Final Exam

Structural models (Spring 2019)

MAE4111 Structural Models
Centre for Educational Measurement at the University of Oslo (CEMO)

Date and time: 04 June 2019, 09:00-13:00

Welcome to the MAE4111 Structural Models exam!

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. describe something).
- You may simplify subscripts wherever appropriate (e.g., Y1 instead of Y_1).
- Keep your explanations and descriptions brief.
- Partial credits will be given.

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

Best regards,
Ronny Scherer

Name:	SUGGESTED SOLUTIONS + SCORING GUIDE

Results

Task	Credits	Max. credits
C1		2
C2		2
C3		3
A1		17
A2		9
A3		6
A4		4
D1		6
D2		8
D3		6
TOTAL:		**63**

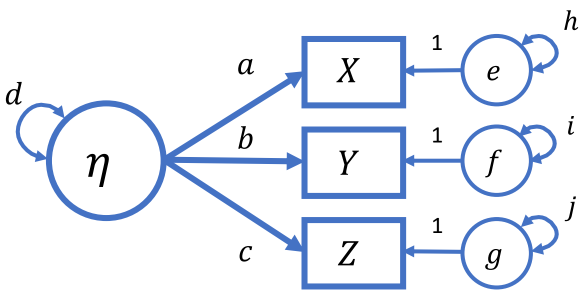
Expected time spent on the exam:
Instructor's time spent on the exam under exam conditions*3 = 80 min*3 = 240 min

Grading

Grade	Credits threshold	% Correct threshold
A	56	89 %
B	50	79 %
C	44	70 %
D	38	60 %
E	31.5	50 %
F	$\leq$ 31	$\leq$ 49 %

C1. Measurement model identification

In the following, a typical measurement model is shown that contains a latent (unobserved) variable η and three manifest (observed) indicator variables X, Y, and Z.



Describe two ways to identify this model and set the scale of the latent variable.

Option 1:
Constrain one factor loading to 1, for instance, $a = 1$. All other parameters, including the remaining two factor loadings b and c, the factor variance d, and the residual variances h, i, and j, are freely estimated. This way, one parameter in the model is fixed so that $df = 0$. This method is referred to as the "reference variable method".
Option 2:
Constrain the factor variance to 1, $d = 1$. All other parameters, including the factor loadings a, b, and c, and the residual variances h, i, and j are freely estimated. This way, one parameter in the model is fixed so that $df = 0$. This method is referred to as the "unit variance identification method".

Alternative option (option 3):
Constrain the average of the three factor loadings to one, $(a + b + c)/3 = 1$. Knowing two factor loadings is equivalent to knowing all three. All other parameters are freely estimated. This method is referred to as the "effects coding method".

Scoring notes:

- One credit per named option.
- Naming the methods is not sufficient; a specific description must be provided, including the parameter constraint.
- Option 1: 1 credit
- Option 2: 1 credit
- **Total: 2 credits**

C2. Mediation and moderation

Define the term "moderation" and describe how to test it statistically.

Aspect	Moderation
Definition	Moderation occurs when the relation between two variables X and Y depends on another variable M, the moderator. X usually denotes the predictor (independent variable), Y the outcome variable (dependent variable)
Statistical test(s)	If X and Z denote predictors of Y, then moderation is tested by including a product term XZ as another predictor in the model. The regression coefficient of this product term is tested against zero, $b/SE(b)$. *Note:* The product term is usually created after centering X and Z, for instance through mean- or residual-centering.

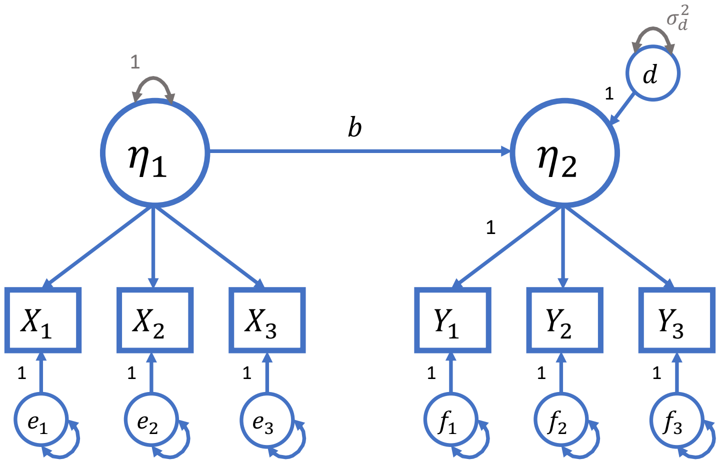
Scoring notes:

- Definition:
    - 1 credit
    - Definitions as provided above or similar
- Statistical tests:
    - 1 credit
    - The descriptions must clarify the statistical criterion used to determine the existence of moderation (as noted above)
- **Total: 2 credits**

C3. Small-sample structural equation models

Researchers are sometimes faced with the challenge of small sample sizes. Structural equation models, however, require relatively large samples for the model parameters to be estimated accurately. To still be able to estimate a structural equation model, several procedures to reduce the complexity of the model are available.

Name and describe **one procedure** to estimate the following structural equation model with the latent variables η_1 (eta1) and η_2 (eta2) if the sample size is small.



Name of the procedure:
e.g., Factor score regression
Description of the procedure:
In factor score regression, two steps are taken: 1. Estimate the two measurement models for eta1 and eta2 separately and extract the factor scores from them. 2. Use the factor scores of eta1 and eta2 as manifest variables in the model instead of eta1 and eta2 and estimate the regression coefficient b.

Alternative procedure: Fixed-parameter approach

Two steps are taken:

1. Estimate the measurement models for eta1 and eta2 separately and extract the model parameters (i.e., factor loadings, residual variances, factor variances) from them.
2. Specify the structural equation model with model parameters, except for b, fixed to the values obtained from step 1 and estimate the coefficient b.

Alternative procedure: Single-indicator latent variables

Two steps are taken:

1. Estimate the measurement models for eta1 and eta2 separately and calculate the reliabilities of eta1 and eta2 as McDonald’s Omega.

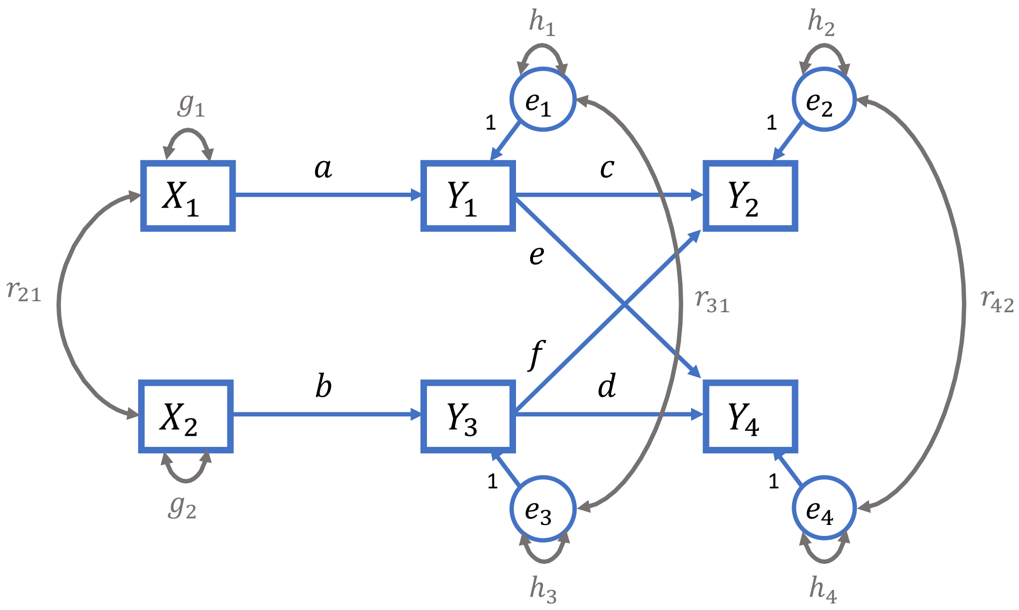
2. Specify the structural equation model with single-indicator latent variables of eta1 and eta2. These latent variables are indicated by the mean or sum scores of the original indicators (e.g., mean of X1-X3 and mean of Y1-Y3) with factor loadings fixed to 1 and the residual variances fixed to $(1 - Var) * Rel$.

Scoring notes:

- Name of the procedure: 1 credit
- Description:
  - 1 credit for describing step 1
  - 1 credit for describing step 2 (and step 3 if applicable)
- For the single-latent variable approach, the specification of the values to which the factor loadings and residual variances are fixed is not needed.
- **Total: 3 credits**

A1. Path model specification

The following path model contains several manifest variables and a set of relations among them.



a) *Divide-and-conquer:* Specify the roles of all the variables included in the model as independent, dependent, and residual variables.

Check the appropriate boxes below.

Variable name	Independent variable	Dependent variable	Residual
X_1	X	☐	☐
X_2	X	☐	☐
Y_1	☐	X	☐
Y_2	☐	X	☐
Y_3	☐	X	☐
Y_4	☐	X	☐
e_1	☐	☐	X
e_2	☐	☐	X
e_3	☐	☐	X
e_4	☐	☐	X

Scoring notes:
- 1 credit per correct column:
  - Independent variable column: 1 credit
  - Dependent variable column: 1 credit
  - Residual column: 1 credit
- **Total: 3 credits**

b) Write out the model equations for the dependent variables.

Equations for the dependent variables:
$Y_1 = 0 \cdot Y_1 + 0 \cdot Y_2 + 0 \cdot Y_3 + 0 \cdot Y_4 + a \cdot X_1 + 0 \cdot X_2 + e_1 = a \cdot X_1 + e_1$ $Y_2 = c \cdot Y_1 + 0 \cdot Y_2 + f \cdot Y_3 + 0 \cdot Y_4 + 0 \cdot X_1 + 0 \cdot X_2 + e_2 = c \cdot Y_1 + f \cdot Y_3 + e_2$ $Y_3 = 0 \cdot Y_1 + 0 \cdot Y_2 + 0 \cdot Y_3 + 0 \cdot Y_4 + 0 \cdot X_1 + b \cdot X_2 + e_3 = b \cdot X_2 + e_3$ $Y_4 = e \cdot Y_1 + 0 \cdot Y_2 + d \cdot Y_3 + 0 \cdot Y_4 + 0 \cdot X_1 + 0 \cdot X_2 + e_4 = e \cdot Y_1 + d \cdot Y_3 + e_4$

Scoring notes:
- 1 credit per correct equation for Y1, Y2, Y3, and Y4
- **Total: 4 credits**

c) Complete the entries of the B-matrix and the Γ-matrix below.

Matrix	Entries
B-matrix	$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ c & 0 & f & 0 \\ 0 & 0 & 0 & 0 \\ e & 0 & d & 0 \end{bmatrix}$
Γ-matrix	$\Gamma = \begin{bmatrix} a & 0 \\ 0 & 0 \\ 0 & b \\ 0 & 0 \end{bmatrix}$

Scoring notes:
- 1 credit per correct matrix column
- **Total: 6 credits**

d) Determine the number of available pieces of information without means (p), the number of parameters that need to be estimated in the model (q), and the resultant degrees of freedom of the model (df_M). Conclude whether or not the model is identified.

Indices	Numbers
# Observed pieces of information (p)	$p = \frac{6(6+1)}{2} = 21$

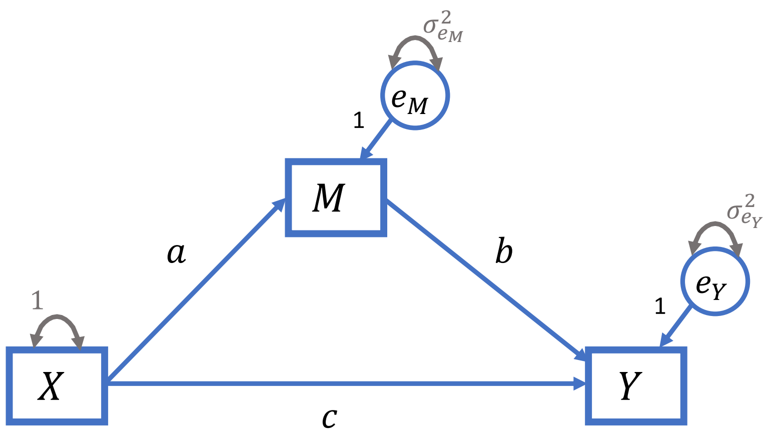
# Parameters to be estimated (q)	$q = 4 + 2 + 3 + 6 = 15$
Degrees of freedom of the model (df_M)	$df_M = 21 - 15 = 6$
Conclusion	The model is over-identified, $df_M > 0$.

Scoring notes:

- 1 credit per correct number (p, q, and df_M)
- 1 credit for the correct conclusion that the model is identified or over-identified.
- **Total: 4 credits**

A2. Indirect effect models

Many theories and hypotheses in educational and psychological science can be translated into statistical models with indirect effects. The figure below shows a simple indirect effects model hypothesizing a direct effect of a variable X on Y and an indirect effect via M.



a) Suppose all variables in the model are fully standardized (and so are the regression coefficients). Applying Wright's tracing rules, provide the formulas for the correlations among the three variables in the model r_{XM}, r_{MY}, and r_{XY}.

Provide the formulas for the three correlations here.

Correlation	Formula
$r_{XM} =$	a
$r_{MY} =$	$b + ac$
$r_{XY} =$	$c + ab$

Scoring notes:

- 1 credit per correct correlation
- **Total: 3 credits**

b) Provide the lavaan code to specify the indirect effects model (labelled as `ind.model`) and to estimate the indirect effect (labelled as `ind`) and the total effect (labelled as `tot`).

Lavaan code for model specification:

```
ind.model <- '

# Structural model
M ~a*X
```

```
Y ~b*M + c*X

# Indirect and total effect
ind:=a*b
tot:=ind+c

'
```

Scoring notes:

- 1 credit per correct specification line (4 lines needed: 2 for the structural model, 1 for defining the indirect effect and 1 for the total effect)
- If the indirect and/or total effect were labelled different but the calculations are correct, full credit is given.
- Note: The total effect can be specified as either `tot:=ind+c` or `tot:=a*b+c`.
- Effects can be labelled differently in the structural model without a loss of credits.
- **Total: 4 credits**

c) Name **two procedures** to estimate the standard error of the indirect effect.

Two procedures to estimate the SE:
- Delta method (Sobel test or other versions of it) - Bootstrapping - Monte Carlo method

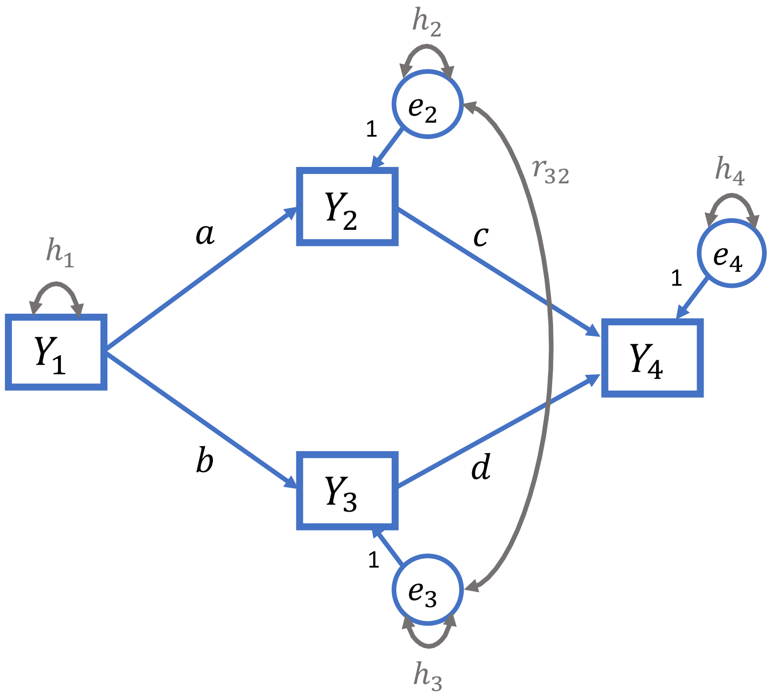
Scoring notes:

- Two of the above-mentioned approaches must be named.
- No detailed explanations required.
- 1 credit per procedure
- **Total: 2 credits**

A3. Model-implied covariance matrices

Structural equation models are based on covariance matrices. For a given model, researchers can derive a model-implied covariance matrix and compare this matrix to the observed covariance matrix.

In his book on test theory, McDonald (1999) presented the so-called "Double Chain Model", which is shown in the figure below.



a) Write out the model equations for the dependent variables Y_2, Y_3, and Y_4.

Fill in the questions here.

Dependent variables	Model equation
$Y_2 =$	$a \cdot Y_1 + e_2$
$Y_3 =$	$b \cdot Y_1 + e_3$
$Y_4 =$	$c \cdot Y_2 + d \cdot Y_3 + e_4$

Scoring notes:

- 1 credit per correct equation for Y2, Y3, and Y4
- **Total: 3 credits**

b) Making use of the covariance laws and the assumptions behind path models, provide the formulas for the model-implied formulas for the following variances and covariances: $Var(Y_1)$, $Var(Y_2)$, and $Cov(Y_2, Y_3)$.

Note. This covariance law may be helpful:
$Cov(aX + bY, cU + dV) = acCov(X, U) + adCov(X, V) + bcCov(Y, U) + bdCov(Y, V)$.

Fill in the three equations here.

Component	Formula
$Var(Y_1) =$	h_1
$Var(Y_2) =$	$a^2 \cdot h_1 + h_2$
$Cov(Y_2, Y_3) =$	$ab \cdot h_1 + r_{32}$

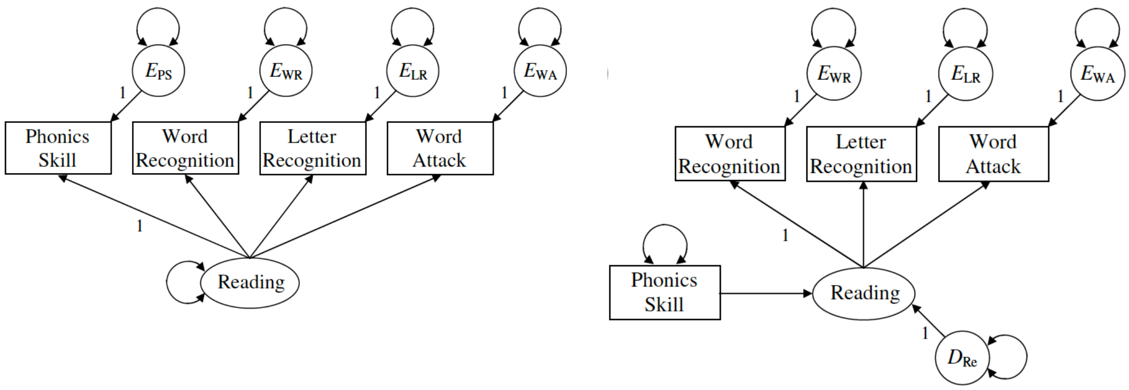
Notes:
In all calculations, we assume that the covariances between any independent variable Y and residual e is zero, $Cov(Y, e) = 0$.

- The variance of Y_1 equals h_1, because it represents an independent variable.
- Using the variance-covariance laws, we can derive the variance of Y_2 as follows:
$$\begin{aligned} Var(Y_2) = Cov(Y_2, Y_2) &= Cov(a \cdot Y_1 + e_2, a \cdot Y_1 + e_2) \\ &= a^2Cov(Y_1, Y_1) + aCov(Y_1, e_2) + aCov(e_2, Y_1) + Cov(e_2, e_2) \\ &= a^2Var(Y_1) + Var(e_2) = a^2 \cdot h_1 + h_2 \end{aligned}$$
- Making use of the covariance law described the task stimulus, we get:
$$\begin{aligned} Cov(Y_2, Y_3) &= Cov(a \cdot Y_1 + e_2, b \cdot Y_1 + e_3) \\ &= abCov(Y_1, Y_1) + aCov(Y_1, e_3) + bCov(e_2, Y_1) + Cov(e_2, e_3) \\ &= abVar(Y_1) + r_{32} = ab \cdot h_1 + r_{32} \end{aligned}$$

Scoring notes:
- 1 credit for each correct formula
- Notes on how to arrive at the formula are not considered in the scoring.
- **Total: 3 credits**

A4. The same but different?

Kline (2016) notices that some models are statistically equivalent but represent different theories or hypotheses. The following two models of reading comprehension (Reading) exemplify this.



Model 1 Model 2

Consider the relation between the latent variable Reading and the manifest variable Phonics Skill in both models. How can this relation be interpreted in Model 1 and in Model 2?

Note: Use the terms "measurement model" and "structural model".

Interpretation of the relation between Reading and Phonics Skill	
Model 1	Model 2
This relation is quantified by a factor loading. In this model, the latent variable Reading explains variance in the manifest indicator variable Phonics Skill. The variable Phonics Skill is part of the measurement model, yet not of the structural model. *Note:* This variable is the reference variable in the model due to the factor loading being constrained to 1. Model 1 represents a model of confirmatory factor analysis.	This relation is quantified by a regression coefficient. In this model, the manifest variable Phonics Skill explains variance in the latent variable Reading. The variable Phonics Skill is not part of the measurement model but of the structural model. Model 2 represents a multiple-causes-multiple-indicators (MIMIC) model.

Scoring notes:

- 1 credit for assigning Phonics Skill to either the measurement part (Model 1) or the structural part (Model 2) → 1+1 credits
- 1 credit for describing how the effects/relations concerning Phonics Skill can be interpreted → 1+1 credits
- **Total: 4 credits**

D1. Lavaan syntax

The R package lavaan provides several options to specify and estimate structural equation models.

Explain the following snippets of lavaan syntax to specify structural equation models.

Lavaan syntax	Explanation
`X~~Y`	The variables X and Y covary (or: are correlated).
`G=~X1+X2+X3`	The latent variable G is measured by three indicators X1, X2, and X3. This command defines a measurement model of G.
`G~X1+X2+X3`	The variable G is regressed on (or: predicted by) three variables X1, X2, and X3. This command defines a structural model. G is the dependent variable, X1-X3 are independent variables in this model.
`Y~a*X+a*Z`	The variable Y is regressed on two predictor variables X and Z with regression coefficients that are constrained to be equal (same label a). Y is the dependent variable, X and Z are independent in this model.
`G~~1*G`	The variable G covaries with itself with a regression coefficient of 1. This represents that the variance of the variable G is constrained to 1.
`sem(model1, data=PISA)`	This command estimates the structural equation model "model1" based on the data "PISA". Both model1 and PISA must have been specified and defined beforehand. By default, lavaan uses maximum-likelihood estimation to obtain the model parameters.

Scoring notes:

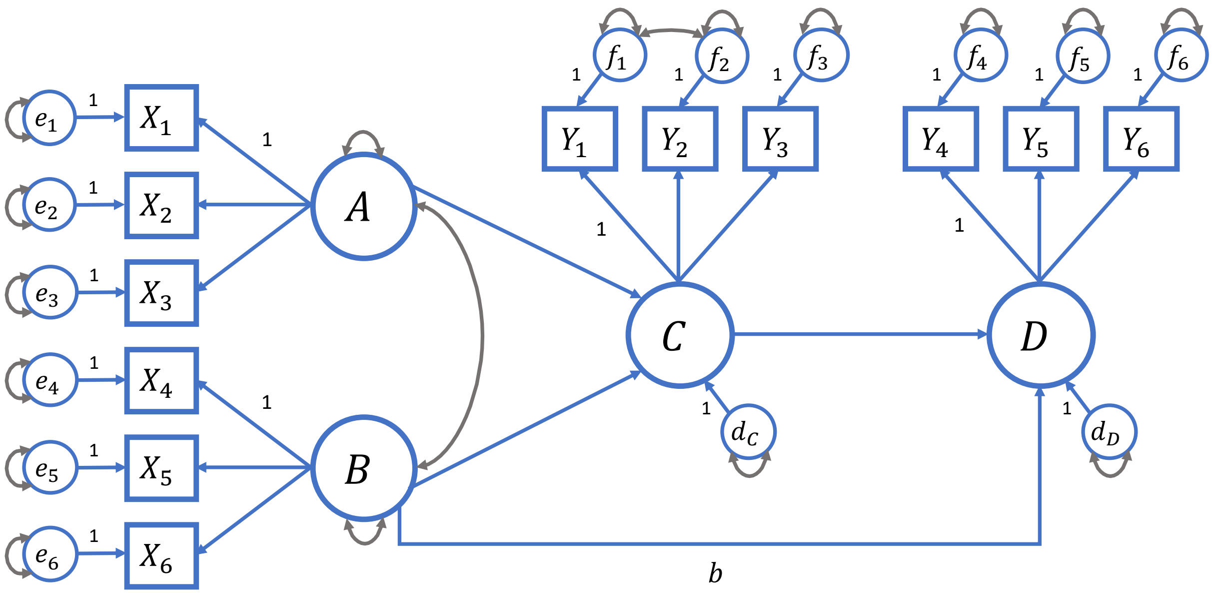
- 1 credit for each correct description

- No detailed descriptions for the sem() function needed (e.g., what exactly `model1` or `PISA` or the default estimator are).
- **Total: 6 credits**

D2. Complex structural equation models

When structural equation models are complex, researchers are advised to dissect the full model into the measurement part and the structural part.

The following model represents a complex structural equation model.



Provide the lavaan syntax to specify this model (labelled as `complex.sem`).

Lavaan code for model specification:

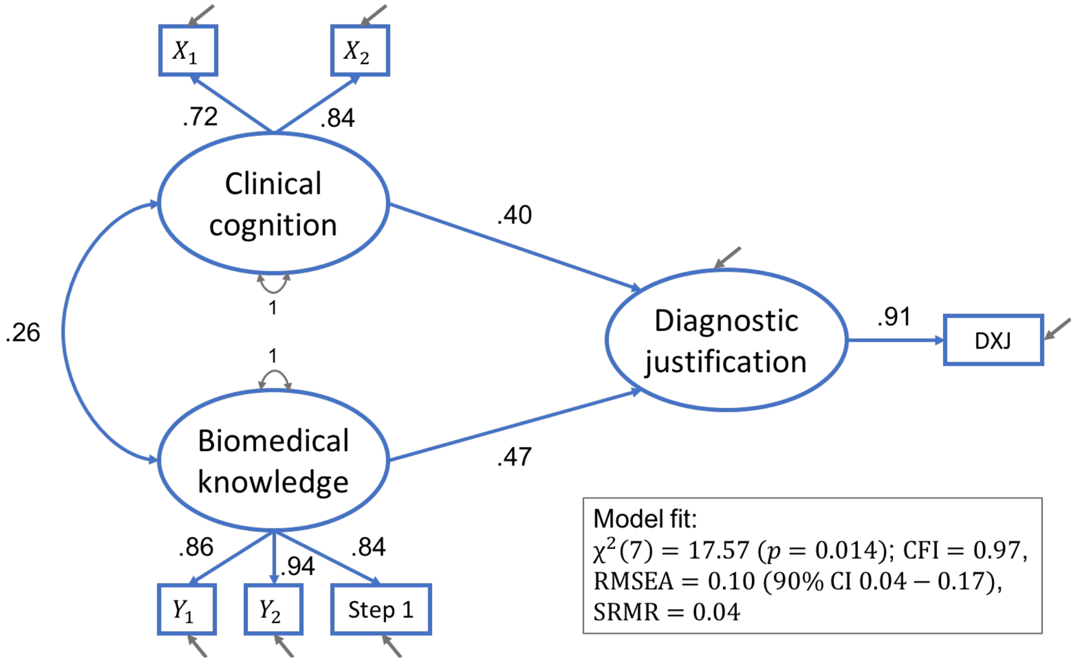
```
complex.sem <- '
# Measurement part
A =~X1+X2+X3
B =~X4+X5+X6
C =~Y1+Y2+Y3
D =~Y4+Y5+Y6
Y1 ~~Y2

# Structural part
C ~ A + B
D ~ C + b*B

# Covariances among independent variables
A ~~B
'
```

Scoring notes:

- 1 credit for each correct line of syntax (4 credits for the measurement models, 1 for the residual correlation between Y1 and Y2, 2 for the structural part, and 1 for the covariance between A and B)
- The path coefficient B→D does not have to be labelled as “b” in the syntax.
- **Total: 8 credits**

D3. Clinical reasoning in medical education

In their study of 133 medical students, Cianciolo et al. (2013) tested a structural equation model that connected students' biomedical knowledge, clinical cognition, and diagnostic justification:



Source: Adapted from Cianciolo et al. (2013), DOI: 10.1111/medu.12096, p. 313

The model shows the standardized parameter estimates.

a) Which of the following statements are true?

Statement	True	False
(1) The chi-square test suggests that the model-implied and observed covariance matrices do not differ significantly.	☐	X
(2) Diagnostic justification is represented as a single-indicator latent variable.	X	☐
(3) The regression coefficients may be incorrect due to multicollinearity issues with the two predictor variables.	☐	X
(4) The structural equation model exhibits exact fit to the data.	☐	X

Scoring notes:

- 1 credit for each correct selection
- **Total: 4 credits**

b) Use Wright's tracing rules to trace the correlations between Y_1 and Y_2 (of biomedical knowledge), and Y_1 (biomedical knowledge) and DXJ (diagnostic justification). Provide the calculations below.

Note. There is no need to actually calculate the final values of these correlations.

Correlation	Calculation
$r_{Y1,Y2} =$	0.86×0.94
$r_{Y1,DXJ} =$	$0.86 \times 0.47 \times 0.91 + 0.86 \times 0.26 \times 0.40 \times 0.91$

Scoring notes:

- 1 credit for each correct calculation
- **Total: 2 credits**