
Final Exam (*Teacher Version*)

Measurement Models Course (Spring 2021)

MAE4101 Measurement Models

Centre for Educational Measurement at the University of Oslo (CEMO)

Release date: 11 June 2021, 1:00 PM, Oslo time

Submission deadline: 18 June 2021, 3:00 PM, Oslo time

Welcome to the MAE4101 Measurement Models home exam!

This exam covers the structural *and* the measurement model part of the course.

Before you begin, please make sure to consider the following:

- **Read** the questions carefully.
- Notice which **task operators** are used (e.g., *name* something vs. *explain/derive* something).
- You may **simplify subscripts** or Greek symbols wherever appropriate (e.g., Y1 instead of Y_1 , lambda1 instead of λ_1).
- Keep your explanations and descriptions **to the point**.
- **Partial credits** will be given.
- Starting with this home exam, you declare that you will work on the tasks **without any help of others**.
- You will be presented with **three tasks**, each of which is accompanied by a data set. Download the data sets here: <https://cemouio.shinyapps.io/mae4101/>.
- **Scheduling:** We recommend spending at least one day on each of the tasks.
- Submit the **responses to the tasks in the Inpera system**.
- Submit the **R code files for each of the three main tasks**. Create one R code file for each of the three tasks and label them "PathModels.R", "FactorAnalysis.R", and "SEM.R".
- **Help!** If needed, please contact us via email or the exam forum on the course Canvas page, and we shall get back quickly.

We wish you all the best for the exam and great success in working on the tasks!

Best regards,

Denise, Jelena, Kseniia, Jarl, and Ronny

Name:	SUGGESTED SOLUTIONS + GRADING
-------	--------------------------------------

Results

Task	Credits	Max. credits	Grader(s)
Path Models		40	Ronny
Factor Analysis: Conspiracy Theory Scale (CTS)		40	Denise
Structural Equation Models		40	Kseniia & Jarl
TOTAL:		120	

Expected time on task

- Estimated time: 1.5 days per task + 1 day for inserting the responses in the Inopera system → 3 tasks*1.5 days + 1 day = **5.5 days**
- Time provided: **7 days**

Grading

Grade	Credits range	% Correct threshold
A	108–120	90 %
B	96–107	80 %
C	84–95	70 %
D	72–83	60 %
E	60–71	50 %
F	0–60	< 50 %

Data Sets

Using your personal access code, please download the three datasets for this exam from the following page: <https://cemouio.shinyapps.io/mae4101/>

Question sets

Before entering your responses in the Inpera system, you may save your responses in a Word file. If you wish to use a file including the tasks and response forms, you may use this one: **< Word file >**.

Declaration

Before starting the exam, confirm the following:

Declaration	Yes, I confirm.
I hereby confirm that I will work on this home exam without the help of others.	<input type="checkbox"/>
I will not exchange any responses, material, or ideas with the other examinees during the exam.	<input type="checkbox"/>
I confirm that I am the sole author of the responses provided in this exam.	<input type="checkbox"/>
I am aware of the possible consequences any form of cheating during this exam may have (https://www.uio.no/english/studies/examinations/cheating/).	<input type="checkbox"/>

Data Check

Check the **PathModel data set** you have downloaded by indicating its sample size and the mean of the variable X_1 below.

(a) Which **sample size** does your PathModel data set have?

500 625 750

(b) What is the **mean of the variable X_1** in the PathModel data set?

-0.045 0.033 0.038

NOTES:

The two data check questions are used to identify which data batch students have used.

- Question (a): 500 → Batch A, 625 → Batch C, 750 → Batch B
- Question (b): -0.045 → Batch A, 0.033 → Batch C, 0.038 → Batch B

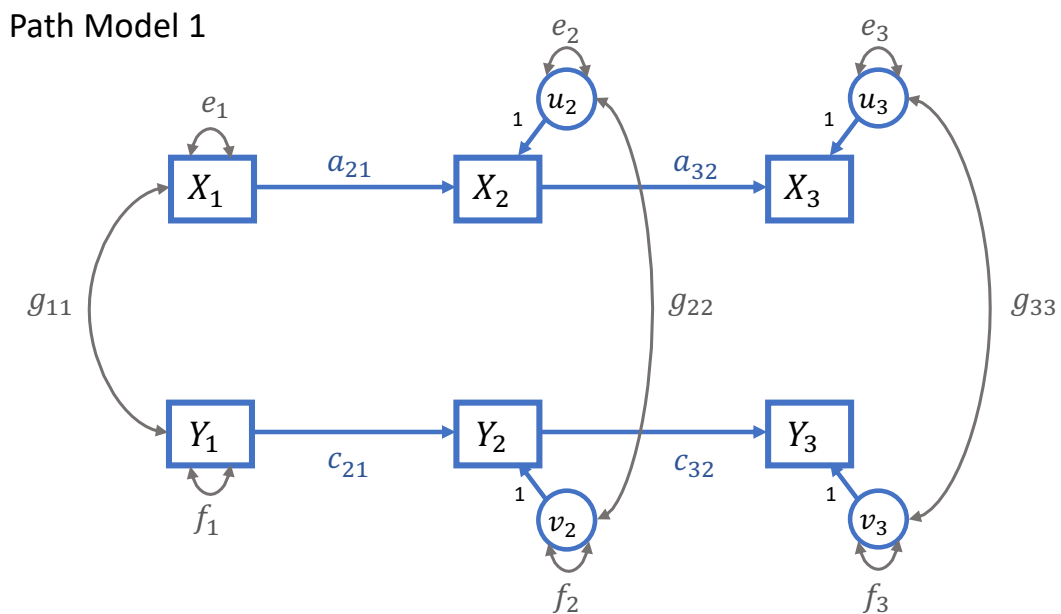
The uploaded R files will be used to check the responses to the data-analytic tasks. In case students may have specified or estimated models or model parameters incorrectly in one tasks, subsequent tasks might be affected. To avoid double taxing on errors, follow-up credits can be given if the responses are consistent with the even erroneous inputs.

Path Models

A group of researchers followed university students from their first year until their last year of a Bachelor program. At three time points, these students reported their interest in the subject they were studying and their life satisfaction. The variables indicating students' **interest** are labelled X_1 , X_2 , and X_3 , and the variables indicating students' **life satisfaction** are labelled Y_1 , Y_2 , and Y_3 .

The researchers hypothesized that the interest and life satisfaction variables are related as shown in **Path Model 1**.

Note: This model does not include a mean structure.



(P1) Specify **Path Model 1**, providing the model equations for the dependent variables and the covariance matrix of the residuals. *Note: For the latter, order the variables as follows: $X_1, Y_1, X_2, Y_2, X_3, Y_3$. You may write subscripts in normal letters.*

Fill in the model equations here.

Dependent variables	Model equation
$X_2 =$	$a_{21}X_1 + u_2$
$X_3 =$	$a_{32}X_2 + u_3$
$Y_2 =$	$c_{21}Y_1 + v_2$
$Y_3 =$	$c_{32}Y_2 + v_3$

SCORING:

- 1 credit per correct equation
- **Total: 4 credits**

Covariance matrix of the residuals:
$\begin{bmatrix} e_2 & g_{22} & 0 & 0 \\ g_{22} & f_2 & 0 & 0 \\ 0 & 0 & e_3 & g_{33} \\ 0 & 0 & g_{33} & f_3 \end{bmatrix}$

SCORING:

- 1 credit for the correct zeros in the covariance matrix
- 1 credit for the correct non-zero entries
- **Total: 2 credits**

(P2) Using the information about the model specification, **derive** the following two elements of the model-implied covariance matrix underlying **Path Model 1**: $Var(X_2)$ and $Cov(X_1, Y_2)$.

Fill in the equations here.

Independent variables	Model equation
$Var(X_2) =$	$\begin{aligned} &Var(a_{21}X_1 + u_2) \\ &= Var(a_{21}X_1) + Var(u_2) + 2Cov(a_{21}X_1, u_2) \\ &= a_{21}^2Var(X_1) + Var(u_2) + 2a_{21}Cov(X_1, u_2) \\ &= a_{21}^2Var(X_1) + Var(u_2) + 0 \\ &= a_{21}^2e_1 + e_2 \end{aligned}$
$Cov(X_1, Y_2) =$	$\begin{aligned} &Cov(X_1, c_{21}Y_1 + v_2) \\ &= Cov(X_1, c_{21}Y_1) + Cov(X_1, v_2) \\ &= c_{21}Cov(X_1, Y_1) + 0 \\ &= c_{21}g_{11} \end{aligned}$

SCORING:

- 1 credit for the correct result for each of the two elements
- 1 credit for the correct derivation for each of the two elements
- *Note: Deriving the end results is not necessary.*
- **Total: 4 credits**

(P3) Specify and estimate **Path Model 1 in lavaan**, using the **PathModel data set** and maximum-likelihood estimation. Provide the R code for the model specification and the estimation.

Note: You may read the data file using the read.csv2() function in R. Decimals are indicated by commas, and the data file contains the variable names in the first row.

Example: read.csv2(file = "PathModel.csv", header = TRUE, dec = ",")

Model specification in lavaan:

```
PathModel1 <- ' # Structural model
                X2 ~ a21*X1
                X3 ~ a32*X2
                Y2 ~ c21*Y1
                Y3 ~ c32*Y2

                # Covariances and variances
                X1 ~~ X1
                Y1 ~~ Y1
                X2 ~~ X2
                X3 ~~ X3
                Y2 ~~ Y2
                Y3 ~~ Y3
                X1 ~~ Y1
                X2 ~~ Y2
                X3 ~~ Y3 '
```

SCORING:

- 1 credit for each correct line in the structural model (in bold; sub-total: 4)
- 1 credit for the covariances between exogenous and residual variables (in bold; sub-total: 1)
- *Note: The model specification elements in bold are minimally needed to gain full credits.*
- **Total: 5 credits**

Model estimation in lavaan:

```
PathModel1.fit <- sem(PathModel1, data = dat, meanstructure = FALSE, estimator = "ML")
```

SCORING:

- 1 credit for choosing an appropriate function and for the correct setup (i.e., all options are included)
- *Note: The options estimator="ML" and meanstructure=FALSE can be left out.*
- **Total: 1 credit**

(P4) Request a **model summary** which contains the standardized model parameters, the variance explanations of all dependent variables (R^2), and the model fit indices. Provide the R code for calling this model summary.

Model summary in lavaan:

```
summary(PathModel1.fit, standardized = TRUE, rsquare = TRUE, fit.measures = TRUE)
```

SCORING:

- 0.5 credit for each of the options in the summary function (total: 4 options)
- **Total: 2 credits**

(P5) The chi-square test statistic of Path Model 1 is accompanied by **8 degrees of freedom**. Explain why there are 8 degrees of freedom in this model.

Model identification:

- The six observed variables provide $6+5+4+3+2+1 = 21$ pieces of information (these are: 6 variances and $5+4+3+2+1$ covariances) $\rightarrow p = 21$
- Path Model 1 contains 13 model parameters (i.e., 4 path coefficients + 3 covariances + 6 variances) $\rightarrow q = 13$
- The difference between these two elements gives the degrees of freedom.

SCORING:

- 1 credit for the correct p
- 1 credit for the correct q
- **Total: 2 credits**

(P6) Determine whether **Path Model 1** exhibits a **reasonably good fit** (Hu & Bentler, 1999), focusing on the chi-square statistic, CFI, RMSEA, and SRMR. Provide the values of these indices with three digits after the comma.

Fit index	PathModel-A	PathModel-B	PathModel-C
Chi-square statistic	188.737	192.794	152.648
CFI	0.868	0.885	0.901
RMSEA	0.213	0.175	0.170
SRMR	0.192	0.149	0.156

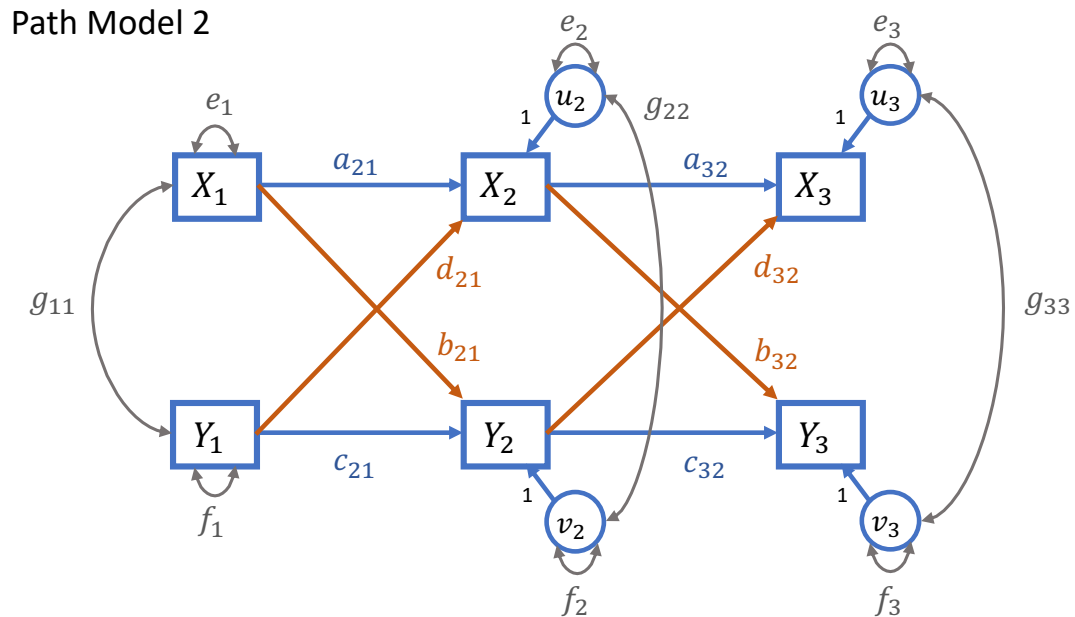
Model fit evaluation:

For all data sets, Path Model 1 does not fit the data well. Criteria for an acceptable fit (e.g., Hu & Bentler, 1999): insignificant chi-square statistic ($p > .05$), CFI greater than or equal to .95, RMSEA smaller than or equal to 0.06, SRMR smaller than or equal to 0.08. See the details below.

SCORING:

- 1 credit for each correct fit index and interpretation (chi-square statistic, CFI, RMSEA, SRMR; sub-total: 4)
- 1 credit for the overall conclusion
- **Total: 5 credits**

Having evaluated the fit of Path Model 1, the researchers decided to revise their model. The revised model is labelled **Path Model 2** and is shown below.



Path Model 2 contains so-called “cross-lagged paths” (i.e., b - and d -paths). These paths connect the interest and life satisfaction variables at different time points.

(P7) Specify and estimate **Path Model 2 in lavaan**, using the same data and estimation procedure as for Path Model 1. Provide the R code for the **model specification**. Compute the **indirect effects** for $X_1 \rightarrow X_2 \rightarrow X_3$ and $Y_1 \rightarrow Y_2 \rightarrow Y_3$.

Model specification in lavaan:

```
PathModel2 <- '
# Structural model
X2 ~ a21*X1 + d21*Y1
X3 ~ a32*X2 + d32*Y2
Y2 ~ c21*Y1 + b21*X1
Y3 ~ c32*Y2 + b32*X2

# Covariances and variances
X1 ~~ X1
Y1 ~~ Y1
X2 ~~ X2'
```



```

X3 ~~ X3
Y2 ~~ Y2
Y3 ~~ Y3
X1 ~~ Y1
X2 ~~ Y2
X3 ~~ Y3

# Indirect effects
indX := a21*a32
indY := c21*c32

```

SCORING:

- 1 credit for each of the correct lines in the structural model (in bold; sub-total: 4)
- 1 credit for all the covariances between exogenous and residual variables (in bold; sub-total: 1)
- 1 credit per correct indirect effect (sub-total: 2)
- *Note: The model specification elements in bold are minimally needed.*
- **Total: 7 credits**

(P8) Provide and describe the evidence supporting the **reasonably good fit of Path Model 2**, focusing on the chi-square statistic, CFI, RMSEA, and SRMR. Provide the values of these indices with three digits after the comma.

Fit index	PathModel-A	PathModel-B	PathModel-C
Chi-square statistic	1.688	4.162	1.870
CFI	1.000	1.000	1.000
RMSEA	0.000	0.007	0.000
SRMR	0.007	0.010	0.009

Model fit evaluation:
For all data sets, Path Model 2 fits the data well. Criteria for an acceptable fit (e.g., Hu & Bentler, 1999): insignificant chi-square statistic ($p > .05$), CFI greater than or equal to .95, RMSEA smaller than or equal to 0.06, SRMR smaller than or equal to 0.08. See the details below.

SCORING:

- 1 credit for each correct fit index and their interpretation (chi-square statistic, CFI, RMSEA, SRMR; sub-total: 4)
- 1 credit for the overall conclusion
- **Total: 5 credits**

(P9) Provide the unstandardized estimates and their standard errors of the **two indirect effects** with three digits after the comma and interpret them.

Parameters	PathModel-A	PathModel-B	PathModel-C
Indirect effect (indX) $X_1 \rightarrow X_2 \rightarrow X_3$	0.332	0.415	0.192
SE	0.035	0.032	0.027
Indirect effect (indY) $Y_1 \rightarrow Y_2 \rightarrow Y_3$	0.660	0.354	0.426
SE	0.050	0.031	0.038

Indirect effects:

Both indirect effects are positive, and there is evidence that they differ from zero. See the estimates and the confidence intervals below. This observation indicates that the relation between interest at t1 and interest at t3 is at least partially mediated via interest at t2. The same applies to life satisfaction.

SCORING:

- 0.5 credit for each correct parameter estimate and standard error (sub-total: 2 credits)
- 1 credit for the correct interpretation
- **Total: 3 credits**

File upload

Please upload the R code file for the path model task and label it " PathModels.R ".	File upload (max. 2 GB):	
--	-----------------------------	--

Conspiracy Theory Scale (CTS)

Exploratory Factor Analysis (EFA)

A group of researchers wants to evaluate the dimensionality of the 15-item “Conspiracy Theory Scale” (CTS) for a sample of $N = 2400$ students. Half of the students were sampled from Statistics and half from Psychology courses.

The following table describes the 15 items (variables) in the dataset (**Q1-Q15**). Participants rated these items on a 5-point Likert-type scale: *1 = Definitely not true, 2 = Probably not true, 3 = Not sure/cannot decide, 4 = Probably true, 5 = Definitely true.*

Variable	Description
Q1	The government is involved in the murder of innocent citizens and/or well-known public figures, and keeps this a secret.
Q2	The power held by heads of state is second to that of small unknown groups who really control world politics.
Q3	Secret organizations communicate with extraterrestrials, but keep this fact from the public.
Q4	The spread of certain viruses and/or diseases is the result of the deliberate, concealed efforts of some organization.
Q5	Groups of scientists manipulate, fabricate, or suppress evidence in order to deceive the public.
Q6	The government permits or perpetrates acts of terrorism on its own soil, disguising its involvement.
Q7	A small, secret group of people is responsible for making all major decisions, such as going to war.
Q8	Evidence of alien contact is being concealed from the public.
Q9	Technology with mind-control capacities is used on people without their knowledge.
Q10	New and advanced technology which would harm current industry is being suppressed.
Q11	The government uses people as patsies to hide its involvement in criminal activity.
Q12	Certain significant events have been the result of the activity of a small group who secretly manipulate world events.
Q13	Some UFO sightings and rumors are planned or staged in order to distract the public from real alien contact.
Q14	Experiments involving new drugs or technologies are routinely carried out on the public without their knowledge or consent.
Q15	A lot of important information is deliberately concealed from the public out of self-interest.

The **FactorAnalysis data set** contains the raw data of these variables. In the following, you will be asked to perform analyses with these variables. Please make sure to perform these analyses and submit the respective code.

(F1) **Correlation:** Obtain the Pearson correlation between variables Q3 and Q8 with two digits after the comma. Interpret this value and indicate the assumptions behind this correlation (i.e., level of measurement, type of relationship).

Measure	Value	Interpretation
Correlation between Q3 and Q8	0.79	The variables Q3 and Q8 are positive related, suggesting that there is a tendency for student who score high in "Q3" will also score high in "Q8".
Assumptions		
Linear relationship between two continuous variables.		

SCORING:

- 1 credit per correct answer/plausible interpretation/explanation
- **Total: 3 credits**

(F2) **Factor selection:** Using the **principal axis factor method**, how many factors can be extracted from the variables Q1-Q15? Apply the Kaiser-Guttman rule and parallel analysis, provide the respective eigenvalues for both methods, and explain your decision for the number of factors.

Method	Number of factors	Eigenvalues	Explanation
Kaiser-Guttman rule	2	7.844 and 1.460	Two factors are suggested by this method since only two eigenvalues from the unreduced solution have value greater than 1.
Parallel analysis	5	eigenvalues of factors from the original data: [1] 7.35 0.91 0.44 0.14 0.08 and eigenvalues of simulated factors (example): [1] 0.23 0.11 0.09 0.07 0.06	Five factors are suggested by this method since five eigenvalues from the original data have value greater than the eigenvalues from the simulated data.

SCORING:

- 1 credit per correct answer/plausible explanation
- **Total: 6 credits**

(F3) **Factor rotation:** A researcher wants to compare the results of the **maximum-likelihood factor extraction** for three factors of the CTS without rotation to the extraction with the Promax rotation.

Using the Promax and no rotation in two separate analyses, provide the following estimates and their interpretations. Explain briefly why one would choose the results for the rotated solution.

Element	Without rotation	With Promax rotation	Interpretation
Factor loading of Q9 for the first factor	0.69	0.34 or 0.36 ("Promax" or "promax" in fa())	It represents the strength of the relationship between the factor and the observed measure.
Communality of Q12	0.82	0.82	82 % of the variance in "Q12" is explained by the three factors
Residual variance of Q10	0.66	0.66	66 % of the variance in "Q12" is not explained by the three factors
Explanation for the use of the rotated solution			
Since none of the solutions reached a simple structure, one may choose the rotated one if the assumption that the factors are intercorrelated is plausible for the study.			

SCORING:

- 1 credit per correct answer/plausible interpretation/explanation
- **Total: 10 credits**

Confirmatory Factor Analysis (CFA)

(F4) **Model specification and estimation:** Using the CTS data, run a CFA model with maximum likelihood estimation with two factors. This model is named “**CFA1**” and contains two factors: “GOV” and “UFO”. GOV is measured by the variables Q1, Q6, and Q11. UFO is measured by the variables Q3, Q8, and Q13. This model does not include a mean structure.

Provide the R code for the specification and the estimation of model CFA1 in lavaan.

Model specification in lavaan:

```
CFA1<- 'GOV=~Q1+Q6+Q11
        UFO=~Q3+Q8+Q13'
```

SCORING:

- 1 credit for each of the correct lines
- **Total: 2 credits**

Model estimation in lavaan:

```
cfa1.fit <- cfa(CFA1, data = CTS, estimator = "ML")
```

SCORING:

- 1 credit for choosing the appropriate function for model estimation (cfa, sem, or lavaan)
- 1 credit for the correct setup
- *Note: The option estimator="ML" can be left out.*
- **Total: 2 credits**

(F5) **Model identification:** A researcher claims that model **CFA1** is **over-identified**. Test his or her claim, indicating all the elements that support your conclusion (i.e., number of available pieces of information, model parameters, degrees of freedom). *Note: Assume that the factor variances are freely estimated.*

Model identification:

For this model, we have that:

- Number of pieces of information in the input variance-covariance matrix: 6 variances and 15 covariances ($p = 21$)
- For the unstandardized solution, the number of model parameters is: 4 factor loadings + 2 factor variances + 1 factor covariance + 6 residual variances ($q = 13$)
- Degrees of freedom: $p-q=8$ ☑ Since the degree of freedom is greater than zero, the statement is correct (i.e., model 1 is overidentified).

SCORING:

- 1 credit for mentioning the degrees of freedom

- 1 credit for the correct p
- 1 credit for the correct q
- **Total: 3 credits**

(F6) **Goodness-of-fit indices:** Provide the RMSEA and CFI of model **CFA1**. Evaluate these two indices and decide whether model CFA1 fits the data.

Element	Value	Interpretation
RMSEA	0.051	Following Hu and Bentler (1999), CFA1 fits the data well. Criteria for an acceptable fit: CFI greater than or equal to .95, RMSEA smaller than or equal to 0.06.
CFI	0.994	

SCORING:

- 1 credit per correct value and interpretation
- **Total: 2 credits**

(F7) **Areas of local misfit:** Provide and interpret the following indices.

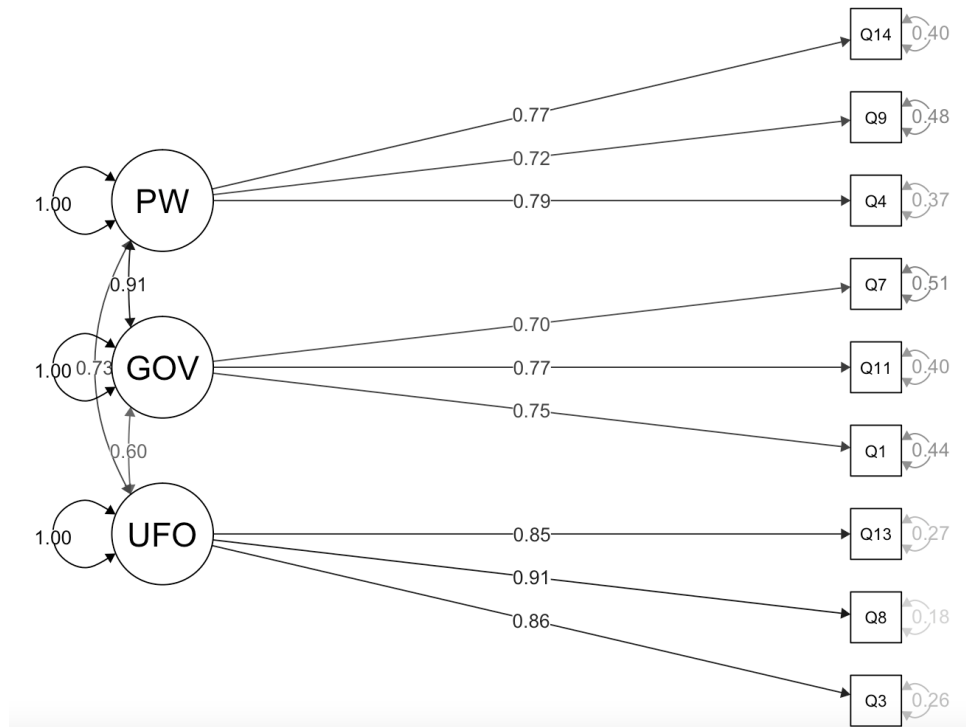
Element	Value	Interpretation
Modification indices Q3 \sim Q13	8.124	A modification index reflects an approximation of how much the overall χ^2 will decrease if a fixed or constrained parameter is freely estimated. If we freely estimate the covariance between Q3 and Q13, the overall chi square will decrease by 8.124.
Modification indices UFO \sim Q6	14.696	If we introduce a cross-loading so that Q6 also loads to the UFO latent factor the overall chi square will decrease by 14.696.
Standardized residual for Q1 and Q11	3.694	It indicates the discrepancy between the observed covariance between Q1 and Q11 and the model implied covariance (usual cut-off points: [-2;2]).

SCORING:

- 1 credit per correct answer/plausible interpretation

- **Total: 6 credits**

(F8) **Path diagram and model specification:** Another researcher has included more CTS variables and specified a CFA model with three factors (model **CFA2**). Model **CFA2** is shown in the figure below.



Using the information from the path diagram, extract the following **standardized parameter estimates of model CFA2** from the path diagram and interpret them.

Element	Estimate	Explanation
PW= \sim Q9	0.72	This is the standardized factor loading that indicates that item Q9 loads to the first latent dimension PW. It represents the correlation between Q9 indicator and latent factor.
GOV \sim UFO	0.60	This element represent correlation between two latent factors GOV and UFO, and it indicates positive, moderate association among them.
Residual variance of Q11	0.40	This is the standardized residual for item Q11 that represents amount of variance in Q11 that is not explained by GOV latent factor

SCORING:

- 1 credit per correct answer/plausible explanation

- **Total: 6 credits**

File upload

Please upload the R code file for the factor analysis task and label it " FactorAnalysis.R ".	File upload (max. 2 GB):	
---	-----------------------------	--

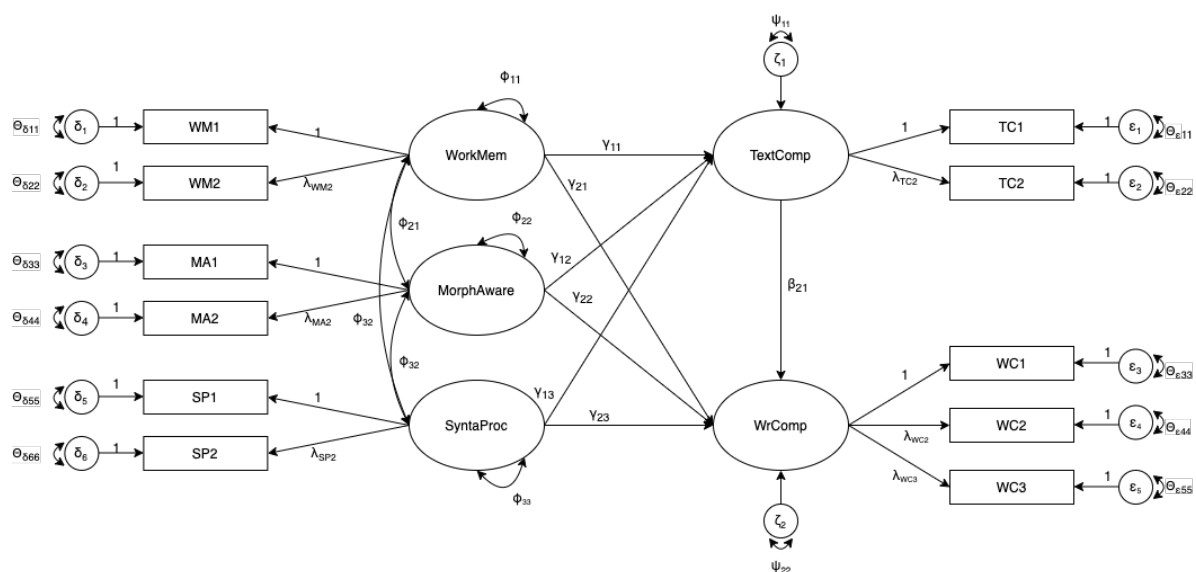
Structural Equation Models

A team of researchers wants to investigate the predictors of individual differences in written composition (**WrComp**) for grade 4 and grade 5 students. The **predictors** are:

- Morphological awareness (**MorphAware**)
- Syntactic processing (**SyntaProc**)
- Working memory (**WorkMem**)

The researchers further hypothesize that the observed predicted relations are **mediated by** text comprehension (**TextComp**). They posited a structural equation model, named “**SEM1**”, which is shown below. Notice that this model does not contain a mean structure.

Model SEM1:



(S1) Specify the **measurement models** for the latent variables **WorkMem**, **SyntaProc**, and **WrComp** by providing their model equations. Fill in the equations here.

Manifest indicators	Model equation
$WM1 =$	$1 \times WorkMem + 0 \times MorphAware + 0 \times SyntaProc + \delta_1$
$WM2 =$	$\lambda_{WM2} \times WorkMem + 0 \times MorphAware + 0 \times SyntaProc + \delta_2$
$SP1 =$	$0 \times WorkMem + 0 \times MorphAware + 1 \times SyntaProc + \delta_5$
$SP2 =$	$0 \times WorkMem + 0 \times MorphAware + \lambda_{SP2} \times SyntaProc + \delta_6$
$WC1 =$	$0 \times TextComp + 1 \times WrComp + \varepsilon_3$

$WC2 =$	$0 \times TextComp + \lambda_{WC2} \times WrComp + \varepsilon_4$
$WC3 =$	$0 \times TextComp + \lambda_{WC3} \times WrComp + \varepsilon_5$

SCORING:

- 1 credit per correct equation in the measurement part
- **Total: 7 credits**

(S2) Specify the model further by providing the model equations for the **structural part**.

Dependent latent variables	Model equation
$TextComp =$	$0 \times WrComp + 0 \times TextComp + \gamma_{11} \times WorkMem + \gamma_{12} \times MorphAware + \gamma_{13} \times SyntaProc + \zeta_1$
$WrComp =$	$\beta_{21} \times TextComp + 0 \times WrComp + \gamma_{21} \times WorkMem + \gamma_{22} \times MorphAware + \gamma_{23} \times SyntaProc + \zeta_2$

- 1 credit per correct equation in the structural part
- **Total: 2 credits**

(S3) Specify model **SEM1 in lavaan**. Provide the R code for the model specification.

```

Model specification in lavaan:
SEM1 <- '
    ## Measurement models
    WorkMem =~ WM1 + WM2
    MorphAware =~ MA1 + MA2
    SyntaProc =~ SP1 + SP2
    TextComp =~ TC1 + TC2
    WrComp =~ WC1 + WC2 + WC3

    ## Structural model
    TextComp ~ WorkMem + MorphAware + SyntaProc
    WrComp ~ WorkMem + MorphAware + SyntaProc + TextComp

    ## Covariance structure of the independent latent variables
    WorkMem ~~ WorkMem + MorphAware
  
```

```
SyntaProc ~ SyntaProc + WorkMem
MorphAware ~ MorphAware + SyntaProc
```

SCORING:

- 1 credit for each of the correct lines of measurement models and structural models parts
- 1 credit for each of the covariances between WorkMem, MorphAware, and SyntaProc
- *The variances of the predictors can be both specified or not specified, and the students still could get full credits.*
- **Total: 10 credits**

(S4) Estimate model **SEM1 in lavaan** using the **SEM dataset** and maximum-likelihood estimation. Provide the R code for the model estimation.

Note: You may read this data file using the read.csv2() function in R. Decimals are indicated by commas, and the data file contains the variable names in the first row. Example: data <- read.csv2(file = "SEM.csv", header = TRUE, dec = ","). The data provided is raw and consists of 11 observed variables and a grouping variable.

Model estimation in lavaan:

```
SEM1.fit <- sem(SEM1, data = data, estimator = "ML")
```

SCORING:

- 1 credit for choosing the appropriate function for model estimation (e.g., sem(), lavaan())
- 1 credit for the correct setup
- *ML estimator can be dropped, and students could still gain full credits.*
- **Total: 2 credits**

(S5) Request a **SEM1 model summary** which contains the standardized model parameters, the variance explanations of all dependent variables (R^2), and the model fit indices. Provide the R code for calling this model summary.

Model summary in lavaan:

```
summary(SEM1.fit, standardized = TRUE, rsquare = TRUE, fit.measures = TRUE)
```

SCORING:

- 0.5 credit for each of the options in the summary function
- **Total: 2 credits**

(S6) Determine whether **SEM1** exhibits **acceptable fit**. Inspect the chi-square statistic, CFI, RMSEA, and SRMR. Provide the values of these indices with three digits after the comma. *Note: As an intermediate check, make certain your chi-square statistic has 34 degrees of freedom.*

Fit index	Value
Chi-square statistic	43.808
CFI	0.997
RMSEA	0.024
SRMR	0.020

Model fit evaluation:

The criteria for an acceptable fit: insignificant chi-square statistic, CFI greater than or equal to .95, RMSEA smaller than or equal to 0.08, SRMR smaller than or equal to 0.08. The model appears to exhibit excellent fit as evidenced by the fit indices below.

SCORING:

- 1 credit for each correct fit index and their interpretation (chi-square statistic, CFI, RMSEA, SRMR) (sub-total: 4)
- 1 credit for the overall conclusion
- **Total: 5 credits**

(S7) In their study, the researchers tested the **mediation effect of text comprehension (TextComp)**. Describe **two procedures** which could be employed to test for the presence of mediation.

Mediation Testing

- **Model comparisons via the likelihood-ratio test:** 1. Run SEM with all effects; 2. Run SEM with restriction of direct effects = 0; 3. Compare the models using likelihood-ratio test.
- **Tests of the indirect effect against zero (e.g., bootstrapping):** 1. Run SEM with all effects; 2. Test the estimated parameters of indirect effects against zero using the Wald test.

SCORING:

- 1 credit for each description of the procedure (steps, elements)
- **Total: 2 credits**

(S8) Name two procedures to deal with possible **multicollinearity** in model **SEM1**.

Multicollinearity procedures

- Marsh's constrained approach
- Second-order factor approach
- Phantom-variable approach

SCORING:

- 1 credit for each correct naming of the procedure
- **Total: 2 credits**

The study includes students from two grades: grade 4 and grade 5. This brings the questions of whether **the measures and their structural relations are the same across groups**. Hence, the researchers would like to test for invariance.

To complete the following tasks, please use the grouping variable provided in the 12th column of the raw data. This variable “group” is a factor with two levels.

- (S9) Using the specification of model **SEM1** from *task S3*, extend this model to a multi-group structural equation model and test the following **levels of measurement invariance**:
- a. Configural invariance
 - b. Metric invariance
 - c. Regression (structural relations) invariance

Provide the R code for each model estimation.

Model estimation in lavaan:	
<code>configural.fit <-</code>	<code>sem(SEM1, data = data, group = "group")</code>
<code>metric.fit <-</code>	<code>sem(SEM1, data = data, group = "group", group.equal = c("loadings"))</code>
<code>structural.fit <-</code>	<code>sem(SEM1, data = data, group = "group", group.equal = c("loadings", "regressions"))</code>

SCORING:

- 1 credit for each correct (configural/metric/structural)
- **Total: 3 credits**

- (S10) Using the estimated models from *task S9*, compare the invariance models via chi-square difference tests. Specifically, compare the configural model to the metric model, and the metric model to the structural invariance model. Provide the results of these comparisons and the respective chi-square values below.

Note: You may use `anova()`-function to perform the tests.

Measurement invariance level	Df	Chi2	Chi2 diff	Df diff	p
Configural	68	61.405	-	-	-

Metric	74	88.366	26.961	6	0.0001473
Regression	81	103.829	15.464	7	0.03049

Explain briefly which level(s) of measurement invariance is (are) supported.

Explanation of invariance level

The chi-square difference concludes that there is no evidence of metric or structural (regression) invariance. That is, the chi-square diff for both metric and structural are significant.

SCORING:

- 3 credits for each correct row in the table with invariance models
- 1 credit for correctly identified level of measurement invariance (i.e., configural invariance)
- 1 credits for the correct explanation (i.e., significant chi2 values of the differences)
- **Total: 5 credits**

File upload

Please upload the **R code** file for the structural equation modeling task and label it "**SEM.R**".

File upload
(max. 2 GB):