

SEM 2: Parameter und Metrik



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.

Eigenschaften eines Strukturgleichungsmodells

- Ein Strukturgleichungsmodell ist eine **formale Repräsentation** einer (kausalen) Theorie oder eines wissenschaftlichen Modells
- In Strukturgleichungsmodellen müssen (unter anderem) unterschieden werden:
 - manifeste und latente Variablen
 - gerichtete und ungerichtete Pfade
 - endogene und exogene Variablen
 - Messmodelle und Strukturmodelle

Ziele der Analyse von Strukturgleichungsmodellen

- Quantifizierung unbekannter Parameter:
u. a. Pfadkoeffizienten, Varianzen, Kovarianzen
- Überprüfung einer Theorie

Parameter eines SEMs

- Die (un)gerichteten Zusammenhänge der latenten und manifesten Variablen im Modell definieren dessen Struktur.
- Nun soll diese mithilfe der **Parameter** quantifiziert werden.
- Welche Parameter gibt es allgemein (u.a.) im Modell?
 1. Varianzen von Variablen
 2. Kovarianzen, Korrelationen, Regressionsgewichte zwischen Variablen.
- Die Ausprägung einzelner Variablen kann nun
 1. frei (unbekannt) sein, sie muss also später aus Daten geschätzt werden
 2. fixiert sein. Es gibt also bspw. Annahmen (Constraints) darüber wie stark ein Zusammenhang genau ist, oder ob mehrere Parameter dieselbe Ausprägung bekommen sollen o.ä.
Hinweis: Dass ein Zusammenhang 0 ist, gilt ebenfalls als Fixierung. Am einfachsten zu erreichen, indem der entsprechende Pfad gar nicht existiert

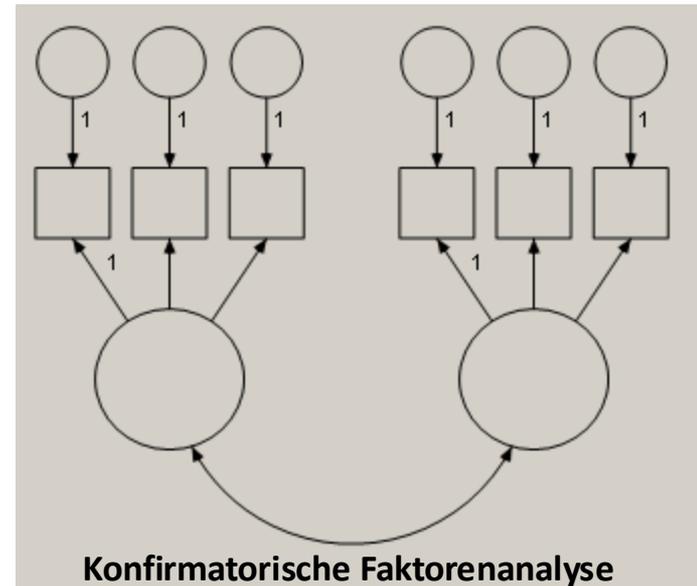
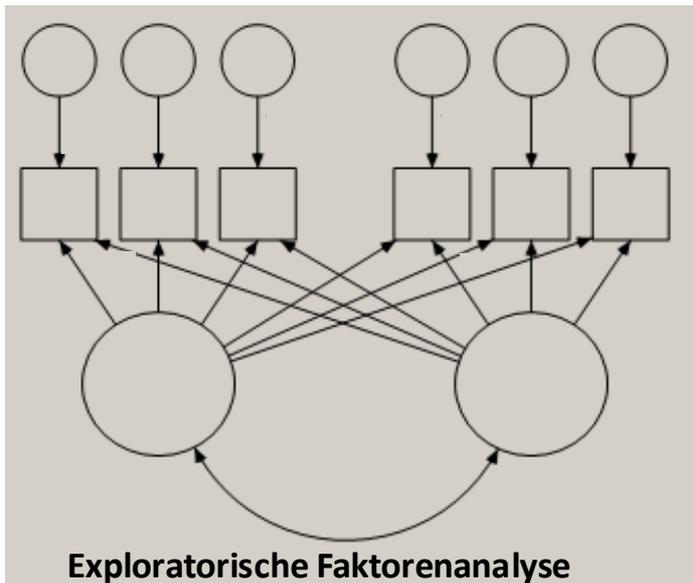
3 Gründe, Parameter zu fixieren

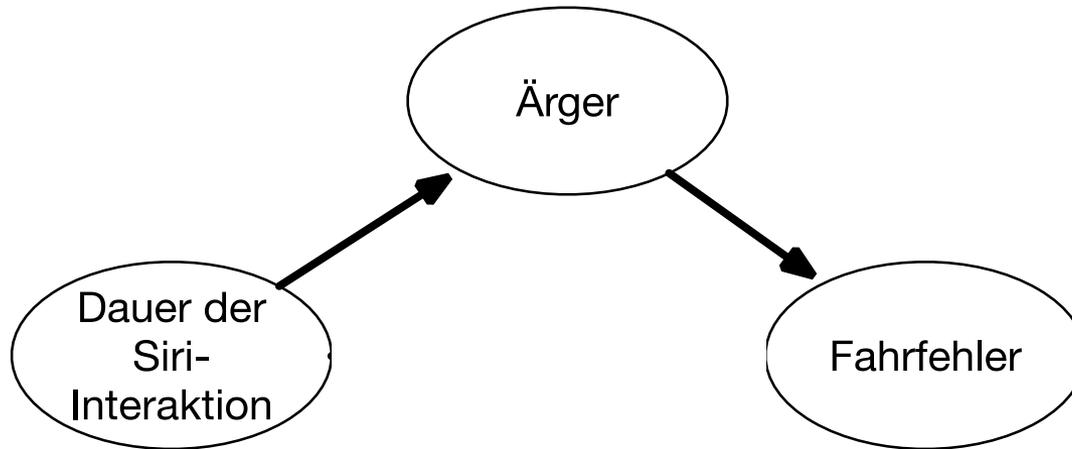
- 1. Theoretische Überlegungen, z.B.**
 - EFA vs. CFA
 - Bekannte Informationen aus Voruntersuchungen / Theorie
- 2. Metrik von latenten Variablen festlegen**
- 3. Identifikation sicherstellen**

Was wird geschätzt?

Bei der **exploratorischen** Faktorenanalyse (**E**xploratory **F**actor **A**nalysis; EFA) wird für jedes Item auf jedem Faktor eine Ladung zugelassen und geschätzt. Es kommt also auch zu „Nebenladungen“.

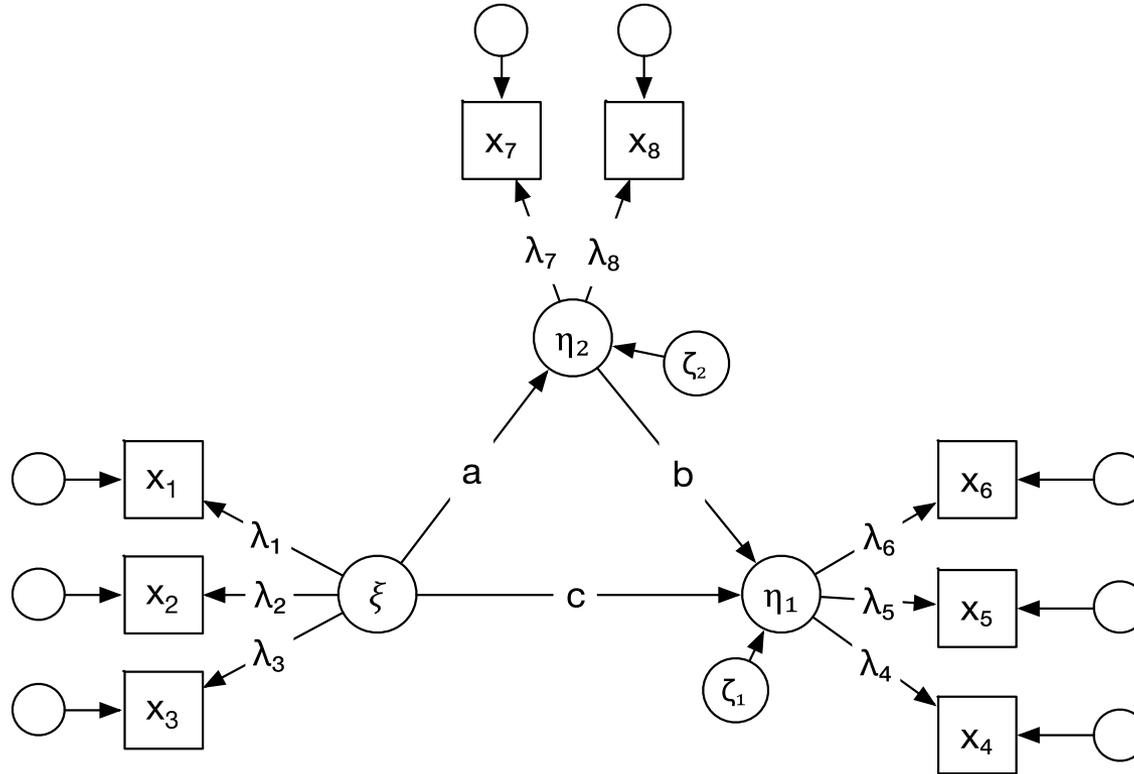
Im Rahmen der **konfirmatorischen** Faktorenanalyse (**C**onfirmatory **F**actor **A**nalysis; CFA) werden nur theoretisch begründete Ladungen geschätzt



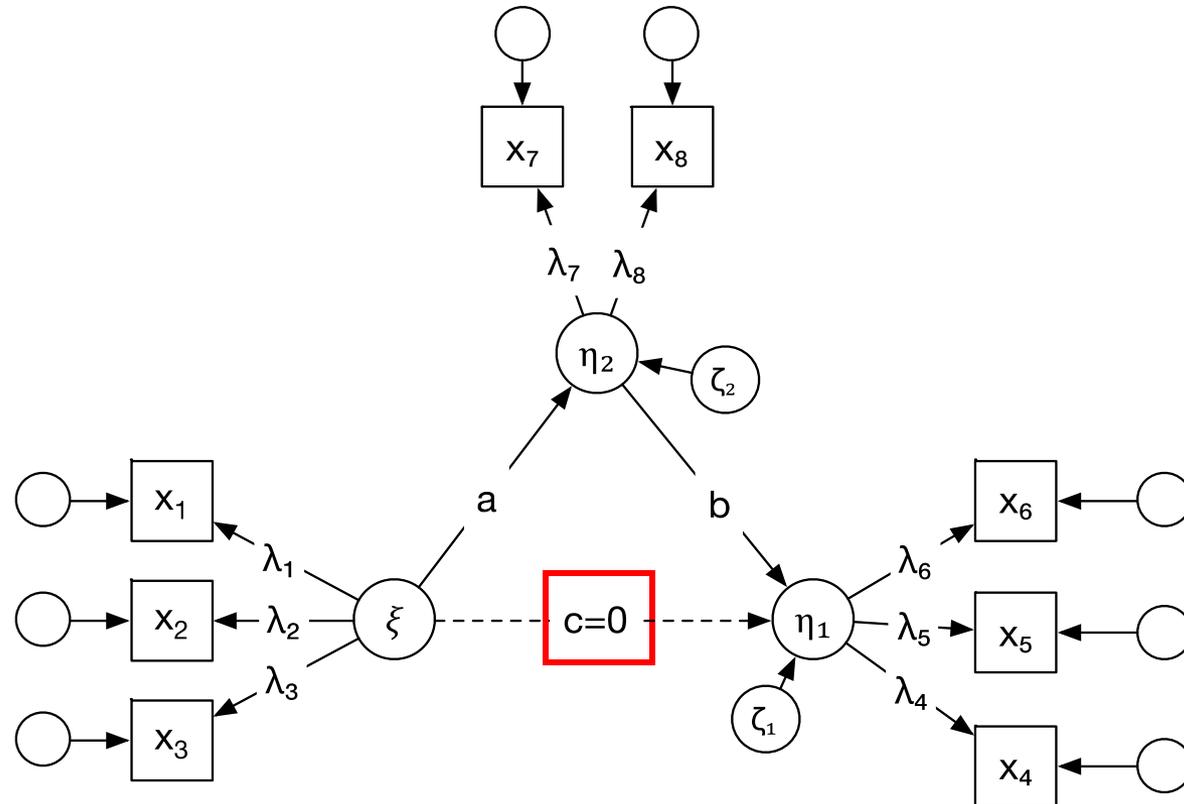


- Hypothese: Die Fahrfehler werden nur über den erlebten Ärger verursacht – es gibt keinen direkten kausalen Effekt von „Siri-Interaktion“ auf „Fahrfehler“, wenn man für Ärger kontrolliert

Bsp 2: Hypothese einer vollständigen latenten Mediation



Bsp 2: Hypothese einer vollständigen latenten Mediation



- Hypothese: Die Fahrfehler werden nur über den erlebten Ärger verursacht – es gibt keinen direkten kausalen Effekt von „Siri-Interaktion“ auf „Fahrfehler“, wenn man für Ärger kontrolliert
- Parameterfixierung: $c = 0$

3 Gründe, Parameter zu fixieren

1. Theoretische Überlegungen, z.B.
 - EFA vs. CFA
 - Bekannte Informationen aus Voruntersuchungen (Hypothesen)
2. **Metrik von latenten Variablen festlegen**
3. Identifikation sicherstellen

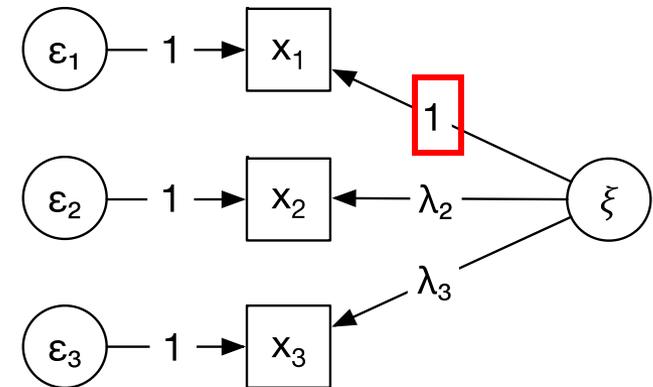
- Manifeste Variablen haben durch das jeweilige Messinstrument bedingt, eine Einheit. Z.B. Punkte auf der 5-stufigen Likert-Skala, oder Meter, oder Jahre, o.ä.
- Nachdem eine latente Variable nicht gemessen wird, hat sie auch keine direkte Einheit (vgl. Testtheorie – Normorientierte Interpretation von latenten Variablenwerten)
 - Wir wissen also bspw. nicht, was ein Unterschied von 1 (anders bei manifesten Variablen in LMMs) zu bedeuten hat.
 - Die Varianz einer Variable hängt von ihrer Einheit ab. Ohne Einheit ist völlig unklar, wie groß die Varianz einer Variable sein soll.

Lösung: Wir könnten also eins von beiden festlegen.

1. Wir definieren, dass wir bei einem Unterschied von 1 in der latenten Variable auch einen Unterschied von 1 in einer bestimmten manifesten Variable erwarten.
 2. Wir definieren, dass die Varianz der latenten Variable einen bestimmten Wert erhält (z.B. 1).
- Dabei sollten wir uns nur sicher sein, dass die wahre Ausprägung des Zusammenhangs bzw. der Varianz nicht gerade 0 ist.
 - **Jede** latente Variable braucht eine Metrik.

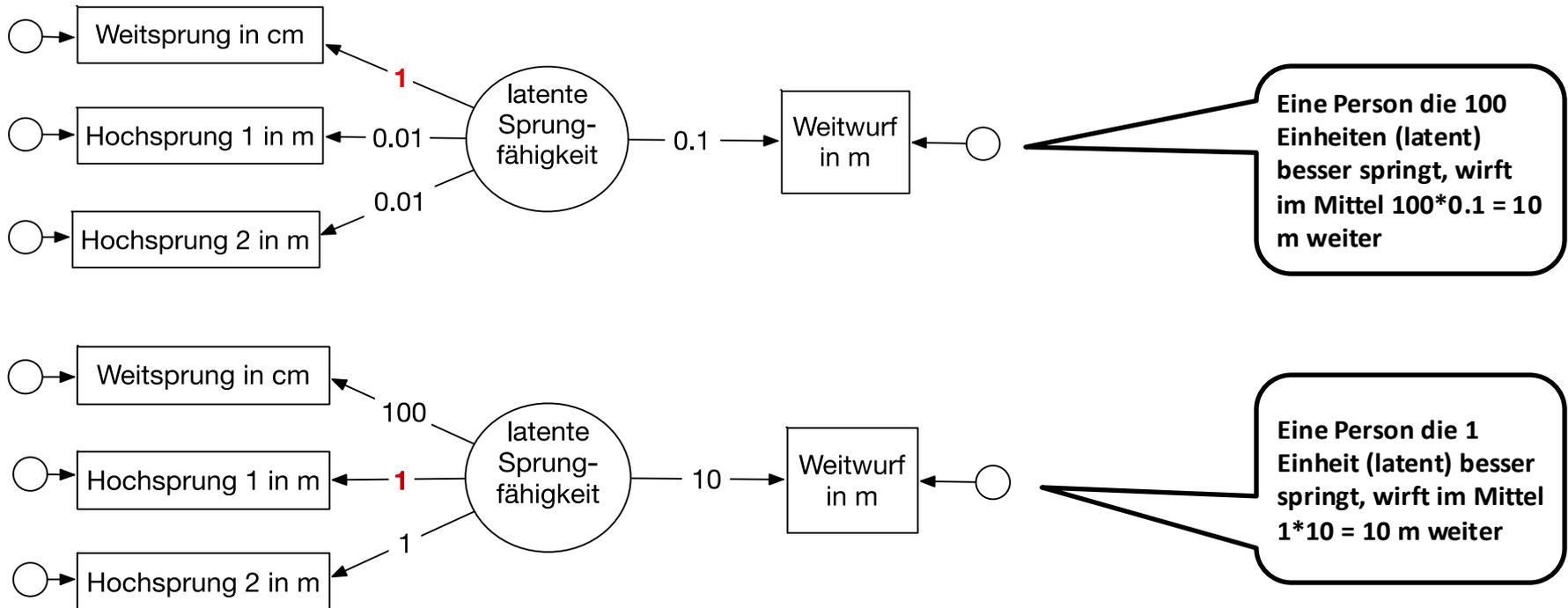
Die Metrik der latenten Variablen wird z. B. dadurch festgelegt, dass für **jede** latente Variable die Ladung *einer* manifesten Variablen auf Eins gesetzt wird

- Ladung von X_1 (oder irgendeines anderen Indikators) auf der latenten Variable ξ wird Eins gesetzt (Unit Loading Identification).

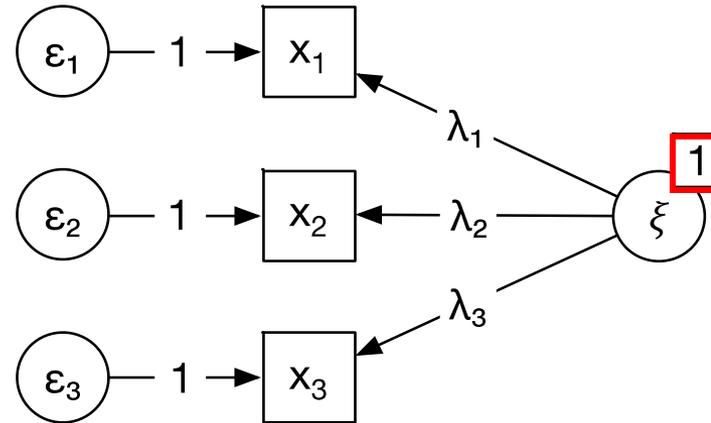


- Variable, deren Ladung auf Eins fixiert wird, heißt Referenzvariable (Normierung)
- Referenzvariable sollte „bester“ Indikator (höchstes Regressionsgewicht) für die latente Variable darstellen, oder sich durch eine hohe Reliabilität auszeichnen, oder inhaltlich die wichtigste Indikatorvariable sein.

- Je nach dem, welcher Indikator als Referenzvariable genommen wird (rote 1 im Bild), ist die latente Variable in dieser Metrik – im Beispiel entweder Meter oder Zentimeter.
- Die Wahl der Metrik beeinflusst die Interpretation der Pfadkoeffizienten
- Beide Modelle sind äquivalent – aber man muss bei der Interpretation aufpassen.



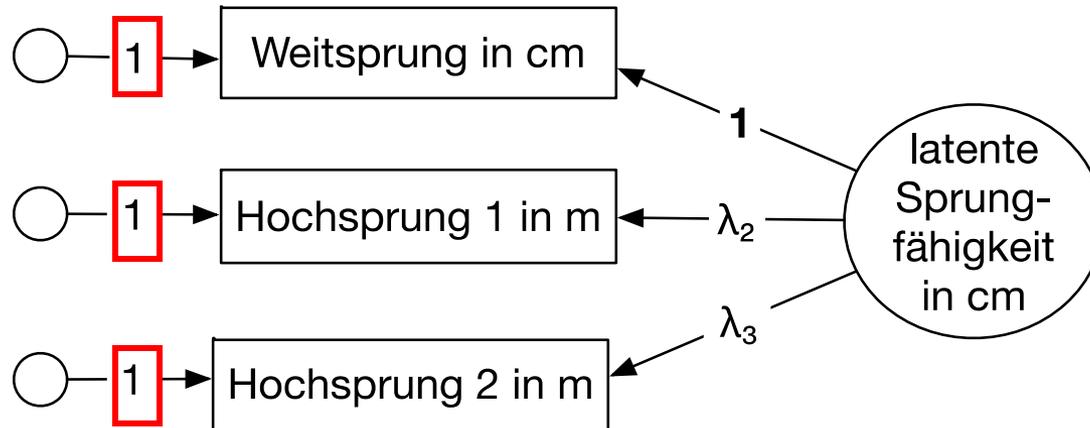
Eine andere Möglichkeit, die Metrik der latenten Variable festzulegen besteht darin, die **Varianz der latenten Variablen auf Eins zu fixieren** (Unit Variance Identification)



- Alle Ladungen auf einer latenten Variablen werden frei geschätzt, dadurch kann für jede Ladung ein Signifikanztest durchgeführt werden
- Die Varianz einer latenten Variablen kann nicht mehr einem Signifikanztest unterzogen werden
- Latente Variable ist dann standardisiert (MW = 0, SD = 1)

Welche Fixierung erfolgt bei Fehlertermen?

- Fehler sind latente Variablen → Schätzung notwendig!
- Fehlern muss ebenfalls eine Metrik zugewiesen werden
- Bei Fehlertermen wird immer ULI genommen → Fehlerterme haben dieselbe Metrik wie die beobachtete Variable.
- D.h., selbst wenn man in einem Modell bei endogenen latenten Variablen die UVI verwendet, hat der Pfad der zugehörigen Fehlervariablen die Fixierung auf 1 (ULI).
- Es wird dann lediglich die Varianz der Fehler geschätzt



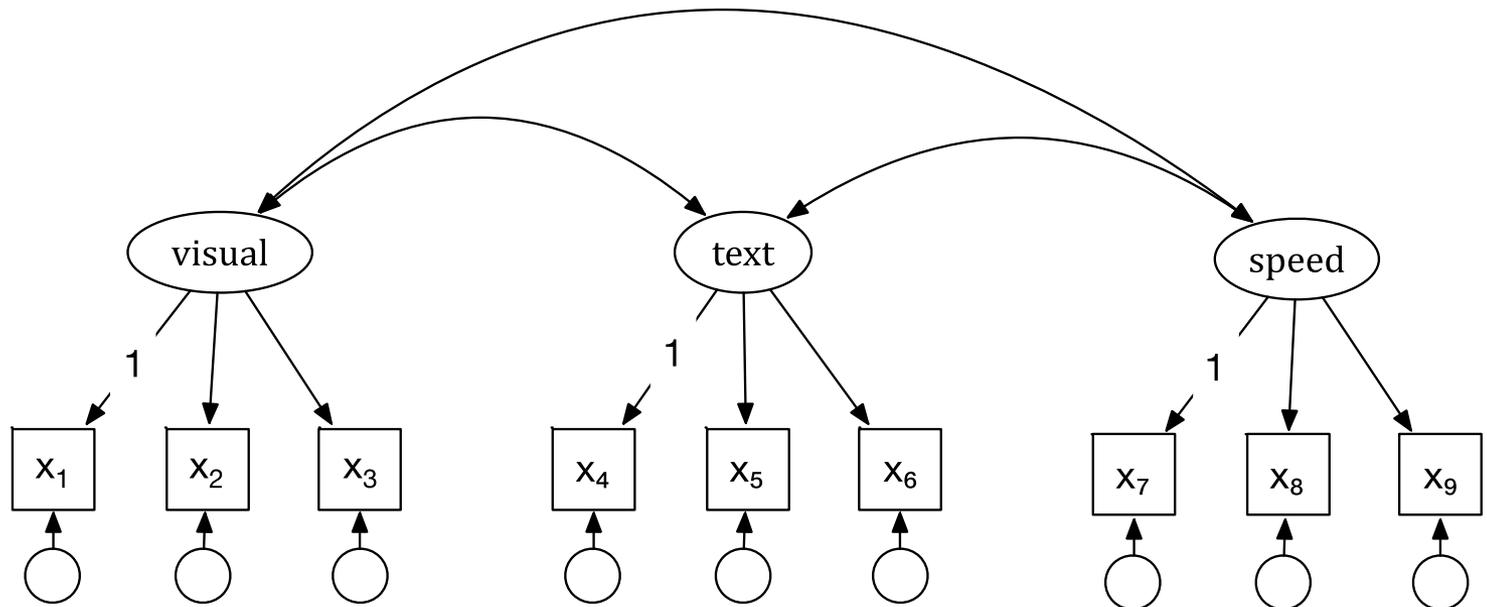
- Bekannter Datensatz, basierend auf Holzinger & Swineford (1939)
- 9 beobachtete Indikatoren messen 3 latente Faktoren:
 - Faktor „visual“ durch x1 (Visual perception), x2 (Cubes) und x3 (Lozenges)
 - Faktor „text“ durch x4 (Paragraph comprehension), x5 (Sentence completion) und x6 (Word meaning)
 - Faktor „speed“ durch x7 (Speeded addition), x8 (Speeded counting of dots) und x9 (Speeded discrimination straight and curved capitals)
- N = 301
- Annahme: 3 korrelierte Faktoren

Beispiel CFA: HolzingerSwineford1939

lavaan Operator =~ : einen latenten Faktor
definieren

```
HS.model <- '  
  visual =~ x1 + x2 + x3  
  textual =~ x4 + x5 + x6  
  speed  =~ x7 + x8 + x9  
,  
fit <- cfa(model = HS.model,  
  data = HolzingerSwineford1939)  
  
summary(fit)
```

Wenn nichts anderes angegeben ist, dann wird
automatisch die Ladung des ersten Indikators
auf 1 fixiert (ULI)



Pfadkoeffizienten zwischen manifesten Indikatoren und latenten Variablen (Ladungen)

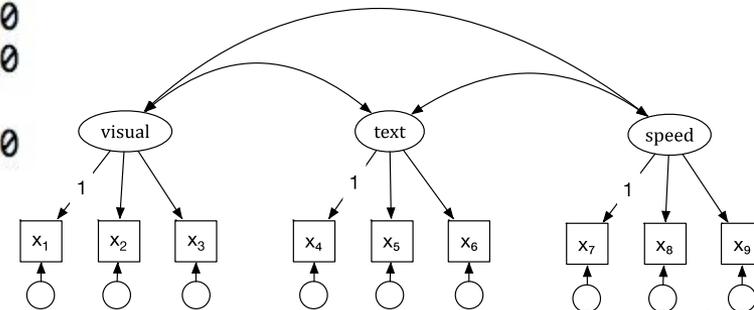
Latent Variables:

	Estimate	Std.Err	Z-value	P(> z)
visual =~				
x1	1.000			
x2	0.554	0.100	5.554	0.000
x3	0.729	0.109	6.685	0.000
textual =~				
x4	1.000			
x5	1.113	0.065	17.014	0.000
x6	0.926	0.055	16.703	0.000
speed =~				
x7	1.000			
x8	1.180	0.165	7.152	0.000
x9	1.082	0.151	7.155	0.000

Kovarianzen zwischen latenten Variablen

Covariances:

	Estimate	Std.Err	Z-value	P(> z)
visual ~~				
textual	0.408	0.074	5.552	0.000
speed	0.262	0.056	4.660	0.000
textual ~~				
speed	0.173	0.049	3.518	0.000



Latent Variables:

visual =~

	Estimate	Std.Err	Z-value	P(> z)
x1	1.000			
x2	0.554	0.100	5.554	0.000
x3	0.729	0.109	6.685	0.000

textual =~

x4	1.000			
x5	1.113	0.065	17.014	0.000
x6	0.926	0.055	16.703	0.000

speed =~

x7	1.000			
x8	1.180	0.165	7.152	0.000
x9	1.082	0.151	7.155	0.000

Standardfehler und p-Wert
der Schätzwerte

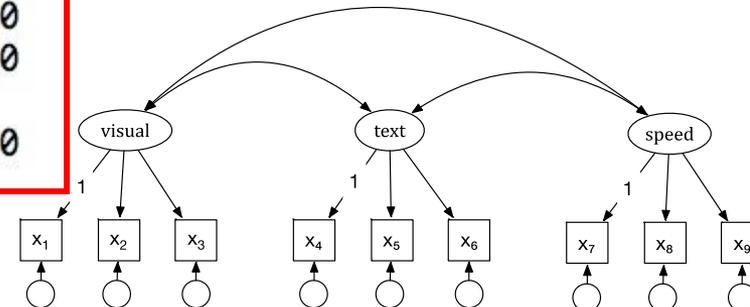
Covariances:

visual ~~

textual	0.408	0.074	5.552	0.000
speed	0.262	0.056	4.660	0.000

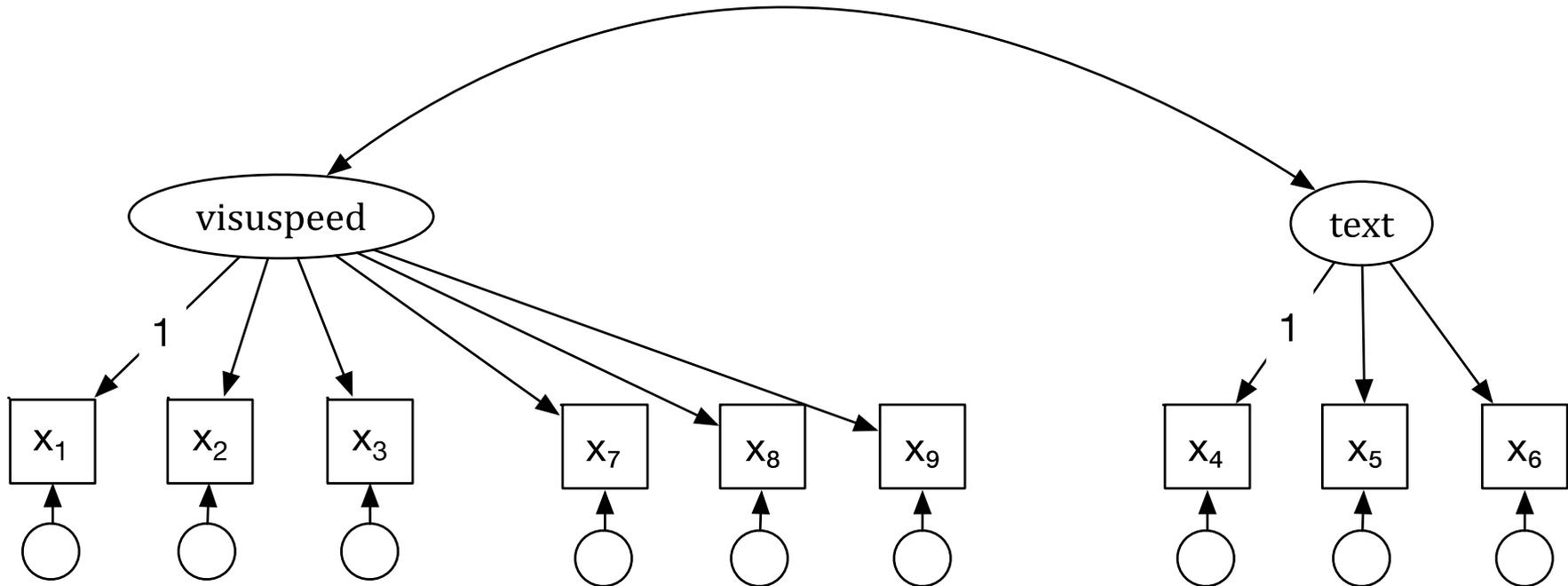
textual ~~

speed	0.173	0.049	3.518	0.000
-------	-------	-------	-------	-------



Beispiel CFA: Alternatives 2-Faktor-Modell

```
HS.model2 <- '  
  visuspeed =~ x1 + x2 + x3 + x7 + x8 + x9  
  textual =~ x4 + x5 + x6  
,
```



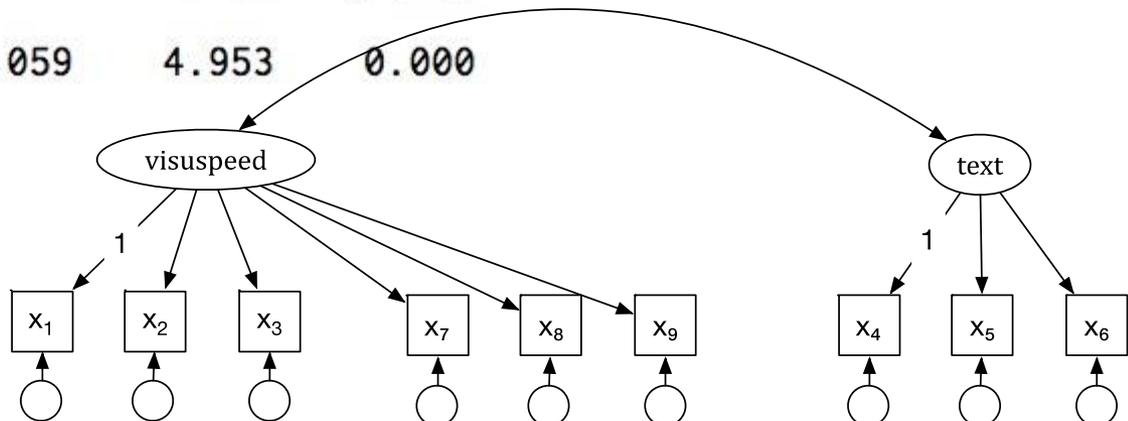
Beispiel CFA: Alternatives 2-Faktor-Modell

Latent Variables:

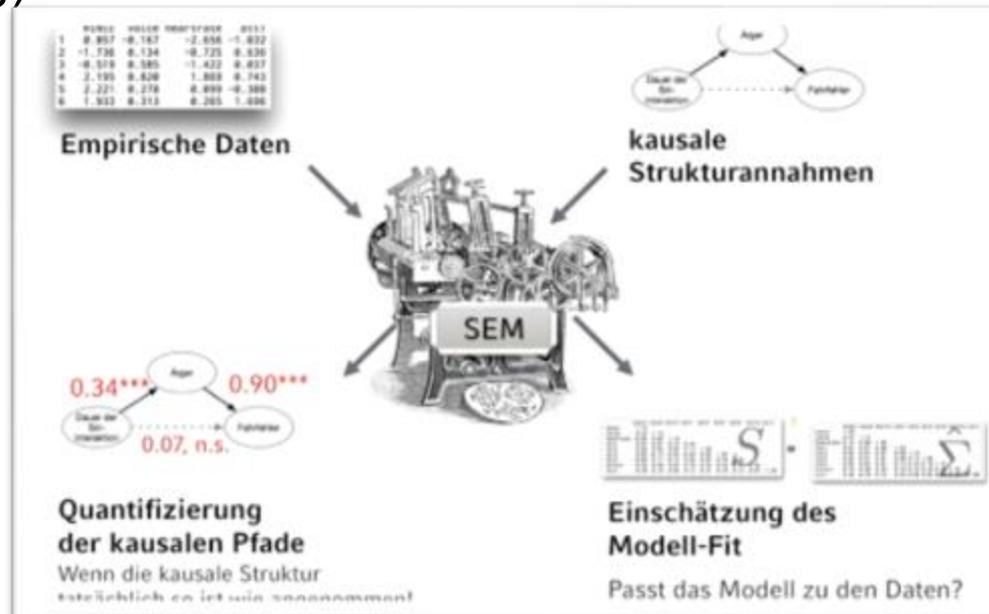
	Estimate	Std.Err	Z-value	P(> z)
visuspeed =~				
x1	1.000			
x2	0.582	0.121	4.819	0.000
x3	0.825	0.125	6.585	0.000
x7	0.583	0.113	5.161	0.000
x8	0.771	0.114	6.787	0.000
x9	1.008	0.129	7.808	0.000
textual =~				
x4	1.000			
x5	1.122	0.066	17.012	0.000
x6	0.927	0.056	16.599	0.000

Covariances:

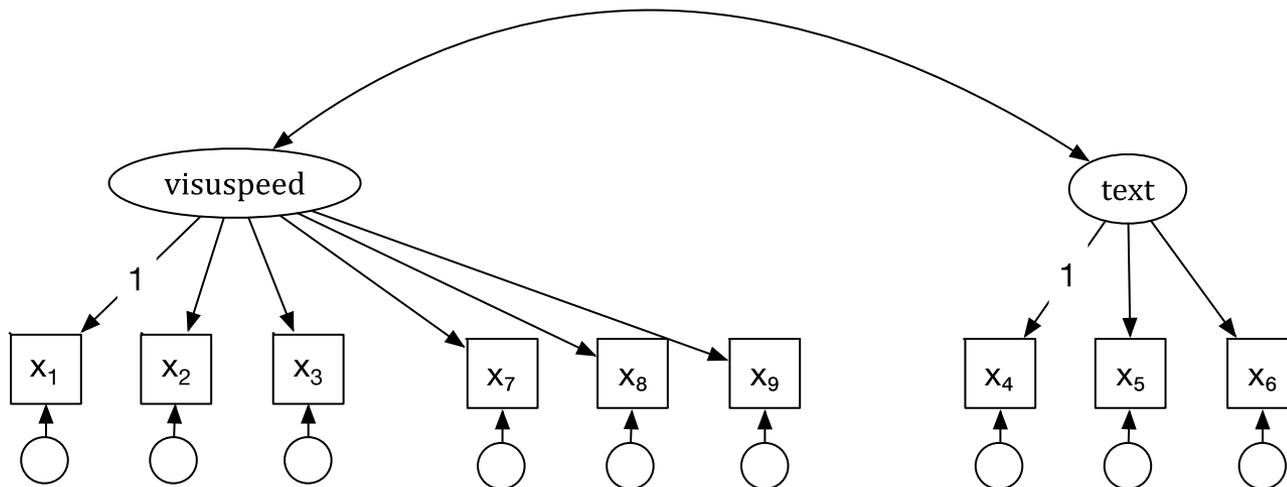
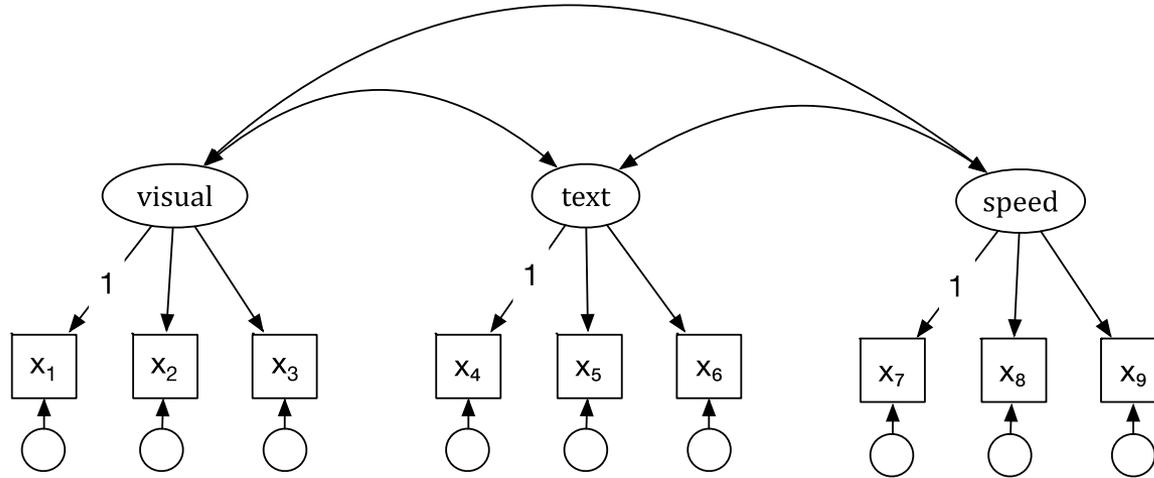
	Estimate	Std.Err	Z-value	P(> z)
visuspeed ~~ textual	0.294	0.059	4.953	0.000



- **Wenn** die Kausalstruktur so ist wie ich annehme, **dann** haben die Pfade diese Gewichte
- Man kann die SEM-Maschine mit beliebigen hypothetischen Kausalstrukturen füttern – es werden immer die Pfade quantifiziert (außer, das Modell ist ungültig)
- Modellvergleich: Welches der Modelle passt besser zu den Daten? (spätere Sitzung)



Vorschau: Was ist das sparsamere Modell?



3 Gründe, Parameter zu fixieren

- 1. Theoretische Überlegung
 - EFA vs. CFA; welche Pfade sind Null?
 - Mediation
- 2. Metrik von latenten Variablen festlegen
- **3. Identifikation sicherstellen**



- Grundkonzepte von SEM:
 - manifeste vs. latente Variable
 - endogene vs. exogene Variablen
 - gerichtete vs. ungerichtete Zusammenhänge
 - Messmodell vs. Strukturmodell
- Parameter: Frei geschätzt oder fixiert?
- Gründe für Parameterfixierung
 - Theorie (z.B. EFA vs. CFA, c-Pfad im Mediationsmodell)
 - Metrik von latenten Variablen festlegen
 - Unit Loading Identification (ein Pfad wird auf 1 fixiert)
 - Unit Variance Identification (eine Varianz wird auf 1 fixiert)
- Verstehen: Welche Metrik hat eine latente Variable?