Grundlagen der Testtheorie
WS 2020/21

4. Klassische Testtheorie
23.11.2020

Prof. Dr. Eunike Wetzel

Semesterplan

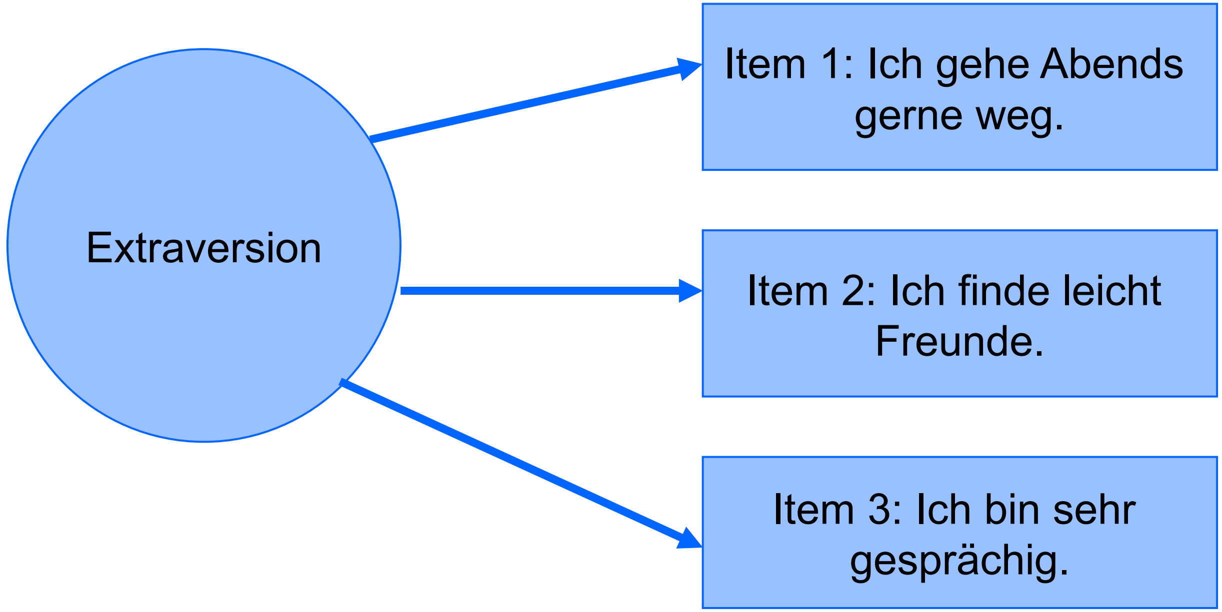
Sitzung	Termin	Thema
1	02.11.	Grundlagen & Gütekriterien
2	09.11.	Schritte der Testkonstruktion: Übersicht Konstruktdefinition & Itemgenerierung
3	16.11.	Erstellung eines Testentwurfs
4	**23.11.**	**Klassische Testtheorie**
5	07.12.	Item Response Theorie
6	14.12.	Exploratorische Faktorenanalyse 1

Was ist Testtheorie?

- Testtheorie beschäftigt sich mit dem Zusammenhang zwischen dem Antwortverhalten im Test und dem zu erfassenden Konstrukt
- Theoretischer Hintergrund zur Konstruktion und Interpretation von Testverfahren

Warum Testtheorie?

- Entspricht das Antwortverhalten direkt dem interessierenden Merkmal, benötigt man keine Testtheorie
- Bsp.: Treiben Sie regelmäßig Sport?
- Bei psychologischen Konstrukten ist eine Testtheorie vonnöten, da von dem Antwortverhalten im Test auf das latente Konstrukt geschlossen wird



Testtheorien

- Eine Testtheorie beschäftigt sich mit dem Zusammenhang zwischen dem Antwortverhalten im Test und dem zu erfassenden Konstrukt
- Wichtige Testtheorien in der Psychologie:

1. Klassische Testtheorie (KTT)
    - Älter, daher „klassisch“
    - Darauf basieren die meisten der auf dem Markt erhältlichen psychologischen Tests
2. Item Response Theorie (IRT)
    - Neuere Ergänzung und Weiterentwicklung der KTT

Klassische Testtheorie

1. Theoretische Konzepte der KTT
   1. Messfehler
   2. Wahrer Wert
2. Lokale Unabhängigkeit
3. Grundgleichung der KTT
4. Folgerungen aus der Grundgleichung
5. Messmodelle

Datenmatrix

		Items j =					y_i =
		1	2	3	4	5	
Personen i =	1	0	0	0	0	0	0
	2	0	0	0	0	1	1
	3	0	0	0	0	0	0
	4	0	0	0	1	1	2
	5	0	0	1	1	1	3
	6	0	1	0	0	1	2
	7	1	1	1	1	1	5
	8	0	1	1	1	1	4
	9	1	1	1	1	1	5
	10	0	0	1	1	1	3

1.1 Messfehler

- Bei mehreren Messungen mit demselben Gerät/Test schwanken die Messwerte
  - Physikalische Messungen: Waage, Thermometer, Radargerät
  - Psychologische Messungen: Testwerte im Intelligenztest, Persönlichkeitstest
    →Es treten Messfehler auf
- Mögliche Ursachen für Messfehler
  - Instruktion falsch verstanden
  - Antwortkästchen verwechselt
  - Elektrode verrutscht
- In der KTT werden unter dem Messfehler nur unsystematische Einflüsse verstanden

1.1 Messfehler

- Systematische Einflüsse werden dagegen nicht berücksichtigt
  - z.B. Antwortstile, sozial erwünschtes Antworten
- Auswirkungen von Messfehlern
  - Unterschätzung der Korrelation zweier Variablen

1.2 Wahrer Wert

- Engl. *true score* (T oder τ)
- Definiert als der Erwartungswert (theoretischer Mittelwert) der intraindividuellen Verteilung der beobachteten Messwerte einer Person

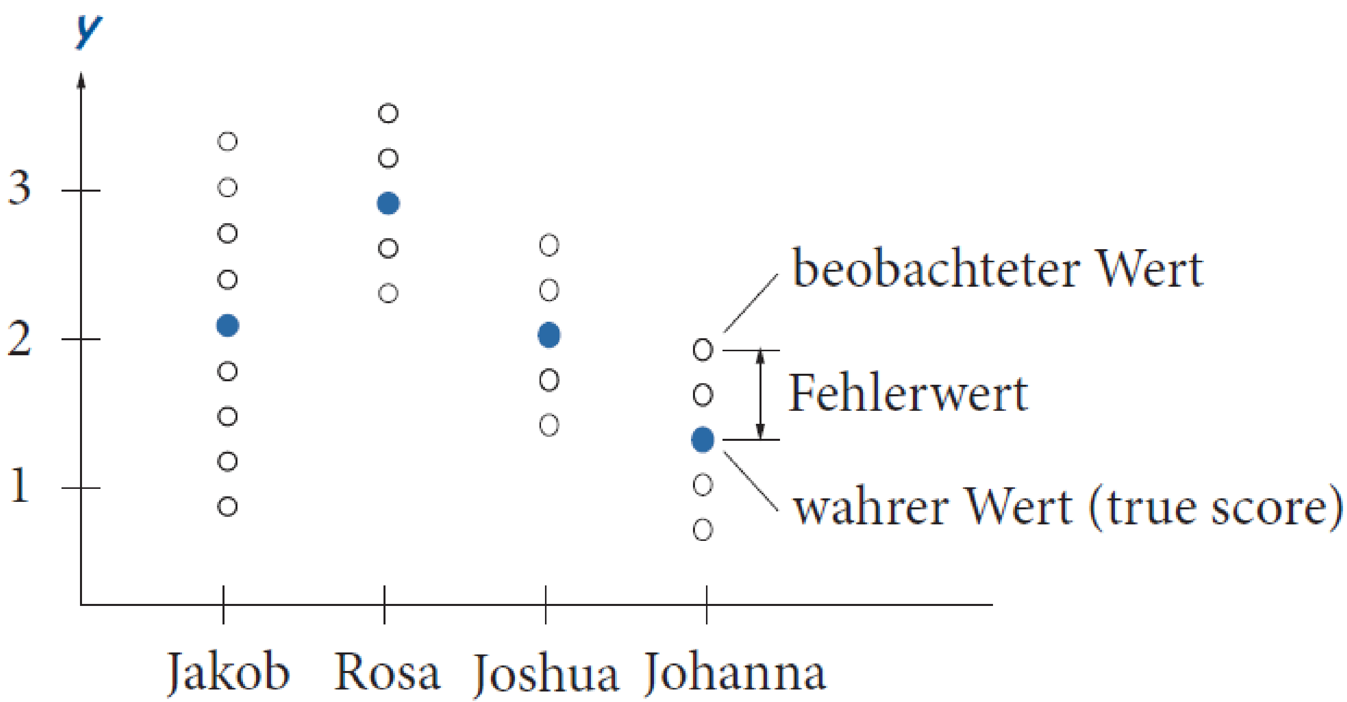


Abb. 22.1 aus Eid et al. (2010)

1.2 Wahrer Wert

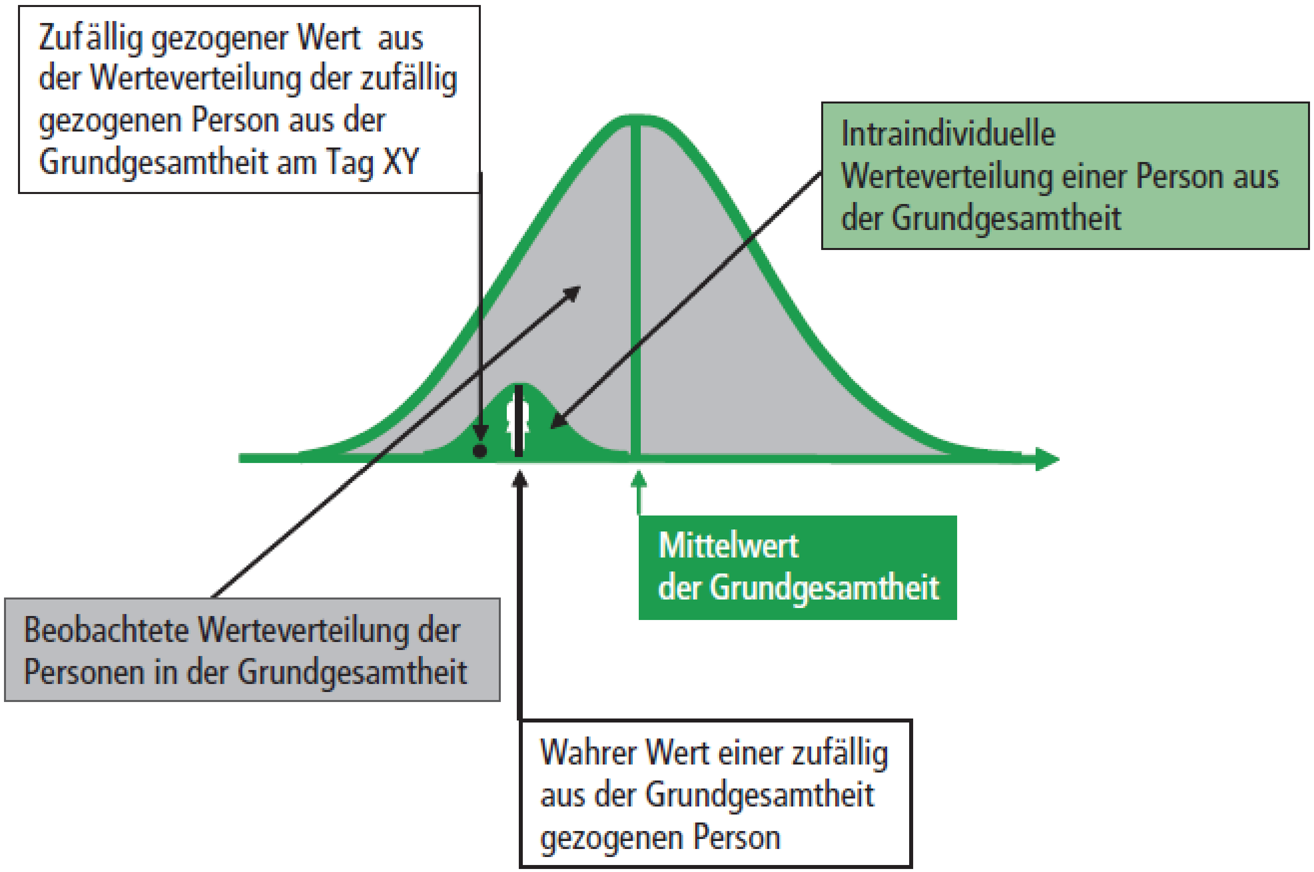
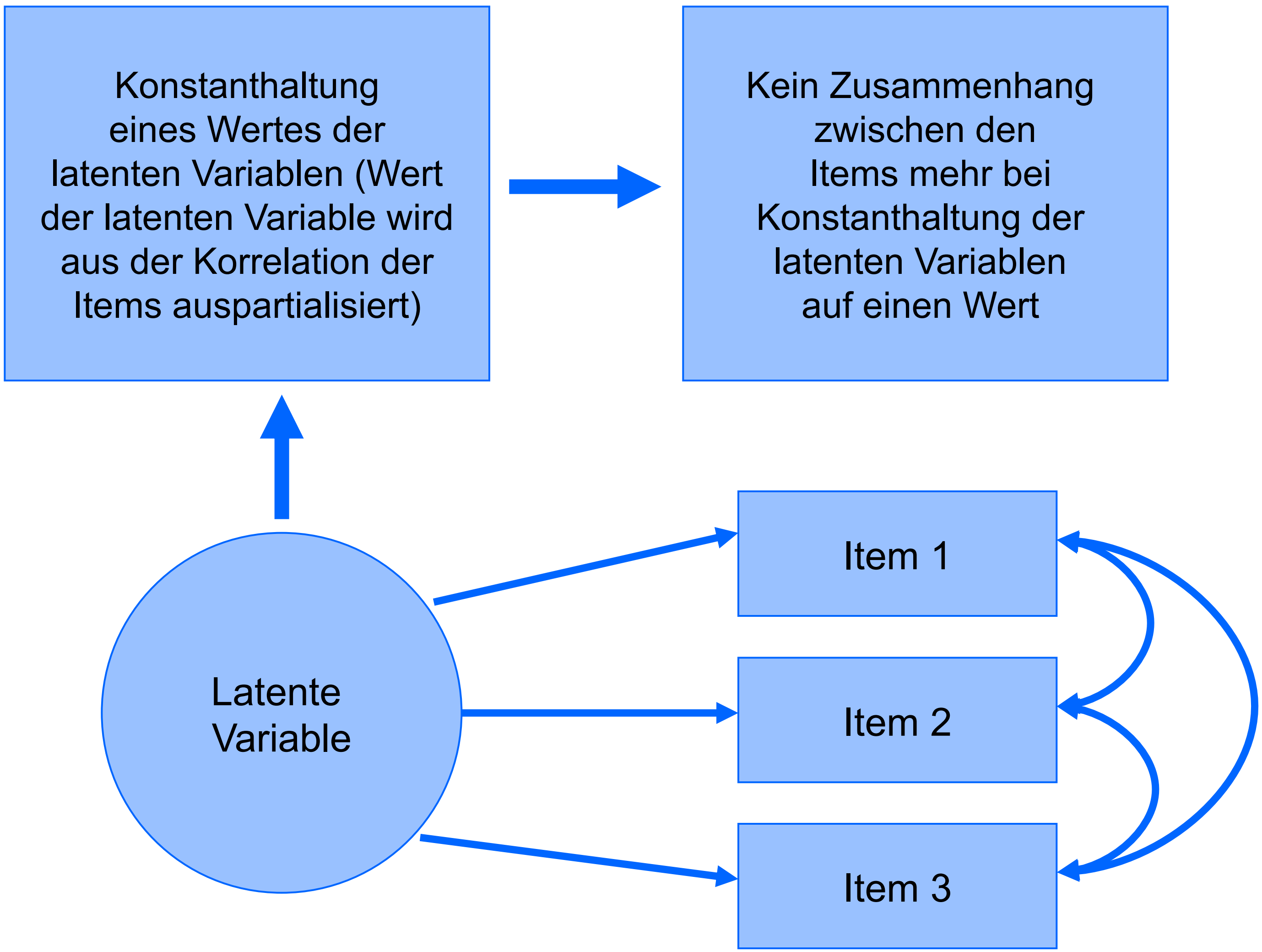


Abb. 2.5 aus Bühner (2011)

2. Lokale Unabhängigkeit

- Lokale Unabhängigkeit liegt vor, wenn die Itemantworten unter Kontrolle der Traitausprägung unabhängig voneinander sind
- Bei einem Test, der ein eindimensionales Konstrukt erfasst, müssen die Zusammenhänge zwischen Items bei lokaler Unabhängigkeit allein durch das zugrundeliegende Konstrukt erklärbar sein
- Ist eine Annahme der KTT und IRT
- Ist verletzt bei logischen Abhängigkeiten zwischen Items und systematischen Messfehlereinflüssen wie Antwortstilen und Konsistenzeffekten

2. Lokale Unabhängigkeit



Klassische Testtheorie

1. Theoretische Konzepte der KTT
    1. Messfehler
    2. Wahrer Wert
2. Lokale Unabhängigkeit
3. **Grundgleichung der KTT**
4. **Folgerungen aus der Grundgleichung**
5. Messmodelle

3. Grundgleichung der KTT

- Der beobachtete Wert einer Person *i* im Test *j* setzt sich zusammen aus ihrem wahren Wert und dem Messfehler:

$$y_{ij} = \tau_{ij} + \varepsilon_{ij}$$

- Über alle Personen *i* (*i* = 1,…,*n*) hinweg:

$$Y_j = \tau_j + \varepsilon_j$$

4. Folgerungen aus der Grundgleichung

Aus der Grundgleichung der KTT folgen 4 Eigenschaften der Messfehler- und True-Score-Variablen:

1. Der Erwartungswert einer Messfehlervariablen ist für jede Ausprägung der True-Score-Variablen gleich 0.
2. Der unbedingte Erwartungswert einer Messfehlervariablen ist gleich 0.
3. Messfehler- und True-Score-Variablen sind unkorreliert.
4. Die Varianz einer beobachteten Messwertvariablen lässt sich additiv zerlegen in die Varianz der True-Score-Variablen und die Varianz der Messfehlervariablen.

- Beachte: Diese Eigenschaften gelten nur für per Zufall aus der Population gezogene Personen.

4.1 Bedingter Erwartungswert von $\varepsilon_j = 0$

- Zwei Personen mit identischem True Score können unterschiedliche beobachtete Werte erhalten, aber über viele Personen mit identischen True Scores mitteln sich die Messfehler aus
- Der Erwartungswert einer Messfehlervariablen ist für jede Ausprägung der True-Score-Variablen = 0
- Z. B.:

$$E(\varepsilon_{Narzissmus} \mid \tau_{Narzissmus} = 10) = 0$$

4.1 Bedingter Erwartungswert von $\varepsilon_j = 0$

- Gilt für die True-Score-Variable des gleichen Merkmals (τ_j) und für die True-Score-Variable τ_k eines beliebigen anderen Merkmals

$$E(\varepsilon_{Narzissmus} \mid \tau_{Intelligenz} = 115) = 0$$

- Daher wird allgemein formuliert:

$$E(\varepsilon_j \mid \tau_k) = 0$$

4.2 Unbedingter Erwartungswert von $\varepsilon_j = 0$

- Aus Eigenschaft 1 folgt, dass der Erwartungswert der Fehlervariable immer 0 ist, unabhängig vom wahren Wert der getesteten Personen auf Merkmal *j*, *k* oder einem anderen Merkmal:

$$E(\varepsilon_j) = 0$$

4.3 Unabhängigkeit von ε_j und τ_k

- Der wahre Wert und die Fehlervariable bei der Messung eines Merkmals *j* sind unkorreliert

$$Cov(\varepsilon_j, \tau_j) = 0$$

- Diese Unabhängigkeit gilt auch für den Messfehler einer Variablen *j* und den wahren Wert einer anderen Variablen *k*:

$$Cov(\varepsilon_j, \tau_k) = 0$$

4.3 Unabhängigkeit von ε_j und τ_k

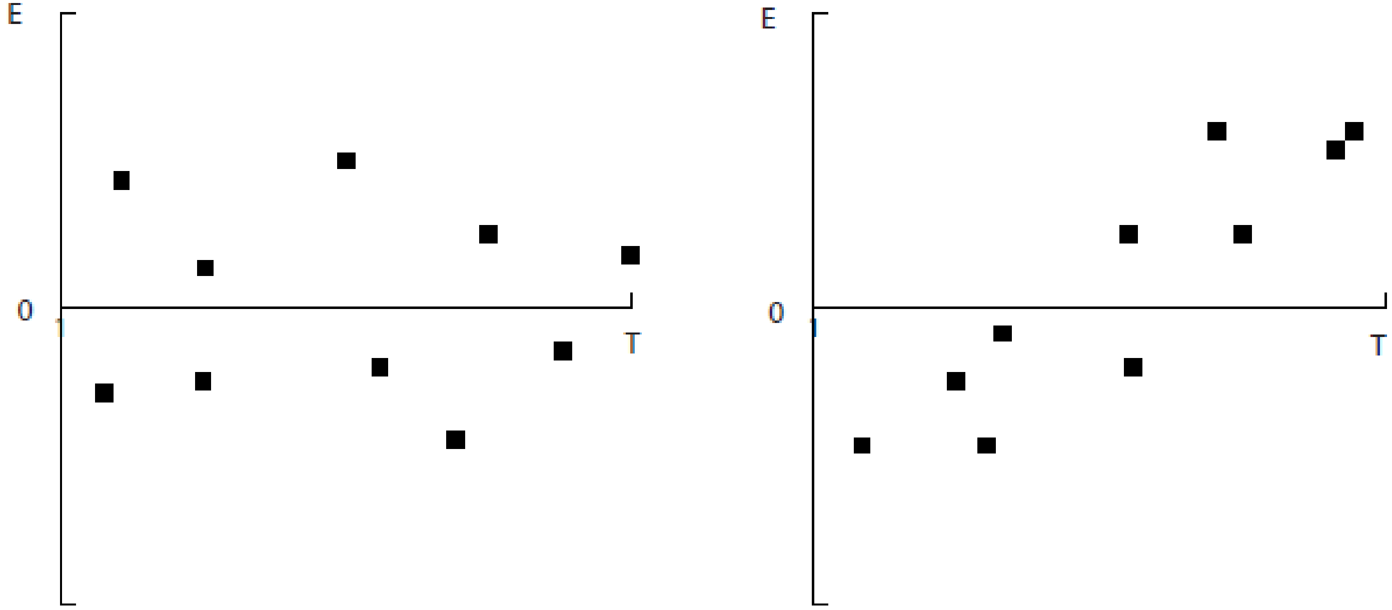


Abb. aus Schuster (2005)

Zusatzannahme zum Messfehler

- Zu 4.3 kommt eine Zusatzannahme hinzu, diese ist keine direkte Folgerung aus der Grundgleichung der KTT
- Die Messfehler bei der Messung zwei verschiedener Merkmale sind unabhängig:

$$Cov(\varepsilon_j, \varepsilon_k) = 0$$

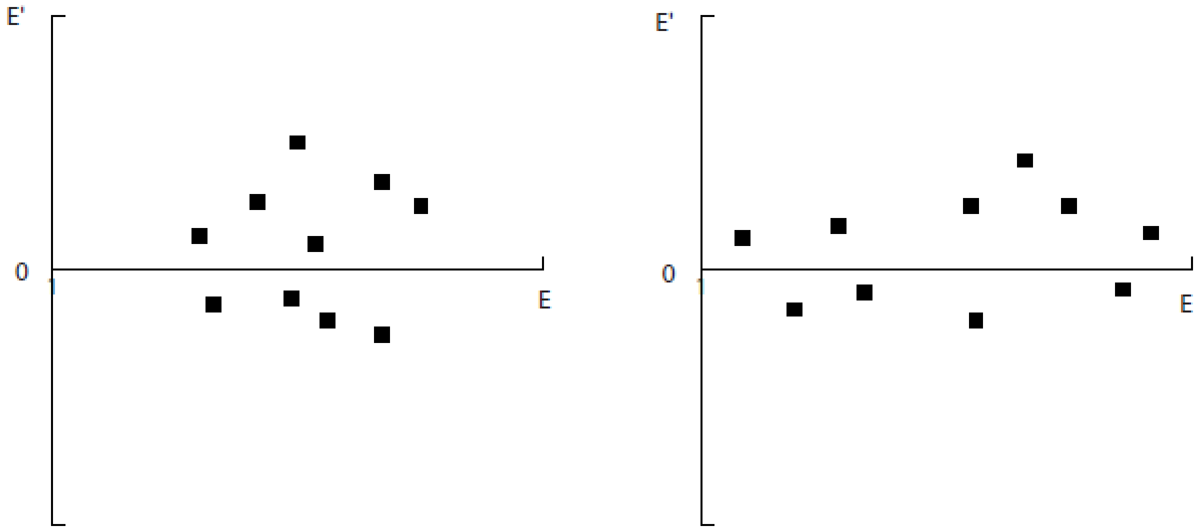


Abb. aus Schuster (2005)

4.4 Additive Varianzzerlegung

- Die Varianz einer beobachteten Messwertvariablen lässt sich additiv zerlegen in die Varianz der True-Score-Variablen und die Varianz der Messfehlervariablen

$$
\begin{aligned}
Var(Y_j) &= Var(\tau_j) + Var(\varepsilon_j) + 2 \cdot Cov(\tau_j, \varepsilon_j) \\
&= Var(\tau_j) + Var(\varepsilon_j)
\end{aligned}
$$

- In den meisten Untersuchungen beeinflussen beide Varianzquellen den beobachteten Wert

Reliabilität

- Die additive Varianzzerlegung ist die Grundlage für die Definition der Reliabilität
- Die Reliabilität wird definiert als Anteil der wahren Varianz an der Gesamtvarianz:

$$Rel(Y_j) = \frac{Var(\tau_j)}{Var(Y_j)} = \frac{Var(\tau_j)}{Var(\tau_j) + Var(\varepsilon_j)}$$

- Die Reliabilität ist also ein Maß für die Messfehlerfreiheit einer Messung

Klassische Testtheorie

1. Theoretische Konzepte der KTT
    1. Messfehler
    2. Wahrer Wert
2. Lokale Unabhängigkeit
3. Grundgleichung der KTT
4. Folgerungen aus der Grundgleichung
5. **Messmodelle**

5. Messmodelle

- Erlauben eine Aussage über das Ausmaß der Vergleichbarkeit (Äquivalenz) von Messungen
- Formulieren Annahmen, die zur Bestimmung der Reliabilität erfüllt sein müssen
- Die Annahmen beziehen sich auf
  - die Unkorreliertheit der Messfehler (siehe Zusatzannahme zu 4.3)
  - den Grad der Übereinstimmung der wahren Werte
  - die Fehlervarianzen

5. Messmodelle

- Die Messmodelle machen unterschiedlich strenge Annahmen bezüglich der wahren Werte und der Fehlervarianzen
- Messmodelle
  1. Parallel
  2. Essenziell parallel
  3. Tau-äquivalent
  4. Essenziell tau-äquivalent
  5. Tau-kongenerisch

5.1 Modell paralleler Messungen

- Annahmen:
  1. Unkorreliertheit der Messfehler

     $$Cov(\varepsilon_j, \varepsilon_k) = 0$$

  2. Identische wahre Werte

     $$\tau_j = \tau_k$$

  3. Identische Fehlervarianzen

     $$\sigma^2_{\varepsilon j} = \sigma^2_{\varepsilon k}$$

- Beispiel: Messung der Körpergröße 2x hintereinander

Legende: τ_j = wahre Werte von Personen *i* (*i* = 1,…,*n*) in Test *j*
$\sigma^2_{\varepsilon j}$ = Messfehlervarianz von Test *j*

5.2 Modell essenziell paralleler Messungen

- Essenziell: im Wesentlichen
- Annahmen:
  1. Unkorreliertheit der Messfehler

     $$Cov(\varepsilon_j, \varepsilon_k) = 0$$

  2. Wahre Werte unterscheiden sich um additive Konstante

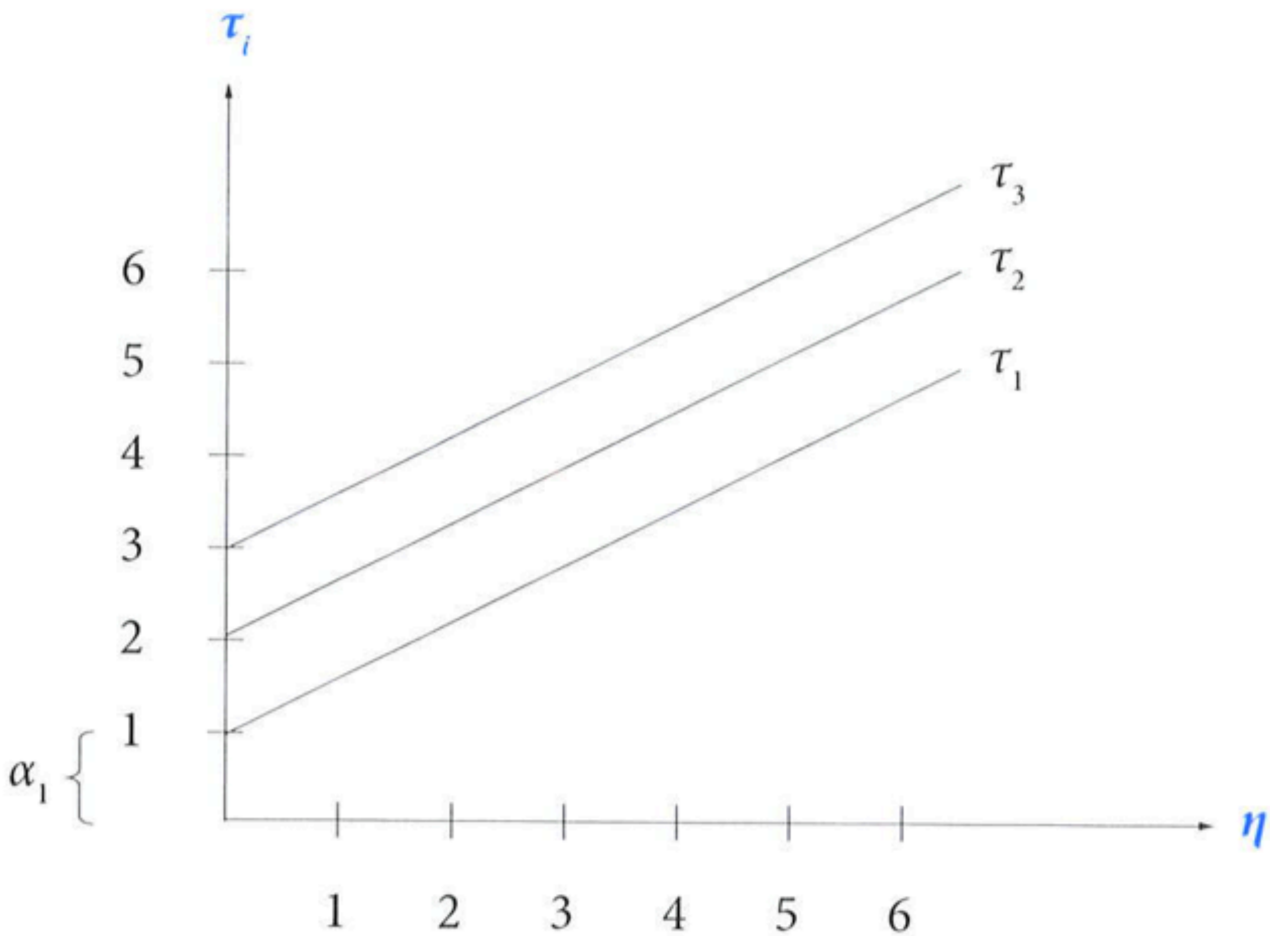
     $$\tau_j = \tau_k + \alpha$$

  3. Identische Fehlervarianzen

     $$\sigma^2_{\varepsilon j} = \sigma^2_{\varepsilon k}$$

- Beispiel: Messung der Körpergröße 2x hintereinander, wobei bei der zweiten Messung ein Maßband verwendet wird, das bei 5 cm anfängt

$$\tau_j = \tau_k + \alpha$$



5.3 Modell tau-äquivalenter Messungen

- Annahmen:
  1. Unkorreliertheit der Messfehler

     $$Cov(\varepsilon_j, \varepsilon_k) = 0$$

  2. Identische wahre Werte

     $$\tau_j = \tau_k$$

  3. Fehlervarianzen unterscheiden sich:

     $$\sigma^2_{\varepsilon j} \neq \sigma^2_{\varepsilon k}$$

- Beispiel: Messung des Gewichts auf einer Waage mit 3 Dezimalstellen und auf einer Waage mit 0 Dezimalstellen

5.4 Modell essenziell tau-äquivalenter Messungen

- Annahmen:
  1. Unkorreliertheit der Messfehler

     $$Cov(\varepsilon_j, \varepsilon_k) = 0$$

  2. Wahre Werte unterscheiden sich um additive Konstante

     $$\tau_j = \tau_k + \alpha$$

  3. Fehlervarianzen unterscheiden sich:

     $$\sigma^2_{\varepsilon j} \neq \sigma^2_{\varepsilon k}$$

- Beispiel: Messung des Gewichts einer Schokoladentafel bei der ersten Messung ohne Verpackung auf einer Waage mit 3 Dezimalstellen und bei der zweiten Messung mit Verpackung auf einer Waage mit 0 Dezimalstellen

5.5 Modell tau-kongenerischer Messungen

- Annahmen:
  1. Unkorreliertheit der Messfehler

     $$Cov(\varepsilon_j, \varepsilon_k) = 0$$

  2. Wahre Werte stehen in einer linearen Beziehung zueinander

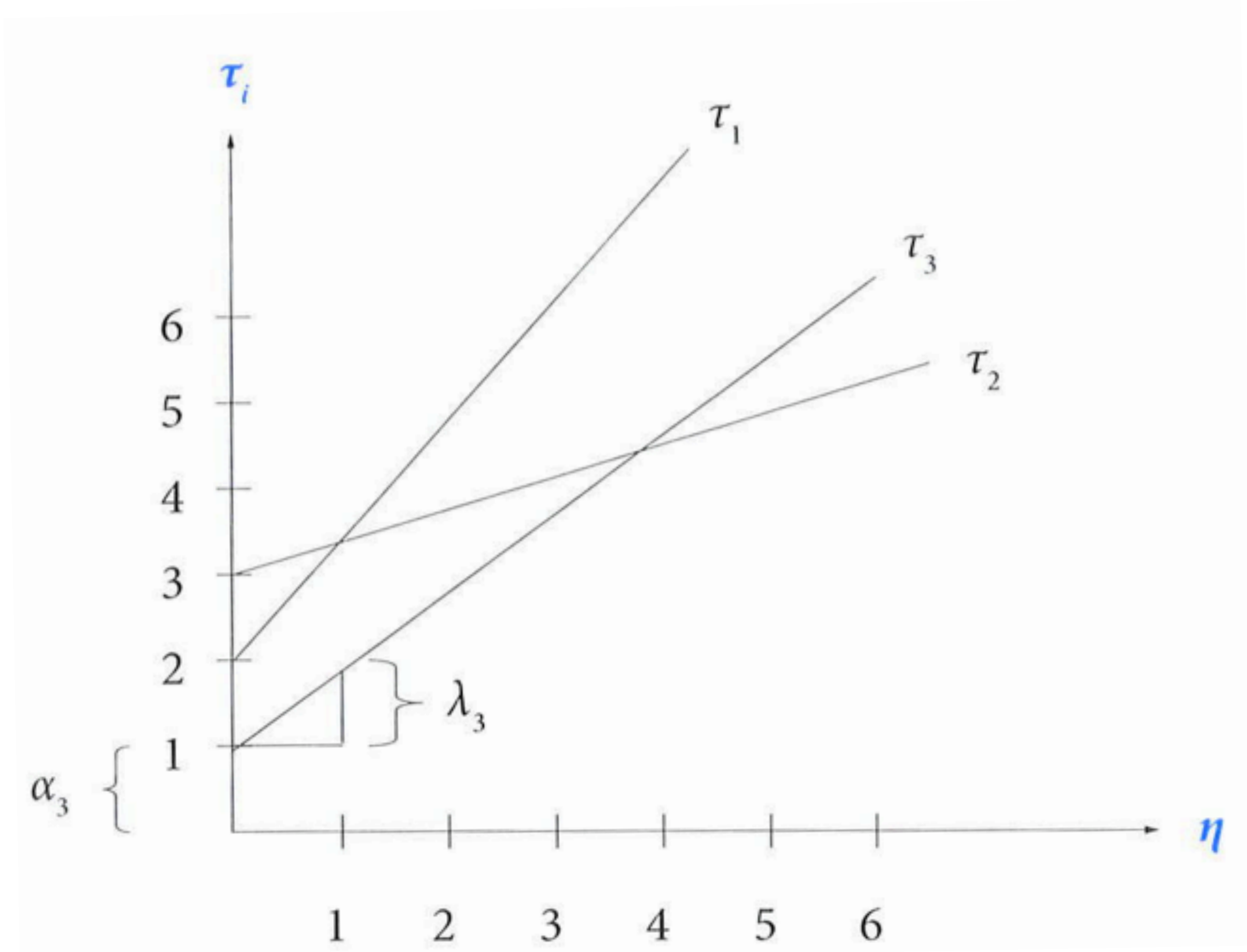
     $$\tau_j = \lambda \cdot \tau_k + \alpha$$

  3. Fehlervarianzen unterscheiden sich:

     $$\sigma^2_{\varepsilon j} \neq \sigma^2_{\varepsilon k}$$

- Tau-kongenerische Messungen bilden dieselbe Eigenschaft ab, allerdings mit verschiedenen Skalen
- Beispiel: Messung der Körpergröße in cm und m

$$\tau_j = \lambda \cdot \tau_k + \alpha$$



5. Messmodelle

- Die Messmodelle sind hierarchisch geordnet und ineinander geschachtelt von dem allgemeinsten Modell (kongenerisches Modell) bis zum restriktivsten Modell (paralleles Modell)
- Die empirische Gültigkeit der Modelle kann getestet werden
- Da die Modelle geschachtelt sind, können sie auch direkt gegeneinander getestet werden

Beispiel NARQ

3 Subskalen des NARQ:

1) Charmingness (Y_1)
   „Ich verhalte mich im Umgang mit anderen meist überaus gewandt."
2) Devaluation (Y_2)
   „Die meisten Menschen sind ziemliche Versager."
3) Supremacy (Y_3)
   „Es freut mich insgeheim, wenn meine Gegner scheitern."

Beispiel Messmodelle

Mittelwerte, Varianzen, Kovarianzen und Korrelationen (kursiv) der beobachteten Variablen

	y_1	y_2	y_3
Mittelwerte	3,18	3,06	3,15
y_1	0,47	*0,71*	*0,71*
y_2	0,37	0,56	*0,71*
y_3	0,34	0,37	0,49

Diese und folgende Tabellen aus Eid et al. (2010)

Beispiel Messmodelle

Wie müssten modellkonforme Mittelwerte und Kovarianzmatrizen für die verschiedenen Messmodelle aussehen?

1. Parallel

Beispiel Messmodelle

Wie müssten modellkonforme Mittelwerte und Kovarianzmatrizen für die verschiedenen Messmodelle aussehen?

1. Parallel

	y_1	y_2	y_3
Mittelwerte	3,13	3,13	3,13
y_1	0,51		
y_2	0,36	0,51	
y_3	0,36	0,36	0,51

Beispiel Messmodelle

Wie müssten modellkonforme Mittelwerte und Kovarianzmatrizen für die verschiedenen Messmodelle aussehen?

2. Essenziell parallel

	y_1	y_2	y_3
Mittelwerte	3,18	3,06	3,15
y_1	0,51		
y_2	0,36	0,51	
y_3	0,36	0,36	0,51

Beispiel Messmodelle

Wie müssten modellkonforme Mittelwerte und Kovarianzmatrizen für die verschiedenen Messmodelle aussehen?

3. Tau-äquivalent

	y_1	y_2	y_3
Mittelwerte	3,14	3,14	3,14
y_1	0,49		
y_2	0,36	0,54	
y_3	0,36	0,36	0,49

Beispiel Messmodelle

Wie müssten modellkonforme Mittelwerte und Kovarianzmatrizen für die verschiedenen Messmodelle aussehen?

4. Essenziell tau-äquivalent

	y_1	y_2	y_3
Mittelwerte	3,18	3,06	3,15
y_1	0,49		
y_2	0,36	0,53	
y_3	0,36	0,36	0,49

Beispiel Messmodelle

Wie müssten modellkonforme Mittelwerte und Kovarianzmatrizen für die verschiedenen Messmodelle aussehen?

5. Tau-kongenerisch

	y_1	y_2	y_3
Mittelwerte	3,18	3,06	3,15
y_1	0,47		
y_2	0,37	0,56	
y_3	0,34	0,37	0,49

Anwendungen von Messmodellen

Beispiele für Fragestellungen, in denen die Äquivalenz der Messungen sichergestellt werden muss:

- Sind Frauen ängstlicher als Männer?
- Unterscheiden sich Deutsche und Italiener in ihrem Narzissmus?
- Gibt es über die Lebensspanne Veränderungen in der Gewissenhaftigkeit?
- Bleibt die Reduktion der Depressionsscores, die direkt nach einer Therapie gefunden wurde, mit zunehmendem zeitlichen Abstand zur Therapie stabil?
- Haben sich deutschen Schüler*innen zwischen PISA 2000 und PISA 2018 in ihrer Lesekompetenz verbessert?

Verletzung der Äquivalenz von Messungen

Kann auftreten, wenn sich die psychometrischen Eigenschaften der Items verändern, z.B. ihre Schwierigkeit:

Instrument 2000

Frage 5: SPINNEN UNTER DROGEN

Welche der [redacted] Drogen ([redacted]) hat keine unmittelbare Wirkung auf die Spinne?

Instrument 2009

Frage 11: SPINNEN UNTER DROGEN

Welche der [redacted] Drogen ([redacted]) hat nicht sofort eine Wirkung auf die Spinne?

Grenzen und Schwächen der KTT

- Die Grundgleichung der KTT ist nicht empirisch überprüfbar
- Die KTT berücksichtigt nur unsystematische Einflüsse, allerdings gibt es eine Erweiterung der KTT, die *Generalisierbarkeitstheorie*, mit der auch systematische Einflüsse wie Situationseffekte oder Beurteilereffekte berücksichtigt werden können
- Es besteht keine Möglichkeit, die Homogenität der Items bezüglich des Merkmals (Eindimensionalität) zu testen
- Die Kennwerte der KTT (z. B. Reliabilitäten) sind stichprobenabhängig

Literatur zu dieser Sitzung

Eid, Gollwitzer & Schmitt (2010). *Statistik und Forschungsmethoden.* Weinheim: Beltz.
Kapitel 22.1 bis 22.3 ohne Modellgeltungstests, Translation, pfadanalytische Darstellung.