Psychometrics & Statistics II

Factor Analysis as a part of Structural Equation Modeling

Daria Gracheva, Sergei Tarasov

Plan of the course in brief

Module 2: Confirmatory Factor Analysis. Measurement Invariance. Structural Equation Modeling: moderation & mediation.
Module 3: Multilevel regression analysis: random and fixed effects, interaction term. Explanatory IRT models: Linear Logistic Test Model (LLTM), latent regression model. Discrete latent variables: Latent Class Analysis, Cognitive Diagnosis Modeling.

It is okay if you do not know these words :)

Assessment in the course

R Exercises * 0.12 + 2 Written HW * 0.24 + 2 Final Tests * 0.24 + Exam * 0.4

R Exercises during or after each seminar (deadline - next class), submission in SmartLMS - we will discuss it at the beginning of each class. Assessment criteria: 1 (more than 50% correct), 0 otherwise.

Written HW (two person in groups) - Final work of the module, conduct analysis in R and create a report.

Final Test - interpretations of the statistical output (end of Module 2 and 3)

Exam contains both R Exercises and interpretations of the statistical output

Structural Equation Modeling (SEM)

Structural equation modeling (SEM) uses various types of models to depict **relationships among variables**. It is a combination of **factor analysis and multiple regression analysis**.

- **Latent variables** (constructs or factors) are variables that are not directly observable, and hence are inferred from a set of variables that we do measure using tests, surveys, and so on (intelligence, confidence of consumers)
- **Observed (indicators) variables** are a set of variables that we use to define the latent variable or construct (for example, items from Wechsler Intelligence Scale defined a construct of intelligence)

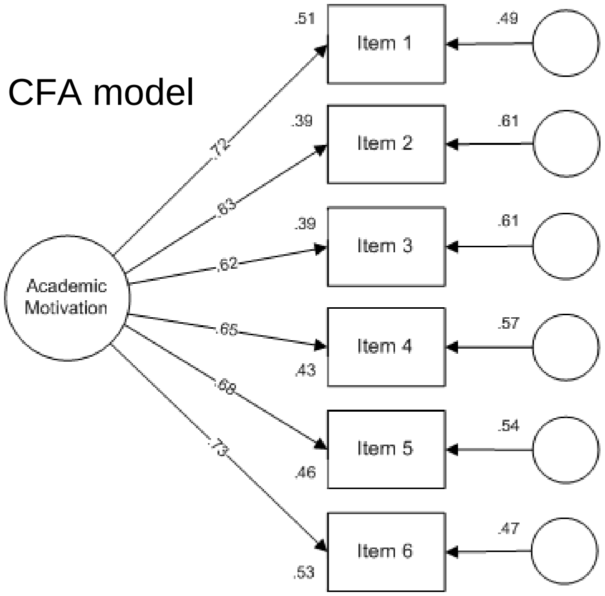
Types of models

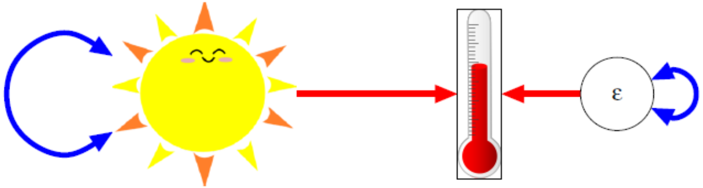
Measurement model (factor model) specifies the number of factors, how the **various indicators are related to the factors**, and the relationships among indicator errors. Measurement models are usually evaluated using **Confirmatory Factor Analysis (CFA).**

Structural model, which specifies how the various **latent factors are related to one another.**

*Path analysis is a special case of SEM, examine the relationships between *observed variables*

CFA model



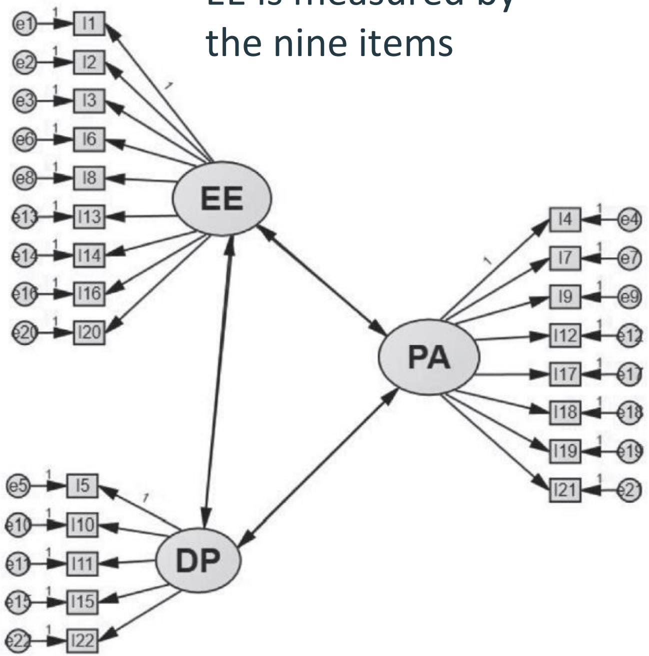


Example of CFA model

MBI measures **three dimensions** of burnout which include the Emotional Exhaustion (EE), the Depersonalization (DP), and the Reduced Personal Accomplishment (PA).

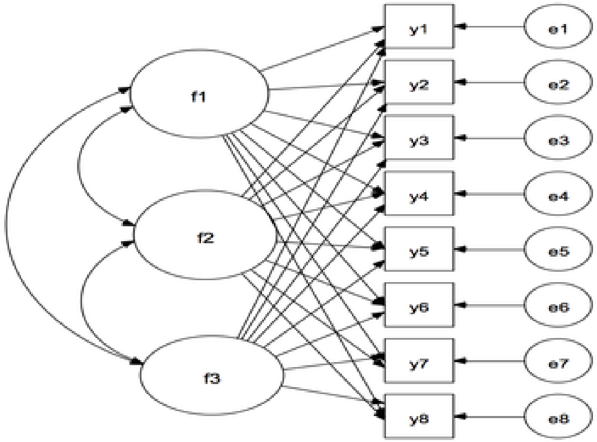
Researcher suggest the structure of the instrument based on theory

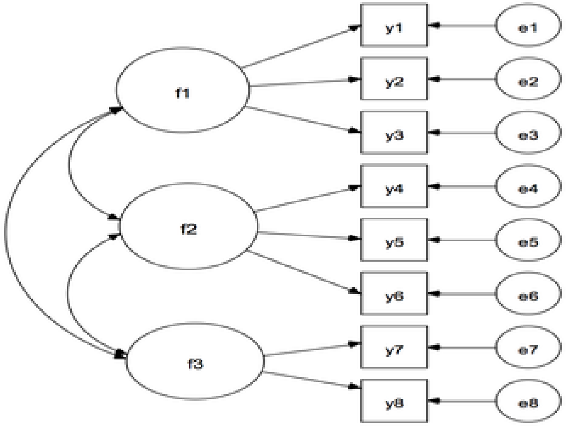
EE is measured by the nine items



CFA is a type of latent modelling

Exploratory FA vs Confirmatory FA





All items load on all factors
Explore in nature - how many factors I have?
How I can explain it?

Researcher creates a structure based on the theory
Structure is a testable hypothesis.
Does my structure fits the data?

Purposes of CFA

- Construct Validation of measurement instrument
- Testing Method Effects
- Testing Measurement Invariance Across Groups or Populations

…

Recommendation to read Introduction chapter

Confirmatory Factor Analysis (Pocket Guide to Social Work Research Methods)

Measurement model as regression

Measurement model is a **linear regression model** where the main predictor (factor) is latent and dependent variables are observed.

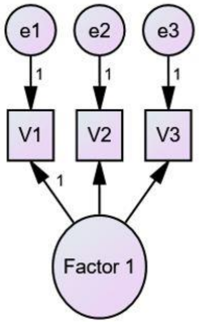
Linear regression equation: $y = a + b * x + e$

Can rewrite as a measurement model for one item: $y_1 = \tau_1 + \lambda_1 * \eta_1 + e_1$

where *τ_1* is the **intercept** of the first item (predicted value of the indicator when the factor is zero), *λ_1* is the **loading** of the first factor on the first item, and *e* is the **residual** for the first item, η - **latent factor**.

For several items and one latent variable:

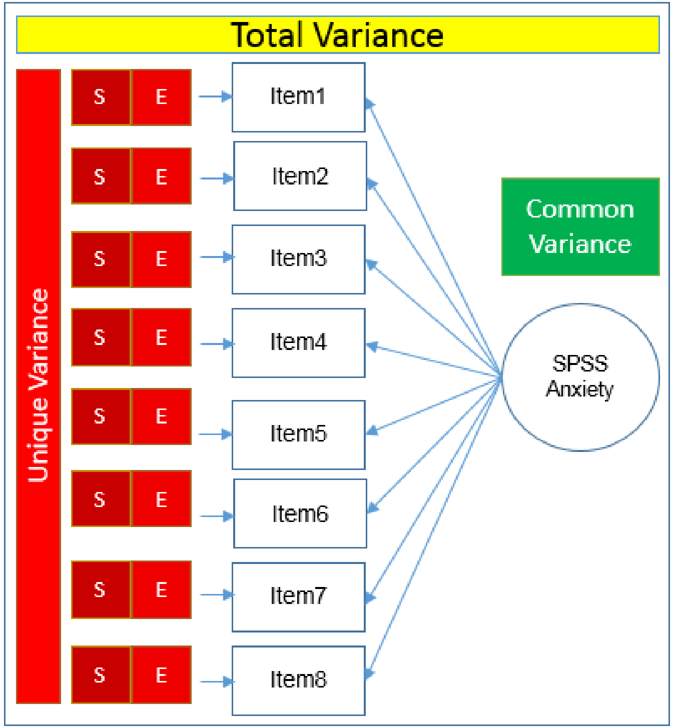
$$
\begin{aligned}
y_1 &= \tau_1 + \lambda_1\eta_1 + \epsilon_1 \\
y_2 &= \tau_2 + \lambda_2\eta_1 + \epsilon_2 \\
y_3 &= \tau_3 + \lambda_3\eta_1 + \epsilon_3
\end{aligned}
\qquad
\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} + \boxed{\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} (\eta_1) + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}}
$$



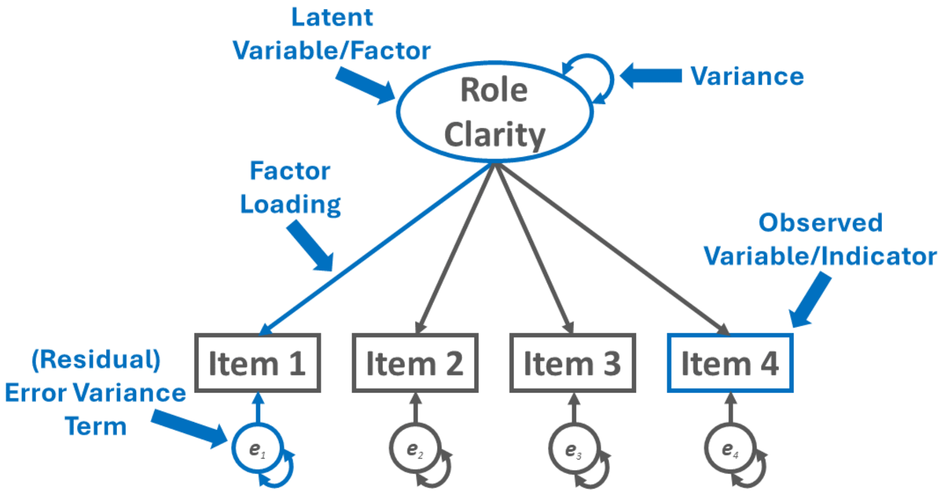
Idea of common and unique variance

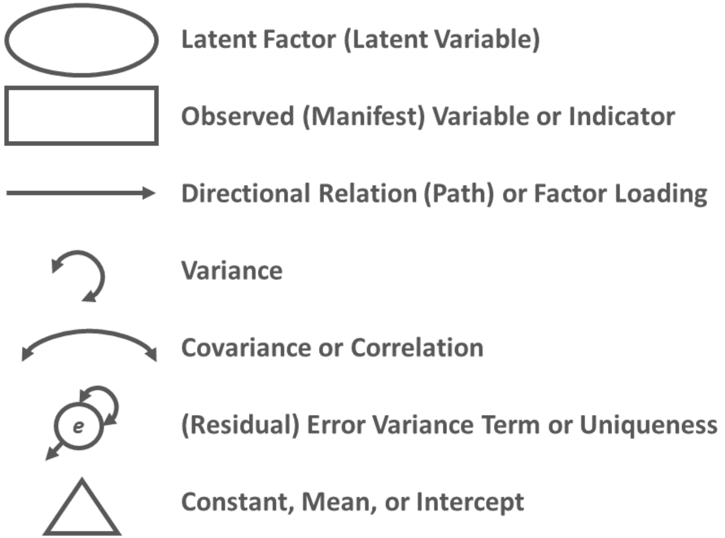
Variance of all items can be decomposed into **common part** (latent factor) and **unique part.**

Toy example:
I feel anxiety at work
I feel anxiety at home
I feel anxiety when I talk with my colleagues



Model Legend





https://rforhr.com/cfa.html

Steps of latent modeling

1. **Model specification** - what is structure of the model? (theoretical model)
2. **Model identification** - check: can we found a unique set of parameter estimates?
3. **Model estimation** - estimate model parameters
4. **Model testing** - test model quality (does the model fit the data or not?)
5. **Model improvement** - improve model if needed

Model specification: Variance-covariance matrix

Researcher specifies a specific model that should be confirmed with **variance-covariance matrix** Illustrate the idea how items are related to each other.

The **sample covariance matrix** implies some *unknown theoretical model* or structure and the researcher's goal is to find the model that most closely fits that variable covariance structure.

	A	B	C
A	8.2	5.8	3.5
B	5.9	4.8	2.1
C	3.5	2.1	1.7

	A	B	C
A	1.00	0.92	0.93
B	0.92	1.00	0.75
C	0.93	0.75	1.00

Diagonal - variance of variable in Variance-Covariance matrix and 1 in Correlation matrix

We have only sample variance–covariance matrix that we used as an estimation of **observed population covariance matrix** - input for CFA model

Variance-covariance matrix

The objective of CFA is to obtain **estimates for each parameter** to produce a predicted (model) variance–covariance matrix that *resembles* the sample variance–covariance matrix as closely as possible. We want to **minimize the difference** between model and sample variance-covariance matrix.

Predicted (model) variance–covariance matrix for three items and one latent factor can be calculated via formula:

$$\Sigma(\theta) = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \overset{\text{variance of factor}}{(\psi_{11})} \overset{\text{factor loadings}}{(\lambda_1 \quad \lambda_2 \quad \lambda_3)} + \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{pmatrix} \quad \text{variance-covariance matrix of the residuals}$$

It is likely that one's theoretical model may not fit the data or have model identifications problems.

Model identification

One aspect of identification specific to the analysis of latent variables is scaling the latent variable. By nature, latent variables are unobserved and thus have ***no defined metrics (units of measurement).*** Two possibilities (first by default in R):

- **Marker or reference indicator** - fix the metric of the latent variable to be the same as one of its indicators (usually fix the factor loading of first indicator to 1)
- **Fix the variance of the latent variable** to a specific value, usually 1.00 - produce standardized solution

Model estimation

The estimation process involves the use of a particular **fitting function to minimize the difference between model and sample variance-covariance matrix** Several fitting functions or estimation procedures are available.

- maximum likelihood (ML)
- maximum likelihood restricted (MLR)
- ordinary least squares (ULS or OLS)
- weighted least square mean and variance adjusted (WLSMV)

…

Model testing

Model fit with **Chi-square test** - how close the model-implied covariance matrix matches the observed covariance matrix?

Accept H0 ($p > 0.05$) - great!

Reject H0 ($p < 0.05$) - model estimates do not sufficiently reproduce the sample variances and covariances (i.e., the model does not fit the data well)

$$H_0 : \Sigma(\theta) = \Sigma$$

$$H_1 : \Sigma(\theta) \neq \Sigma$$

BUT it is rarely used in applied research as a sole index of model fit (very strict hypothesis, significant for sample size > 100)

```
Model Test User Model:

  Test statistic                               554.191
  Degrees of freedom                                20
  P-value (Chi-square)                           0.000
```

Model fit indexes

Index	Description	Recommended values
CFI	*Comparative Fit Index*	> 0.9 / 0.95
TLI	*Tucker Lewis Index*	> 0.9 / 0.95
RMSEA	*Root mean square error of approximation*	< 0.08 / 0.06 / 0.05
SRMR	*Standardized root mean square residual*	< 0.06 / 0.05

```
User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.871
  Tucker-Lewis Index (TLI)                       0.819
```

```
Root Mean Square Error of Approximation:

  RMSEA                                          0.102
  90 Percent confidence interval - lower         0.095
  90 Percent confidence interval - upper         0.109
  P-value RMSEA <= 0.05                        0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.055
```

Read it to know more: https://stats.oarc.ucla.edu/r/seminars/rcfa/

Factor loadings

Unstandardized or standardized factor loadings. Standardized factor loadings can be from -1 to 1 (as correlations) and shows how strong indicators are related to latent construct?

From psychometrics point of view, we can think about **factor loadings as discrimination index**. Stronger (standardized) factor loadings indicate better, more discriminating items.

Recommendations: factor loadings > |0.3 |(strict > |0.5|), significant ($p < 0.05$), in accordance with theoretical expectations

```
Latent Variables:Unstandardized                    Significance        Standardized
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  f =~
    q03               1.000                                0.583    0.543
    q04              -1.139    0.073  -15.652    0.000   -0.665   -0.701
    q05              -0.945    0.056  -16.840    0.000   -0.551   -0.572
```

Model improvement

Modification index (MI) for a particular nonfree parameter indicates that if this parameter were allowed to become free, then the chi-square goodness-of-fit value would be predicted to decrease by at least the value of the index.

On practice we pay attention to large modification indices that differ from others (do not have specific cut-offs)

- MI “give advices” and highlight if something wrong
- Try to minimize the difference between model and sample variance-covariance matrix

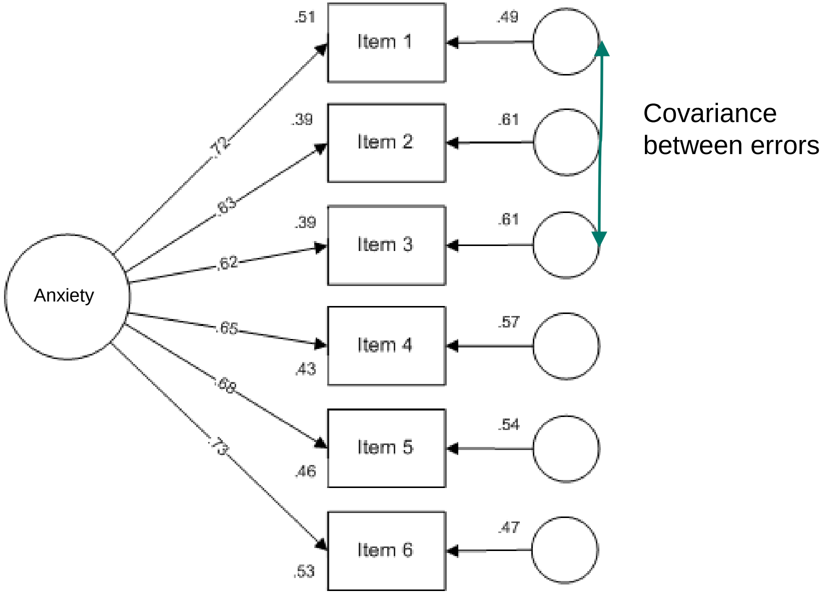
Typical suggestion of MI (1)

Correlation between error terms of items (~~)

Error term is a unique part of item variance.

Toy example:
I feel anxiety at work
I feel anxiety at home
I feel anxiety when I talk with my colleagues

We want to take into account real relationships in the data (become close to variance-covariance matrix)

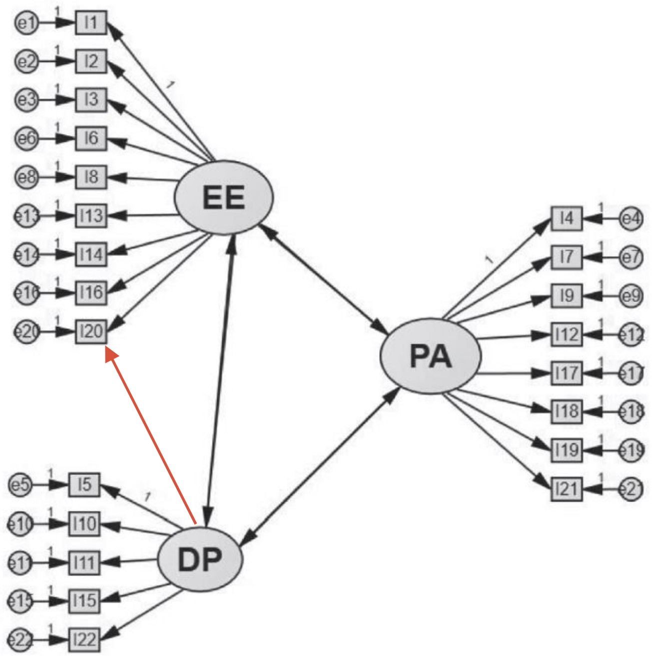


Typical suggestion of MI (2)

Cross-loadings (=~) - items loaded on several latent factors. For example, I20 also loaded on DP factor.

Be caution with MI. On practice, if you make all improvements, you definitely improve the model☺

But can you explain all improvements, how they are related to theoretical structure?



Home reading (read & try) - optional

Tindle, R., Castillo, P., Doring, N., Grant, L., & Willis, R. (2022). Developing and validating a university needs instrument to measure the psychosocial needs of university students. *British Journal of Educational Psychology*, *92*(4), 1550-1570. (in SmartLMS)

Data and R code for this article (for smaller sample)

https://data.mendeley.com/datasets/4x8f48b6jb/1

+ R Exercises as a part of assessment

Reading / watching (optional)

*In SmartLMS

CONFIRMATORY FACTOR ANALYSIS (CFA) IN R WITH LAVAAN - video / theory / code

Most cited textbook ever about CFA:

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Very good explanation:

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. psychology press.

Brief but useful: Confirmatory Factor Analysis (Pocket Guide to Social Work Research Methods)

Textbook with examples in R: Chen, D. G., & Yung, Y. F. (2023). *Structural Equation Modeling Using R/SAS: A Step-by-Step Approach with Real Data Analysis*. CRC Press.

Evaluating Measurement Models Using Confirmatory Factor Analysis
https://rforhr.com/cfa.html

Thank you for your attention!

Model identification

Model identification pertains in part to the difference between the number of freely estimated model parameters and the number of known values in the input variance-covariance matrix.

number of known values can be found by: (p - number of indicators) $p(p+1)/2$

number of free parameters = number of unique parameters - number of fixed parameters

df = number of known values - number of free parameters

- df negative, known < free (*under-identified*, bad)
- df = 0, known = free (*just identified or saturated*, neither bad nor good)
- df positive, known > free (*over-identified*, good)

Example for model identification

number of known values for 3 items? $p(p+1)/2$

number of free parameters = number of unique parameters - number of fixed parameters

What is fixed parameter? Parameter that is fixed to a specific value. Usually covariances between errors are fixed to 0:

$$\Sigma(\theta) = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} (\psi_{11}) \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \end{pmatrix} + \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{pmatrix} \longrightarrow \begin{pmatrix} \theta_{11} & 0 & 0 \\ 0 & \theta_{22} & 0 \\ 0 & 0 & \theta_{33} \end{pmatrix}$$

Also, for scaling the latent variable we fix first factor loading to 1 or factor variance to 1
How many free parameters do we have?

Example for model identification

number of known values for 3 items? = 6 $p(p+1)/2$

number of free parameters = number of unique parameters - number of fixed parameters = 10 - 4 = 6

df = number of known values - number of free parameters

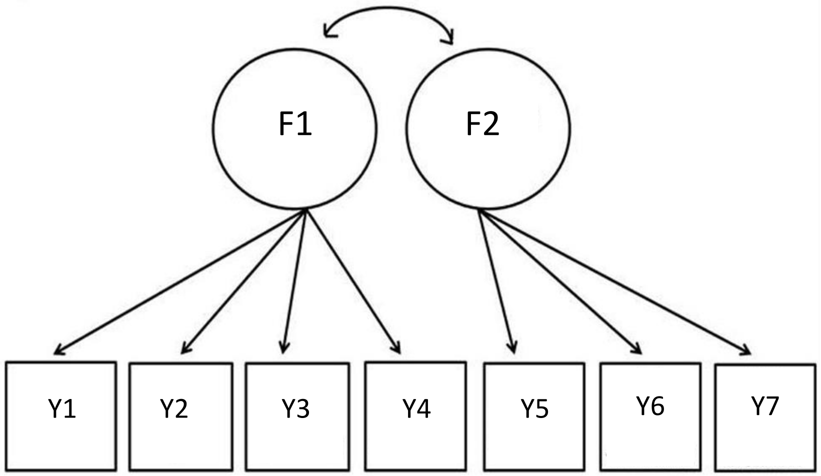
df = 6 - 6 = 0 **-** *just identified or saturated model (that is okay)*

Confirmatory Factor Analysis: Different Models

Daria Gracheva, Sergei Tarasov

Repeat the material

- Where is latent variables?
- Where is observed variables?
- What are arrows (paths)?
- What is dependent and independent variables?
- What is missing from the diagram?



Find model

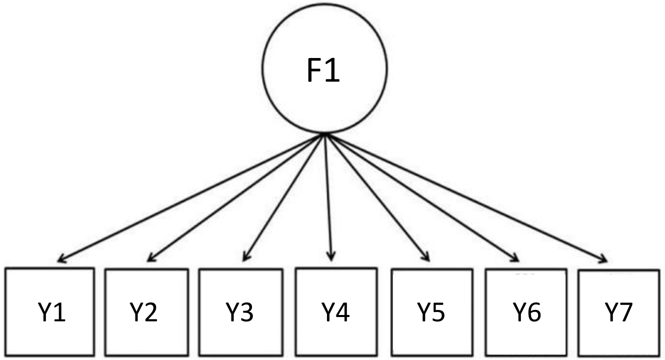
a) F1 =~ Y1 + Y2 +
Y3 + Y4 + Y5 + Y6 + Y7

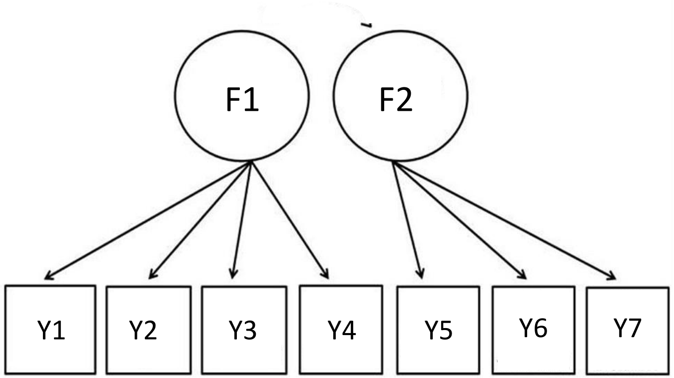
b) F1 =~ Y1 + Y2 + Y3 + Y4
F2 =~ Y5 + Y6 + Y7
Y3 ~~ Y5
Y6 ~~ Y7

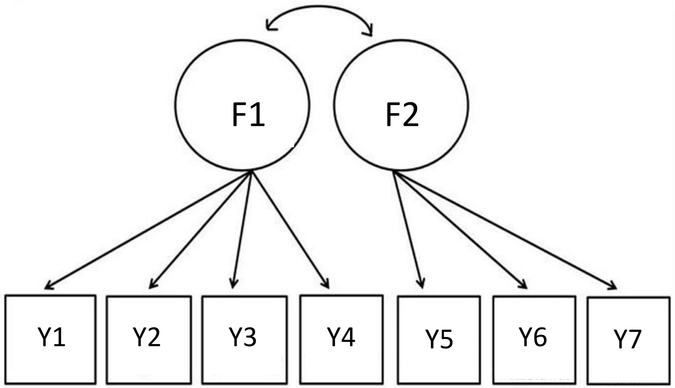
c) F1 =~ Y1 + Y2 + Y3 + Y4
F2 =~ Y5 + Y6 + Y7
F1 ~~ 0*F2

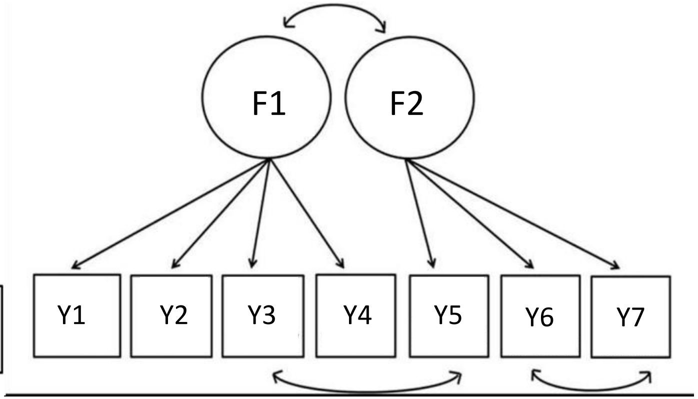
d) F1 =~ Y1 + Y2 + Y3 + Y4
F2 =~ Y5 + Y6 + Y7

e) F1 =~ Y1 + Y2 + Y3 + Y4
F2 =~ Y5 + Y6 + Y7
F1 ~~ 1*F2









Factor loadings

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  self_concept =~
    sest_1            1.000                               0.499    0.732
    sest_2            0.947    0.026   36.070    0.000    0.472    0.662
    sest_3           -1.178    0.035  -33.658    0.000   -0.588   -0.609
    sest_5           -0.777    0.038  -20.625    0.000   -0.388   -0.362
    sest_6            0.941    0.031   30.407    0.000    0.469    0.554
    sest_7           -1.161    0.035  -33.653    0.000   -0.579   -0.609
  les_inv =~
    r_les1            1.000                               0.393    0.582
    r_les2            1.235    0.041   29.855    0.000    0.485    0.629
    r_les3            0.919    0.041   22.606    0.000    0.361    0.432
    r_les4           -0.629    0.049  -12.745    0.000   -0.247   -0.227
    r_les5            0.997    0.036   27.332    0.000    0.391    0.553
    r_les6            0.982    0.032   30.757    0.000    0.385    0.660
    r_les7            1.391    0.043   32.457    0.000    0.546    0.730
    sest_6            0.429    0.035   12.241    0.000    0.168    0.199

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  self_concept ~~
    les_inv           0.061    0.004   14.153    0.000    0.313    0.313
```

What values should we focus on?

Why estimate of first item is equal to 1?

What does negative factor loadings mean?

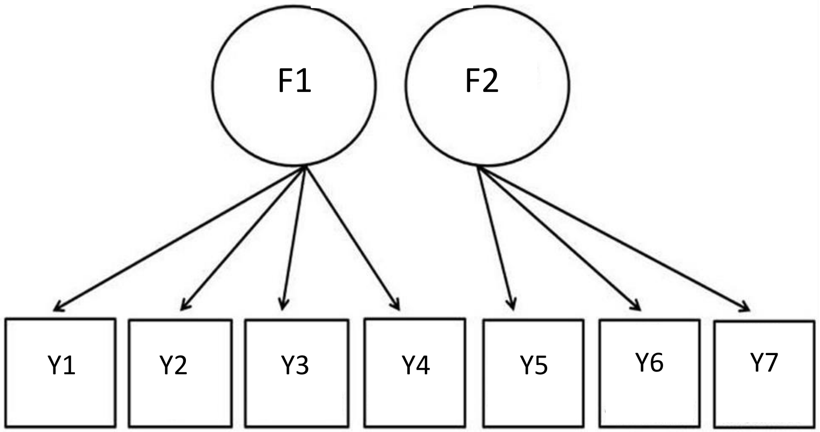
What is estimate = 0.061?

Improve the model

Do we need to improve the model?

F1 =~ Y1 + Y2 + Y3 + Y4
F2 =~ Y5 + Y6 + Y7

What is minimum.value = 20 in the script?




```
> modindices(fit_mod4, standardized = T, sort. = T, minimum.value = 20)
        lhs op  rhs      mi    epc sepc.lv sepc.all sepc.nox
68       Y1 ~~  Y3 522.436  0.256   0.256    0.443    0.443
110      Y5 ~~  Y6 149.380  0.087   0.087    0.283    0.283
46       Y2 ~~  Y3 147.253  0.084   0.084    0.277    0.277
34       F2 =~  Y4 138.708 -0.236  -0.236   -0.217   -0.217
```

Problems with items: delete or not delete?

- **Non-significant or weak (< 0.3) loadings**: *Does item measure the latent trait?*
- **Negative loadings:** Check reverse-coded items. Otherwise, it is a problem. The higher the score for the item, the lower the latent trait.
- **High extra covariances (correlations) between items** - if items are so similar, why you need both?
- If an item appears **frequently in modification indexes**, it is better to remove it.

Not delete: You can keep items with weakly **significant** factor loadings if this is ***consistent with theory*** or not to ***reduce reliability. We do not delete items without good reason!***

Problems with model fit

- **Check Modification Indexes** - MI (make improvement using the model suggestion with the highest MI - make sure this suggestion makes sense! or choose another)

There is no cut-off for MI. Usually we observe the difference between the highest MI and other MI. If model have good fit, there is no need to improve the model.

- **Check factor loadings** (delete non-significant ones one by one)
- **Check data & structure of model & R script** - is everything alright?

Add improvements consistently. Stop when model fit has reached recommended values.

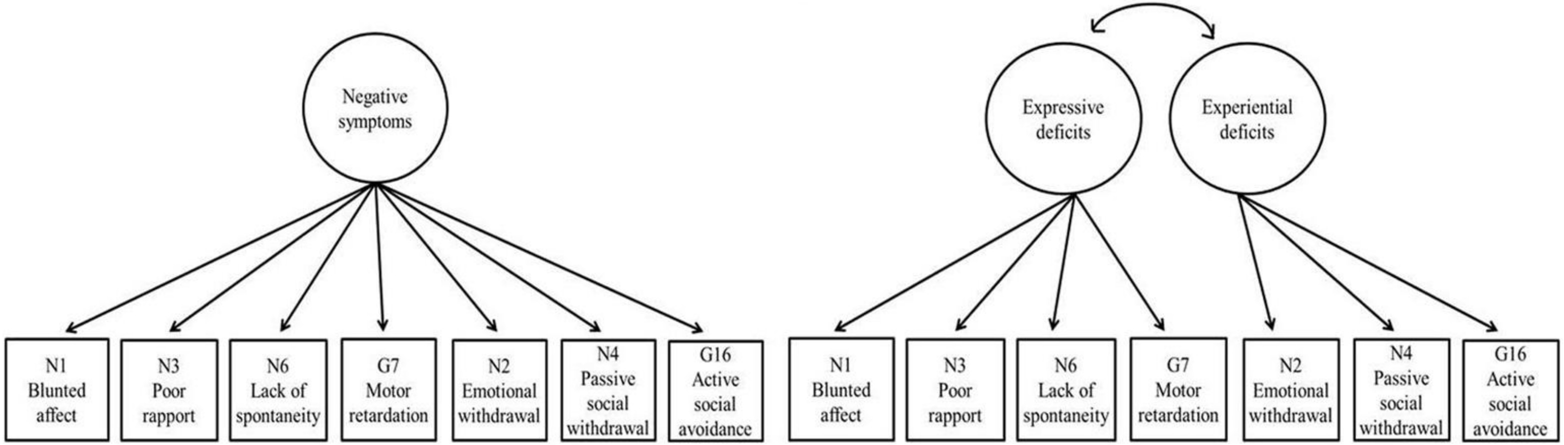
Other problems

- Model identification (too few items per factor OR too complex structure of the model)
- Non-positive definite matrix (often happens with multi-factor models). Can’t use this model!

Lecture

Uni- & Multi-dimensional models

How to figure out which is better?



Steps of latent modeling

1. **Model specification** - what is structure of the model (theoretical model)
2. **Model identification** - check: can we found a unique set of parameter estimates?
3. **Model estimation** - estimate model parameters
4. **Model testing** - test model quality (does the model fit the data or not?)
5. **Model improvement** - improve model if needed

\+ **Model comparison**

Comparison of models

- **Information criterion** (AIC, BIC) - lower is better!
- **Fit indexes** (CFI, TLI etc)
- **Likelihood ratio test (**LR-test) - Chi Square Difference Test - compare **nested models**. Nested models use the same observed variables, but differ in estimated parameters.

H0: there is no difference in the models => simple model is better

Likelihood ratio test

```
model_hw2 <- "
enjoy  =~ enjoy_r1r + enjoy_r2 + enjoy_r3 +
enjoy_r4r + enjoy_r5 + enjoy_r6
reason =~ reason1 + reason2 + reason3 + reason4
+ reason5 + reason6
reason2 ~~  reason5"
```

+1 parameter to the model

```
model_hw <- "
enjoy  =~ enjoy_r1r + enjoy_r2 + enjoy_r3 +
enjoy_r4r + enjoy_r5 + enjoy_r6
reason =~ reason1 + reason2 + reason3 +
reason4 + reason5 + reason6"
```

```
anova(fit_hw2, fit_hw))
Chi-Squared Difference Test

        Df    AIC    BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
fit_hw2 52 103235 103400 587.56
fit_hw  53 107296 107456 724.88     137.32 0.17809       1  < 2.2e-16 ***
---
```

The chi-square fit index assesses the fit between the hypothesized model and data from a set of measurement items (the observed variables)

If models significantly differ, the more complicated model is better.

Steps of latent modeling

1. **Model specification** - what is structure of the model (theoretical model)
2. **Model identification** - check: can we found a unique set of parameter estimates?
3. **Model estimation** - estimate model parameters
4. **Model testing** - test model quality (does the model fit the data or not?)
5. **Model improvement** - improve model if needed
6. **Model comparison** - choose the best model

\+ **Reliability of factor model**

Reliability

When we talk about any measurement, we want it to be **reliable**!

Reliability refers to the precision or consistency of measurement (i.e., the overall proportion of true-score variance to total observed variance of the measure). **Should be 0,7 and higher for each factor.**

semTools::reliability(model_fit): Omega = Raykov's omega, omega2 = bentler's omega, omega3 = McDonald's omega

semTools::compRelSEM(model_fit): McDonald's omega

Construct validity

Internal structure of the instrument as validity evidence. How well theoretical expectations about construct structure fit the data?

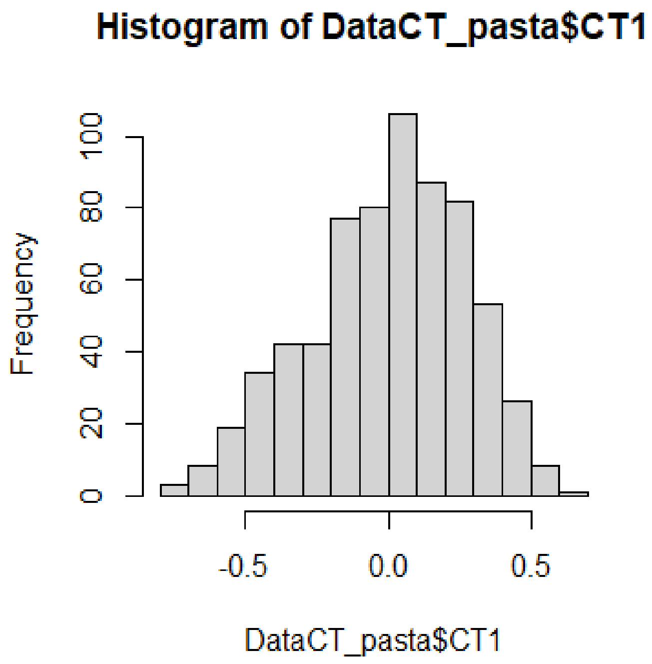
Convergent validity. High values of factor loadings for items related to latent trait (as was expected by theory)

Discriminative validity. Indicators of theoretically distinct constructs are not highly intercorrelated & No high correlation between different latent construct (if corr = 0.9, can we really talk about different latent constructs?)

Factor scores

You can get **factor scores - ** latent variable to determine a participant's **relative standing on the latent dimension.**

You get an **interval scale with mean = 0.** If you fix variance to 1 for identification purposes, standard deviation of factor scores will be close to 1 (z-scores).



Factor scores

Participants with same raw score can have different factor score

Factor scores are more precise estimation of ability (and it is on internal scale!) than sum of raw scores

FS		Raw Enjoy	Raw Problem
enjoy	**reason**	**sum**	**sumpr**
3.181673	3.4255732	24	15
3.193988	4.0885699	21	19
3.201882	5.5761986	23	22
3.226910	3.6612984	24	17
3.241204	3.9353915	23	17
3.336924	4.8334008	21	19
3.418954	5.2608544	21	20
3.469870	5.1269552	23	20
3.609374	5.6542977	24	22

CFA for non-normal data

- If indicators can be considered interval, not **non-normal** (a special case is Likert scale with number of categories => 5), we can use special estimators.
- The most commonly used estimator for non-normal continuous data is **robust ML (MLR)** (Bentler, 1995; Satorra & Bentler, 1994).
- Notes about MLR: when compare models with MLR, testing cannot be conducted by simply calculating the difference in chi-square values. Scaled difference in chi-square should be used (Satorra & Bentler, 2001, 2010).

Consequences of using ML for not normal data

The **consequences of using ML** under conditions of severe non-normality:

- spuriously inflated model chi-square values (i.e., overrejection of solutions);
- difference in fit indices such as the TLI and CFI
- moderate to severe underestimation of the standard errors of the parameter estimates (risk of Type I error - concluding that a parameter is significantly different from zero when that is not the case in the population)

Estimator for categorical data

When at least one factor indicator is categorical (i.e., dichotomous, polytomous, ordinal), ordinary estimators should not be used to estimate CFA model

The most popular estimator is **WLSMV** (weighted least square mean and variance adjusted). It is a robust variant of DWLS (diagonally weighted least squares) that correctly handles non-normal and discrete variables.

The framework and procedures of ordered CFA differ considerably from those of normal-theory CFA.

Think about modeling in practice

A teacher constructs 20 items for a math test that covers algebra and geometry, collect the data. What is next?

- Should there be one score or two scores for math ability? (separate for geometry and algebra?)
- Does the text contains items that require both algebra and geometry?
- How accurate is that single score as a measure of math ability? How accurate would two scores be?

Think about modeling in practice

- Are 20 items sufficient to give a reasonably accurate determination of each student's knowledge? Should more be used? Could fewer have been used?
- Are all items equally good measures of math ability, or are some items better than others?
- Are there other ways of getting the right answer besides ability?
- Could some students have irrelevant problems with certain items because of differences in their background and experience?

Psychometrics & Statistics II

Confirmatory Factor Analysis: Different Models 2

Daria Gracheva, Sergei Tarasov

Think about modeling in practice

A teacher constructs 20 items for a math test that covers algebra and geometry, collect the data. What is next?

- Should there be one score or two scores for math ability? (separate for geometry and algebra?)
- Does the test contains items that require both algebra and geometry?
- How accurate is that single score as a measure of math ability? How accurate would two scores be?

Think about modeling in practice

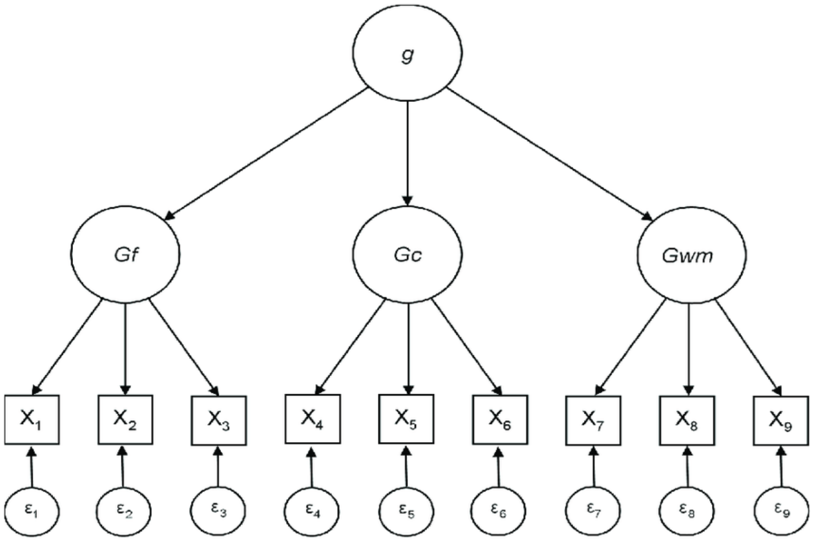
- Are 20 items sufficient to give a reasonably accurate determination of each student’s knowledge? Should more be used? Could fewer have been used?
- Are all items equally good measures of math ability, or are some items better than others?

Second-order factor models

It is possible to specify **higher-order structure** of the measurement model (Hierarchical CFA)

Often used for theory testing

Typical example - intelligence where more specialized factors (e.g., verbal comprehension, memory) are influenced by a broader dimension of general intelligence (g)



Second-order factor models

- Develop a well-behaved first-order CFA solution;
- Examine the magnitude and pattern of correlations among factors in the first-order solution;
- Fit the second-order factor model*, as justified on conceptual and empirical grounds

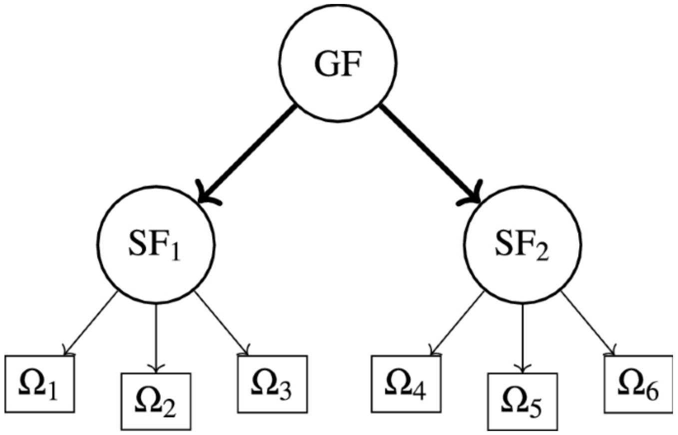
Second-order factor model should be **identified**! (read chapter to know more about it)

Brown textbook. Chapter 8

Note on second-order model identification

Second-order model with 1 second-order factor and two first-order factors is **underidentified.**

Second-order model with 1 second-order factor and three first-order factors is **just-identified** (higher-order solution will produce the same fit as the first-order model, in which the three factors are allowed to covary freely)



Second-order models example (textbook)

	PROBSLV	COGRES	EXPREMOT	SOCSPT
	--------	--------	--------	--------
PROBSLV	1.0000			
COGRES	0.6624	1.0000		
EXPREMOT	0.2041	0.2005	1.0000	
SOCSPT	0.1959	0.2018	0.6699	1.0000

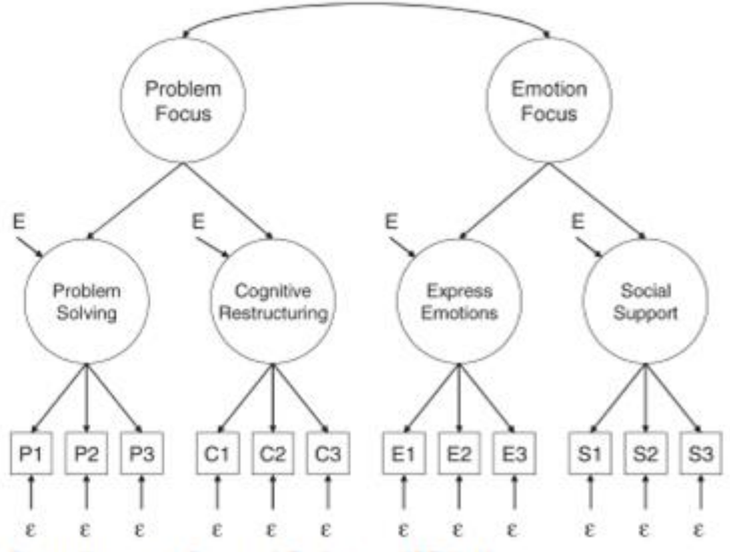
- Examine the pattern of correlations among factors in the first-order solution;

- What second-order factor can we create?

Good youtube video with R-script
https://www.youtube.com/watch?v=spq2l5C2vlk

Second-order models example (textbook)

	PROBSLV	COGRES	EXPREMOT	SOCSPT
	--------	--------	--------	--------
PROBSLV	1.0000			
COGRES	0.6624	1.0000		
EXPREMOT	0.2041	0.2005	1.0000	
SOCSPT	0.1959	0.2018	0.6699	1.0000



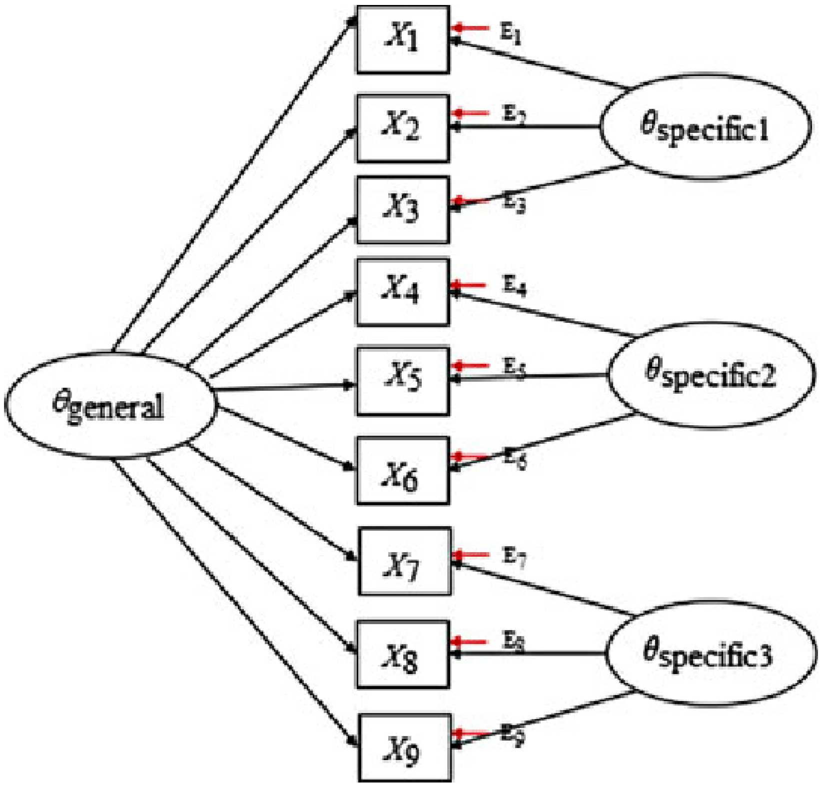
- Examine the pattern of correlations among factors in the first-order solution;
- Probslv and Cogres are more related (0.66), Socspt and Expermot are more related (0.66) -> two second order factors **(in practice - need theory!)**

Good youtube video with R-script
https://www.youtube.com/watch?v=spq2l5C2vlk

Bi-factor models

In a bifactor model, there exists a **general factor** that accounts for significance covariance in all the observed measures, and **domain-specific factors** that accounts for unique variance in the indicators of a specific domain over and beyond the general factor.

General factors are orthogonal to specific & specific are orthogonal to each other (correlation = 0)

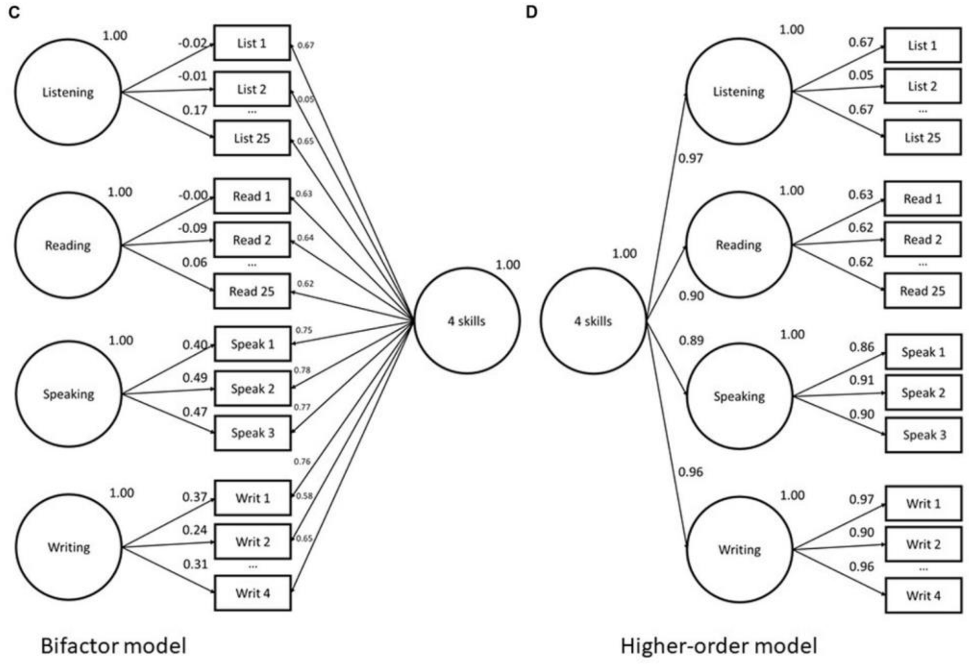


Bi-factor models

Testing method effects (test format)

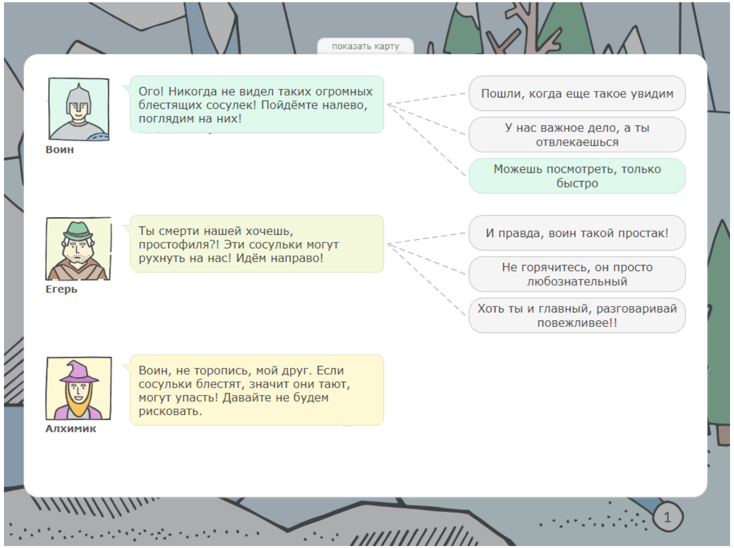
Testing research hypotheses about **General factor and specific factors**

Usually needed to improve model fit.



Method effects

Scenario-based tasks for measuring communication skills. The context of the scenario creates method effect (local dependence between items): items within one scenario are closely related



Different factor models: real research

Confirmatory factor analysis of Project Spectrum activities. A second-order *g* factor or multiple intelligences?

This paper compares different theoretical models of the structure of intelligence. The data were analyzed using confirmatory factor analysis. The models compared were: a) a model with six first-order uncorrelated factors, b) a model with six first-order factors and one second-order general factor, *g*; c) a model with two correlated second-order general factors, in which the cognitive intelligences load on a "cognitive" general factor and the non-cognitive intelligences load on a "non-cognitive" general factor, and d) a model with six first-order correlated factors.

https://www.sciencedirect.com/science/article/pii/S0160289610000899?casa_token=yQ9oWeoJoLYAAAAA:sTmz02rrNuvB1o3ObOGpCflxScqPumOgmr1U6O1cdXy7v8b0J8PGH0EX96mXhONpuyYXRzlK_FMM

*Decomposition of variance - related to G-factor or first-order factors (discuss in later classes)

Different factor models: real research

The Place of the Bifactor Model in Confirmatory Factor Analysis Investigations Into Construct Dimensionality in Language Testing

This paper compares a range of CFA model structures with the bifactor model in terms of theoretical implications and practical considerations, framed for the language testing audience. The models are illustrated using primary data from the British Council's Aptis English test.

https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01357/full

Categorical data in CFA

When at least one factor indicator is categorical (i.e., dichotomous, polytomous, ordinal), ordinary estimators should not be used to estimate CFA model

The most popular estimator is **WLSMV** (weighted least square mean and variance adjusted).

The framework and procedures of ordered CFA differ considerably from those of normal-theory CFA.

In ordered CFA we use correlation matrix instead of covariance matrix.

Brown, page 353 Categorical data

Types of correlations

<table>
<tr><th colspan="2" rowspan="2">Variables X/Y</th><th rowspan="2">Interval X</th><th colspan="2">Ordinal X</th><th rowspan="2">Nominal X</th></tr>
<tr><th>Dichotomous</th><th>Polytomous</th></tr>
<tr><td colspan="2">Interval Y</td><td>Linear (Pearson)</td><td>Biserial</td><td>Polyserial</td><td>Point-biserial</td></tr>
<tr><td rowspan="2">Ordinal Y</td><td>Dichotomous</td><td>Biserial</td><td>Tetrachoric</td><td>Polychoric</td><td rowspan="2">Rank-biserial</td></tr>
<tr><td>Polytomous</td><td>Polyserial</td><td>Polychoric</td><td>Polychoric</td></tr>
<tr><td colspan="2">Nominal Y</td><td>Point-biserial</td><td colspan="2">Rank-biserial</td><td>ϕ</td></tr>
</table>

Tetrachoric correlation

	0	1
0	A	B
1	C	D

Correlation between two dichotomous items: Item 1 and Item 2

What is the correlation in this case?

	0	1
0	30	0
1	0	30

What is the correlation in this case?

	0	1
0	0	30
1	30	0

https://www.youtube.com/watch?v=SPTM5dzG8kw&t=524s

Tetrachoric correlation formulas

- When AD>BC (+)

$$r_{tr} = Cos\left(\frac{180^{0} \times \sqrt{BC}}{\sqrt{AD} + \sqrt{BC}}\right)$$

- When BC>AD (-)

$$r_{tr} = Cos\left(\frac{180^{0} \times \sqrt{AD}}{\sqrt{AD} + \sqrt{BC}}\right)$$

Minus sign will be affixed by the experimenter

- When AD=BC (0)

$$r_{tr} = Cos\left(\frac{180^{0} \times \sqrt{AD}}{2\sqrt{AD}}\right) = Cos90^{0} = 0$$

https://www.youtube.com/watch?v=JNJLj0xkGdY

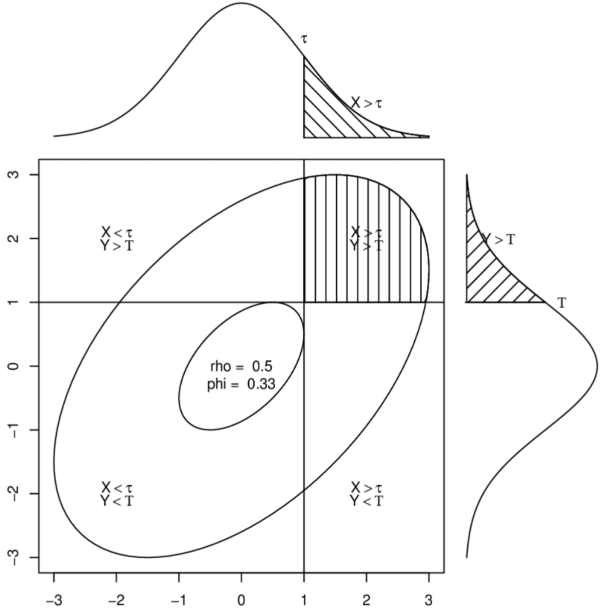
Tetrachoric correlation

Tetrachoric correlation assumes that binary variables are categorical expressions of some **underlying continuous latent processes.**

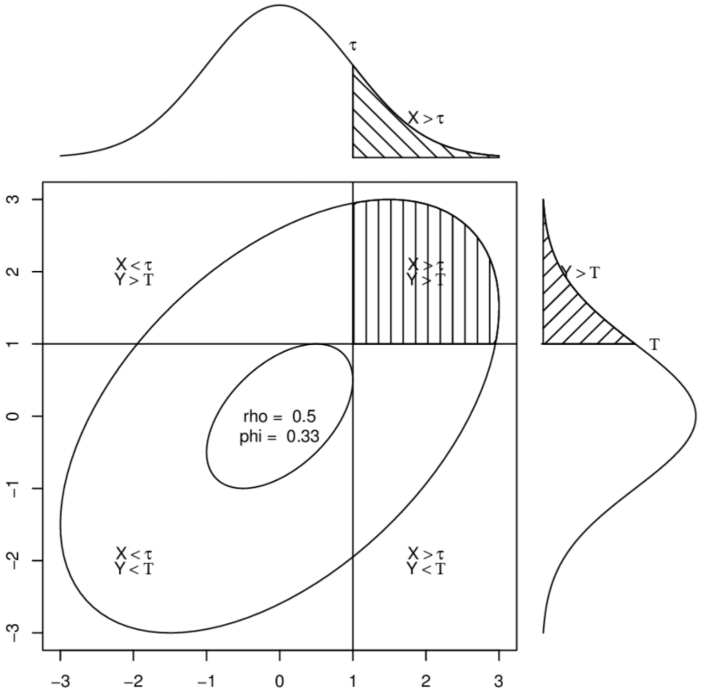
For example, item1 (correct or not correct) might depend on a latent variable x1*
item2 (correct or not correct) might depend on a latent variable x2*

To convert continuous latent variables x1* and x2* into binary ones, thresholds (t\tau) are applied.

A person with a score in the underlying latent variable lower than tau is most likely to answer the first category (0).

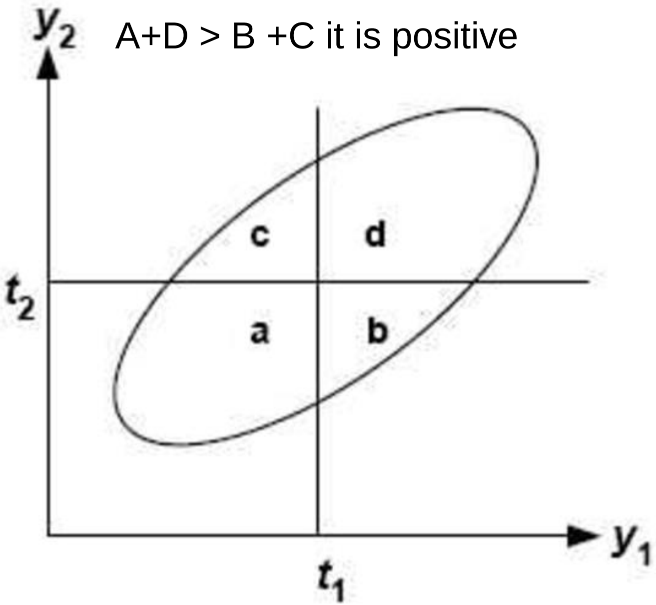


Assumption: continuous latent variables are normally distributed (standard normal distribution)



	0	1
1	A	B
0	C	D

A+D > B +C it is positive



[Example](https://www.john-uebersax.com/stat/tetra.htm)
https://www.john-uebersax.com/stat/tetra.htm

CFA equation

The underlying continuous dependent variable x∗ is now modeled instead of the observed ordinal variable.

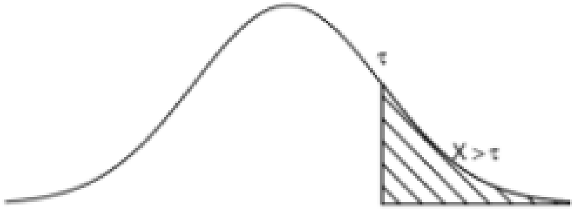
$$\mathbf{x}^* = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta}.$$

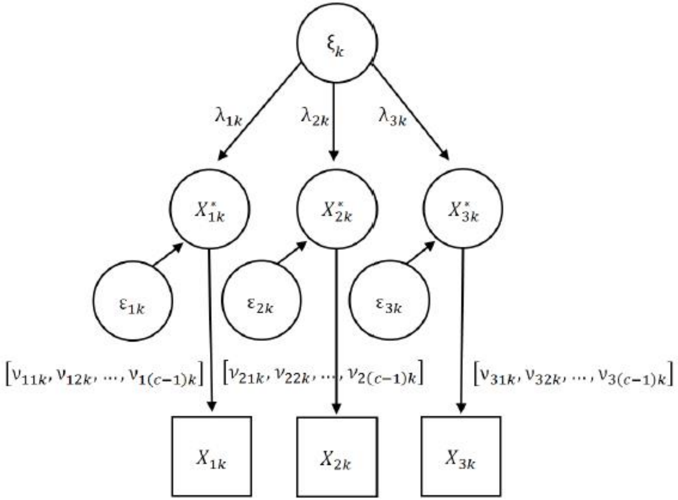
$X = b * L + a \Rightarrow X^* = b * L + a$

Item thresholds

New parameters for each item: tau (thresholds): if we were on standard normal distribution, where would our cut points be to get binary data?

Number of thresholds = number of categories - 1





Each threshold is the score in the latent variable needed for a subject to change answer one category over the other.

```
item_001|t1          0.447
```

A person with a score in the underlying latent variable lower than .447 is most likely to answer the first category (0).

Output in R

New parameters for each observed variable - thresholds.

Thresholds are on the same scale as factor scores

Usually **do not pay attention to significance level** (check the hypothesis if thresholds are differ from zero)

Thresholds:

	Estimate	Std.Err	z-value	P(>|z|)
mrt1|t1	0.153	0.055	2.762	0.006
mrt2|t1	0.347	0.056	6.171	0.000
mrt3|t1	-0.614	0.059	-10.399	0.000
mrt4|t1	-0.236	0.056	-4.250	0.000
mrt5|t1	0.410	0.057	7.214	0.000
mrt6|t1	0.404	0.057	7.127	0.000
mrt7|t1	-0.027	0.055	-0.482	0.630
mrt8|t1	0.163	0.055	2.937	0.003
mrt9|t1	1.466	0.083	17.660	0.000

Thresholds:

	Estimate
enjoy_r1r|t1	-0.811
enjoy_r1r|t2	-0.132
enjoy_r1r|t3	0.355
enjoy_r2|t1	-0.207
enjoy_r2|t2	0.606
enjoy_r2|t3	1.228
enjoy_r3|t1	0.124
enjoy_r3|t2	0.988
enjoy_r3|t3	1.568

A person with a score in the underlying latent variable lower than tau is most likely to answer the first category (0).

For several thresholds

Output in R

Thresholds be converted and interpret (surprise) as a difficulty of items in IRT framework

(Talk about it during next class)

Thresholds:

	Estimate	Std.Err	z-value	P(>|z|)
mrt1|t1	0.153	0.055	2.762	0.006
mrt2|t1	0.347	0.056	6.171	0.000
mrt3|t1	-0.614	0.059	-10.399	0.000
mrt4|t1	-0.236	0.056	-4.250	0.000
mrt5|t1	0.410	0.057	7.214	0.000
mrt6|t1	0.404	0.057	7.127	0.000
mrt7|t1	-0.027	0.055	-0.482	0.630
mrt8|t1	0.163	0.055	2.937	0.003
mrt9|t1	1.466	0.083	17.660	0.000

Thresholds:

	Estimate
enjoy_r1r|t1	-0.811
enjoy_r1r|t2	-0.132
enjoy_r1r|t3	0.355
enjoy_r2|t1	-0.207
enjoy_r2|t2	0.606
enjoy_r2|t3	1.228
enjoy_r3|t1	0.124
enjoy_r3|t2	0.988
enjoy_r3|t3	1.568

For several thresholds

Psychometrics & Statistics II

Confirmatory Factor Analysis: Measurement Invariance

Daria Gracheva, Sergei Tarasov

Lecture, part 1

CFA vs IRT

Ordered CFA & IRT for binary data

Ordered CFA can be compared to IRT models with binary items.

In IRT, the probability of a positive response (solve item correct) is predicted by latent trait (theta), difficulty of item (b) and discrimination of item (a) using the **logistic function**:

$$P(y_{is} = 1 \mid \theta_s, b_i, a_i) = \frac{\exp[a_i(\theta_s - b_i)]}{1 + \exp[a_i(\theta_s - b_i)]}$$

2PL IRT model

*CFA counterpart to a 1PL IRT model can be estimated by holding the factor loadings equal across items.

2PL vs 2PNO

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz,$$

The **normal ogive model** was the first IRT model for measuring psychological and/or educational latent traits. **George Rasch**, came up with a different approach to IRT by using the logistic function to instead of the normal ogive function:

$$P(Y_{ip} = 1) \stackrel{\smile}{=} \frac{\exp\{a_i(\theta_p - b_i)\}}{1+\exp\{a_i(\theta_p - b_i)\}}$$

2PL - logistic IRT model (logistic metric)

In practice D = 1 used

$$P(Y_{ip} = 1) \stackrel{\smile}{=} \frac{\exp\{Da_i(\theta_p - b_i)\}}{1+\exp\{Da_i(\theta_p - b_i)\}}$$

2PNO - normal ogive IRT model (normal metric)

D is a scaling factor whose value is 1.702. This D-value produces **near equivalent values and interpretations between the item parameters in 2PNO and 2PL**

Read more:
https://www.umass.edu/remp/software/simcata/wingen/modelsF.html

Ordered CFA & IRT parameters

In IRT terminology, Item difficulties = item threshold. **Item difficulty parameters in 2PL are analogous to item thresholds (t) in CFA with categorical outcomes.**

Threshold indicates the level of latent variable at which respondents have a 50% chance of transitioning from one response choice on an item to another.

```
Thresholds:

                   Estimate
    mrt1|t1           0.153
    mrt2|t1           0.347
```

A person with a score in the underlying latent variable higher than 0.153 is most likely to answer item 1 correct (higher 0.357 to answer item 2 correct) => item 2 is more difficult

IRT item discrimination parameters are analogous to factor loadings in CFA because they represent the relationship between the latent trait and the item responses.

Transformation of parameters FA ~ IRT

D = 1.702

Δ-FA → 2PNO IRT	Δ-FA → 2PL IRT
$a_i = \lambda_i / \sqrt{1-\lambda_i^2}$ $d_{ik} = \tau_{ik} / \sqrt{1-\lambda_i^2}$ $b_{ik} = \tau_{ik}/\lambda_i$	$a_i = D\,\lambda_i / \sqrt{1-\lambda_i^2}$ $d_{ik} = D\,\tau_{ik} / \sqrt{1-\lambda_i^2}$ $b_{ik} = \tau_{ik}/\lambda_i$
2PNO IRT →Δ-FA	**2PL IRT →Δ-FA**
$\lambda_i = a_i / \sqrt{1+a_i^2}$ $\tau_{ik} = d_{ik} / \sqrt{1+a_i^2} = a_i b_{ik} / \sqrt{1+a_i^2}$	$\lambda_i = a_i / \sqrt{D^2+a_i^2}$ $\tau_{ik} = d_{ik} / \sqrt{D^2+a_i^2} = a_i b_{ik} / \sqrt{D^2+a_i^2}$

CFA & IRT for polytomous data

Ordered CFA can be compared to Graded Response Model (GRM), not Rating Scale (RSM) or Partial Credit Model (PCM) in Item Response Theory framework.

Ordered CFA & IRT

Both IRT and ordered CFA can be used to:

- Explore the latent dimensionality of categorical outcomes (for multidimensional IRT)
- Evaluate the psychometric properties of a test
- Conduct differential item functioning analysis (part of measurement invariance)
- Assign sample participants a latent trait level estimate (akin to a factor score).

Bean, G. J., & Bowen, N. K. (2021). Item response theory and confirmatory factor analysis: complementary approaches for scale development. Journal of Evidence-Based Social Work, 18(6), 597-618.

Nice article

Lecture, part 2

Brown, 2015
Chapter 7. Starting from CFA IN MULTIPLE GROUPS (page ~ 241)

Multigroup CFA model

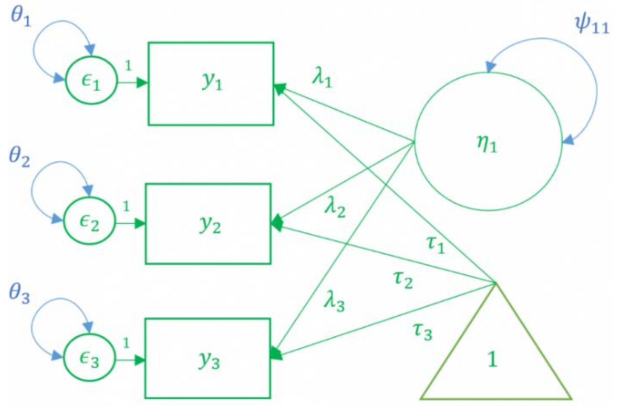
In CFA it is possible to examine the equivalence of all model parameters of the factor model across multiple groups.

Do the items **measure the same constructs** (same factor structure) and evidence **equivalent relationships to these constructs** (equal factor loadings) in different subgroups? Does **people with same ability level from different groups** have the same scores for indicators?

To answer this question, we need to test the equivalence of different model parameters across groups = test of ***measurement invariance*** in multigroup CFA model.

Measurement Invariance Models

- **Configural invariance**: number of factors and pattern of indicator–factor loadings are identical across groups.
- **Metric invariance** (weak invariance): unstandardized factor loadings of items are equal across groups
- **Scalar invariance** (strong invariance): intercepts or thresholds of items are equal across groups
- **Strict invariance**: residual or error variance are equal across groups (optional only for interval CFA)



Configural Invariance

Are there gender differences that prevent comparable questionnaire responses for some latent trait? Two groups: males and females

First check **model fit on separate models** for males and females. Does both models fit the data? Does all factor loadings are significant?

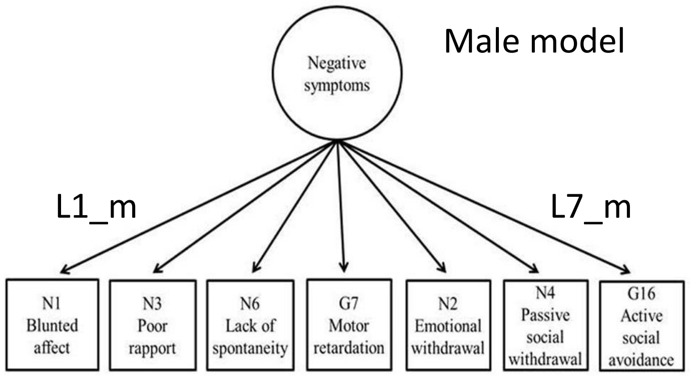
- ➔ If yes: **create multigroup CFA model** with free parameters = configural model, go next to metric model
- ➔ If no: **construct structure is differ across groups** (ex: different number of factors, cross-loadings), no sense to test more strict levels of measurement invariance

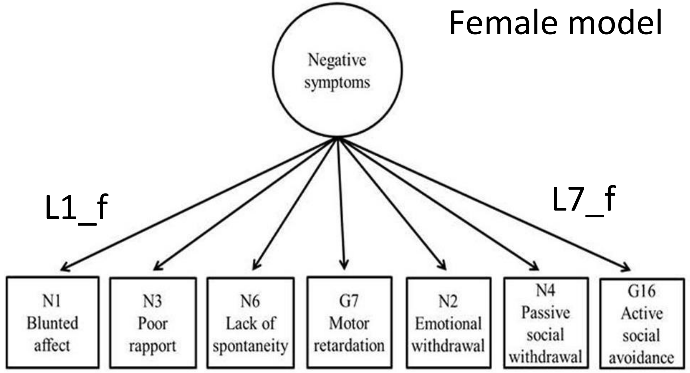
Metric Invariance

Set the same constraints on the **unstandardized factor loadings** for all groups. L1_m = L1_f …
L7_m = L7_f

Is the **model fit significantly worse than the configural model** when factor loadings are freely estimated for both groups?

How we can answer the question?





Compare models

Is the **model fit significantly worse than the configural model** when factor loadings are freely estimated for both groups? Compare models with LR test!

- Configural model is **more complex (less df)** = we estimate more parameters
- Metric model is **more simple (more df)** = we constrain some parameters to be equal = estimate less parameters

```
Chi-Squared Difference Test

                Df   AIC   BIC  Chisq Chisq diff    RMSEA Df diff Pr(>Chisq)
fit_configural 18 64787 65016 186.44
fit_metric     23 64789 64987 198.67     12.238 0.025815       5    0.03167 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

What does the result mean?

Compare models

- ➔ **LR test is significant** = complex model (configural better). Not all factor loadings of items are equal across groups! Try to find problematic items..
- ➔ **LR test is non-significant** = simple model (with constraints) better. Factor loadings are equal across groups. Go next to scalar model!

```
Chi-Squared Difference Test

               Df   AIC   BIC  Chisq Chisq diff    RMSEA Df diff Pr(>Chisq)
fit_configural 18 64787 65016 186.44
fit_metric     23 64789 64987 198.67     12.238 0.025815       5    0.03167 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

What does the result mean?

Compare models using fit indexes

The results of the Chi-squared difference test are influenced by sample size & can be very strict. That is why we can use **alternative way to compare models - via the difference in fit indexes** (CFI, TLI, RMSEA)

Use recommendations from **Chen, F. F**. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural equation modeling: a multidisciplinary journal, 14(3), 464-504.

- With sample size N > 300, for **testing loadings invariance**, a change of 0.010 in CFI / 0.015 in RMSEA / 0.030 in SRMR would indicate noninvariance.
- For testing **intercept or residual invariance**, a change of 0.010 in CFI / 0.015 in RMSEA / 0.010 in SRMR would indicate noninvariance.

Scalar invariance

Set the same constraints on the indicator **<u>intercepts (or thresholds)</u>** for all groups.

Is the **model fit significantly worse than the metric model** when intercepts (thresholds) are freely estimated for both groups? Compare metric and scalar model.

Group differences on the indicator's thresholds is similar to differential item functioning (in IRT).

- ➔ **LR test is significant** = complex model (metric better). Not all intercepts of items are equal across groups! Try to find problematic items..
- ➔ **LR test is non-significant** = simple model (with constraints) better. Intercepts are equal across groups. **<u>Measurement invariance holds!</u>**

Comparison of means

Study “true differences” in latent trait between groups.

Group comparisons of latent means are meaningful only if the factor loadings and indicator intercepts (thresholds) have been found to be invariant* (full scalar invariance).

Why not use t-test for comparison? **Measurement invariance tests ensure that group comparisons are appropriate**.CFA may reveal some non-invariance (in loadings, intercepts), making comparisons of latent means impossible.

Group comparisons have more precision in CFA because the structural parameters (factor means, variances, covariances) have been adjusted for measurement error.

Comparison of means: tech issues

Due to identification problems, **we need to fix the mean of one latent group to 0**. This group becomes the reference group.The **latent means in the remaining groups are freely estimated** and represents **deviations from the reference group's latent mean.**

if Group 2's latent mean = -0.379, on average, this group scores 0.379 units lower than the reference group on the latent dimension (factor scores).

```
Intercepts:
                   Estimate
   .enjy_r1 (.14.)    2.774
   .enjy_r2 (.15.)    2.059
   .enjy_r3 (.16.)    1.783
   .enjy_r4 (.17.)    1.774
   .enjy_r5 (.18.)    2.124
   .enjy_r6 (.19.)    1.832
    enjoy             0.000
```

Means for reference group (left) and another group (right)

```
Intercepts:
                   Estimate
   .enjy_r1 (.14.)    2.774
   .enjy_r2 (.15.)    2.059
   .enjy_r3 (.16.)    1.783
   .enjy_r4 (.17.)    1.774
   .enjy_r5 (.18.)    2.124
   .enjy_r6 (.19.)    1.832
    enjoy            -0.379
```

Partial Measurement Invariance

If you have difficulties in achieving a certain level of measurement invariance (e.g. metric, scalar), it is possible to free some parameters to achieve partial measurement invariance. Use **modification indexes** to find out which item parameter is a problem. Still not recommended to compare latent means in case of partial MI

Scaler Invariance	Significant difference	1,853.85	94	1,977.15	30	<0.001	0.762
Partial Scaler Invariance	Significant difference	348.26	82	169.94	18	<0.001	

Measurement invariance for ordered CFA

Chi-square test for model comparison usually get biased results for ordered CFA

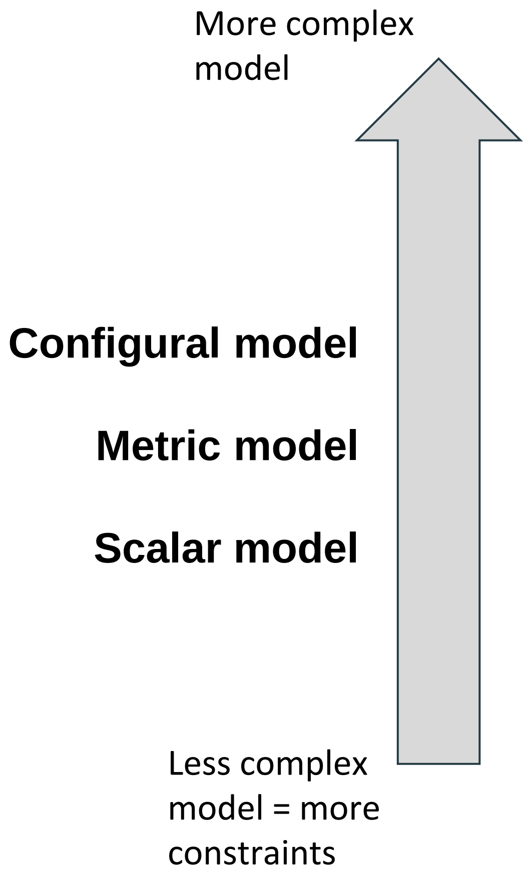
In practice researchers use difference in fit indexes by Chen, but event these cut-offs were investigated based on classical CFA

Summary

The idea of measurement invariance testing is to find out - can we provide **comparable measurements for people in different groups?**

- Is the construct structure the same?
- Does people with same ability level from different groups have the same scores for indicators?

Compare models with constraints step by step (Configural VS Metric, Metric VS Scalar).



HW: Real example for reading

*Google “measurement invariance” and find mooooore

Popular approach to study differences in cultures (does the same instrument can be used in different cultures?)

Gender invariance / SES invariance also very popular

Samavi, A., Hajializadeh, K., Javdan, M., & Farshad, M. R. (2022). Psychometric validation of teacher empathy scale: Measurement invariance in gender. *Frontiers in Psychology*, *13*, 1042993.

Psychometrics & Statistics II

Confirmatory Factor Analysis: Measurement Invariance and Reporting Results

Daria Gracheva, Sergei Tarasov

Repeat the material

Measurement invariance

- What is the goal of measurement invariance?
- What steps of measurement invariance do you know?
- How to test for measurement invariance?
- What is partial invariance?

Steps of measurement invariance

1. **Test the CFA model separately in each group**
2. **Configural model (both groups together)**
3. **Metric model - test the equality of factor loadings.**
4. **Scalar model - test the equality of intercepts.**
5. **Strict model - test the equality of indicator error variance**

+ **Test the equality of factor variances**
+ **Test the equality of factor covariances**
+ **Test the equality of latent means**

Steps 1-5 tests measurement invariance. Extra steps are tests of population heterogeneity according to Brown, 2015.

Example of MI results

Step	Invariance test	Model comparison	x^{2a}	DF	$\Delta\ x^{2b,\ c}$	Δ DF	p value	CFI	ΔCFI
1	Configural Invariance	–	178.63	36	–	–	–	0.981	–
2	Metric Invariance	Non-significant difference	202.62	64	20.91	28	0.829	0.981	0
3	Scaler Invariance	Significant difference	1,853.85	94	1,977.15	30	<0.001	0.762	0.219

What can you tell about the results?

Example of partial MI results

- *Measurement invariance results & can also include chi square difference test*

Table 2
Fit Statistics of Various Invariance Models

Model	χ^2	*df*	CFI	ΔCFI	RMSEA	ΔRMSEA
Configural invariance	119.862	6	0.997		0.057	
Weak invariance	136.377	10	0.997	0.000	0.047	−0.010
Partial weak invariance (freed λ_5 and all τs and θs)	124.963	9	0.997	0.000	0.047	−0.010
Partial strong invariance (freed λ_5, τ_5, and all θs)	310.578	12	0.993	−0.004	0.065	0.018
Partial strong invariance (freed λ_5, τ_2, τ_4, τ_5, and all θs)	125.906	10	0.997	0.000	0.045	−0.002
Partial strict invariance (freed λ_5, τ_2, τ_4,τ_5, θ_4, θ_5)	128.181	13	0.997	0.000	0.039	−0.006
Partial strict invariance (freed λ_5, τ_2, τ_4, τ_5, θ_4, θ_5,θ_{12})	129.045	14	0.997	0.000	0.038	−0.007

Note. The changes in fit statistics (ΔCFI and ΔRMSEA) are computed between configural and partial weak invariance models, partial weak and partial strong invariance models, partial strong and partial strict invariance models.

Measurement invariance with intercepts

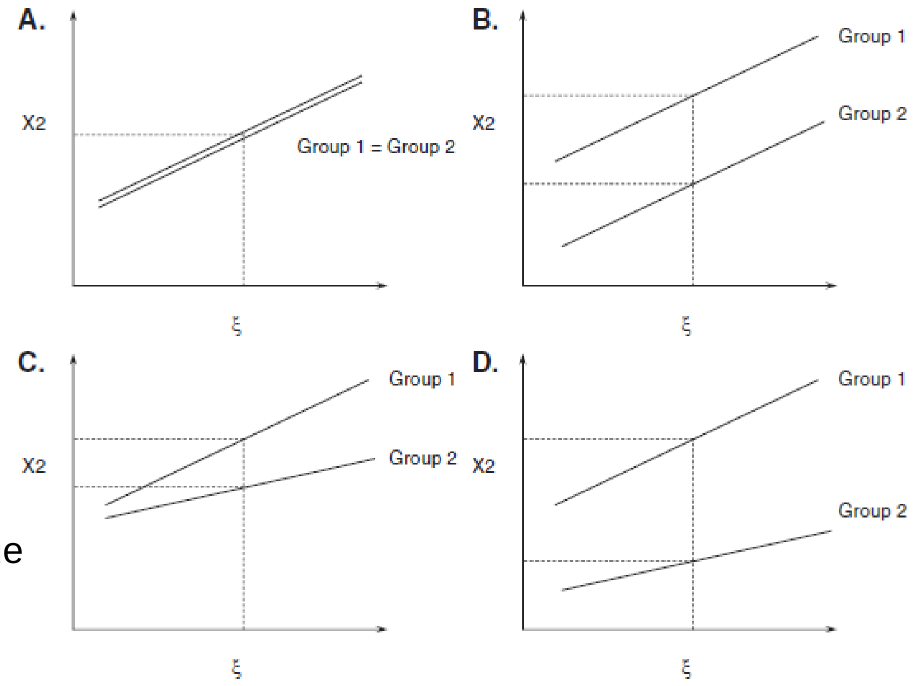
With default estimators (ML, MLR), we test invariance of factor loadings and **intercepts.**

What is intercept in factor model?

$$y_1 = \tau_1 + \lambda_1 \eta_1 + \epsilon_1$$
$$y_2 = \tau_2 + \lambda_2 \eta_1 + \epsilon_2$$
$$y_3 = \tau_3 + \lambda_3 \eta_1 + \epsilon_3$$

In graphs

- **A) Factor loadings and intercepts are invariant**
- B) Factor loadings are invariant, intercepts differ
- C) Intercepts are invariant, factor loadings differ
- D) Both differ



Idea of non-invariance: with the same level of true score (latent trait) predicted scores on indicators will differ

Group size

Many aspects of CFA are sensitive to sample size. Measurement invariance is not exception.

For multi-group CFA, it is preferable for the **size of the groups to be as balanced as possible.** In instances where the group sizes differ considerably, it is harder to reach full measurement invariance.

Any questions?

About article:

Samavi, A., Hajializadeh, K., Javdan, M., & Farshad, M. R. (2022). Psychometric validation of teacher empathy scale: Measurement invariance in gender. *Frontiers in Psychology*, *13*, 1042993.

Or measurement invariance?

Or CFA?

Exercises

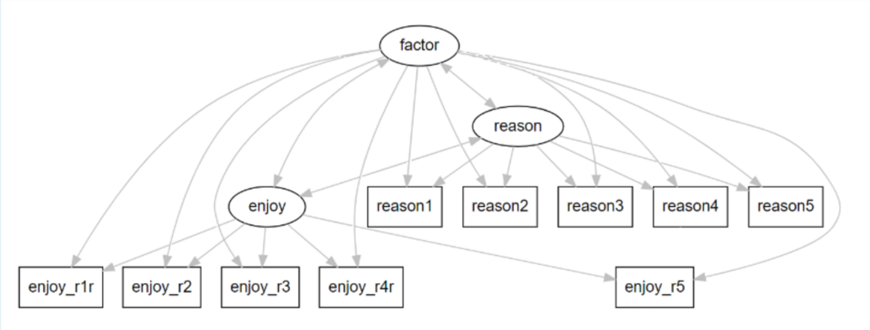
Exercise 1

1) Choose the model that is identified.

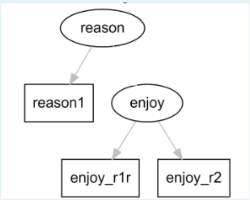
2) How many items should be in one factor (the min number)?

3) What is the name of the Model c?

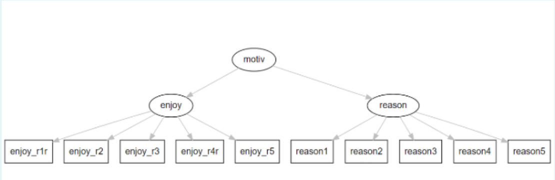
○ a.



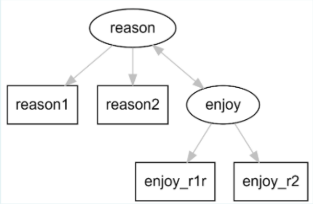
○ b.



○ c.



○ d.



Exercise 2

You have created a CFA model based on a dataset with several dichotomous (1/0) items. Since data is categorical, you used the **WLSMV estimator.**

You get the following results about the model fit. In your answer, write values of most popular fit indexes and compare them to recommended cut-offs, interpret the results. Make an overall conclusion about the model fit.

```
## Model Test User Model:
##
##                                               Standard      Robust
##   Test Statistic                              1495.744    1540.553
##   Degrees of freedom                               252         252
##   P-value (Chi-square)                           0.000       0.000
##   Scaling correction factor                                  1.003
##   Shift parameter                                           49.957
##     simple second-order correction
##
## Model Test Baseline Model:
##
##   Test statistic                             15374.629   10642.046
##   Degrees of freedom                               276         276
##   P-value                                        0.000       0.000
##   Scaling correction factor                                  1.457
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.918       0.876
##   Tucker-Lewis Index (TLI)                       0.910       0.864
##
##   Robust Comparative Fit Index (CFI)                            NA
##   Robust Tucker-Lewis Index (TLI)                               NA
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.055       0.056
##   90 Percent confidence interval - lower         0.053       0.054
##   90 Percent confidence interval - upper         0.058       0.059
##   P-value RMSEA <= 0.05                          0.001       0.000
##
##   Robust RMSEA                                                  NA
##   90 Percent confidence interval - lower                        NA
##   90 Percent confidence interval - upper                        NA
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.086       0.086
##
```

Exercise 3

Latent Variables:

		Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
CM3 =~							
	Item 1	1.000				0.409	0.409
	Item 2	1.067	0.146	7.318	0.000	0.436	0.436
	Item 3	0.929	0.143	6.503	0.000	0.380	0.380
	Item 4	1.049	0.143	7.345	0.000	0.428	0.428
	Item 5	-1.362	0.161	-8.462	0.000	-0.556	-0.556
	Item 6	1.021	0.141	7.267	0.000	0.417	0.417
	Item 7	0.391	0.119	3.284	0.001	0.160	0.160
	Item 8	1.439	0.166	8.661	0.000	0.588	0.588
	Item 9	1.492	0.169	8.815	0.000	0.609	0.609
	Item 10	1.003	0.155	6.468	0.000	0.410	0.410
	Item 11	-0.047	0.119	-0.397	0.692	-0.019	-0.019
	Item 12	1.405	0.160	8.808	0.000	0.574	0.574
	Item 13	1.973	0.215	9.163	0.000	0.806	0.806

Exercises 4

1) What does error covariance between two items mean? Give at least one reason why there might be error covariances between two items in the test.

2) Describe the situation where you need to apply bi-factor model.

Exercise 5

What is the difference between EFA and CFA?

Module 2 Assessment

Final test 1

The final test will take place on **24th December from 18:10 to 19:30**.

It will contain questions related to CFA, measurement invariance, structural equation modelling (covered in the next class).

There will be both open and closed questions close to the exercises. Most questions will require you to interpret R output and understand the R syntax of the models.

- You will need to have your cameras on during the exam.
- You can use any materials during the exam

HW 1

Deadline for submission on SmartLMS is 12th January at 23:59.

Divide into groups of 2-3 students. It is allowed to do the work alone, but it is not desirable.

Need to send R code and solution (can be markdown file)

Description of the HW task

Reporting CFA

How to report CFA results?

In methodology section: description of method

In this study we used **Confirmatory factor analysis**… to check the internal structure of the instrument

We used ML / MLR / WLSMV **estimator** because we used interval data that can be considered normal / interval non-normal data / categorical data

We use the following **fit indexes** to assess the quality of the model: CFI > 0.95, TLI > 0.95, RMSEA < 0.05, SRMR < 0.05 *(usually we put reference here)*

We calculate **reliability estimates** of factors using the Omega coefficient (McDonald, 1999). Values above 0.7 are satisfactory *(usually we put reference here)*

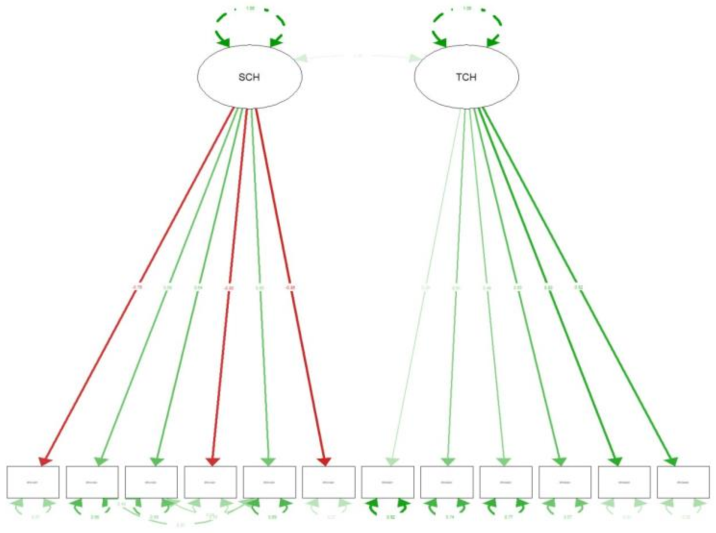
The **analysis was conducted** in soft R, package ‘lavaan’

Results section

- Model description (can be earlier)

According the theory, **construct motivation consists of two correlated factors** - external and internal motivation (or you can describe other structure: second-order, bi-factor here). Some error covariances were added to the model to address the issue of local item dependence due to close wording **(Picture 1 - usually show structure)**

Next - report model fit of the final model



Results section

- *Model fit*

The model with one factor demonstrated acceptable *(or poor)* fit ($\chi2$ (df = 9) = 181.43, p = 0.00; CFI = 0.974, TLI = 0.956, RMSEA = 0.066 (90% CI 0.058; 0.075), SRMR = 0.031)

For ML estimator!

```
Model Test User Model:

  Test statistic                              181.433
  Degrees of freedom                                9
  P-value (Chi-square)                          0.000

Model Test Baseline Model:

  Test statistic                             6600.532
  Degrees of freedom                               15
  P-value                                       0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                   0.974
  Tucker-Lewis Index (TLI)                      0.956

Root Mean Square Error of Approximation:

  RMSEA                                         0.066
  90 Percent confidence interval - lower        0.058
  90 Percent confidence interval - upper        0.075
  P-value H_0: RMSEA <= 0.050                   0.001
  P-value H_0: RMSEA >= 0.080                   0.004

Standardized Root Mean Square Residual:

  SRMR                                          0.031
```

The model with one factor demonstrated acceptable *(or poor)* fit ($\chi 2$ (df = 9) = 183.813, p = 0.00; CFI = 0.955, TLI = 0.925, RMSEA = 0.067 (90% CI 0.059; 0.075), SRMR = 0.031)

```
Root Mean Square Error of Approximation:

  RMSEA                                          0.042       0.067
  90 Percent confidence interval - lower         0.034       0.059
  90 Percent confidence interval - upper         0.051       0.075
  P-value H_0: RMSEA <= 0.050                    0.929       0.000
  P-value H_0: RMSEA >= 0.080                    0.000       0.006

  Robust RMSEA                                               0.044
  90 Percent confidence interval - lower                     0.038
  90 Percent confidence interval - upper                     0.049
  P-value H_0: Robust RMSEA <= 0.050                         0.967
  P-value H_0: Robust RMSEA >= 0.080                         0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.031       0.031
```

```
Model Test User Model:
                                              Standard      Scaled
  Test Statistic                                78.293     183.813
  Degrees of freedom                                 9           9
  P-value (Chi-square)                           0.000       0.000
  Scaling correction factor                                  0.428
  Shift parameter                                            0.704
    simple second-order correction

Model Test Baseline Model:

  Test statistic                              6475.341    3916.099
  Degrees of freedom                                15          15
  P-value                                        0.000       0.000
  Scaling correction factor                                  1.656

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.989       0.955
  Tucker-Lewis Index (TLI)                       0.982       0.925

  Robust Comparative Fit Index (CFI)                         0.988
  Robust Tucker-Lewis Index (TLI)                            0.981
```

Example of WLSMV estimator

Results section

- *Factor loadings.* Report standardized loadings + significance value

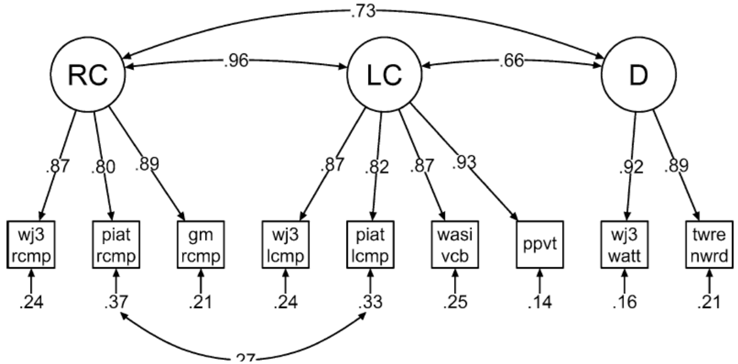
Two options:

➔ Report factor loadings via picture
➔ Report factor loadings in table

Also, give a brief description:

All factor loadings were significant ($p < 0.01$) with mean size of standardized factor loadings = 0.4 (sd = 0.1).

- Correlation between factors
- Reliability for all factors




```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  enjoy =~
    enjoy_r1r         0.341    0.018   18.638    0.000    0.341    0.294
    enjoy_r2          0.492    0.017   29.745    0.000    0.492    0.488
    enjoy_r3          0.552    0.016   34.800    0.000    0.552    0.626
    enjoy_r4r         0.613    0.016   38.664    0.000    0.613    0.661
    enjoy_r5          0.669    0.014   49.336    0.000    0.669    0.726
    enjoy_r6          0.694    0.014   50.928    0.000    0.694    0.807
```

Item	Std. factor loading
Item1	0.294**
Item2	0.488**
**p<0.01; * p<0,05	

Results section

- *Model comparison*

Table 5
Chi-square test for the change in fit between nested models.

Nested models	χ^2 difference	df difference	p
Model 1. Six first-order uncorrelated factors	–	–	–
Model 2. Six first-order factors, and one second-order general factor	287.743	6	.001
Model 3. Six first-order factors, and two correlated second-order general factors	4.869	1	.04
Model 4. Six first-order correlated factors	19.788	8	.02

Psychometrics & Statistics II

Structural Equation Modeling

Daria Gracheva, Sergei Tarasov

Lecture

Studying relationship between variables

The **regression model** permits the prediction of dependent observed variable scores (Y) given a linear weighting of a set of independent observed scores (X's)

How many dependent variables are there in regression models? What types of variables we can use in regression? What types of regression models exist?

Give an example of research question / hypothesis for regression model.

Regression results

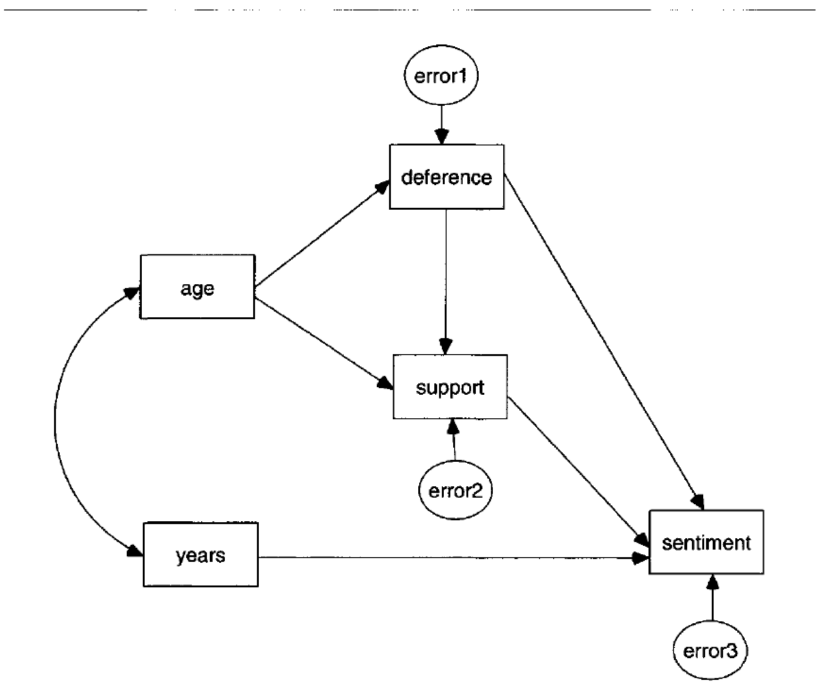
```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               395.303      1.758  224.92   <2e-16 ***
GENDERMale                 16.109      1.218   13.22   <2e-16 ***
BOOK_HOME11-25 books       34.472      2.100   16.41   <2e-16 ***
BOOK_HOME26-100 books      75.300      1.948   38.66   <2e-16 ***
BOOK_HOME101-200 books     94.500      2.227   42.44   <2e-16 ***
BOOK_HOME201-500 books    114.986      2.421   47.50   <2e-16 ***
BOOK_HOME>500 books       109.338      2.879   37.97   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Path Model

Path models use correlation coefficients and regression analysis to model more **complex relationships among observed variables**. For example, several dependent and independent variables.

Lines directed from one observed variable to another observed variable denotes **direct effects (regression).**

Double-headed line between two independent observed variables indicates **covariance (correlation)**



Each dependent variable has an **error term**
Some portion of deference will be predicted or explained by age and some will not.

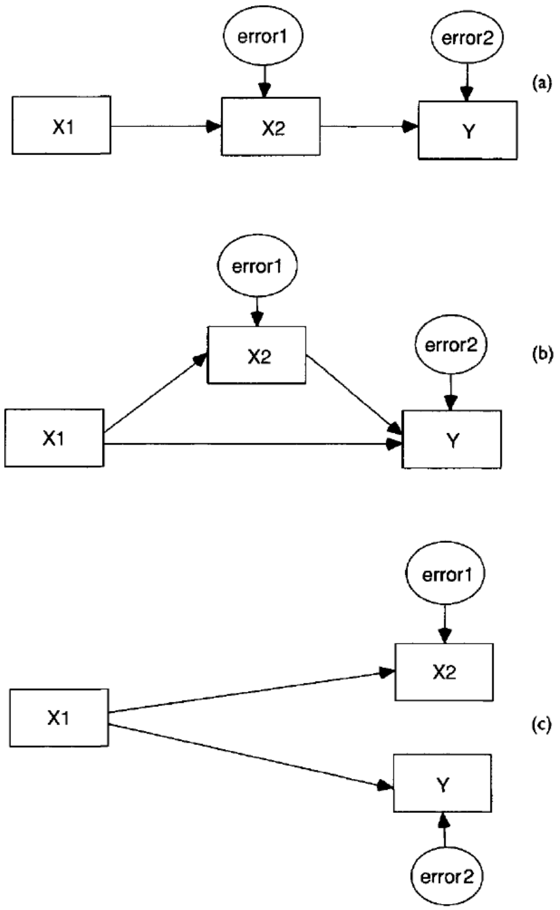
Path models

Three variables – different relationships

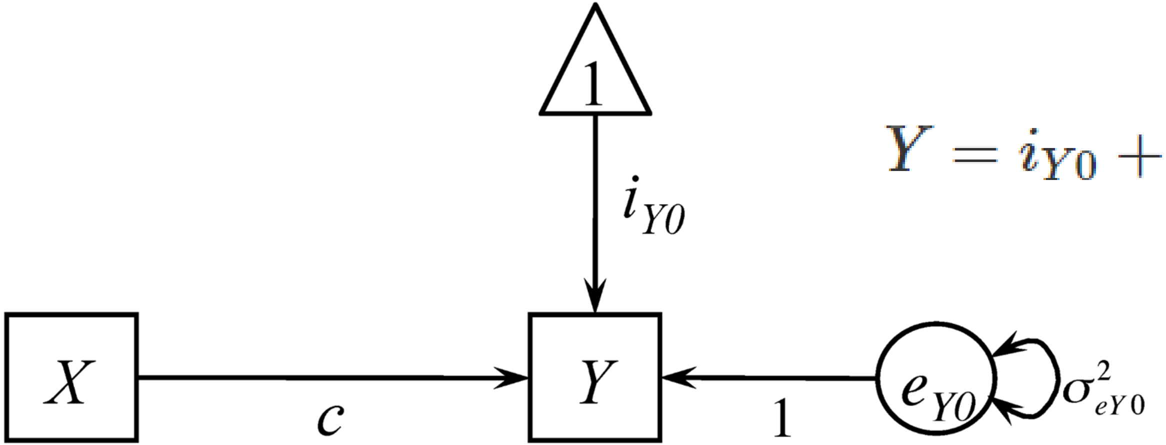
We also can create simple regression model

Y = a + B1*X1 + B2*X2 – another model!

Path models are more complex



Regression as path diagram

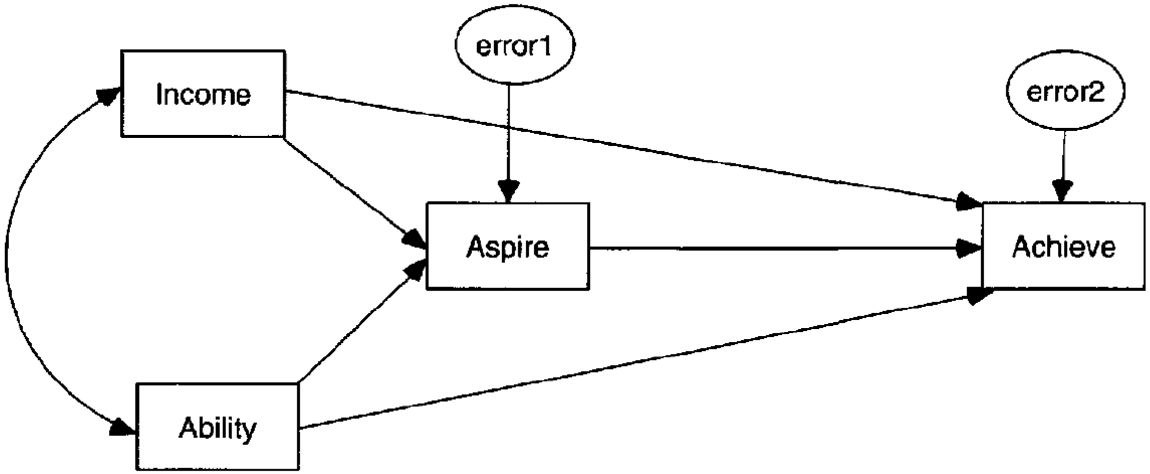


$$Y = i_{Y0} + cX + e_{Y0}$$

Exercise

Write a regression equations for path model

How many dependent variables we have here?



Structural Equation Modelling

SEM models combine path models and confirmatory factor models. SEM models incorporate **both latent and observed variables**

Variables, whether they are observed or latent, can also be defined as either *independent variables* or *dependent variables*

The goal of SEM analysis is to determine the extent to which the theoretical model is supported by sample data

Regression

Path Analysis

CFA

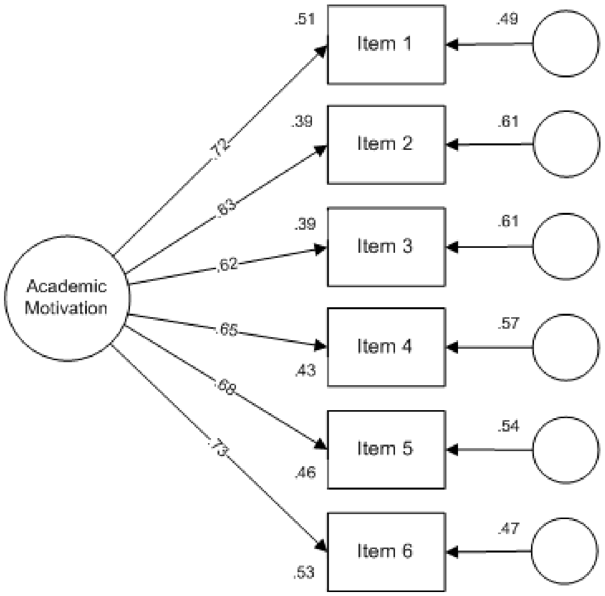
Types of models

Measurement model (factor model) specifies the number of factors, how the **various indicators are related to the factors**, and the relationships among indicator errors. Measurement models are usually evaluated by **Confirmatory Factor Analysis (CFA).**

Structural model, which specifies how the various **latent factors are related to one another.**

*Path analysis is a special case of SEM, examine the relationships between *observed variables*

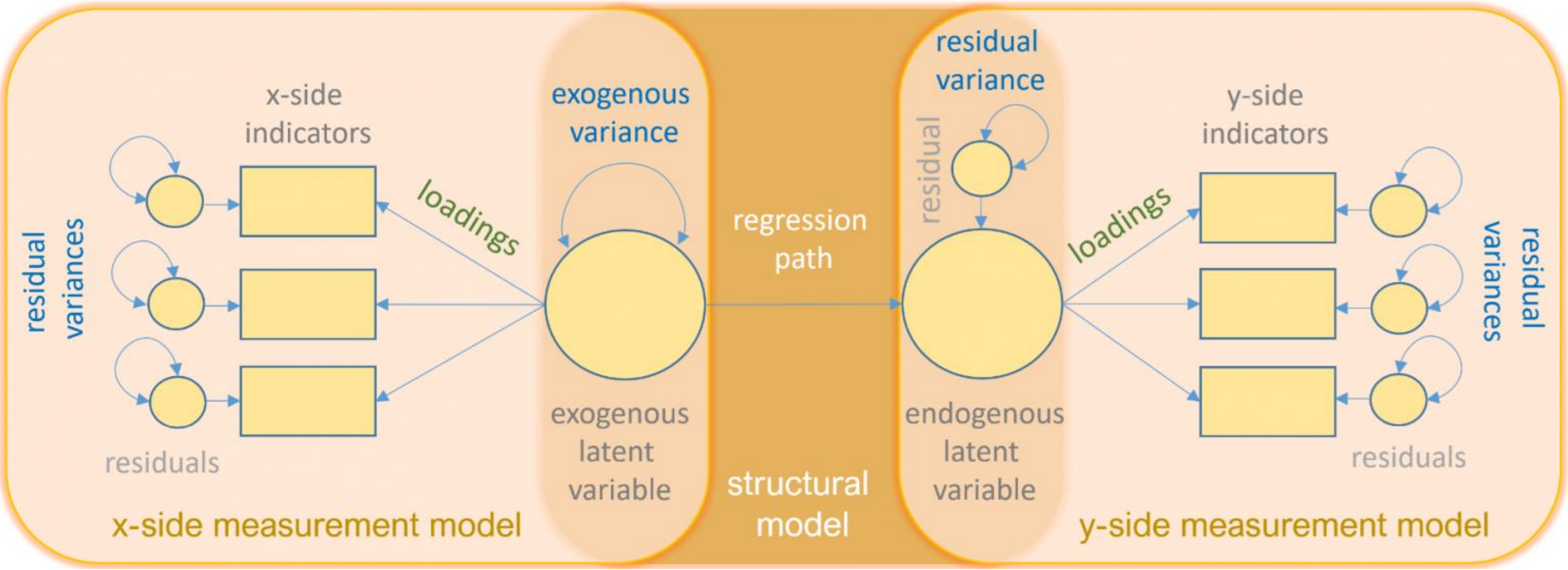
CFA model = Measurement model



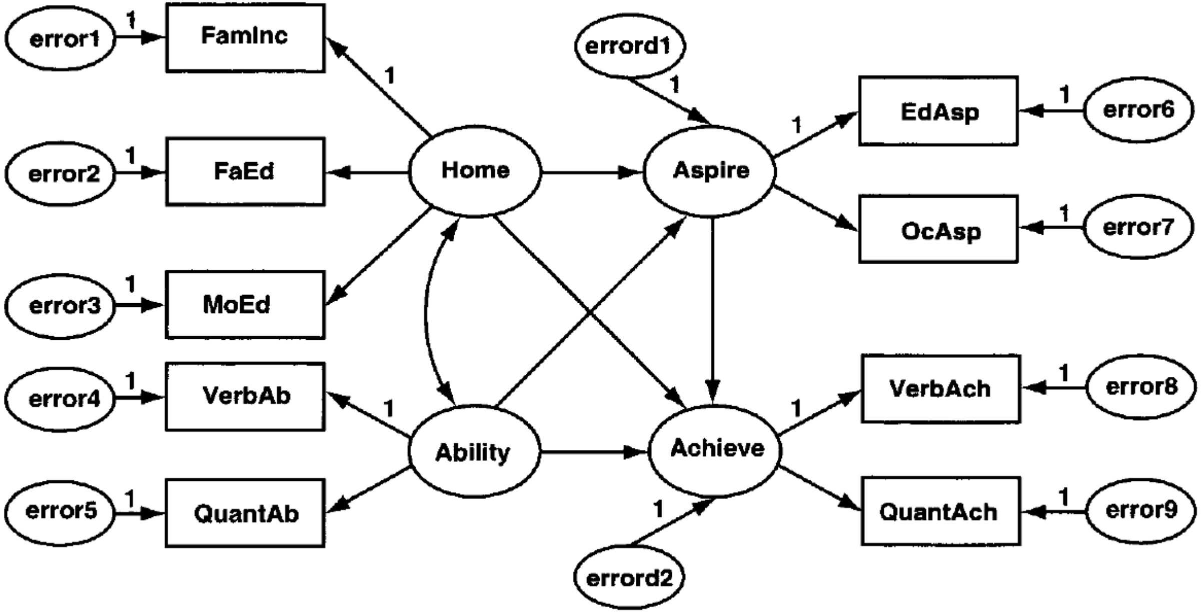
Structural model

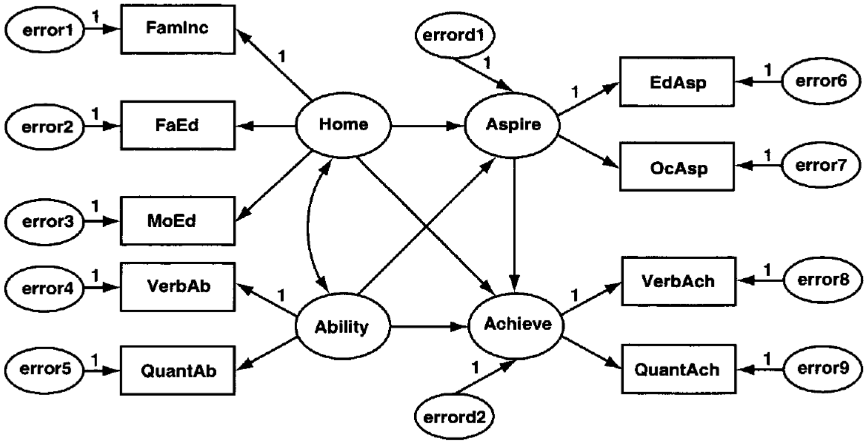
[INTRODUCTION TO STRUCTURAL EQUATION MODELING (SEM) IN R WITH LAVAAN]

Exogenous = independent
Endogenous = dependent



Find dependent variables





$$\text{EdAsp} = \text{function of Aspire(1)} + \text{measurement error}$$
$$\text{OcAsp} = \text{function of Aspire} + \text{measurement error}$$
$$\text{VerbAch} = \text{function of Achieve(1)} + \text{measurement error}$$
$$\text{QuantAch} = \text{function of Achieve} + \text{measurement error}$$
$$\text{FamInc} = \text{function of Home(1)} + \text{measurement error}$$
$$\text{FaEd} = \text{function of Home} + \text{measurement error}$$
$$\text{MoEd} = \text{function of Home} + \text{measurement error}$$
$$\text{VerbAb} = \text{function of Ability(1)} + \text{measurement error}$$
$$\text{QuantAb} = \text{function of Ability} + \text{measurement error.}$$

Measurement part

Structural part

$$\begin{aligned}\text{Aspire} &= \text{structure coefficient} * \text{Home} + \text{structure coefficient} * \text{Ability} \\ &\quad + \text{prediction error} \\ \text{Achieve} &= \text{structure coefficient} * \text{Home} + \text{structure coefficient} * \text{Ability} \\ &\quad + \text{structure coefficient} * \text{Aspire} + \text{prediction error.}\end{aligned}$$

Measurement & Structural models

"The testing of the structural model, i.e., the testing of the initially specified theory, may be meaningless unless it is **first established that the measurement model holds**. If the chosen indicators for a construct do not measure that construct, the specified theory must be modified before it can be tested. Therefore, the measurement model should be tested before the structural relationships are tested."

Why SEM?

- Various more **complex theoretical models** can be tested in SEM that hypothesize how sets of variables define constructs and how these constructs are related to each other.
- Structural equation modeling techniques explicitly **take measurement error into account** when statistically analyzing data.

Moderation & Mediation models

The relationship between X and Y is influenced by a **third variable** (moderator Z or mediator M).

"**when**" or "**for whom**" X most strongly predicts Y

Moderator

Moderator is a variable that affects the direction and/or strength of the relation between an independent variable and a dependent variable

why X can predict Y

Mediator

A given variable may be said to function as a **mediator** to the extent that it accounts for the relation between X and Y

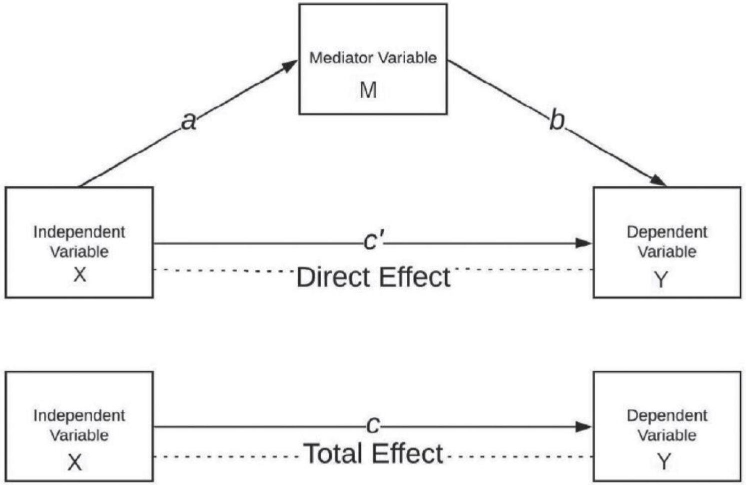
Mediation model

When we hypothesize an effect of X on Y in a statistical model, we might not assume such an effect occurs instantly. The effect might be realized **through a process or mechanism** (“black box”).

Total effect of X on Y - do not know what might have happened in the black box.

Mediation model:

- Indirect effect - the pathway from X to Y through
- Direct effect – direct pathway from X to Y – the effect of X on Y that is not through the mediation process of M.

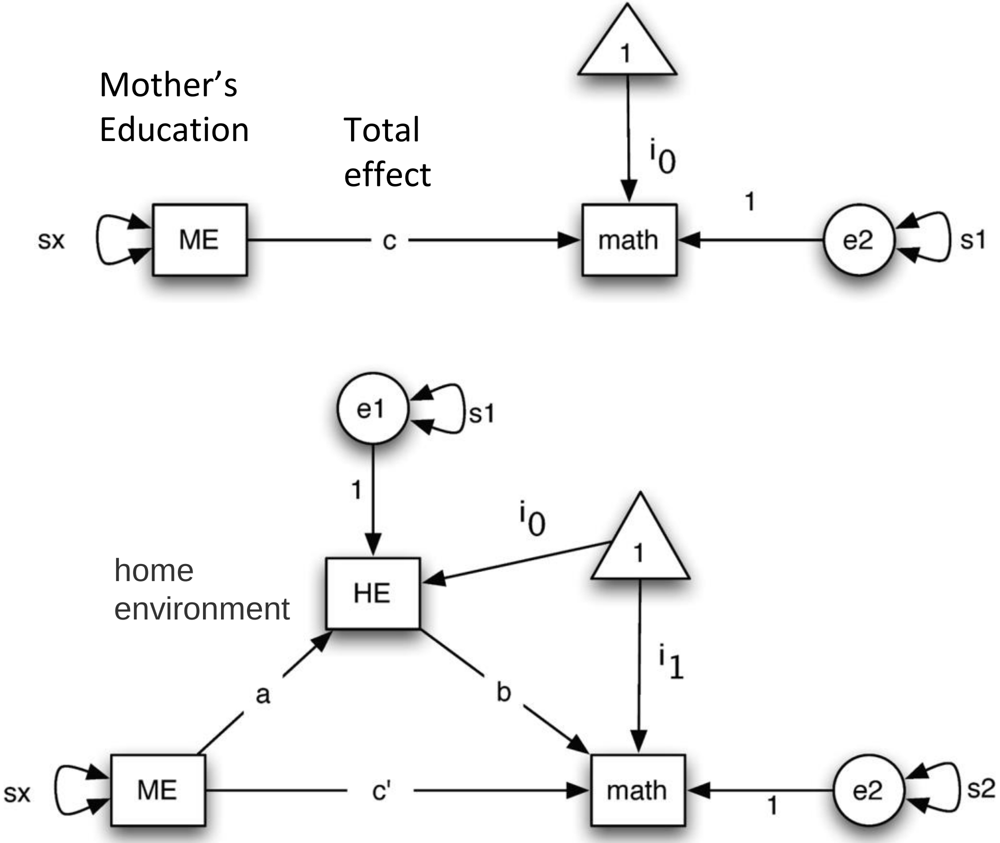


Mediation

Parents' education levels can relate to math **directly and indirectly.**

Hypothesis: **Home environment is a mediator** in the relation between mothers' education and children's mathematical achievement

The effect might be realized through a <u>process or mechanism</u>. **ME might first enhance home environment** (mediator), which in turn boosts the academic performance (Y) in math.



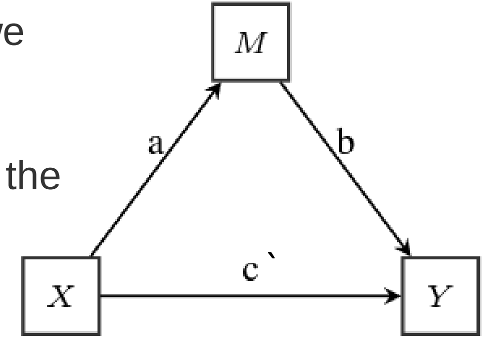
Hypothesis: **Home environment is a mediator** in the relation between mothers' education and children's mathematical achievement

c' - effect of X on Y
a - effect of X on M
b - effect of M on Y

$$M = a_0 + a \times X + e_M$$
$$Y = b_0 + b \times M + c' \times X + e_Y$$

To test these specifically, we need to define new parameters in the lavaan model that are a product of the individual paths.

"Open the black box and understand the mechanism"



$$\text{Direct effect} = c'$$
$$\text{Indirect effect} = b \times a$$
$$\text{Total effect} = c' + b \times a$$

```
myModel <- '
            Y ~ b*M + c*X
            M ~ a*X

            indirect := a*b
            total    := c + (a*b)
          '

fit <- sem(model = myModel,
           data  = myData,
           se    = "bootstrap")

summary(fit)
```

Full & Partial Mediation

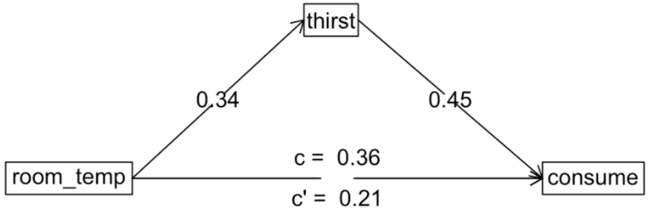
Full mediation:

- Variable X does not have a significant direct effect on Y (c` = 0)
- Mediator accounts for all of the relationship between the X and Y .
- Variable X does have a significant impact on moderator M, which also has a significant impact on response variable Y (a not equal 0; b not equal 0)

Partial mediation in case:

- Variable X have a significant direct effect on Y (c` not equal 0)
- Mediator accounts for some, but not all, of the relationship between the X and Y .
- Variable X does have a significant impact on moderator M, which also has a significant impact on response variable Y (a not equal 0; b not equal 0)

Output in R



room temperature is associated with water drinking indirectly through thirstiness.

Total effect = ab + c`
Indirect effect = a*b
Direct effect = cp (c`)

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  thirst ~
    room_temp  (a)    0.339    0.101    3.362    0.001    0.339    0.371
  consume ~
    thirst     (b)    0.451    0.149    3.025    0.002    0.451    0.413
    room_temp (cp)    0.208    0.130    1.599    0.110    0.208    0.208

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .thirst            0.911    0.153    5.966    0.000    0.911    0.862
   .consume           0.912    0.155    5.892    0.000    0.912    0.723

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
    ab                0.153    0.064    2.396    0.017    0.153    0.153
    total             0.360    0.116    3.098    0.002    0.360    0.361
```

Non significant direct effect (cp), significant a and b

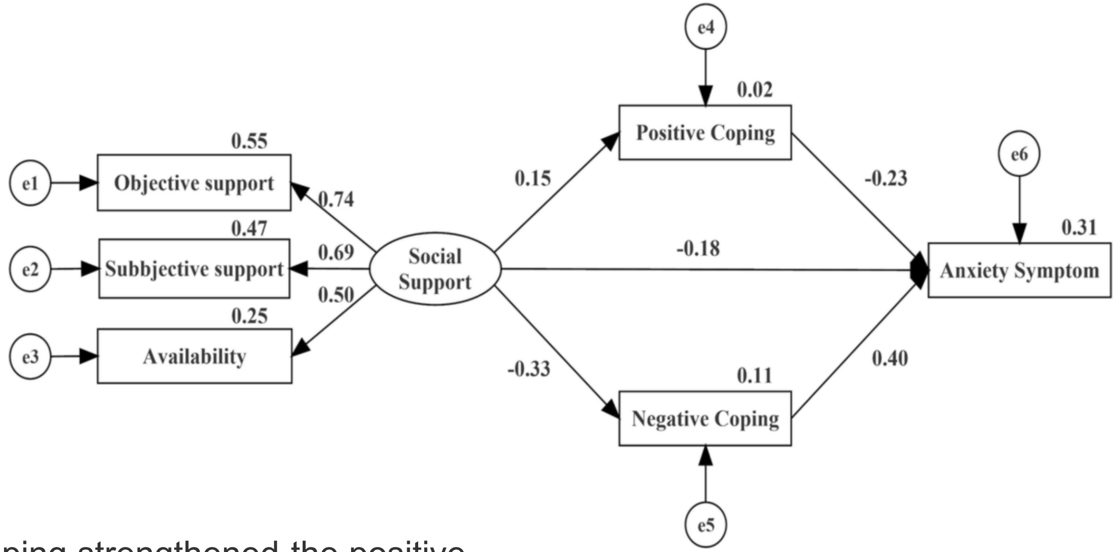
=> full mediation

https://nmmichalak.github.io/nicholas_michalak/blog_entries/2018/nrg01/nrg01.html

Real example

Sufficient social support can reduce anxiety directly.

Meanwhile, social support affected anxiety through positive coping and negative coping paths indirectly.



“We demonstrated that positive coping strengthened the positive effect of social support on anxiety”;

«negative coping negatively influenced the effect of social support on anxiety. When individuals received sufficient social support, they were more likely to use positive coping strategies to achieve a lower level of anxiety. If there was not enough social support, individuals tended to take negative actions facing problematic life events, resulting in higher anxiety.»

Zhu, W., Wei, Y., Meng, X., & Li, J. (2020). The mediation effects of coping style on the relationship between social support and anxiety in Chinese medical staff during COVID-19. *BMC Health Services Research*, *20*, 1-7.

Text in rus: Гордеева, Т. О., & Осин, Е. Н. (2010). Позитивное мышление как фактор учебных достижений старшеклассников. *Вопросы психологии*, *1*, 24-33.

Mediation summary

- Goal of mediation analysis is to estimate **total effect, direct and indirect effects**, and open “black box”
- Mediation model usually requires **some hypotheses about indirect effects**
- Opportunity to get new results using **more complex modelling**

Exercises

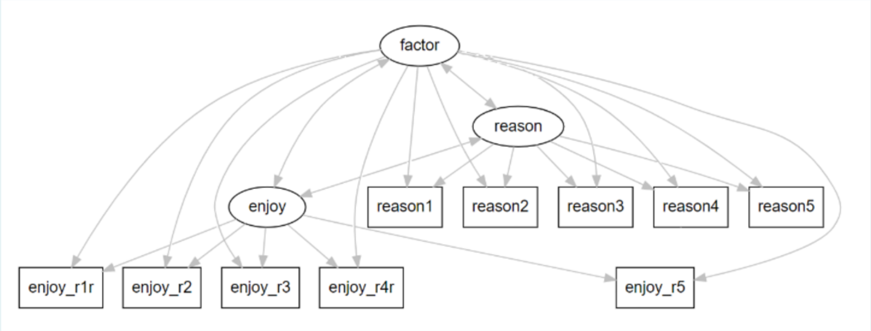
Exercise 1

1) Choose the model that is identified.

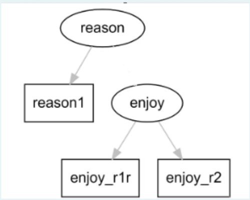
2) How many items should be in one factor (the min number)?

3) What is the name of the Model c?

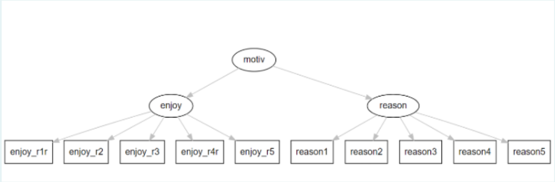
a.



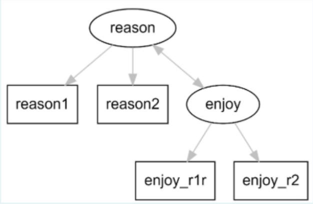
b.



c.



d.



Exercise 2

You have created a CFA model based on a dataset with several dichotomous (1/0) items. Since data is categorical, you used the **WLSMV estimator.**

You get the following results about the model fit. In your answer, write values of most popular fit indexes and compare them to recommended cut-offs, interpret the results. Make an overall conclusion about the model fit.

```
## Model Test User Model:
##                                               Standard      Robust
##   Test Statistic                              1495.744    1540.553
##   Degrees of freedom                               252         252
##   P-value (Chi-square)                           0.000       0.000
##   Scaling correction factor                                  1.003
##   Shift parameter                                           49.957
##     simple second-order correction
##
## Model Test Baseline Model:
##
##   Test statistic                             15374.629   10642.046
##   Degrees of freedom                               276         276
##   P-value                                        0.000       0.000
##   Scaling correction factor                                  1.457
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.918       0.876
##   Tucker-Lewis Index (TLI)                       0.910       0.864
##
##   Robust Comparative Fit Index (CFI)                            NA
##   Robust Tucker-Lewis Index (TLI)                               NA
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.055       0.056
##   90 Percent confidence interval - lower         0.053       0.054
##   90 Percent confidence interval - upper         0.058       0.059
##   P-value RMSEA <= 0.05                          0.001       0.000
##
##   Robust RMSEA                                                  NA
##   90 Percent confidence interval - lower                        NA
##   90 Percent confidence interval - upper                        NA
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.086       0.086
##
```

Exercise 3

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  CM3 =~
            Item 1    1.000                               0.409    0.409
            Item 2    1.067    0.146    7.318    0.000    0.436    0.436
            Item 3    0.929    0.143    6.503    0.000    0.380    0.380
            Item 4    1.049    0.143    7.345    0.000    0.428    0.428
            Item 5   -1.362    0.161   -8.462    0.000   -0.556   -0.556
            Item 6    1.021    0.141    7.267    0.000    0.417    0.417
            Item 7    0.391    0.119    3.284    0.001    0.160    0.160
            Item 8    1.439    0.166    8.661    0.000    0.588    0.588
            Item 9    1.492    0.169    8.815    0.000    0.609    0.609
           Item 10    1.003    0.155    6.468    0.000    0.410    0.410
           Item 11   -0.047    0.119   -0.397    0.692   -0.019   -0.019
           Item 12    1.405    0.160    8.808    0.000    0.574    0.574
           Item 13    1.973    0.215    9.163    0.000    0.806    0.806
```

Exercises 4

1) What does error covariance between two items mean? Give at least one reason why there might be error covariances between two items in the test.

2) Describe the situation where you need to apply bi-factor model.

Exercise 5

What is the difference between EFA and CFA?

Reporting CFA

How to report CFA results?

In methodology section: description of method

In this study we used **Confirmatory factor analysis**… to check the internal structure of the instrument

We used ML / MLR / WLSMV **estimator** because we used interval data that can be considered normal / interval non-normal data / categorical data

We use the following **fit indexes** to assess the quality of the model: CFI > 0.95, TLI > 0.95, RMSEA < 0.05, SRMR < 0.05 *(usually we put reference here)*

We calculate **reliability estimates** of factors using the Omega coefficient (McDonald, 1999). Values above 0.7 are satisfactory *(usually we put reference here)*

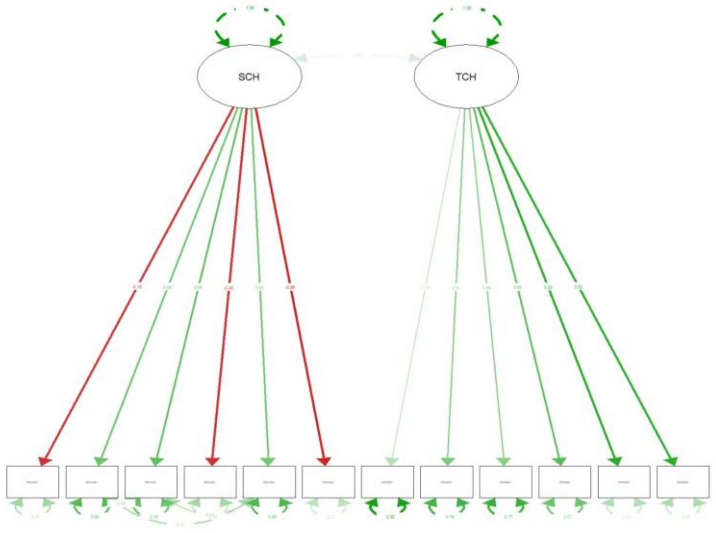
The **analysis was conducted** in soft R, package ‘lavaan’

Results section

- Model description (can be earlier)

According the theory, **construct motivation consists of two correlated factors** - external and internal motivation (or you can describe other structure: second-order, bi-factor here). Some error covariances were added to the model to address the issue of local item dependence due to close wording **(Picture 1 - usually show structure)**

Next - report model fit of the final model



Results section

- *Model fit*

The model with one factor demonstrated acceptable *(or poor)* fit ($\chi 2$ (df = 9) = 181.43, p = 0.00; CFI = 0.974, TLI = 0.956, RMSEA = 0.066 (90% CI 0.058; 0.075), SRMR = 0.031)

For ML estimator!

```
Model Test User Model:

  Test statistic                              181.433
  Degrees of freedom                                9
  P-value (Chi-square)                          0.000

Model Test Baseline Model:

  Test statistic                             6600.532
  Degrees of freedom                               15
  P-value                                       0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                   0.974
  Tucker-Lewis Index (TLI)                      0.956

Root Mean Square Error of Approximation:

  RMSEA                                         0.066
  90 Percent confidence interval - lower        0.058
  90 Percent confidence interval - upper        0.075
  P-value H_0: RMSEA <= 0.050                   0.001
  P-value H_0: RMSEA >= 0.080                   0.004

Standardized Root Mean Square Residual:

  SRMR                                          0.031
```

The model with one factor demonstrated acceptable *(or poor)* fit ($\chi 2$ (df = 9) = 183.813, p = 0.00; CFI = 0.955, TLI = 0.925, RMSEA = 0.067 (90% CI 0.059; 0.075), SRMR = 0.031)

```
Root Mean Square Error of Approximation:

  RMSEA                                          0.042       0.067
  90 Percent confidence interval - lower         0.034       0.059
  90 Percent confidence interval - upper         0.051       0.075
  P-value H_0: RMSEA <= 0.050                    0.929       0.000
  P-value H_0: RMSEA >= 0.080                    0.000       0.006

  Robust RMSEA                                               0.044
  90 Percent confidence interval - lower                     0.038
  90 Percent confidence interval - upper                     0.049
  P-value H_0: Robust RMSEA <= 0.050                         0.967
  P-value H_0: Robust RMSEA >= 0.080                         0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.031       0.031
```

```
Model Test User Model:
                                              Standard      Scaled
  Test Statistic                                78.293     183.813
  Degrees of freedom                                 9           9
  P-value (Chi-square)                           0.000       0.000
  Scaling correction factor                                  0.428
  Shift parameter                                            0.704
    simple second-order correction

Model Test Baseline Model:

  Test statistic                              6475.341    3916.099
  Degrees of freedom                                15          15
  P-value                                        0.000       0.000
  Scaling correction factor                                  1.656

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.989       0.955
  Tucker-Lewis Index (TLI)                       0.982       0.925

  Robust Comparative Fit Index (CFI)                         0.988
  Robust Tucker-Lewis Index (TLI)                            0.981
```

Example of WLSMV estimator

Results section

- *Factor loadings.* Report standardized loadings + significance value

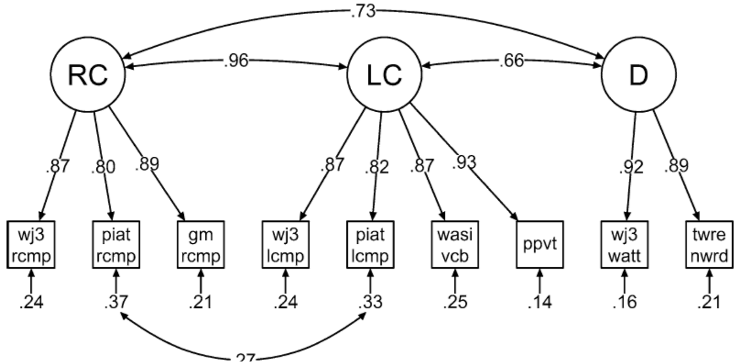
Two options:

➔ Report factor loadings via picture
➔ Report factor loadings in table

Also, give a brief description:

All factor loadings were significant ($p < 0.01$) with mean size of standardized factor loadings = 0.4 (sd = 0.1).

- Correlation between factors
- Reliability for all factors




```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  enjoy =~
    enjoy_r1r         0.341    0.018   18.638    0.000    0.341    0.294
    enjoy_r2          0.492    0.017   29.745    0.000    0.492    0.488
    enjoy_r3          0.552    0.016   34.800    0.000    0.552    0.626
    enjoy_r4r         0.613    0.016   38.664    0.000    0.613    0.661
    enjoy_r5          0.669    0.014   49.336    0.000    0.669    0.726
    enjoy_r6          0.694    0.014   50.928    0.000    0.694    0.807

Variances:
```

Item	Std. factor loading
Item1	0.294**
Item2	0.488**
**p<0.01; * p<0,05	

Results section

- *Model comparison*

Table 5
Chi-square test for the change in fit between nested models.

Nested models	χ^2 difference	df difference	p
Model 1. Six first-order uncorrelated factors	–	–	–
Model 2. Six first-order factors, and one second-order general factor	287.743	6	.001
Model 3. Six first-order factors, and two correlated second-order general factors	4.869	1	.04
Model 4. Six first-order correlated factors	19.788	8	.02

Psychometrics & Statistics II

Structural Equation Modelling. Part 2

Daria Gracheva, Sergei Tarasov

Moderation & Mediation models

The relationship between X and Y is influenced by a **third variable** (moderator Z or mediator M).

"**when**" or "**for whom**" X most strongly predicts Y

Moderator

Moderator is a variable that affects the direction and/or strength of the relation between an independent variable and a dependent variable

why X can predict Y

Mediator

A given variable may be said to function as a **mediator** to the extent that it accounts for the relation between X and Y

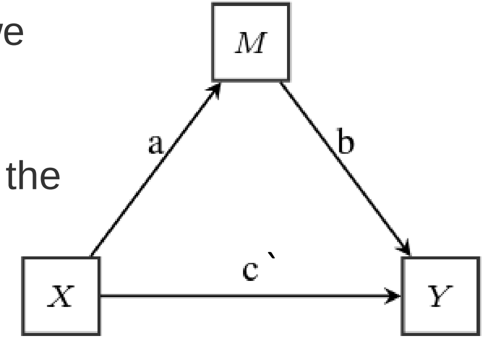
Hypothesis: **Home environment is a mediator** in the relation between mothers' education and children's mathematical achievement

c' - effect of X on Y
a - effect of X on M
b - effect of M on Y

$$M = a_0 + a \times X + e_M$$
$$Y = b_0 + b \times M + c' \times X + e_Y$$

To test these specifically, we need to define new parameters in the lavaan model that are a product of the individual paths.

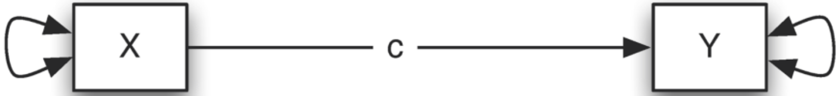
"Open the black box and understand the mechanism"



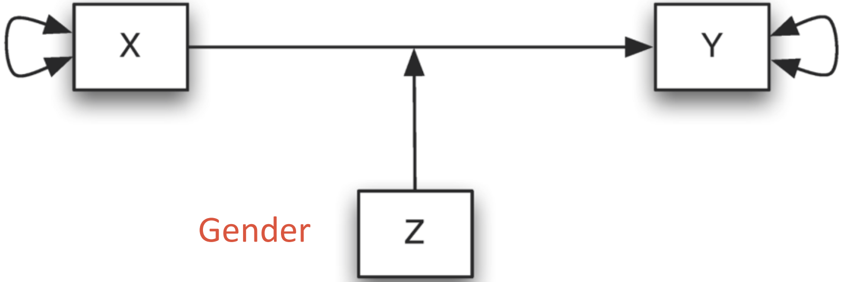
$$\text{Direct effect} = c'$$
$$\text{Indirect effect} = b \times a$$
$$\text{Total effect} = c' + b \times a$$

```
myModel <- '
             Y ~ b*M + c*X
             M ~ a*X

             indirect := a*b
             total    := c + (a*b)
           '

fit <- sem(model = myModel,
           data  = myData,
           se    = "bootstrap")

summary(fit)
```

Moderation



number of training sessions X predicts math test performance Y



For instance, if male students (Z=0) benefit more (or less) from training than female students (Z=1), then gender can be considered as a moderator.

A **moderator variable Z** is a variable that alters the strength of the relationship between X and Y. In other words, the effect of X on Y depends on the levels of the moderator Z.

Questions involving moderators address "**when**" or "**for whom**" a variable most strongly predicts or causes an outcome variable.

In statistics, moderation is expressed as **interaction** between predictor X and third variable Z.

$$Y_i = \text{int} + a_1 X_i + a_2 Z_i + \boxed{a_3 X_i Z_i} + \varepsilon_i .$$

Interaction term with binary variable

Since interaction term is significant (xz) there is a significant **moderation effect**

Relationship between X and Y depends on gender (Z)

In model with interaction we do not have "main" effects any more. Coef training - effect of training on math (Y) when another var (gender) in interaction = 0

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	4.98999	0.27499	18.146	< 2e-16	***
training	-0.33943	0.05387	-6.301	8.70e-09	***
gender	-2.75688	0.37912	-7.272	9.14e-11	***
xz	0.50427	0.06845	7.367	5.80e-11	***

When Z=0 (males), the estimated effect of training intensity on math performance is −0.34.
When Z=1 (female students), the estimated effect is -0.34 + 0.5 = 0.16.
The moderation analysis tells us that the effects of training intensity on math performance for males (-.34) and females (.16) are significantly different for this example.

https://statisticsbyjim.com/regression/interaction-effects/

Interaction term with continuous data

We need to specify **zero on continuous variables to interpret interaction effects**. It is recommended to center variables (0 = mean of the variable)

For instance, mcx denotes the change in Y by a 1 unit change in X for individuals **with average Z** (as opposed to Z = 0 in the uncentered case).

But we can calculate **slope of X for different levels of Z (+/- std Z)**

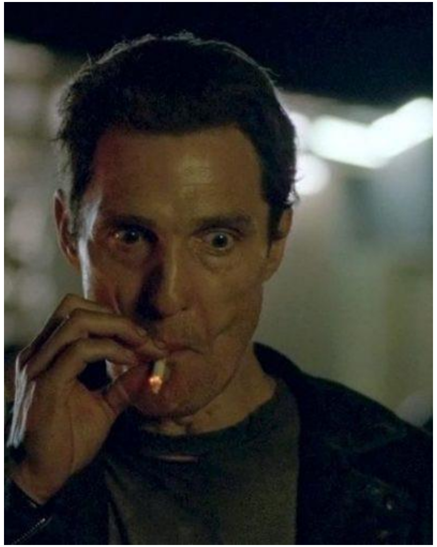
```
##                Estimate Std. Error  t value Pr(>|t|)
## (Intercept)     3.7097     0.0227 163.0954   0.0000
## mcx             0.5003     0.0241  20.7653   0.0000
## mcz            -0.1790     0.0257  -6.9765   0.0000
## mcx:mcz        -0.0663     0.0236  -2.8094   0.0051
```

```
##                INT  Slope     SE    LCL    UCL
## at zHigh 3.5492 0.4407 0.0313 0.3793 0.5022
## at zMean 3.7097 0.5003 0.0241 0.4530 0.5475
## at zLow  3.8703 0.5598 0.0328 0.4953 0.6242
```

If Z = 0, Slope = 0.5003
If Z = 1, Slope = 0.5003 + 1* (-0.0663)
If Z = -1, Slope = 0.5003 + (-1) * (-0.0663) = 0.5666

But what if we use **latent factors** in interaction?

Be ready for that.. Not so simple

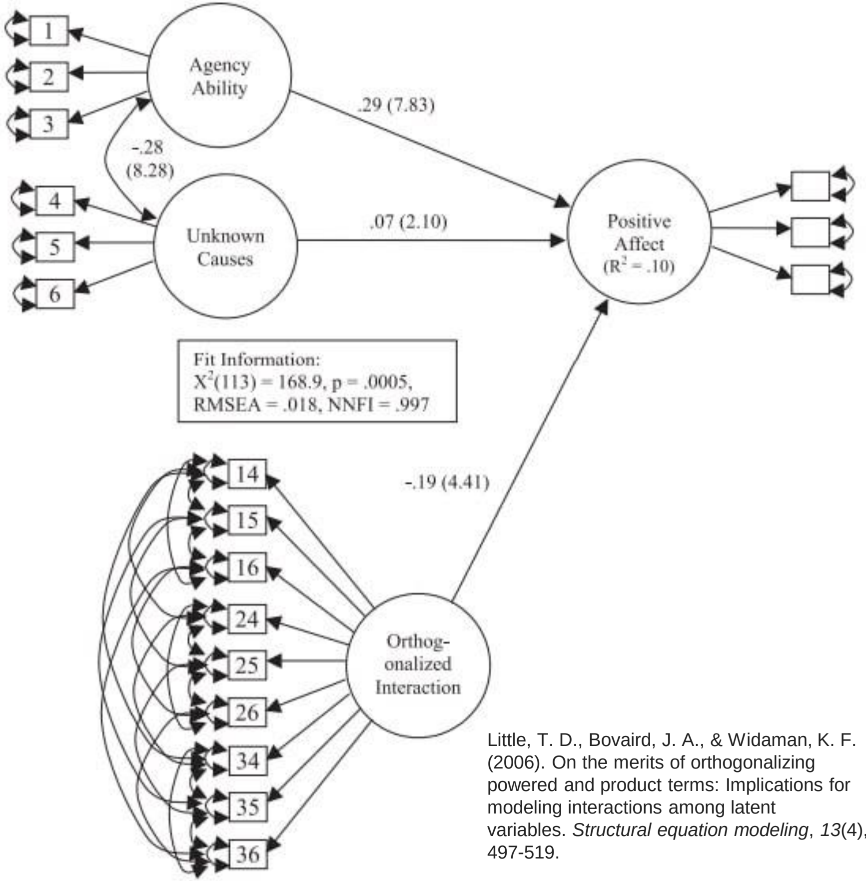


Interaction between latent variables

We want to estimate the interaction between latent factors = agency ability and unknown cases.

We create an interaction factor which is a product of all observed variables of two factors.

Also add error covariances in the interaction factor between the same items.



Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural equation modeling*, *13*(4), 497-519.

FIGURE 2 A latent variable interaction using orthogonalized indicators. *Note*. The *t* value is listed in the parentheses next to the standardized estimate.

How to do it in R

Function indProd in SemTools package

Product of all observed variables of two factors

```
mydata_dmc <- indProd (mydata , var1 = c("IV1",
"IV2", "IV3"),
var2 = c("MOD1", "MOD2", "MOD3"),
match = FALSE , meanC = TRUE ,
residualC = FALSE , doubleMC = TRUE)
```

https://regorz-statistik.de/en/lavaan_sem_latent_interaction.html

```
int_model <- '
# Loadings
iv =~ IV1 + IV2 + IV3
mod =~ MOD1 + MOD2 + MOD3
dv =~ DV1 + DV2 + DV3
int =~ IV1.MOD1 + IV1.MOD2 + IV1.MOD3 + IV2.MOD1
+ IV2.MOD2 +
IV2.MOD3 + IV3.MOD1 + IV3.MOD2 + IV3.MOD3

# Regression
dv ~ iv + mod + int

# Error covariances
IV1.MOD1 ~~ IV1.MOD2 + IV1.MOD3 + IV2.MOD1 +
IV3.MOD1
IV1.MOD2 ~~ IV1.MOD3 + IV2.MOD2 + IV3.MOD2
IV1.MOD3 ~~ IV2.MOD3 + IV3.MOD3
IV2.MOD1 ~~ IV2.MOD2 + IV2.MOD3 + IV3.MOD1
IV2.MOD2 ~~ IV2.MOD3 + IV3.MOD2
IV2.MOD3 ~~ IV3.MOD3
IV3.MOD1 ~~ IV3.MOD2 + IV3.MOD3
IV3.MOD2 ~~ IV3.MOD3
'

int_fit <- sem(int_model, data = mydata_dmc,
estimator="MLM")
```