \theta_C)$
  - ...
- *Hinweis*: Je nach inhaltlicher Interpretation der latenten Variablen und inhaltlicher Theorie, werden auch negative Korrelationen erwartet!

## 2.3. Zusammenhänge zwischen den latenten Variablen und messfehlerfreien Variablen



- Fragestellung: Sind die Korrelationen zwischen den latenten Variablen  $\theta_{A1}, \theta_{A2}, \theta_{A3}$  des Tests A mit manifesten Variablen (sogenannten „Kriterien“)  $K_1, K_2$  mit der psychologischen Theorie vereinbar?



- Fragestellung: Sind die Korrelationen zwischen den latenten Variablen  $\theta_{A1}, \theta_{A2}, \theta_{A3}$  des Tests A mit manifesten Variablen (sogenannten „Kriterien“)  $K_1, K_2$  mit der psychologischen Theorie vereinbar?
- Mögliche Hypothesen:
  - $COR(\theta_{A1}, K_1) = 0$
  - $COR(\theta_{A1}, K_2) > 0$
  - $COR(\theta_{A1}, K_2) > 0.4$
  - $COR(\theta_{A1}, K_1) > COR(\theta_{A2}, K_1)$
  - $COR(\theta_{A1}, K_1) > COR(\theta_{A1}, K_2)$
  - ...
- *Hinweis*: Je nach inhaltlicher Interpretation der latenten Variablen und inhaltlicher Theorie, werden auch negative Korrelationen erwartet!

# 3. Überprüfung der Validität in der Praxis

- Eine formale Testung der Hypothesen, die zur Überprüfung der Validität eines psychologischen Tests aufgestellt wurden, erfolgt am besten mithilfe von **Strukturgleichungsmodellen** (eine Erweiterung der CFA; siehe Mastervorlesung)
- Hier werden wir mithilfe von **Faktorenanalysen** und **beobachteten Korrelation von Summenwerten** zumindest beurteilen, ob unsere Daten mit den aufgestellten Hypothesen konform sind
  - In der Regel weisen wir die Validität dabei nicht über die Hypothesentests nach
  - Gründe sind u.a. die Inflation des Fehlers erster Art und die Komplexität der Hypothesenprüfung in konfirmatorischen Faktorenanalysen
  - **Stattdessen nehmen wir vorher ein bestimmtes Muster von Zusammenhängen an und inspizieren deskriptiv, ob dieses Muster so eingetreten ist**

Zusammenhänge zwischen den latenten Variablen eines oder mehrerer Tests:

Ansatz 1: Gemeinsame **Faktorenanalyse** der Items aller beteiligten Tests

- Passt die Anzahl der latenten Variablen und die Zuordnung der Items zur Theorie?
  - Optimalfall 1: Theoretisch abgeleitete Anzahl entspricht dem Ergebnis der Parallelanalyse (oder anderer Methode zur Bestimmung der Faktorenanzahl) und oblique rotierte Faktoren in der **exploratorischen Faktorenanalyse** können im Sinne der Theorie interpretiert werden (z.B. nahe an Einfachstruktur)
  - Optimalfall 2: Es ist bekannt, welche Items welche latenten Variablen messen, das Modell wird im Rahmen einer **konfirmatorischen Faktorenanalyse** spezifiziert und ist mit den Daten vereinbar (z.B. Hypothesentest, Fit-Indizes, Informationskriterien)
- Entsprechen die geschätzten Korrelationen zwischen den latenten Variablen den aus der psychologischen Theorie abgeleiteten Hypothesen?

Ansatz 2: **Korrelation** der Summenwerte der Items der Skalen des/der Test(s)

Zusammenhänge zwischen der latenten Variable eines Tests und einer manifesten Variable (Kriterium):

Ansatz 1: Gemeinsame **konfirmatorische Faktorenanalyse** der Items des Tests und des Kriteriums

- Entspricht die geschätzte Korrelation zwischen der latenten Variable und dem Kriterium den aus der psychologischen Theorie abgeleiteten Hypothesen?

Ansatz 2: **Korrelation** des Summenwertes der Items der Skala des Tests mit der manifesten Variable

Im Rahmen der Validitätsprüfung gelten weiterhin die **Modellannahmen des mehrdimensionalen  $\tau$ -kongenerischen Modells**. Und zwar auch wenn die Validierung mithilfe von Korrelationen statt Faktorenanalysen erfolgt.

Teilweise werden implizit sogar zusätzliche Modellannahmen getroffen.

- Zusammenhänge zwischen latenten Variablen eines Tests:
  - Beachte die folgende Annahme aus dem mehrdimensionalen  $\tau$ -kongenerischen Modell:  $COV(\varepsilon_i, \varepsilon_j) = 0$  für alle Itempaare  $i, j$  gilt auch über verschiedene latente Variablen hinweg
- Zusammenhänge zwischen latenten Variablen mehrerer Tests:
  - Zusätzliche Annahme: Die Fehler müssen nicht nur innerhalb, sondern auch über beide Tests A und B hinweg unabhängig sein, sodass  $COV(\varepsilon_{Ai}, \varepsilon_{Bj}) = 0$  für alle Itempaare  $i, j$  ( $i = 1, \dots, k_A$  und  $j = 1, \dots, k_B$ )

- Da die **strenger**en Testmodelle (parallel, essentiell parallel,  $\tau$ -äquivalent, essentiell  $\tau$ -äquivalent) **Spezialfälle** des (mehrdimensionalen)  $\tau$ -kongenerischen Modells sind, funktioniert das Vorgehen auch, wenn für die einzelnen Tests sogar eines der strenger



## 4. Verschiedene Validitätslabels

Die **Validität** gibt an, ob ein psychologischer Test auch wirklich das misst, was er zu messen beansprucht.

Arbeitsdefinition zur Überprüfung der Validität eines Tests A:

- Testung von **Hypothesen** über den Zusammenhang der latenten Variable(n)  $\theta_{A1}, \theta_{A2}, \theta_{A3}, \dots$  des Tests A **untereinander**, sowie mit **anderen** Variablen (z.B. latente Variable(n) anderer Tests B, C, ...; manifeste Variablen wie Alter)
- Je nach **Art dieser Hypothesen** gibt man den verschiedenen Ansätzen zur Prüfung der Validität häufig **verschiedene Labels**:
  - Konstruktvalidität
    - Faktorielle Validität
    - Konvergente Validität
    - Divergente Validität
  - Kriteriumsvalidität
  - (Inhaltsvalidität)

- Hypothesen zum Zusammenhang zwischen mehreren latenten Variablen eines Tests werden in der Praxis häufig unter dem Begriff der **faktoriellen Validität** zusammengefasst
- Hypothesen zu Zusammenhängen zwischen den latenten Variablen mehrerer Tests werden häufig unter dem Begriff der **Konstruktvalidität** zusammengefasst
  - Hypothesen zu Zusammenhängen von latenten Variablen mit *gleichem Gültigkeitsbereich* fallen häufig unter den Begriff der **konvergenten Validität**
  - Hypothesen zu Zusammenhängen von latenten Variablen mit *unterschiedlichem Gültigkeitsbereich* fallen häufig unter den Begriff der **divergenten Validität**
- Hypothesen zu Zusammenhängen zwischen den latenten Variablen eines Tests mit messfehlerfreien Variablen werden häufig unter dem Begriff der **Kriteriumsvalidität** zusammengefasst



- Validitätslabels werden **nicht immer gleich verwendet**. Zum Beispiel ...
    - ... versteht man unter konvergenter Validität manchmal allgemein Hypothesen, bei denen *hohe Zusammenhänge* zwischen latenten (oder anderen) Variablen erwartet werden
    - ... versteht man unter divergenter Validität manchmal allgemein Hypothesen, bei denen *niedrige Zusammenhänge* zwischen latenten (oder anderen) Variablen erwartet werden
- Ab welcher Höhe ein Zusammenhang hoch oder niedrig ist, ist nicht eindeutig!



- Validitätslabels werden nicht immer gleich verwendet. Zum Beispiel ...
  - ... versteht man unter Kriteriumsvalidität auch Hypothesen zum Zusammenhang mit anderen latenten Variablen, die in praktischen Anwendungen vorhergesagt werden sollen
- Wann eine Variable den Status eines Kriteriums verdient hat, ist nicht eindeutig!



- Eine eindeutige Definition verschiedener Labels von Validität ist daher leider nicht möglich, aber auch nicht unbedingt notwendig (oder sogar hilfreich)
- Letztendlich sind diese Begriffe nur Labels, um die vielfältigen Hypothesen innerhalb eines nomologischen Netzwerks zu strukturieren und in zusammenfassender Weise zu berichten, welche Teile des Netzwerks sich in empirischen Untersuchungen bestätigt haben. Z.B. verdeutlichen die Labels konvergente und divergente Validität die Wichtigkeit (und Schwierigkeit) der Eingrenzung und Abgrenzung psychologischer Konstrukte in der Praxis
- Wenn im Einzelfall nicht genau klar ist, was mit einem Label gemeint ist, sollte es immer möglich sein, sich auf der Ebene der konkreten Hypothesen ein umfassendes Bild über die Validität eines Tests zu verschaffen
- Eine **aussagekräftige Validierung** eines Tests sollte nicht durch eine zu starre Orientierung an Validitätslabels beeinträchtigt werden. In der Regel gilt:  
Je spezifischer und umfassender die logisch aus einer Theorie abgeleiteten Hypothesen, desto höher der Erkenntnisgewinn (wenn diese geprüft werden)

- Arbeitsdefinition: Hypothesen über die inhaltliche Beschaffenheit eines Konstrukts (einer latenten Variable) und damit seiner Indikatoren (Items)
- Keine statistische Frage und daher nicht direkt Teil der psychologischen Testtheorie
  - Inhaltliche Fächer: Persönlichkeitspsychologie, Klinische Psychologie, etc.
  - Vorlesung/UK Diagnostik II
- Bilden die Items das Konstrukt sind seiner ganzen Breite repräsentativ ab? Wenn ich alle Items des Konstrukts kennen würde und in eine Urne lege, könnte ich eine repräsentative Itemmenge daraus ziehen? → In der Praxis meist nicht möglich
- Muss durch eine gute Theorie oder Definition bzw. Experten oder technischen Hilfsmitteln (z.B. automatische Itemgenerierung) erzielt werden
- Kann durch statistische Verfahren der Itemanalyse unterstützt werden
  - UK Fragebogenentwicklung

## 5. Beispiel



## Beispiel: Fragebogen irrationaler Einstellungen (FIE) von Klages (1989)

- Der Test misst vier Skalen:
  - Negative Selbstbewertung (nsb)
  - Abhängigkeitskognitionen (abk)
  - Irritation und Gefühlsexternalisierung (ivm)
  - Internalisierung von Misserfolgen (ige)
- Erwartungen an die Zusammenhänge:
  - Alle Korrelationen (s. rechts) signifikant
  - Höhere Korrelationen innerhalb des FIE als mit dem Beck-Depressions-Inventar (BDI; Hautzinger et al., 1994) → höhere faktorielle als divergente Korrelationen

### Hypothesen zur faktoriellen Validität:

- $COR(\theta_{nsb}, \theta_{abk}) > 0$
- $COR(\theta_{nsb}, \theta_{ivm}) > 0$
- $COR(\theta_{nsb}, \theta_{ige}) > 0$
- $COR(\theta_{abk}, \theta_{ivm}) > 0$
- $COR(\theta_{abk}, \theta_{ige}) > 0$
- $COR(\theta_{ivm}, \theta_{ige}) > 0$

### Hypothesen zur divergenten Validität:

- $COR(\theta_{nsb}, \theta_{bdi}) > 0$
- $COR(\theta_{abk}, \theta_{bdi}) > 0$
- $COR(\theta_{ivm}, \theta_{bdi}) > 0$
- $COR(\theta_{ige}, \theta_{bdi}) > 0$

### Hypothesen zur Kriteriumsvalidität:

- $COR(\theta_{nsb}, \text{Alter}) > 0$
- $COR(\theta_{abk}, \text{Alter}) > 0$
- $COR(\theta_{ivm}, \text{Alter}) > 0$
- $COR(\theta_{ige}, \text{Alter}) > 0$

- Außerdem aus der Theorie und Empirie abgeleitet (Bühner & Müller, 2006, S. 2 & 3):

„Welche Skalen im FIE sollten mit dem Beck Depressionsinventar (BDI; Beck, Ward, Mendelson, Mock & Erbaugh, 1961) hoch positiv korrelieren? Vor allem zwischen Internalisierung von Misserfolgen (gemessen mit dem Attributional Style Questionnaire, ASQ, von Stiensmeier, Kammer, Pelster & Niketta, 1985) und BDI-Werten fanden sich immer wieder signifikante Zusammenhänge wie z. B. bei Försterling, Bühner und Gall (1998). Überträgt man diese Befunde auf den FIE, kann erwartet werden, dass auch **zwischen der FIE-Skala „Internalisierung von Misserfolgen“ und dem BDI hohe positive Korrelationen** bestehen.

Zudem sollten **hohe positive Zusammenhänge zwischen der Skala „Negative Selbstbewertung“ und dem BDI-Gesamtwert** auftreten, da eine negative Selbstbewertung als eine von mehreren Ursachen für das Entstehen einer Depression diskutiert wird (Beck, 1967) und auch der ATQ eine inhaltähnliche Skala enthält.“

„Für die **Skala „Abhängigkeitskognitionen“ wird in Analogie zu Befunden von Joormann (2004) eine geringe Korrelation mit dem BDI** angenommen. Sie fand lediglich Korrelationen von  $r = .22$  zwischen einer inhaltlich ähnlichen Skala der DAS „Anerkennung durch andere“ und einem Depressionsfragebogen (Allgemeine Depressionsskala, ADS; Hautzinger & Bailer, 1993). Zum **Zusammenhang zwischen der FIE-Skala „Irritierbarkeit“ und dem BDI bestehen keine Hypothesen**. Ihr Zusammenhang mit dem BDI und ihre potenzielle Nützlichkeit zur Trennung von Depressiven und nicht Depressiven soll explorativ untersucht werden.“

- ➔ nsb und ivm sollten höher mit bdi korrelieren als abk ( $r$  ungefähr 0.22)
- ➔ für ige gibt es keine Vorannahmen

- Annahmen zur Konstruktvalidität:
  - nsb, abk, ivm und ige korrelieren untereinander positiv ✓
  - nsb, abk, ivm und ige korrelieren positiv mit bdi ✓
    - nsb und ivm korrelieren höher mit bdi als abk ✓
    - keine Hypothese für ige
  - nsb, abk, ivm und ige korrelieren untereinander höher als mit bdi ✗
- Beobachtete Pearson-Korrelationen der Skalen bzw. Fragebögen (Summenwerte):

	fie_nsb	fie_abk	fie_ivm	fie_ige	bdi_ges
fie_nsb		0.42	0.74	0.65	0.72
fie_abk	0.42		0.58	0.61	0.37
fie_ivm	0.74	0.58		0.63	0.59
fie_ige	0.65	0.61	0.63		0.54
bdi_ges	0.72	0.37	0.59	0.54	

- Problem: Drei der FIE-Skalen weisen untereinander **niedrigere Korrelationen auf als jeweils mit dem BDI!**

→ **Das erwartete Muster passt nicht, was gegen die Konstruktvalidität spricht!**

- Annahmen zur Kriteriumsvalidität:
  - nsb, abk, ivm und ige korrelieren positiv mit Alter **✗**
- Beobachtete Pearson-Korrelation zwischen den FIE-Skalen (Summenwerte) und Alter:

	fie_nsb	fie_abk	fie_ivm.	fie_ige
Alter	0.046 n.s.	0.034 n.s.	0.005 n.s.	0.009 n.s.

- Problem: Alle vier FIE-Skalen sind **nicht mit Alter korreliert!**  
→ **Das erwartete Muster passt nicht, was gegen die die Kriteriumsvalidität spricht!**

### Zusammenfassend:

- Konstruktvalidität nicht gegeben → *partielle Itemüberlappung mit dem BDI?*
- Kriteriumsvalidität nicht gegeben → *problematische Altersverteilung der Stichprobe?*
- Der FIE sollte überdacht werden (?)

- Die Ergebnisse der Korrelationen der Skalen-Summenwerte können mit den Korrelationen der latenten Variablen aus der Faktorenanalyse verglichen werden:

Pearson-Korrelationen der Summenwerte:

	fie_nsb	fie_abk	fie_ivm	fie_ige	bdi_ges
fie_nsb	1.00	0.42	0.74	0.65	0.72
fie_abk	0.42	1.00	0.58	0.61	0.37
fie_ivm	0.74	0.58	1.00	0.63	0.59
fie_ige	0.65	0.61	0.63	1.00	0.54
bdi_ges	0.72	0.37	0.59	0.54	1.00

Korrelationen aus der Faktorenanalyse (CFA):

Covariances	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
NSB ~~						
ABK	0.468	0.070	6.692	0.000	0.468	0.468
IVM	0.843	0.035	24.020	0.000	0.843	0.843
IGE	0.812	0.045	18.007	0.000	0.812	0.812
ABK ~~						
IVM	0.624	0.064	9.760	0.000	0.624	0.624
IGE	0.777	0.057	13.709	0.000	0.777	0.777
IVM ~~						
IGE	0.789	0.054	14.599	0.000	0.789	0.789

- Korrelationen der Summenwerte und in Klammern die Korrelationen aus der CFA:

	fie_nsb	fie_abk	fie_ivm.	fie_ige
fie_nsb				
fie_abk	0.42 (0.468)			
fie_ivm	0.74 (0.843)	0.58 (0.624)		
fie_ige	0.65 (0.812)	0.61 (0.777)	0.63 (0.789)	
OMEGA	0.87	0.74	0.74	0.70

- Die im Rahmen der konfirmatorischen Faktorenanalyse geschätzten Korrelationen sind höher als die Pearson-Korrelationen der Summenwerte!
- Wie lassen sich diese Unterschiede in den Korrelationen erklären?
  - Die Korrelationen in konfirmatorischen Faktorenanalysen sind Korrelationen zwischen den geschätzten wahren Werten der latenten Variablen
  - Die Pearson-Korrelationen sind Korrelationen zwischen messfehlerbehafteten Summen- oder Mittelwerten der manifesten Variablen (d.h., der Items)

## 8. Minderungskorrektur

- In der Praxis wird zur Überprüfung der **Konstruktvalidität** häufig nicht die Korrelation zwischen den latenten Variablen zweier Tests A und B betrachtet, sondern einfach die Korrelation zwischen dem Summenwert (oder Mittelwert) des einen Tests A und dem Summenwert (oder Mittelwert) des anderen Tests B
- Dies ist nicht optimal, da die Items beider Tests Messfehler aufweisen und somit die Korrelation ihrer Summenwerte (oder Mittelwerte) die Korrelation zwischen den latenten Variablen unterschätzt
- Häufig wird daher folgende Formel zur Schätzung von  $COR(\theta_A, \theta_B)$  verwendet:

$$COR(\theta_A, \theta_B) = \frac{COR(\sum_{i=1}^{k_A} X_{iA}, \sum_{j=1}^{k_B} X_{jB})}{\sqrt{REL(\sum_{i=1}^{k_A} X_{iA}) \cdot REL(\sum_{j=1}^{k_B} X_{jB})}}$$

- Bei dieser **doppelten Minderungskorrektur** wird die Korrelation der Summenwerte um die **Messungenauigkeit beider Summenwerte** korrigiert
- Je niedriger die Reliabilitäten der Summenwerte, desto stärker die Korrektur und desto höher die geschätzte Korrelation zwischen den latenten Variablen (bei gleichbleibendem Zähler)



- Korrelationen der Summenwerte und in Klammern die Korrelationen aus der CFA:

	fie_nsb	fie_abk	fie_ivm.	fie_ige
fie_nsb				
fie_abk	0.42 (0.468)			
fie_ivm	0.74 (0.843)	0.58 (0.624)		
fie_ige	0.65 (0.812)	0.61 (0.777)	0.63 (0.789)	
OMEGA	0.87	0.74	0.74	0.70

- Beispiel: Doppelt-minderungskorrigierte Korrelation zwischen nsb und ige

$$COR(\theta_{nsb}, \theta_{ige}) = \frac{0.65}{\sqrt{0.87 \cdot 0.70}} = 0.83$$

- Es handelt sich um eine Schätzung der um den Messfehler bereinigten wahren Korrelation der latenten Variable Negative Selbstbewertung (nsb) mit Irritation und Gefühlsexternalisierung (ige)
- Solche Korrelationen werden manchmal in Metaanalysen verwendet, um die wahren messfehlerfreien Korrelationen zwischen Konstrukten zu ermitteln

- In der Praxis wird zur Überprüfung der **Kriteriumsvalidität** häufig nicht die Korrelation zwischen der latenten Variablen eines Tests A und dem manifesten Kriterium K betrachtet, sondern einfach die Korrelation des Summenwerts (oder Mittelwerts) des Tests und dem Kriterium
- Dies ist nicht optimal, da die Items des Tests Messfehler aufweisen und somit die Korrelation des Summenwerts (oder Mittelwertes) mit dem Kriterium die Korrelation zwischen der latenten Variable und dem Kriterium unterschätzt
- Häufig wird daher folgende Formel zur Schätzung von  $COR(\theta_A, K)$  verwendet:

$$COR(\theta_A, K) = \frac{COR(\sum_{i=1}^{k_A} X_{iA}, K)}{\sqrt{REL(\sum_{i=1}^{k_A} X_{iA})}}$$

- Diese **einfache Minderungskorrektur** ergibt sich für den Spezialfall, dass  $REL(K) = 1$  („das manifeste Kriterium K wird ohne Messfehler erfasst“)
- Je niedriger die Reliabilität des Summenwerts, desto stärker die Korrektur und desto höher die geschätzte Korrelation mit der latenten Variable (bei gleichbleibendem Zähler)

- Pearson-Korrelation zwischen den FIE-Skalen (Summenwerte) und Alter:

	fie_nsb	fie_abk	fie_ivm.	fie_ige
Alter	0.046 n.s.	0.034 n.s.	0.005 n.s.	0.009 n.s.

- Beispiel: Einfach-minderungskorrigierte Korrelation zwischen nsb und Alter:

$$COR(\theta_{nsb, alter}) = \frac{0.046}{\sqrt{0.87}} = 0.049$$

- Es handelt sich um eine Schätzung der um den Messfehler bereinigten wahren Korrelation von der latenten Variable Negative Selbstbewertung (nsb) mit dem Kriterium Alter

- Beide Formeln der Minderungskorrektur nehmen an, dass für jeden Test jeweils das  **$\tau$ -kongenerische Modell** oder ein strengeres Modell (parallel, essentiell parallel,  $\tau$ -äquivalent, essentiell  $\tau$ -äquivalente) gilt.
- Wichtig: Die doppelte Minderungskorrektur trifft eine zusätzliche Annahme:
  - Die Fehler der Items sind nicht nur innerhalb der Tests (Annahme aller bisher besprochenen Testmodelle), sondern auch über beide Tests hinweg unkorreliert:  $COV(\varepsilon_{Ai}, \varepsilon_{Bj}) = 0$  für alle Itempaare  $i, j$  ( $i = 1, \dots, k_A$  und  $j = 1, \dots, k_B$ )
  - Diese Annahme kann nur im Rahmen von konfirmatorischen Faktorenanalysen (oder deren Erweiterung: Strukturgleichungsmodelle, siehe Mastervorlesung) sinnvoll überprüft werden
  - In diesem Fall ist jedoch auch die gesamte Korrekturformel unnötig, da das Strukturgleichungsmodell direkt einen Schätzwert für  $COR(\theta_A, \theta_B)$  ausgibt

- Um die Validität eines Tests A zu untersuchen, prüfen wir Hypothesen über den Zusammenhang der latenten Variable(n)  $\theta_{A1}, \theta_{A2}, \theta_{A3}$  des Tests untereinander, sowie mit anderen Variablen (z.B. latente Variablen anderer Tests B, C, ... oder manifeste Variablen wie Alter)
- Zur Schätzung dieser Zusammenhänge können wir verschiedene statistische Methoden verwenden:
  - Strukturgleichungsmodelle (siehe Mastervorlesung)
  - Faktorenanalysen (CFA & EFA) der beteiligten Items
  - Minderungskorrigierte Korrelationen der Summenwerte
    - Nachteil: Modellannahmen nicht automatisch mit-geprüft
  - Einfache Korrelationen der Summenwerte
    - Nachteil: Modellannahmen nicht automatisch mit-geprüft
    - Nachteil: Messfehler der manifesten Summenwerten nicht berücksichtigt

- *Ausblick:* In der nächsten Vorlesung starten wir mit der Einzelfalldiagnostik, d.h., der Frage wie psychologische Tests in der Praxis zur Diagnose einzelner Personen eingesetzt werden können
- *Aber zuerst:*
  - **Gibt es offene Fragen zur heutigen Vorlesung?**
  - Zur Vertiefung:
    - Übungsblatt 8 zur Validität auf Moodle
    - R Code zu den Beispielen der heutigen VL im Zusatzmaterial auf Moodle
    - Bühner (2021, S. 115-124)

- Klages, U. (1989). *Fragebogen irrationaler Einstellungen:FIE*. Verlag für Psychologie Dr. CJ Hogrefe.
- Hautzinger, M., Bailer, M., Worall, H., & Keller, F. (1994). *Beck-depressions-inventar (BDI)*. Bern: Huber.
- Bühner, M., & Müller, J. (2006). Validität des Fragebogens irrationaler Einstellungen (FIE). *Diagnostica*, 52(1), 1-10. <https://doi.org/10.1026/0012-1924.52.1.1>